STAT 8320 Spring 2015 Assignment 5

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,

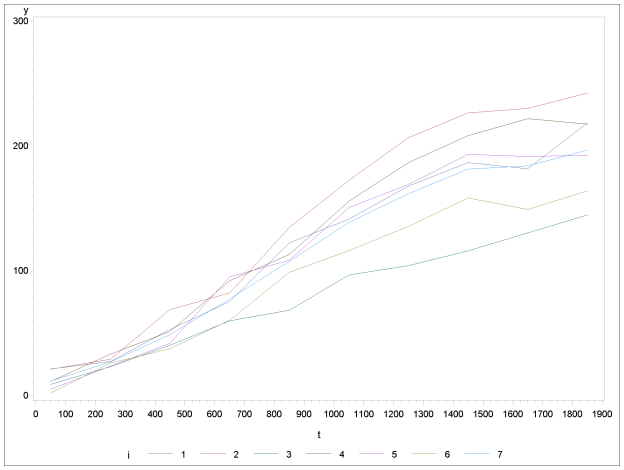
$$\begin{aligned} f(y_i) &= \int_0^\infty f(y_i|\lambda)f(\lambda)d\lambda = \frac{f(y_i|\lambda)f(\lambda)}{f(\lambda|y_i)} \\ &= \frac{\frac{\lambda^{y_i+a-1}}{\Gamma(a)b^a y_i!}e^{-\lambda(1+1/b)}}{\frac{\lambda^{y_i+a-1}}{\Gamma(y_i+a)\left(\frac{b}{1+b}\right)^{y_i+a}}e^{-\lambda(1+1/b)}} \\ &=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

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to fixed the over-dispersion problem of count data when Poisson distribution assumption or independence assumption are no longer valid.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). Before we fit this nonlinear mixed model, we should firstly plot the profile of the data, trying to acquire some intuitive result from the plot. As the Figure 1 shows, different plants are label as 1 to 6, and number 7 represents the average profile. We can approximately know that the max value is between 150 and 250, and the inflection point should be between 500 and 1000. Without loss of generality, we set the initial value of (β_1, β_2) as (200, 850). Then we can solve the β_3 based on the data, and it is about 350. Next, by using PROC MEANS in SAS, we can get the standard deviations at different time points. Because $1 + e^{-(t_{ij}-\beta_2)/\beta_3}$ is relatively large, it is reasonable to assume that the variance of Y is mostly from σ^2. So we set the initial value of σ^2 as 40. As t_{ij} grows, the denominator becomes smaller and smaller, then the proportion of σ_u^2 in variance of Y becomes larger, and it seems that σ_u^2 should be between 400 to 1600, so we set the initial value of σ_u^2 as 900.

Figure 1: Profile of Plant Growth



Then we use the PROC NLIN and PROC NLMIXED to fit the nonlinear model and the nonlinear mixed model respectively. The results about parameters are listed as Figure 2 and Figure 3.

Figure 2: Parameters of Nonlinear Models

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits		Skewness
beta1	199.7	10.3827	178.9	220.4	0.5330
beta2	797.8	55.1103	687.4	908.1	0.3725
beta3	300.7	42.8631	214.8	386.5	0.4782

Figure 3: Parameters of Nonlinear Mixed Models

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
beta1	199.41	15.2372	5	13.09	<.0001	0.05
beta2	797.42	14.6250	5	54.52	<.0001	0.05
beta3	298.48	11.4146	5	26.15	<.0001	0.05
resvar	49.8315	9.5902	5	5.20	0.0035	0.05
varu	1346.95	784.96	5	1.72	0.1468	0.05

Parameter Estimates

Parameter	Lower	Upper	Gradient
beta1	160.24	238.58	4.83E-7
beta2	759.82	835.01	-3.07E-6
beta3	269.14	327.82	2.884E-6
resvar	25.1791	74.4838	2.602E-6
varu	-670.87	3364.76	-1.06E-8

From these output, we can answer the questions like,

1. To test
$$H_0 : \beta_3 = 350 \qquad \text{v.s.} \qquad H_A : \beta_3 \neq 350$$
we can easily reject the null hypothesis because the Wald-type confidence intervals of both models do not contain 350, which is equivalent to a Wald test. We also can use the ESTIMATE statement in PROC NLMIXED to estimate $\beta_3 - 350$, as Figure 4

Figure 4: Additional Estimates

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Beta_3=350?	-51.5207	11.4146	5	-4.51	0.0063	0.05

Additional Estimates

Label	Lower	Upper
Beta_3=350?	-80.8627	-22.1786

It is obviously that we should reject the null hypothesis, which is the same result as above. So β_3 does not equal 350.

2. To test whether the random effect is necessary. Because the method of parameter estimate of mixed model is not based on likelihood, we cannot use likelihood ratio test. So we still use the Wald-type confidence interval. Because the interval contains zero, we cannot reject the null hypothesis, that is, the random effect is not significant. Furthermore, we can see that the estimate of parameters of fixed effect does not change too much, so this also indicates that it is not necessary to introduce random effect into model.

