Singapore Med J 2004 Vol 45(10) : 456

CME Article

Biostatistics 301A. Repeated measurement analysis (mixed models)

Y H Chan

In our last article, I discussed the use of the general linear model (GLM)[1] to analyse repeated measurement data and mentioned two major disadvantages:

1. Lost of subjects due to missing data in any of the time points (Table I).
2. The limitation of the availability of variance-covariance structure (only have two choices).

Table I. Subjects 2 and 3 are "lost to analysis".

Subject	Time 1	Time 2	Time 3
1	xxxx	Xxxx	xxxx
2	xxxx	missing	xxxx
3	xxxx	Xxxx	missing

To overcome the above two disadvantages, the Mixed Model technique can be used. We have to transform the usual longitudinal data form for repeated measurement (Table I) to the relational form (Table II) by using the SPSS Restructure option discussed in the last article[1].

Table II. Relational form of Table I.

Subject	Time	Score
1	1	Xxxx
1	2	Xxxx
1	3	Xxxx
2	1	Xxxx
2	2	missing
2	3	Xxxx
3	1	Xxxx
3	2	Xxxx
3	3	missing

In this case, only two data points are "lost", and the other information for subjects 2 and 3 are still included in the analysis.

Faculty of Medicine
National University of Singapore
Block MD11
Clinical Research Centre #02-02
10 Medical Drive
Singapore 117597

Y H Chan, PhD
Head
Biostatistics Unit

Correspondence to:
Dr Y H Chan
Tel: (65) 6874 3698
Fax: (65) 6778 5743
Email: medcyh@nus.edu.sg

Table III. Relational form of anxiety data set.

Subject	Anxiety	Trial	Score
1	Low	1	18
1	Low	2	14
1	Low	3	12
1	Low	4	6
2	Low	1	19
2	Low	2	12
2	Low	3	8
2	Low	4	4
Etc			

Table III shows the relational data form for the first two of the 12 subjects from our last article's anxiety example[1].

VARIANCE-COVARIANCE/CORRELATION STRUCTURES

For the GLM **Univariate** approach, the assumption for the within-subject variance-covariance is a Type H structure (or circular in form – correlation between any two levels of within-subject factor has the same constant value). The **Compound symmetry (CS)/ Exchangeable** structure would be appropriate. Table IV shows the structure for a 4 time-point study.

Table IV. Compound symmetry/exchangeable structure.

Variance-covariance Correlation

$$\begin{bmatrix} \sigma^2 & \alpha & \alpha & \alpha \\ \alpha & \sigma^2 & \alpha & \alpha \\ \alpha & \alpha & \sigma^2 & \alpha \\ \alpha & \alpha & \alpha & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

This structure is overly simplistic: the variance at all time points are the same and the correlation between any two measurements is the same – i.e. only need to estimate two parameters (σ^2 & ρ).

For the GLM **Multivariate** approach, the assumption that the correlation for each level of within-subject factor is different is modeled by an **Unstructured covariance** structure, see Table V.

Table V. Unstructured correlation structure.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

This structure is overly complex: the variance at all time points and the correlation between any two measurements are all different – i.e. need to estimate 4 variances and 6 covariances = 10 parameters! General form for the number of parameters to be estimated is given by [n + n(n-1)/2], where n = number of repeated trials.

Does the variance-covariance/correlation structure of our anxiety data satisfies any of the above 2 structures? Table VI shows the correlation structure of the anxiety data by using the ***Analyze, Correlate, Bivariate*** option.

We observe that the correlation between two time-points are not really similar (which accounts for the p=0.053 value for the sphericity's test shown in our last article, near rejection of sphericity assumption), thus the compound symmetry assumption may not be appropriate. That leaves us with the unstructured option only - but we need to estimate ten unknown parameters with 12 subjects! There would be concern that with such a small sample size (worse still, if we have missing data!), the variance-covariance structure assumed may not be very appropriate and the results would be based on these "could-be" unstable estimates. What other choices do we have? None if we use the GLM technique!

Using the **Mixed Model** technique, we have more variance-covariance choices. Taking a closer analysis on Table VI, the correlation between two adjacent time-points (Trial1 and Trial2, for example) is always higher than that of those between two time-points that are further apart (Trial1 and Trial3, for example). In such a situation, an appropriate structure could be the **1st Order Autoregressive, AR(1)**, which assumes that the correlation between adjacent time-points is the same and the correlation decreases by the power of the number of time intervals between the measures (Table VII).

