

MATH 60604A

Statistical modelling

§ 5d - Compound symmetry model

Léo Belzile

HEC Montréal
Department of Decision Sciences

Covariance structure of the compound symmetry model

- Assume that the observations within a group are interchangeable. That is, assume that the correlation (conditional on the explanatory variables) between two Y observations within a group is always the same, and that the conditional variance of Y is constant.
- In this case, if there are five observations within group i , the associated within-group covariance matrix is

$$\Sigma_i = \begin{pmatrix} \sigma^2 + \tau & \tau & \tau & \tau & \tau \\ \tau & \sigma^2 + \tau & \tau & \tau & \tau \\ \tau & \tau & \sigma^2 + \tau & \tau & \tau \\ \tau & \tau & \tau & \sigma^2 + \tau & \tau \\ \tau & \tau & \tau & \tau & \sigma^2 + \tau \end{pmatrix}.$$

- Note here is that the conditional covariance between two observations in the same group is τ , and that the conditional variance of each observation is $\sigma^2 + \tau$.

Correlation structure of the compound symmetry model

The corresponding correlation matrix for the compound symmetry covariance model is

$$\mathbf{R}_i = \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix},$$

where $\rho = \tau/(\sigma^2 + \tau)$.

- The conditional correlation between two observations within a group is always ρ .
- This covariance structure is called “**compound symmetry**” and has two parameters, σ^2 and τ .

mixed procedure to fit models to correlated data

SAS code

```
/* Copy t */  
data revenge;  
set statmod.revenge;  
tcat=t;  
run;  
  
proc mixed data=revenge method=reml;  
class id tcat;  
model revenge = sex age vc wom t / solution;  
repeated tcat / subject=id type=cs r=1 rcorr=1;  
run;
```

Declaring the dependence structure within `proc mixed`

The command `repeated` allows us to define the dependence structure.

- The first argument of the `repeated` function specifies what order the observations are within each group. This variable must be a categorical variable (created via `class`).
- The option `subject` specifies the variable which identifies the groups.
- The option `type` specifies the model for the within-group correlation.
- The option `r=1` (`rcorr=1`) adds the estimated covariance (correlation) matrix for individual `I` in the output.

We will also use the variable `t` as a continuous variable in the model, which is why we also created a copy of the variable `t` (`tcat` here), in order to use it as an argument for `repeated`.

Technical aside

- The first argument `tcats` in the `repeated` command is ignored here, as the compound symmetry covariance structure does not use the order of the observations within a group.
- However, the order must be specified for other types of structures. It's good to specify the “repeated” argument, even when it's not necessary.

Model specification

Model Information	
Data Set	WORK.TEMP
Dependent Variable	revenge
Covariance Structure	Compound Symmetry
Subject Effect	id
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Class Level Information		
Class	Levels	Values
id	80	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
tcat	5	1 2 3 4 5

Covariance and correlation matrices for individual 1

Estimated R Matrix for id 1					
Row	Col1	Col2	Col3	Col4	Col5
1	0.3858	0.1374	0.1374	0.1374	0.1374
2	0.1374	0.3858	0.1374	0.1374	0.1374
3	0.1374	0.1374	0.3858	0.1374	0.1374
4	0.1374	0.1374	0.1374	0.3858	0.1374
5	0.1374	0.1374	0.1374	0.1374	0.3858

Estimated R Correlation Matrix for id 1					
Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.3563	0.3563	0.3563	0.3563
2	0.3563	1.0000	0.3563	0.3563	0.3563
3	0.3563	0.3563	1.0000	0.3563	0.3563
4	0.3563	0.3563	0.3563	1.0000	0.3563
5	0.3563	0.3563	0.3563	0.3563	1.0000

Since we specified a compound symmetry structure for the covariance, the correlation is the same for all observations within subject 1.

Parameters for the covariance structure

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
CS	id	0.1374
Residual		0.2483

- The compound symmetry covariance structure is
 - $\text{Var}(Y_{ij}) = \sigma^2 + \tau$;
 - $\text{Cov}(Y_{ij}, Y_{ij'}) = \tau$.
- The estimate of the conditional covariance between observations for the same person is $\hat{\tau} = 0.137$.
- The estimated conditional variance of an observation is $\hat{\tau} + \hat{\sigma}^2 = 0.386$.

