Lecture 11: Clustered and Longitudinal Data

Maochao Xu

Department of Mathematics
Illinois State University
mxu2@ilstu.edu

Longitudinal data

Longitudinal data: outcomes are repeated measurements over time on the same subjects; for example, the weights of the same people when they are 30, 40, 50 and 60 years old. Longitudinal data for the same individuals are likely to exhibit correlation between successive measurements.

Example: stroke The data are from an experiment to promote the recovery of stroke patients. There were three experimental groups:

- A was a new occupational therapy intervention;
- B was the existing stroke rehabilitation program conducted in the same hospital where A was conducted;
- C was the usual care regime for stroke patients provided in a different hospital.

There were eight patients in each experimental group. The response variable was a measure of functional ability, the Bartel index; higher scores correspond to better outcomes and the maximum score is 100. Each patient was assessed weekly over the eight weeks of the study.

Table *Functional ability scores measuring recovery from stroke for patients in three experimental groups over 8 weeks of the study.*

		Week							
Subject	Group	1	2	3	4	5	6	7	8
1	A	45	45	45	45	80	80	80	90
2	A	20	25	25	25	30	35	30	50
3	A	50	50	55	70	70	75	90	90
4	A	25	25	35	40	60	60	70	80
5	A	100	100	100	100	100	100	100	100
6	A	20	20	30	50	50	60	85	95
7	A	30	35	35	40	50	60	75	85
8	A	30	35	45	50	55	65	65	70
9	B	40	55	60	70	80	85	90	90
10	B	65	65	70	70	80	80	80	80
11	B	30	30	40	45	65	85	85	85
12	B	25	35	35	35	40	45	45	45
13	B	45	45	80	80	80	80	80	80
14	B	15	15	10	10	10	20	20	20
15	B	35	35	35	45	45	45	50	50
16	B	40	40	40	55	55	55	60	65
17	C	20	20	30	30	30	30	30	30
18	C	35	35	35	40	40	40	40	40
19	C	35	35	35	40	40	40	45	45
20	C	45	65	65	65	80	85	95	100
21	C	45	65	70	90	90	95	95	100
22	C	25	30	30	35	40	40	40	40
23	C	25	25	30	30	30	30	35	40
24	C	15	35	35	35	40	50	65	65