Table VI. Correlation structure of anxiety data.

Correlations					
		Trial 1	Trial 2	Trial 3	Trial 4
Trial 1	Pearson Correlation	1	.488	.246	.223
	Sig. (2-tailed)	.	.107	.442	.487
	N	12	12	12	12
Trial 2	Pearson Correlation	.488	1	.812*	.803*
	Sig. (2-tailed)	.107	.	.001	.002
	N	12	12	12	12
Trial 3	Pearson Correlation	.246	.812*	1	.785*
	Sig. (2-tailed)	.442	.001	.	.003
	N	12	12	12	12
Trial 4	Pearson Correlation	.223	.803*	.785*	1
	Sig. (2-tailed)	.487	.002	.003	.
	N	12	12	12	12

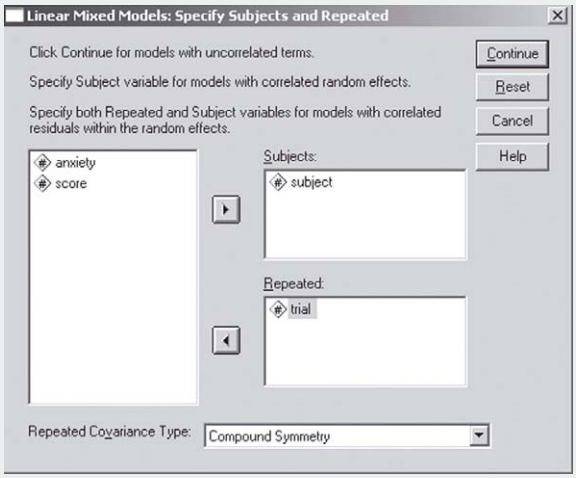
**. Correlation is significant at the 0.01 level (2-tailed).

Table VII. 1st Order Autoregressive, AR(1) structure.

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

We shall discuss the analysis of the Anxiety data using the Mixed Model technique with the above three structures (Compound symmetry, Unstructured and 1st Order Autoregressive). To perform the Mixed Model analysis, go to ***Analyze, Mixed Models, Linear*** to get Template I.

Template I. Specifying subjects and repeated measurements.



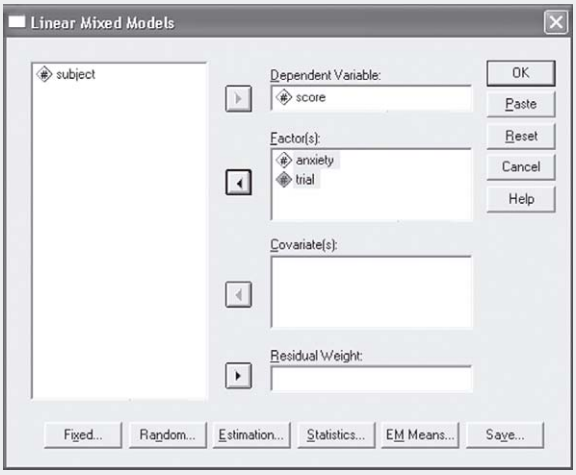
Put the variable "subject" into the Subject option and "trial" into the Repeated option. Choose "Compound Symmetry" for the Repeated Covariance Type option. Table VIII shows all the variance-covariance structures available in SPSS. A brief description for each structure could be obtained from the Help button.

Table VIII. Available variance-covariance structures.

- Ante-dependence: first order
- AR(1)
- AR(1): \heterogeneous
- ARMA(1,1)
- Compound symmetry
- Compound symmetry: correlation metric
- Compound symmetry: heterogeneous
- Diagonal
- Factor analytic: first order
- Factor analytic: first order, heterogeneous
- Huynh-Feldt
- Scaled identity
- Toeplitz
- Toeplitz: heterogeneous
- Unstructured
- Unstructured: correlation metric

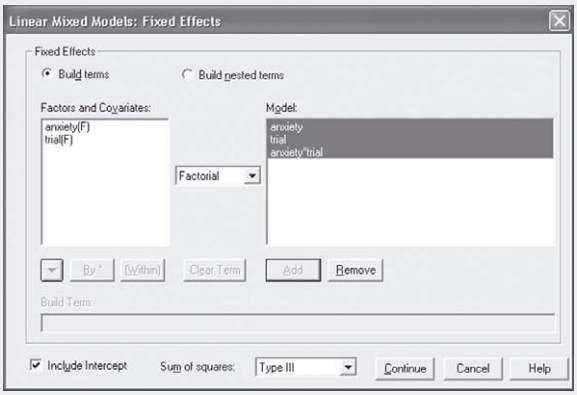
In Template I, click continue to get Template II.

