STAT 8320 Spring 2015 Assignment 4

Peng Shao 14221765

April 8, 2015

▶ **1. Solution.** (a).

$$\begin{aligned} f(\lambda|y_i) &= \frac{f(y_i|\lambda)f(\lambda)}{f(y_i} \\ &= \frac{\frac{\lambda^{y_i+a-1}}{\Gamma(a)b^a y_i!}e^{-\lambda(1+1/b)}}{f(y_i)} \\ &\propto \frac{\lambda^{y_i+a-1}}{\Gamma(a)b^a y_i!}e^{-\lambda(1+1/b)} \\ &\propto \lambda^{y_i+a-1}e^{-\lambda(1+1/b)} \end{aligned}$$

So $\lambda|y_i \sim \text{GAM}(y_i+a-1, \frac{1}{1+1/b})$, and

$$f(\lambda|y_i) = \frac{\lambda^{y_i+a-1}}{\Gamma(y_i+a)\left(\frac{b}{1+b}\right)^{y_i+a}}e^{-\lambda(1+1/b)}$$

Thus,

$$\begin{aligned} f(y_i) = \int_0^\infty f(y_i|\lambda)f(\lambda)d\lambda &= \frac{f(y_i|\lambda)f(\lambda)}{f(\lambda|y_i)} \\ &= \frac{\frac{\lambda^{y_i+a-1}}{\Gamma(a)b^a y_i!}e^{-\lambda(1+1/b)}}{\frac{\lambda^{y_i+a-1}}{\Gamma(y_i+a)\left(\frac{b}{1+b}\right)^{y_i+a}}e^{-\lambda(1+1/b)}} \\ &= \frac{\Gamma(y_i+a)}{\Gamma(a)y_i!}\left(\frac{1}{1+b}\right)^a\left(\frac{b}{1+b}\right)^{y_i} \\ &= \binom{a+y_i-1}{a-1}\left(\frac{b}{1+b}\right)^{y_i} \end{aligned}$$

We can conclude that $y_i \sim \text{NB}(\frac{1}{1+b}, a)$.

(b) From the theories in generalized linear model, we have already known that negative binomial distribution usually is used to fixed the over-dispersion problem of count data when Poisson distribution assumption or independence assumption are no longer valid. And we also know that in most time the over-dispersion may be caused by the dependence of data, like some repeated measurements in student attendance example. The GLMM essentially takes covariates between dependent data into model, so it also can model the over-dispersed count data. Or in other words, the derivation in part (a) just shows that negative binomial distribution can work well with over-dispersed count data.

▶ **2.** **Solution.** (a). We have the form of model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\epsilon}$$

where $\boldsymbol{Y} = (y_1, y_2)'$, $\boldsymbol{X} = (X_1, X_2)'$, $\boldsymbol{\beta} = \beta$, $\boldsymbol{b} = (b_1, b_2)'$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2)'$, and

$$\boldsymbol{Z} = \begin{pmatrix} 1 & \\ & 2 \end{pmatrix}$$
$$\boldsymbol{D} = \begin{pmatrix} \tau^2 & \frac{\phi\tau^2}{1+\phi^2} \\ \frac{\phi\tau^2}{1+\phi^2} & \tau^2 \end{pmatrix}$$
$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \\ & \sigma^2 \end{pmatrix}$$

(b). The marginal variance/covariance matrix of $\boldsymbol{Y}$ is that

$$Var(\boldsymbol{Y}) = \boldsymbol{ZDZ}^T + \boldsymbol{\Sigma} = \begin{pmatrix} \tau^2 + \sigma^2 & \frac{2\phi\tau^2}{1+\phi^2} \\ \frac{2\phi\tau^2}{1+\phi^2} & 4\tau^2 + \sigma^2 \end{pmatrix}$$

Then the marginal variance of Y_2 is $4\tau^2 + \sigma^2$ and the marginal covariance between Y_1 and Y_2 is

$$cov(Y_1,\ Y_2) = \frac{2\phi\tau^2}{1+\phi^2}$$

(c). Nothing. Because in the restricted likelihood function there is no parameters other than those from variance and covariance matrix, we can only test the variances and covariances, but no the parameters of fixed and random effects based on REML.

