MATH 60604A
Statistical modelling
§ 5e - Covariance models: first-order autoregressive covariance

HEC Montréal
Department of Decision Sciences

Alternative covariance structure

- We saw how to model the correlation between repeated measures from the same person, while assuming a compound symmetry structure.
- This structure was certainly plausible, since the parameter estimating within-person correlation was significantly different from zero.
- But how do we know if this is the best correlation structure? There may be another more appropriate structure for the correlation.
- There are several other covariance structures; in fact, SAS has a large number of possibilities. We will show several here; an exhaustive list can be found within SAS.

Choosing the covariance structure

- In many cases, the correlation structure is considered to be a nuisance parameter. More precisely, usually the primary interest in a study is the effect of the predictor variables, β.
- In this case, the covariance structure is not particularly interesting, other than the fact that we need to account for correlation to make sure our inference concerning β is valid.
- We could base the choice of covariance structure on information criteria if the models are not nested (provided they have the same covariates if the model is fitted using REML).

Auto-regressive structure

- The compound symmetry structure used before assumes the correlation between two observations is always the same.
- When we have repeated measures taken at different time points, as we do here, it's possible that the magnitude of the correlation depends on the amount of time between observations.
- We might believe that the closer together observations are in time, the more they are correlated. The autoregressive of order 1, or AR(1), structure allows us to do this.
- The AR(1) model has two parameters: a correlation parameter ρ and a variance parameter σ^2.

Auto-regressive structure

- For subject i with five repeated measurements, the correlation matrix is

$$\mathbf{R}_i = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}.$$

- The covariance structure is

$$\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{R}_i.$$

Correlation for autoregressive model

- Thhe (conditional) correlation between two observations separated by one time point (two weeks, in this example) is $\rho \in (-1, 1)$.
- Two observations separated by two time points (four weeks, in this example) is ρ^2, and so on.
- When $0 < \rho < 1$, the sequence ρ, ρ^2, ρ^3, ρ^4, ..., is decreasing. Consequently, the correlation between two observations decreases exponentially as a function of the time difference between them.

Syntax for fitting an AR(1) model

SAS code to fit an AR(1) model

```
proc mixed data=revenge method=reml;
class id tcat;
model revenge = sex age vc wom t / solution;
repeated tcat / subject=id type=ar(1) r=1 rcorr=1;
run;
```

Correlation and covariance matrix for subject 1 in AR(1)

Estimated R Matrix for id 1					
Row	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	0.3770	0.1855	0.09128	0.04492	0.02210
2	0.1855	0.3770	0.1855	0.09128	0.04492
3	0.09128	0.1855	0.3770	0.1855	0.09128
4	0.04492	0.09128	0.1855	0.3770	0.1855
5	0.02210	0.04492	0.09128	0.1855	0.3770

Estimated R Correlation Matrix for id 1					
Row	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	1.0000	0.4921	0.2421	0.1192	0.05863
2	0.4921	1.0000	0.4921	0.2421	0.1192
3	0.2421	0.4921	1.0000	0.4921	0.2421
4	0.1192	0.2421	0.4921	1.0000	0.4921
5	0.05863	0.1192	0.2421	0.4921	1.0000

- We can see that the correlation between two observations decreases the further apart they are in time.
- This is exactly what we want to model when choosing the AR(1) covariance structure.

Parameters in the AR(1) covariance/correlation structure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
AR(1)	id	0.4921
Residual		0.3770

- We can see that the estimate of the parameter ρ is $\widehat{\rho} = 0.492$.
- We can verify in the correlation matrix for subject 1 that the correlation between t_1 and t_2 is 0.492 times the correlation between t_1 and t_3; that is, it is $0.492^2 = 0.24$
- Note: in the compound symmetry model, the estimated correlation between two observations from the same person (regardless of the time between measures) was 0.356.

Mean parameter estimates

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-0.1686	0.3201	75	-0.53	0.6000
sex	0.1562	0.09791	75	1.60	0.1149
age	0.04562	0.006540	75	6.98	<.0001
vc	0.5209	0.02831	75	18.40	<.0001
wom	0.4002	0.03590	75	11.15	<.0001
t	-0.5686	0.02335	319	-24.35	<.0001

- The estimates of the β parameters are very similar to those from the previous model (with the compound symmetry structure), but not identical.
- The explanatory variables are all significant except for `sex`. The conclusions are the same as those from the previous model.