Data analysis

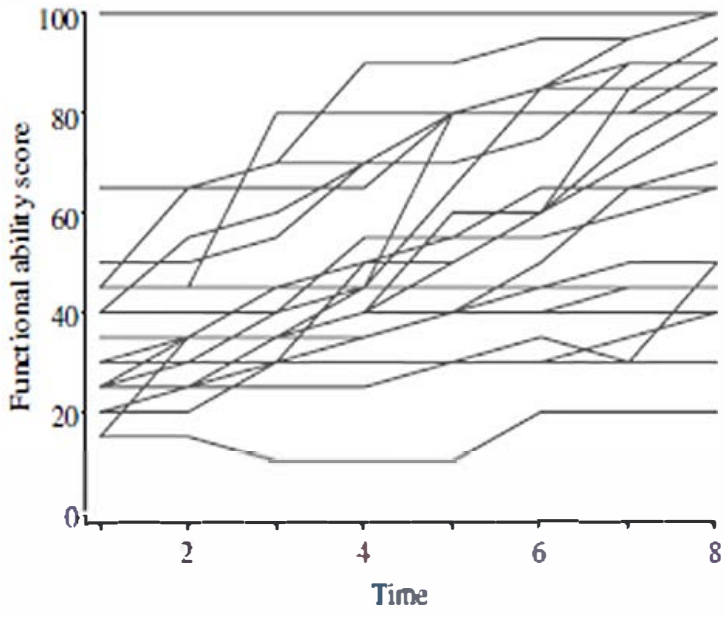


Figure *Stroke recovery scores of individual patients.*

Data analysis

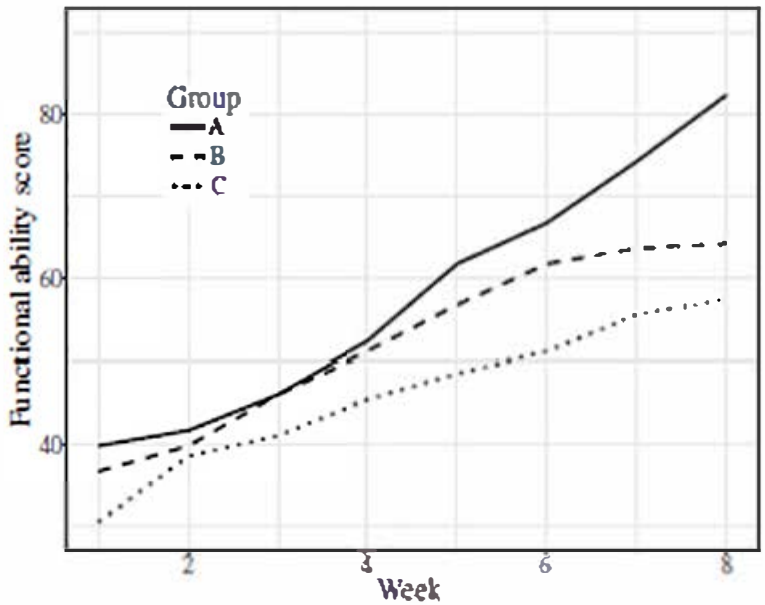


Figure *Average stroke recovery scores for groups of patients.*

Data analysis

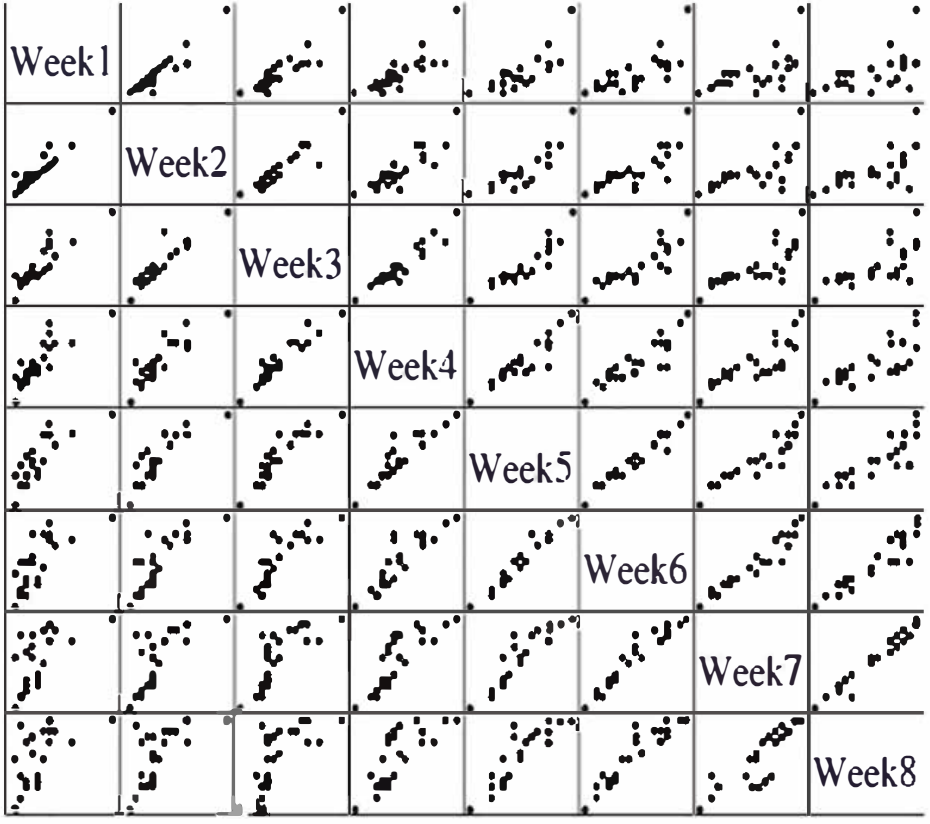


Figure *Scatter plot matrix for stroke recovery scores*

Data analysis

Table *Correlation coefficients for the stroke recovery scores.*

	Week						
	1	2	3	4	5	6	7
Week 2	0.93						
Week 3	0.88	0.92					
Week 4	0.83	0.88	0.95				
Week 5	0.79	0.85	0.91	0.92			
Week 6	0.71	0.79	0.85	0.88	0.97		
Week 7	0.62	0.70	0.77	0.83	0.92	0.96	
Week 8	0.55	0.64	0.70	0.77	0.88	0.93	0.98

These show high positive correlation between measurements made one week apart and decreasing correlation between observations further apart in time.