▶ **3. Solution.** (a). If the intercepts of eight plots are exactly at same point, but the increments are more complicated and a little fluctuant, not just a simple quadratic curve. Then we should consider adding a random component into the coefficient for time.

(b). We have the form of model

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$

where

$$\begin{aligned}
\boldsymbol{Y}_i &= (y_{i1}, \dots, y_{in_i})', \\
\boldsymbol{X}_i &= \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 \end{pmatrix}, \\
\boldsymbol{Z} &= (t_{i1}, \dots, t_{in_i})', \\
\boldsymbol{\beta} &= (\beta_0, \beta_1, \beta_2)', \\
\boldsymbol{b}_i &= b_{1i}, \\
\boldsymbol{e} &= (e_{11}, \dots, e_{in_i})', \\
var(\boldsymbol{e}_i) = \boldsymbol{\Sigma} &= \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix} = \sigma^2 \boldsymbol{I}_{n_i \times n_i}, \\
var(\boldsymbol{b}_i) = \boldsymbol{D} &= \begin{pmatrix} \sigma_b^2 & & \\ & \ddots & \\ & & \sigma_b^2 \end{pmatrix} = \sigma_b^2 \boldsymbol{I}_{n_i \times n_i},
\end{aligned}$$

(c). The marginal variance/covariance matrix of $\boldsymbol{Y}$ is that

$$\begin{aligned}
Var(\boldsymbol{Y}_i) &= \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{\Sigma} \\
&= \begin{pmatrix} t_{i1} \\ \vdots \\ t_{in_i} \end{pmatrix} \begin{pmatrix} \sigma_b^2 & & \\ & \ddots & \\ & & \sigma_b^2 \end{pmatrix} (t_{i1}, \dots, t_{in_i}) + \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix} \\
&= \begin{pmatrix} t_{i1}^2\sigma_b^2 + \sigma^2 & t_{i1}t_{i2}\sigma_b^2 & \cdots & t_{i1}t_{in_i}\sigma_b^2 \\ t_{i2}t_{i1}\sigma_b^2 & t_{i2}^2\sigma_b^2 + \sigma^2 & \cdots & t_{i2}t_{in_i}\sigma_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ t_{in_i}t_{i1}\sigma_b^2 & t_{in_i}t_{i2}\sigma_b^2 & \cdots & t_{in_i}^2\sigma_b^2 + \sigma^2 \end{pmatrix}_{n_i \times n_i}
\end{aligned}$$

(d). Because the marginal covariance $\boldsymbol{Y}$ is

$$cov(Y_{ij}, Y_{ik}) = t_{ij}t_{ik}\sigma_b^2$$

then the correlation of Y_i is that

$$\begin{aligned} corr(Y_{ij}, Y_{ik}) &= \frac{cov(Y_{ij}, Y_{ik})}{\sqrt{var(Y_{ij})}\sqrt{var(Y_{ik})}} \\ &= \frac{t_{ij}t_{ik}\sigma_b^2}{\sqrt{t_{ij}^2\sigma_b^2 + \sigma^2}\sqrt{t_{ik}^2\sigma_b^2 + \sigma^2}} \\ &= \frac{jk}{\sqrt{j^2+1}\sqrt{k^2+1}} \\ &= \frac{1}{\sqrt{1/j^2+1}\sqrt{1/k^2+1}} \end{aligned}$$

The correlations will increase with the increase of time, j and k, but no trend just with temporal separation. This is not so realistic. In common sense, we usually may think that the correlations may be smaller with large temporal separation than the correlations with small small temporal separation, because status of one time point is more likely to affect or to be affected by the status of the near time point. The reason causing this unrealistic result may be we simply assume the conditional independence while the data may not have this property.

(e). There are two advantages of the marginal covariance derived hierarchically. First, compared to the unstructured covariance structure, the hierarchical marginal covariance have less unknown parameters to estimate, so it can reduce the computation, and avoid suffering overfitting problem. Secondly, it easily to understand and interpret the variance components, we can know that which parts of variation come from random effect and which parts come from the violation of conditional independence.