Template II. Defining the variables.



Put "score" in the Dependent Variable option and "anxiety" and "trial" in the Factor option. Click on the Fixed folder to get Template III.

Template III. Defining the Fixed effects.



Highlight both "anxiety(F)" and "trial(F)", the Add button becomes visible. Leave the selection as Factorial and click on the Add button to define the Model (anxiety, trial, anxiety*trial). Click on Continue to return to Template II and click OK.

Table IXa shows the model defined and the covariance structure used – compound symmetry.

Table IXa. Model and covariance structure definition.

Model Dimension[a]						
		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables	Number of Subjects
Fixed Effects	Intercept	1		1		
	anxiety	2		1		
	trial	4		3		
	anxiety * trial	8		3		
Repeated Effects	trial	4	Compound Symmetry	2	Subject	12
Total		19		10		

[a] Dependent Variable: Score.

Table IXb. Covariance structure.

Estimates of Covariance Parameters[a]			
Parameter		Estimate	Std. Error
Repeated Measures	CS diagonal offset	2.5694444	.6634277
	CS covariance	3.6305556	1.9180907

[a] Dependent variable: Score.

Table IXb gives the variance (= 2.57) within each time-point, and the covariance between any two time-points is 3.63. The interest in our model building is not in the variance-covariance structure but in the treatment effects. But it is important to get the appropriate structure to obtain the appropriate standard errors for the inferences of the treatment effects.

Question: How do we know which covariance structure is the most appropriate?

Table IXc. Model selection measures.

Information Criteria[a]	
-2 Restricted Log Likelihood	184.546
Akaike's Information Criterion (AIC)	188.546
Hurvich and Tsai's Criterion (AICC)	188.870
Bozdogan's Criterion (CAIC)	193.924
Schwarz's Bayesian Criterion (BIC)	191.924

The information criteria are displayed in smaller-is-better forms.
[a] Dependent Variable: Score

Table IXc shows some basic measure for model selection which has to be used in comparison with the measures when other covariance structures are being used. The -2 Restricted Log Likelihood (-2RLL) value is valid for simple models and modifications of this value for more complicated models are given by Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (BIC). The BIC measurement is most 'severely adjusted' and is the recommended measure used for comparison. Hurvich and Tsai's Criterion (AAIC) and Bozdogan's Criterion (CAIC) are the adjustments of AIC for small sample sizes.

We want the "smaller is better" comparisons amongst the covariance structures. Table IXd gives the model selection measurements for the three covariance structures (Note: Unstructured and Unstructured correlation metric, see Table VIII, have the same model selection measurements but because of the small sample size, no estimates were obtained for the within-subject effects, trial and trial*anxiety, when the unstructured covariance structure was used!)

The appropriate covariance structure for this anxiety data is AR(1) as it has the smallest BIC among the 3 structures. We can also try the other various covariance structures (Table VIII) to compare their model selection measurements. Since the AR(1)

Table IXd. Model selection measures.

Information Criteria	Compound Symmetry (CS)	Unstructured: correlation metric	1st Order autoregressive, AR(1)
-2 RLL	184.546	168.924	176.828
AIC	188.546	188.924	180.828
AICC	188.870	196.510	181.153
CAIC	193.924	215.813	186.206
BIC	191.924	205.813	184.206

Table IXe. Results for the between and within subjects effects (p-values).

	Compound symmetry (CS)	Unstructured – correlation metric	1st order autoregressive, AR(1)
Anxiety	0.460	0.460	0.465
Trial	<0.001	<0.001	<0.001
Trial*anxiety	0.368	0.067	0.150

structure is chosen, then we should only use the between and within subjects results from this model. For discussion purposes, Table IXe shows the results for all three structures.