Modeling

A naive analysis

All 192 observations (for 3 groups $\times$ 8 subjects $\times$ 8 times) are assumed to be independent with

$$E(Y_{ijk}) = \alpha_i + \beta t_k + e_{ijk}$$

where
Y_{ijk} is the score at time t_k $(k = 1, \ldots, 8)$ for patient j $(j = 1, \ldots, 8)$ in group i (where $i = 1$ for group A, $i = 2$ for group B and $i = 3$ for group C);
α_i is the mean score for group i;
β is a common slope parameter;
t_k denotes time ($t_k = k$ for week k, $k = 1, \ldots, 8$);

The random error terms e_{ijk} are all assumed to be independent. The null hypothesis

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3,$$

and alternative hypothesis

$$H_a : \alpha_1 > \alpha_2 > \alpha_3.$$

Modeling

The other model

The slopes may differ between the three groups so the following model was also fitted

$$E(Y_{ijk}) = \alpha_i + \beta_i t_k + e_{ijk},$$

where the slope parameter β_i denotes the rate of recovery for group i. Models can be compared to test the hypothesis

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

against an alternative hypothesis that the β's differ.

Note : Neither of these naive models takes account of the fact that measurements of the same patient at different times are likely to be more similar than measurements of different patients.

Fitting

Table 11.3 *Results of naive analyses of stroke recovery scores, assuming all the data are independent*

Parameter	Estimate	Standard error
Model (1)		
α_1	36.842	3.971
$\alpha_2 - \alpha_1$	−5.625	3.715
$\alpha_3 - \alpha_1$	−12.109	3.715
β	4.764	0.662
Model (2)		
α_1	29.821	5.774
$\alpha_2 - \alpha_1$	3.348	8.166
$\alpha_3 - \alpha_1$	−0.022	8.166
β_1	6.324	1.143
$\beta_2 - \beta_1$	−1.994	1.617
$\beta_3 - \beta_1$	−2.686	1.617

Conclusion: For model (2), the Wald statistics for the intercepts are very small compared with the standard Normal distribution which suggests that the intercepts are not different (i.e., on average the groups started with the same level of functional ability).

Data summary

For the stroke data, appropriate summary statistics are the intercept and slope of the individual regression lines. The intercept and slope estimates and their standard errors for each of the 24 stroke patients are shown in the following.

Table *Estimates of intercepts and slopes (and their standard errors) for each subject.*

Subject	Intercept	(std. error)	Slope	(std. error)
1	30.000	(7.289)	7.500	(1.443)
2	15.536	(4.099)	3.214	(0.812)
3	39.821	(3.209)	6.429	(0.636)
4	11.607	(3.387)	8.393	(0.671)
5	100.000	(0.000)	0.000	(0.000)
6	0.893	(5.304)	11.190	(1.050)
7	15.357	(4.669)	7.976	(0.925)
8	25.357	(1.971)	5.893	(0.390)
9	38.571	(3.522)	7.262	(0.698)
10	61.964	(2.236)	2.619	(0.443)
11	14.464	(5.893)	9.702	(1.167)
12	26.071	(2.147)	2.679	(0.425)
13	48.750	(8.927)	5.000	(1.768)
14	10.179	(3.209)	1.071	(0.636)
15	31.250	(1.948)	2.500	(0.386)
16	34.107	(2.809)	3.810	(0.556)
17	21.071	(2.551)	1.429	(0.505)
18	34.107	(1.164)	0.893	(0.231)
19	32.143	(1.164)	1.607	(0.231)
20	42.321	(3.698)	7.262	(0.732)
21	48.571	(6.140)	7.262	(1.216)
22	24.821	(1.885)	2.262	(0.373)
23	22.321	(1.709)	1.845	(0.339)
24	13.036	(4.492)	6.548	(0.890)

ANOVA

Table *Analysis of variance of intercept estimates*

Source	d.f.	Mean square	F	p-value
Groups	2	30	0.07	0.94
Error	21	459		

Parameter	Estimate	Std. error
α_1	29.821	7.572
$\alpha_2 - \alpha_1$	3.348	10.709
$\alpha_3 - \alpha_1$	−0.018	10.709