Correlation structure

- The estimate of the conditional correlation between two observations from the same person (within-person correlation) is

$$\hat{\rho} = \frac{\hat{\tau}}{\hat{\tau} + \hat{\sigma}^2} = \frac{0.137}{0.137 + 0.248} = 0.356.$$

- We can recover these values in the covariance/correlation matrices given for the first individual.
- **You need to know how to retrieve the correlation based on output (hence the formulae.)**

Likelihood ratio test for covariance parameter

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
1	67.25	<.0001

- We can test $\mathcal{H}_0 : \tau = 0$ against $\mathcal{H}_1 : \tau \neq 0$ using the **likelihood ratio test**.
- The above table gives the likelihood ratio test for $\mathcal{H}_0 : \tau = 0$, which corresponds to the covariance model of the classic regression model with covariance $\sigma^2 \mathbf{I}$ (reduced model), but adjusted using REML.
- We conclude that the reduced model without a correlation structure is **not an adequate simplification** of the more complex model with the compound symmetry correlation structure.
- **The likelihood ratio test reported by SAS always perform the comparison with the homoscedastic linear model without correlation.**

Likelihood ratio test, by hand

Fit Statistics		Fit Statistics	
-2 Res Log Likelihood	776.7	-2 Res Log Likelihood	709.4
AIC (Smaller is Better)	778.7	AIC (Smaller is Better)	713.4
AICC (Smaller is Better)	778.7	AICC (Smaller is Better)	713.4
BIC (Smaller is Better)	782.6	BIC (Smaller is Better)	718.2

- We could obtain the value of the test statistic manually by comparing the restricted maximum likelihood estimates of the two models, here $-2\ell_{\text{reml}}(\hat{\theta}_0) = 776.7$ and $-2\ell_{\text{reml}}(\hat{\theta}) = 709.4$, so the likelihood ratio test statistic is 67.3.
 - This is the value reported on the previous slide, modulo rounding.
- The null distribution of the likelihood ratio test is χ^2_1 (why?).
- We can compare the value of the test to the 95% quantile of the χ^2_1 , 3.84. Since the value of the statistic is larger than 3.84, we reject \mathcal{H}_0 at level $\alpha = 0.05$.

Mean parameter estimates

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-0.1689	0.3422	75	-0.49	0.6231
sex	0.1357	0.1060	75	1.28	0.2044
age	0.04586	0.007080	75	6.48	<.0001
vc	0.5225	0.03065	75	17.05	<.0001
wom	0.3989	0.03887	75	10.26	<.0001
t	-0.5675	0.01762	319	-32.21	<.0001

Desire for revenge seems to decrease in time, after accounting for the other variables.

Coefficient estimates

- The fitted model is always a linear regression model,

$$\widehat{\text{revenge}} = -0.169 + 0.136\text{sex} + 0.0459\text{age} + 0.523\text{vc} \\ + 0.399\text{wom} - 0.568\text{t}.$$

- It turns out that the estimates $\hat{\beta}$ are exactly the same as we saw in the ordinary linear regression model.
- This is a special case (compound symmetry correlation, and same number of observations in each group) and **will not always be true for other models.**
- However, these estimates will usually be close to those coming from ordinary linear regression.

Model comparison for coefficients

Effect	Estimate	Standard Error	Effect	Estimate	Standard Error
Intercept	-0.1689	0.2249	Intercept	-0.1689	0.3422
sex	0.1357	0.06748	sex	0.1357	0.1060
age	0.04586	0.004507	age	0.04586	0.007080
vc	0.5225	0.01951	vc	0.5225	0.03065
wom	0.3989	0.02474	wom	0.3989	0.03887
t	-0.5675	0.02177	t	-0.5675	0.01762

- The precision of our estimates $\hat{\beta}$ changes (left is independence, right is equicorrelation model).
- The standard errors are greater in the model with non-zero correlation. The conclusions did not change for any of the predictor variables, except for **sex**. It is no longer significant.
- In fact, the correlations make within-person observations redundant to an extent. We actually have less information than we would for independent observations, so parameter estimates are less precise.