Homework 2

Bohao Zou
BST 224
University of California, Davis

May 5, 2020

Question 1

(a) Compute variance $\text{Var}(\epsilon_{ij})$ and correlation $\text{Corr}(\epsilon_{ij}, \epsilon_{ik})$.

Solution

The variance $\text{Var}(\epsilon_{ij})$ is :

$$\begin{aligned}
\text{Var}(\epsilon_{ij}) &= \text{Var}(U_i + W_{ij}) && (1)\\
&= \text{Var}(U_i) + \text{Var}(W_{ij}) && (2)\\
&= \nu^2 + \delta^2 && (3)
\end{aligned}$$

The correlation $\text{Corr}(\epsilon_{ij}, \epsilon_{ik})$ is :

$$\begin{aligned}
\text{Corr}(\epsilon_{ij}, \epsilon_{ik}) &= \frac{\text{Cov}(U_i + W_{ij}, U_i + W_{ik})}{\text{Var}(\epsilon_{ij})} && (4)\\
&= \frac{\text{Var}(U_i) + \text{Cov}(W_{ij}, W_{ik})}{\text{Var}(\epsilon_{ij})} && (5)\\
&= \frac{\text{Var}(U_i) + \text{Corr}(W_{ij}, W_{ik})\dot{\text{V}}\text{ar}(W_{ij})}{\text{Var}(\epsilon_{ij})} && (6)\\
&= \frac{\nu^2 + \alpha^{|t_{ij}-t_{ik}|}\dot{\delta}^2}{\nu^2 + \delta^2} && (7)
\end{aligned}$$

(b)

Solution

$$\begin{aligned}
L &= -\frac{1}{2}\dot{\log}(\sigma^{2\dot{(}N-p)}) - \frac{1}{2\sigma^2}\vec{y}^T A^T \vec{y} && (8)\\
&= -(N-p)\log(\sigma) - \frac{1}{2}\sigma^{-2}\vec{y}^T A^T \vec{y} && (9)
\end{aligned}$$

$$\frac{dL}{d\sigma} = -\frac{1}{\sigma}(N-p) + \sigma^{-3}\vec{y}^T A^T \vec{y} = 0 \tag{10}$$

From the equation (10), the matrix A is symmetric and idempotent and $A^T\vec{y} = \vec{y} - X\hat{\vec{\beta}}$, we have :

$$\hat{\sigma^2} = \frac{\vec{y}^T A^T \vec{y}}{N-p} \tag{11}$$

$$= \frac{\vec{y}^T (AA)^T \vec{y}}{N-p} \tag{12}$$

$$= \frac{\vec{y}^T A^T A^T \vec{y}}{N-p} \tag{13}$$

$$= \frac{\vec{y}^T A A^T \vec{y}}{N-p} \tag{14}$$

$$= \frac{(\vec{y} - X\hat{\vec{\beta}})^T (\vec{y} - X\hat{\vec{\beta}})}{N-p} \tag{15}$$

$$\tag{16}$$

Question 2

(a) Obtain the residuals in a model for anxiety, ignoring longitudinal structure for now. We will consider a very general mean model, consisting of categorical effects of each time point interacted with each group (22 mean model parameters). Examine correlation matrix and auto-correlation function using the residuals obtained. Explain your exploratory results.

Solution

The correlation matrix of this data set is :

	T.4	T.6	T.8	T.10	T.11	T.14	T.16	T.17	T.18	T.19	T.23
T.4	1.00										
T.6	0.58	1.00									
T.8	0.38	0.79	1								
T.10	0.32	0.38	0.61	1.00							
T.11	0.45	0.35	0.38	0.61	1.00						
T.14	0.50	0.58	0.49	0.13	0.51	1.00					
T.16	0.21	0.40	0.36	0.20	0.22	0.44	1.00				
T.17	0.03	0.25	0.28	0.20	0.14	0.26	0.9	1.00			
T.18	-0.05	0.25	0.31	0.30	0.24	0.22	0.85	0.94	1.00		
T.19	0.28	0.49	0.39	0.27	0.40	0.49	0.73	0.70	0.67	1.00	
T.23	0.33	0.23	0.17	0.24	0.58	0.57	0.44	0.40	0.38	0.67	1.00

Table 1: *The correlation matrix of time sequence.*

This table shows that it has a strong correlation when the time is near each other and it has a weak correlation when the lag of time became huge. This phenomenon roughly agrees with our common sense that it will have a weak correlation with long time lag.

Besides using the correlation matrix to explore the correlation structure, we can also use the auto-correlation function to display the correlation structure of AUX variable. We can draw the correlogram with 95% Tolerance limits to give a directive displaying. The correlogram plot is in Figure 1.