▶ **4. Solution.** (a) We have the split-plot design model

$$Y_{ijk} = \mu + \rho_i + \alpha_j + e_{ij} + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$$

where ρ is plot effect, α is pasture effect and β is mineral effect. The random effects are ρ, e and ϵ. From the ANOVA table(Figure 2, Figure 1), we have that

$$\sigma^2_{\text{plot}} = 12.74, \quad \sigma^2_e = 1.05, \quad \sigma^2_\epsilon = 2.25$$

In addition, the test(Figure 3) for the significance for interaction of the pasture and mineral effects yields a P-value of 0.4981, the Factor pasture effect yields a P-value of 0.0377, and the Factor mineral effect yields a P-value of 0.0932. So only pasture effect are significant at $\alpha = 0.05$.

Figure 1: Anova Table

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square
past	3	71.166667	23.722222
min	1	8.166667	8.166667
past*min	3	5.833333	1.944444
plot	2	212.583333	106.291667
plot*past	6	26.083333	4.347222
Residual	8	18.000000	2.250000

Type 1 Analysis of Variance

Source	Expected Mean Square	Error Term	Error DF
past	Var(Residual) + 2 Var(plot*past) + Q(past,past*min)	MS(plot*past)	6
min	Var(Residual) + Q(min,past*min)	MS(Residual)	8
past*min	Var(Residual) + Q(past*min)	MS(Residual)	8
plot	Var(Residual) + 2 Var(plot*past) + 8 Var(plot)	MS(plot*past)	6
plot*past	Var(Residual) + 2 Var(plot*past)	MS(Residual)	8
Residual	Var(Residual)	.	.

Type 1 Analysis of Variance

Source	F Value	Pr > F
past	5.46	0.0377
min	3.63	0.0932
past*min	0.86	0.4981
plot	24.45	0.0013
plot*past	1.93	0.1909

Figure 2: Covariance Estimates

Covariance Parameter Estimates

Cov Parm	Estimate
plot	12.7431
plot*past	1.0486
Residual	2.2500

Figure 3: Type 3 Tests

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
past	3	6	5.46	0.0377
min	1	8	3.63	0.0932
past*min	3	8	0.86	0.4981

(b). Because the data may come from different distribution, so the degree of freedom of the variance of random components may need some modification, like Satterthwaite method. The Kenwardroger method give us a more conservative distribution about t-test or F-test than Satterthwaite when sample size is not large enough, making the assumption seem more appropriate. For example, we assume that $E(\rho)$=0. When we test whether every predict of ρ_i equals zero. For unmodified degree of freedom, the degree of freedom of t-statistic for ρ_1 is 8 and the P-value is 0.0965, but for Kenwardroger method, the degree of freedom is 2.14 and the P-value is 0.1934. The results of not rejecting null hypothesis are same, but which of Kenwardroger method makes us a little harder to reject our assumption.

The method only modified the degree of freedom, so it will leave the result of fixed result unchanged. In this specific problem, the results of random components are also unchanged.

(c). From the output "Differences of Least Squares Means" (Figure 4), we can see that the only significance difference is the difference between pasture 1 and pasture 4 based on the Tukey-Kramer adjustment.

Figure 4: Least Squares Means

Differences of Least Squares Means

Effect	past	_past	Estimate	Standard Error	DF	t Value	Pr > |t|
past	1	2	-1.0000	1.2038	6	-0.83	0.4379
past	1	3	-0.3333	1.2038	6	-0.28	0.7911
past	1	4	-4.3333	1.2038	6	-3.60	0.0114
past	2	3	0.6667	1.2038	6	0.55	0.5997
past	2	4	-3.3333	1.2038	6	-2.77	0.0325
past	3	4	-4.0000	1.2038	6	-3.32	0.0159

Differences of Least Squares Means

Effect	past	_past	Adjustment	Adj P
past	1	2	Tukey-Kramer	0.8385
past	1	3	Tukey-Kramer	0.9918
past	1	4	Tukey-Kramer	0.0427
past	2	3	Tukey-Kramer	0.9421
past	2	4	Tukey-Kramer	0.1135
past	3	4	Tukey-Kramer	0.0587

▶ **5. Solution.** (a). The hypotheses are

$$H_0 : \mu_{11} = \mu_{12} = \mu_{13} \qquad \text{v.s.} \qquad H_a : \text{ at least two means are not equal}$$

or we can write null hypothesis as

$$H_0 : \boldsymbol{C}_1\boldsymbol{\mu}_1 = \boldsymbol{0}$$

where

$$\boldsymbol{C}_1 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

The statistics are

Appendices

A SAS

B Output