Table *Analysis of variance of slope estimates*

Source	d.f.	Mean square	F	p-value
Groups	2	15.56	1.67	0.21
Error	21	9.34		

Parameter	Estimate	Std. error
β_1	6.324	1.080
$\beta_2 - \beta_1$	−1.994	1.528
$\beta_3 - \beta_1$	−2.686	1.528

Repeated measures models for Normal data

Suppose there are N study units or subjects with n_i measurements for subject i (e.g., n_i longitudinal observations for person i or n_i observations for cluster i). Let $\mathbf{y}_i$ denote the vector of responses for subject i and let $\mathbf{y}$ denote the vector of responses for all subjects

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \qquad \text{so } \mathbf{y} \text{ has length } \sum_{i=1}^{N} n_i.$$

A Normal linear model for $\mathbf{y}$ is

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\mu}; \qquad \mathbf{y} \sim \mathrm{MVN}(\boldsymbol{\mu}, \mathbf{V}).$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},$$

$\mathbf{X}_i$ is the $n_i \times p$ design matrix for subject i and $\boldsymbol{\beta}$ is a parameter vector of length p.

The variance–covariance matrix for measurements for subject i is

$$\mathbf{V}_i = \begin{bmatrix} \sigma_{i11} & \sigma_{i12} & \cdots & \sigma_{i1n_i} \\ \sigma_{i21} & \ddots & & \vdots \\ \vdots & & \ddots & \\ \sigma_{in1} & & & \sigma_{in_in_i} \end{bmatrix},$$

and the overall variance–covariance matrix has the block diagonal form

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{O} & & \mathbf{O} \\ \mathbf{O} & \mathbf{V}_2 & & \mathbf{O} \\ & & \ddots & \\ \mathbf{O} & \mathbf{O} & & \mathbf{V}_N \end{bmatrix},$$

assuming that responses for different subjects are independent (where $\mathbf{O}$ denotes a matrix of zeros). Usually the matrices $\mathbf{V}_i$ are assumed to have the same form for all subjects.

If the elements of $\mathbf{V}$ are known constants, then $\boldsymbol{\beta}$ can be estimated from the likelihood function for model (11.3) or by the method of least squares. The maximum likelihood estimator is obtained by solving the score equations

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial l}{\partial \boldsymbol{\beta}} = \mathbf{X}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{X}_i^T \mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) = \mathbf{0}$$

where l is the log-likelihood function. The solution is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} = (\sum_{i=1}^{N} \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}(\sum_{i=1}^{N} \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{y}_i)$$

with

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} = (\sum_{i=1}^{N} \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}$$

and $\hat{\boldsymbol{\beta}}$ is asymptotically Normal.

If the estimate $\widehat{\mathbf{V}}$ is substituted for $\mathbf{V}$ in Equation, the variance of $\boldsymbol{\beta}$ is likely to be underestimated. Therefore, a preferable alternative is

$$\mathbf{V}_s(\widehat{\boldsymbol{\beta}}) = \mathfrak{I}^{-1}\mathbf{C}\mathfrak{I}^{-1},$$

where

$$\mathfrak{I} = \mathbf{X}^T\widehat{\mathbf{V}}^{-1}\mathbf{X} = \sum_{i=1}^{N}\mathbf{X}_i^T\widehat{\mathbf{V}}_i^{-1}\mathbf{X}_i$$

and

$$\mathbf{C} = \sum_{i=1}^{N}\mathbf{X}_i^T\widehat{\mathbf{V}}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})^T\widehat{\mathbf{V}}_i^{-1}\mathbf{X}_i,$$

where $\widehat{\mathbf{V}}_i$ denotes the ith sub matrix of $\widehat{\mathbf{V}}$. $\mathbf{V}_s(\hat{\boldsymbol{\beta}})$ is called the **information sandwich estimator**, because $\mathfrak{I}$ is the information matrix.

It is also sometimes called the **Huber estimator**. It is a consistent estimator of var($\widehat{\boldsymbol{\beta}}$) when $\mathbf{V}$ is not known, and it is robust to mis-specification of $\mathbf{V}$.

Commonly used correlation matrices

1. All the off-diagonal elements are equal so that

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}.$$

This is appropriate for clustered data where it is plausible that all measurements are equally correlated, for example, for elements within the same primary sampling unit such as people living in the same area. The term ρ is called the **intra-class correlation coefficient**.