Using the compound symmetry structure, the results obtained are identical to those given by GLM Univariate analysis provided there is no missing data. GLM and Mixed Model will have different results if there were missing data. The between-subject effect (anxiety) of the Mixed Model is identical to GLM but though both models used the unstructured covariance structured, different results are obtained for the Trial*anxiety (p=0.138 for GLM). This is because both techniques used different estimation methods to derive the results – will not bore you with the details (those interested could refer to any standard statistical text on mixed model for further reading).

From Table IXe, we could see that the p-values are "similar" in terms of significance (not worrying about the exact values), the issue of using the "right covariance structure" arises when we have a difference of opinions in terms of significance for the between and within subjects effects for the different models.

We have only analyzed the Fixed effects aspects of the anxiety data in the above discussions, which means that the anxiety levels selected represented all levels of this factor or the researcher is only specifically interested in these two levels. In Template II, we have a Random folder which allows us to define the Random effects for the model. Factor effects are random if the levels of the factor that are used in the study represent a random sample of a larger set of potential levels. For the extension of the fixed effects to a mixed effect model (having both fixed and random effects), it would be most appropriate to seek the assistance of a biostatistician! Finally, the above analyses could be performed using other statistical software (SAS, S-plus and STATA) which offers more choices of covariance structures and greater flexibility in the modeling aspects for random effects.

Our next article, "Biostatistics 302. Principal component and factor analysis", will discuss the approach to summarising and uncovering any patterns in a set of variables (for example, a questionnaire).

REFERENCE

1. YH Chan. Biostatistics 301. Repeated measurement analysis. Singapore Med J 2004; 45:354-69.

SINGAPORE MEDICAL COUNCIL CATEGORY 3B CME PROGRAMME

Multiple Choice Questions (Code SMJ 200410A)

	True	False
Question 1. The results from the GLM Univariate procedure of repeated measurement analysis is identical to the Mixed Model procedure when:		
(a) The covariance structure is compound symmetry with no missing data.	❑	❑
(b) The covariance structure is compound symmetry with missing data.	❑	❑
(c) Unstructured covariance structure with no missing data.	❑	❑
(d) Unstructured covariance structure with missing data.	❑	❑
(e) As long as there is no missing data.	❑	❑
Question 2. We compare the appropriate covariance structure used for a model by comparing:		
(a) The p-values of the between-subject effects.	❑	❑
(b) The p-values of the within-subjects effects.	❑	❑
(c) The model selection measures between different covariance structures.	❑	❑
(d) The model selection measures within each covariance structure.	❑	❑
(e) All of the above.	❑	❑
Question 3. The Mixed Model technique has the following advantages over the GLM:		
(a) Allows random effects in the model.	❑	❑
(b) Gives faster results - shorter computing time.	❑	❑
(c) More likely to get a significant p-value.	❑	❑
(d) Can select the appropriate variance-covariance structure.	❑	❑
(e) Makes use of data from subjects with incomplete data.	❑	❑
Question 4. The following statements are true:		
(a) The Mixed Model procedure allows us to plot the data.	❑	❑
(b) The smaller-the-better criterion is used to compare the model selection measures for the different covariance structures.	❑	❑
(c) The most severely corrected measurement for the -2RLL is the AIC.	❑	❑
(d) The longitudinal data structure could be used for a Mixed Model analysis.	❑	❑
(e) The unstructured covariance structure gives the best results.	❑	❑

Doctor's particulars:

Name in full: _____

MCR number: _____ Specialty: _____

Email address: _____

Submission instructions:

A. Using this answer form

1. Photocopy this answer form.
2. Indicate your responses by marking the "True" or "False" box ☑
3. Fill in your professional particulars.
4. Either post the answer form to the SMJ at 2 College Road, Singapore 169850 OR fax to SMJ at (65) 6224 7827.

B. Electronic submission

1. Log on at the SMJ website: URL http://www.sma.org.sg/cme/smj
2. Either download the answer form and submit to smj.cme@sma.org.sg OR download and print out the answer form for this article and follow steps A. 2-4 (above) OR complete and submit the answer form online.

Deadline for submission: (October 2004 SMJ 3B CME programme): 25 November 2004

Results:

1. Answers will be published in the SMJ December 2004 issue.
2. The MCR numbers of successful candidates will be posted online at http://www.sma.org.sg/cme/smj by 20 December 2004.
3. Passing mark is 60%. No mark will be deducted for incorrect answers.
4. The SMJ editorial office will submit the list of successful candidates to the Singapore Medical Council.