When the lag of those time sequence is at 14, 15 and 17, the result of ACF function is lower than the 95% tolerance limits. This indicates that at those lag point, we should assume that the truth correlations which have those lag are all zero. Beside those lag points, we can have 95% percent confidence that those correlation which have those time lag are not zero.

The ACF declines strongly with lag, this indicates that we should use the VCC model which will contain *exponential component*. From this plot we can know that the ACF dose not appear to $\to 0$ as $lag \to \infty$. This situation tells us we should use *exchangeable component* in our VCC model. Finally, the plot shows that the ACF appears to $\to 1$ as $lag \to 0$.

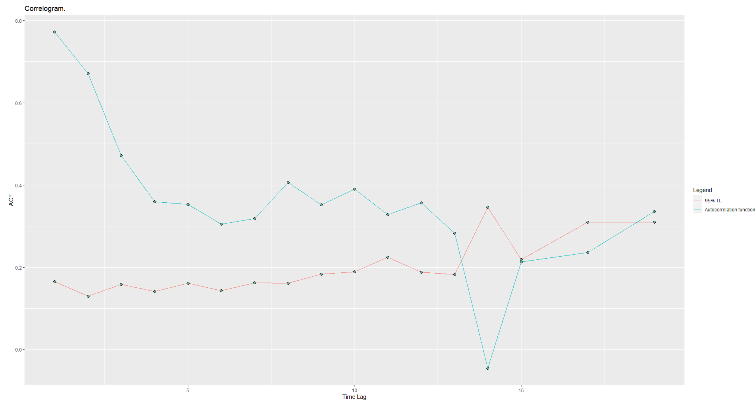

Figure 1: *The correlogram plot.*

This indicates that we should not use *measurement component* in our VCC model. We can get the conclucsion that the best VCC model is *exponential plus exchangeable VCC model*.

(b) Assume now we want to compare two candidate models: eg, exponential v.s. exchangeable plus exponential (or other two nested candidate VCC models selected based on your findings in (a)). Fit the two models with the flexible mean model in (a) using ReML, and report covariance parameter estimates. Do the LRT to compare the two models. Provide the test statistic, degrees of freedom and P-value. Draw conclusions.

Solution

For this question, I selected two different VCC models. The first model is *Exchangeable model*. The second model is *Exchangeable plus Exponential model*.

In the *Exchangeable model*, $\epsilon_{ij} = U_i + Z_{ij}$, Where the $U_i \sim \mathcal{N}(0, \nu^2)$, the $Z_{ij} \sim \mathcal{N}(0, \tau^2)$ and the $\text{Var}(\epsilon_{ij}) = \sigma^2 = \nu^2 + \tau^2$. The parameters which we want to estimate is ν and τ.

In the *Exchangeable plus Exponential model*, $\epsilon_{ij} = U_i + W_{ij}$ Where the $U_i \sim \mathcal{N}(0, \nu^2)$, the $W_{ij} \sim \mathcal{N}(0, \delta^2)$ and $W_{ij} = \alpha W_{ij-1} + E_{ij}$ where $E_{ij} \sim \mathcal{N}(0, \delta^2(1-\alpha^2))$. The parameters which we want to esitimate is α, ν and δ.

The estimated parameters of those two models show in Table 2.

For constructing the LRT test which $H_0 : \alpha = 0$ and $H_1 : \alpha > 0$, we know that $-2\{\log(L_{reduced}) - \log(L_{full})\} \sim \chi^2_{DF}$. From the table we can get the log likelihood of *exchangeable with fixed* ρ model is -823.73 and it is the reduced model.The log likelihood of *AR(1)* model is -775.9203 and it is the full model. So, the test statistic is $T_{statistic} = -2\{\log(L_{reduced}) - \log(L_{full})\} = -2\{-823.73 - (-775.9203)\} = 95.62$.

The degrees of freedom is 1 because there is only one parameter reduced. The P-value of this LRT test is $P - value = \frac{1}{2} \times Pr(\chi^2 > 95.62) = \frac{1}{2} \times 1.39 \times 10^{-22} = 6.95 \times 10^{-23}$.

	Exchangeable Model	Exchangeable plus Exponential Model
ν^2	0.478	0.972
τ^2	3.463	Null
δ^2	Null	2.93
α	Null	0.5574
log likelihood	-823.73	-775.92

Table 2: *The estimated parameters of Exchangeable model and Exchangeable plus Exponential model*

Because this value is too small. We can consider that this P-value is zero. This indicates that the parameter α is significant and this shows that the *Exchangeable plus Exponential model* model is better than *Exchangeable model.* model.