The **exchangeable** matrix is called **equicorrelation** or **spherical**. If the off-diagonal term ρ can be written in the form $\sigma_a^2/(\sigma_a^2 + \sigma_b^2)$, the matrix is said to have **compound symmetry**.

The number of parameters needed for this variance–covariance matrix is $P = 2$, one for the variance (σ^2) and one for the correlation (ρ).

2. The off-diagonal terms decrease with "distance" between observations; for example, if all the vectors $\mathbf{y}_i$ have the same length n and

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & & \rho_{2n} \\ \vdots & & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix},$$

where ρ_{jk} depends on the "distance" between observations j and k. Examples include $\rho_{jk} = |t_j - t_k|$ for measurements at times t_j and t_k (provided these are defined so that $-1 \leq \rho_{jk} \leq 1$), or $\rho_{jk} = \exp(-|j-k|)$. One commonly used form is the first-order **autoregressive model** with $\rho^{|j-k|}$, where $|\rho| < 1$ so that

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & & \rho^{n-2} \\ \rho^2 & \rho & 1 & & \vdots \\ \vdots & & & \ddots & \\ \rho^{n-1} & \cdots & & \rho & 1 \end{bmatrix}.$$

The number of parameters needed for this variance–covariance matrix is $P = 2$.

3. All the correlation terms may be different

$$\mathbf{V}_i = \sigma^2 \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & & \rho_{2n} \\ \vdots & & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix}.$$

This **unstructured correlation matrix** involves no assumptions about correlations between measurements but all the vectors $\mathbf{y}_i$ must be the same length n. It is only practical to use this form when the matrix $\mathbf{V}_i$ is not large relative to the number of subjects because the number of nuisance parameters ρ_{jk} is $P = n(n-1)/2$, which increases quadratically with n and may lead to convergence problems in the iterative estimation process. Sometimes it may be useful to fit a model with an unstructured correlation matrix and examine the estimates $\widehat{\rho}_{jk}$ for patterns that may suggest a simpler model.

Which one to use?

The AIC can be used to choose the best variance-covariance matrix.

To do this, a range of different variance-covariance matrices are tried, from the simplest naive independent matrix to the most complex unstructured matrix.

These models must use the same design matrix so that the number of regression parameters p stays fixed, but the number of nuisance parameters P varies.

Table *Akaike information criteria for the four correlation matrices used to model the stroke recovery data. P is the number of "nuisance" variance–covariance parameters. Every model has $p = 6$ regression parameters.*

Correlation matrix	P	AIC	Difference in AIC from the autoregressive model
Autoregressive	2	1320.3	0.0
Unstructured	29	1338.1	17.8
Exchangeable	2	1452.7	132.4
Independent	1	1703.6	383.3

Repeated measures models for non-Normal data

For the generalized linear model

$$\mathrm{E}(Y_i) = \mu_i, \quad g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i$$

for independent random variables $Y_1, Y_2, \ldots, Y_N$ with a distribution from the exponential family, the scores are

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i)}{\mathrm{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

for parameters β_j, $j = 1, \ldots, p$. The last two terms come from

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij}.$$

Therefore, the score equations for the generalized model (with independent responses $Y_i, i = 1, \ldots, N$) can be written as

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i)}{\mathrm{var}(Y_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0, \qquad j = 1, \ldots, p.$$

Repeated measures models for non-Normal data

For repeated measures, let $\mathbf{y}_i$ denote the vector of responses for subject i with $\mathrm{E}(\mathbf{y}_i) = \boldsymbol{\mu}_i$, $g(\boldsymbol{\mu}_i) = \mathbf{X}_i^T \boldsymbol{\beta}$ and let $\mathbf{D}_i$ be the matrix of derivatives $\partial \boldsymbol{\mu}_i / \partial \beta_j$. To simplify the notation, assume that all the subjects have the same number of measurements n.

The **generalized estimating equations** (GEEs) are

$$\mathbf{U} = \sum_{i=1}^{N} \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}.$$

These are also called the **quasi-score equations**. The matrix $\mathbf{V}_i$ can be written as

$$\mathbf{V}_i = \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}_i \mathbf{A}_i^{\frac{1}{2}} \phi,$$

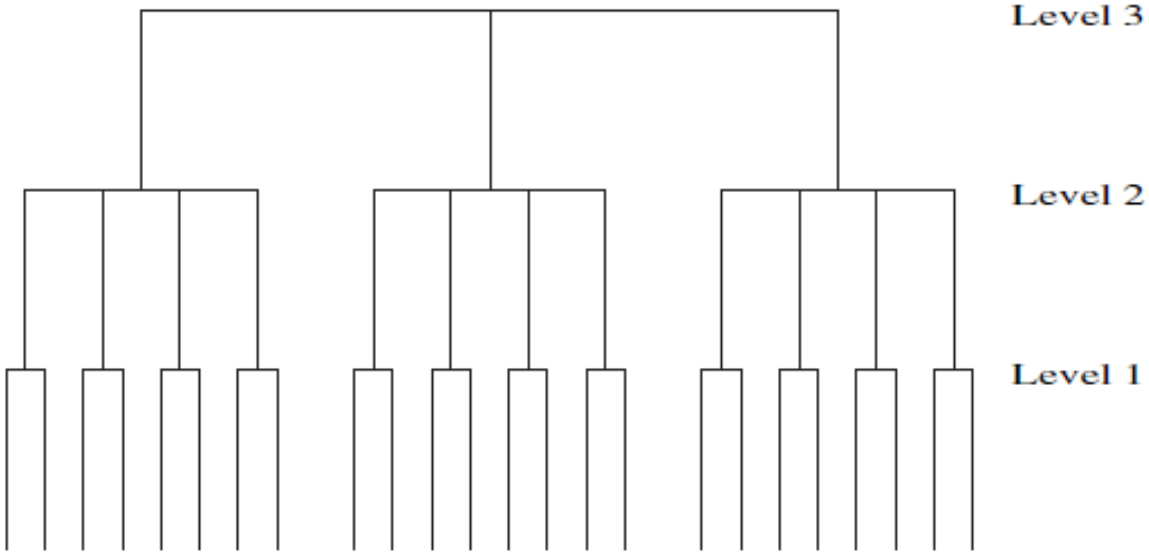
where $\mathbf{A}_i$ is the diagonal matrix with elements $\mathrm{var}(y_{ik})$, $\mathbf{R}_i$ is the correlation matrix for $\mathbf{y}_i$ and ϕ is a constant to allow for overdispersion.

Comments for GEE

- In general, there are no closed-form solutions, so the GEE estimates are obtained by using an iterative algorithm, that is iterative quasi-scoring procedure.
- GEE estimates of model parameters are valid even if the covariance is mis-specified (because they depend on the first moment, e.g., mean).
- If the correlation structure is mis-specified, the standard errors are not good, and some adjustments based on the data(empirical adjustment) are needed to get more appropriate standard errors.
- There is no likelihood function as this is an estimating procedure.
- Empirical based standard errors underestimate the true ones, unless very large sample size.

Multilevel modelling

An alternative approach for modelling repeated measures is based on considering the hierarchical structure of the study design. This is called multilevel modelling.



Mixed effect model

Let Y_{jk} denote the response of the kth subject in the jth cluster. For example, suppose Y_{jk} is the income of the kth randomly selected household in council area j, where council areas, the primary sam-pling units, are chosen randomly from all councils within a country or state. If the goal is to estimate the average household income μ, then a suitable model might be

$$Y_{jk} = \mu + a_j + e_{jk},$$

where a_j is the effect of area j and e_{jk} is the random error term. As areas were randomly selected and the area effects are not of primary interest, the terms a_j can be defined as independent, identically distributed random variables with $a_j \sim \mathrm{N}(0, \sigma_a^2)$. Similarly, the terms e_{jk} are independently, identically distributed random variables $e_{jk} \sim \mathrm{N}(0, \sigma_e^2)$ and the a_j's and e_{jk}'s are independent.

Mixed effect model

In this case

$$\mathrm{E}(Y_{jk}) = \mu,$$

$$\mathrm{var}(Y_{jk}) = \mathrm{E}\left[(Y_{jk}-\mu)^2\right] = \mathrm{E}\left[(a_j+e_{jk})^2\right] = \sigma_a^2+\sigma_e^2.$$

For households in the same area,

$$\mathrm{cov}(Y_{jk},Y_{jm}) = \mathrm{E}\left[(a_j+e_{jk})(a_j+e_{jm})\right] = \sigma_a^2,$$

and for households in different areas,

$$\mathrm{cov}(Y_{jk},Y_{lm}) = \mathrm{E}\left[(a_j+e_{jk})(a_l+e_{lm})\right] = 0.$$

In this model, the parameter μ is a fixed effect, and a_j is a random effect.

Mixed effect model

If $\mathbf{y}_j$ is the vector of responses for households in area j, then the variance–covariance matrix for $\mathbf{y}_j$ is

$$\mathbf{V}_j = \begin{bmatrix} \sigma_a^2+\sigma_e^2 & \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2+\sigma_e^2 & \sigma_a^2 & & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2+\sigma_e^2 & & \\ \vdots & & & \ddots & \\ \sigma_a^2 & & & \sigma_a^2 & \sigma_a^2+\sigma_e^2 \end{bmatrix}$$

$$= \sigma_a^2+\sigma_e^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & & \rho \\ \rho & \rho & 1 & & \\ \vdots & & & \ddots & \\ \rho & & & \rho & 1 \end{bmatrix},$$

where $\rho = \sigma_a^2/(\sigma_a^2+\sigma_e^2)$ is the intra-class correlation coefficient which describes the proportion of the total variance due to within-cluster variance.

If the responses within a cluster are much more alike than responses from different clusters, then σ_e^2 is much smaller than σ_a^2, so ρ will be near unity; thus, ρ is a relative measure of the within-cluster similarity.

Mixed effect model

If $\mathbf{y}_j$ is the vector of responses for households in area j, then the variance–covariance matrix for $\mathbf{y}_j$ is

$$\mathbf{V}_j = \begin{bmatrix} \sigma_a^2+\sigma_e^2 & \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2+\sigma_e^2 & \sigma_a^2 & & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2+\sigma_e^2 & & \\ \vdots & & & \ddots & \\ \sigma_a^2 & & & \sigma_a^2 & \sigma_a^2+\sigma_e^2 \end{bmatrix}$$

$$= \sigma_a^2+\sigma_e^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & & \rho \\ \rho & \rho & 1 & & \\ \vdots & & & \ddots & \\ \rho & & & \rho & 1 \end{bmatrix},$$

where $\rho = \sigma_a^2/(\sigma_a^2+\sigma_e^2)$ is the intra-class correlation coefficient which describes the proportion of the total variance due to within-cluster variance.

If the responses within a cluster are much more alike than responses from different clusters, then σ_e^2 is much smaller than σ_a^2, so ρ will be near unity; thus, ρ is a relative measure of the within-cluster similarity.

Mixed effect model

Consider longitudinal data in which Y_{jk} is the measurement at time t_k on subject j who was selected at random from the popu-lation of interest. A linear model for this situation is

$$Y_{jk} = \beta_0 + a_j + (\beta_1 + b_j)t_k + e_{jk},$$

where β_0 and β_1 are the intercept and slope parameters for the population, a_j and b_j are the differences from these parameters specific to subject j, t_k denotes the time of the kth measurement, and e_{jk} is the random error term. The terms a_j, b_j and e_{jk} may be considered as random variables with $a_j \sim \mathrm{N}(0, \sigma_a^2)$, $b_j \sim \mathrm{N}(0, \sigma_b^2)$, $e_{jk} \sim \mathrm{N}(0, \sigma_e^2)$, and they are all assumed to be independent. For this model

$$\mathrm{E}(Y_{jk}) = \beta_0 + \beta_1 t_k,$$

$$\mathrm{var}(Y_{jk}) = \mathrm{var}(a_j) + t_k^2 \mathrm{var}(b_j) + \mathrm{var}(e_{jk}) = \sigma_a^2 + t_k^2 \sigma_b^2 + \sigma_e^2,$$

$$\mathrm{cov}(Y_{jk}, Y_{jm}) = \sigma_a^2 + t_k t_m \sigma_b^2$$

for measurements on the same subject, and

$$\mathrm{cov}(Y_{jk}, Y_{lm}) = 0$$

for measurements on different subjects.

In this model, β_0 and β_1 are fixed effects, a_j and b_j are random effects and the aim is to estimate $\beta_0, \beta_1, \sigma_a^2, \sigma_b^2$ and σ_e^2.