   But, we should be cautious about the result, because from the parameter estimates of parameter of nonlinear model, skewness of all parameters is much greater than 0.25, which means all parameters have vary apparent skewness. This makes the inferences unreliable.

▶ **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

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

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

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

(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)

$$\boldsymbol{Y} = \begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} -3 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} \right)$$

(b)

$$\begin{aligned} \boldsymbol{Y}|X_2 &\sim \mathrm{N}\left(\begin{pmatrix} -3 \\ 2 \end{pmatrix} + \begin{pmatrix} \frac{x_2}{2} - \frac{1}{2} \\ \frac{x_2}{2} - \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 3.5 & -0.5 \\ -0.5 & 2.5 \end{pmatrix} \right) \\ &= \mathrm{N}\left(\begin{pmatrix} \frac{x_2}{2} - \frac{7}{2} \\ \frac{x_2}{2} + \frac{3}{2} \end{pmatrix}, \begin{pmatrix} 3.5 & -0.5 \\ -0.5 & 2.5 \end{pmatrix} \right) \end{aligned}$$

(c)

$$X_2|\boldsymbol{Y} \sim \mathrm{N}\left(2 + \begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{4} & 0\\ 0 & \frac{1}{3}\end{pmatrix}\begin{pmatrix}x_1+3\\ x_3-2\end{pmatrix}, 2 - \begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{4} & 0\\ 0 & \frac{1}{3}\end{pmatrix}\begin{pmatrix}1\\ 1\end{pmatrix}\right)$$
$$= \mathrm{N}\left(\frac{1}{4}x_1 + \frac{1}{3}x_3 + \frac{13}{12}, \frac{17}{12}\right).$$

(d)

$$Z \sim \mathrm{N}(-3 + 3\cdot 1, 4 + 3^3\cdot 2 + 2\cdot 3\cdot 1) = \mathrm{N}(0, 28)$$

▶ **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

$$T^2 = n_1(\boldsymbol{C}\bar{\boldsymbol{y}})'(\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}')^{-1}(\boldsymbol{C}\bar{\boldsymbol{y}}) = 111.4286$$
$$F = \frac{n_1 - c}{(n_1-1)c}T^2 = 53.3929 \sim f_{c,n_1-c}$$

where $c = 2$ and $n_1 = 25$. The critical value of F statistic is $f_{0.95,2,23} = 3.422$, so $F > f_{0.95,2,23}$ and P-value is $2.28\mathrm{E}^{-09}$. We will reject the null hypothesis, that is, the mean concentrations are significantly different at three time points.

(b).

i The common covariance is

$$\boldsymbol{S}_{pool} = \frac{(n_1-1)\boldsymbol{S} + (n_2-1)\boldsymbol{W}}{(n1-1)+(n_2-1)} = \begin{pmatrix}28.4 & 10.8 & 12.4\\ 10.8 & 15.8 & 5.6\\ 12.4 & 5.6 & 39.4\end{pmatrix}$$

where $n_2 = 17$. The degree of freedom is 25+17-2=40.

ii The hypotheses are

$$H_0 : \boldsymbol{\Sigma}_A = \boldsymbol{\Sigma}_B \qquad \text{v.s.} \qquad H_A : \boldsymbol{\Sigma}_A \neq \boldsymbol{\Sigma}_B$$

The statistic is

$$\begin{aligned} M &= (n_1 + n_2 - 2)\log|\boldsymbol{S}_{pool}| - (n_1 - 1)\log|\boldsymbol{S}| - (n_2 - 1)\log|\boldsymbol{W}| = 1.1948 \\ C^{-1} &= 1 - \frac{2\times 3^2 + 3 \times 3 - 1}{6 \times (3+1) \times (2-1)} \left\{ \frac{1}{n_1 - 1} + \frac{1}{n_2 - 1} - \frac{1}{n_1 + n_2 - 2} \right\} = 0.9142 \\ MC^{-1} &= 1.1948 \times 0.9142 = 1.0923 \sim \chi^2_6 \end{aligned}$$

Because $MC^{-1} = 0.9142 < \chi^2_{0.95,6} = 12.592$ and P-value is 0.9819, we cannot reject the null hypothesis, which means that the assumption of same population covariance are reliable.

iii The squared Mahalanobis distance between $\boldsymbol{y} - \boldsymbol{z}$ is

$$T^2 = (\boldsymbol{y} - \boldsymbol{z})' \left[\boldsymbol{S}_{pool} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{-1} (\boldsymbol{y} - \boldsymbol{z}) = 18.03155$$

iv The hypotheses are

$$H_0 : \boldsymbol{y} - \boldsymbol{z} = \boldsymbol{0} \qquad \text{v.s.} \qquad H_A : \boldsymbol{y} - \boldsymbol{z} = \boldsymbol{0}$$

The statistic can be computed from the Mahalanobis distance from part (iii)

$$F = \frac{n_1 + n_2 - 3 - 1}{(n_1 + n_2 - 2) \times 3} T^2 = 5.70999 \sim f_{3,38}$$

Because $F = 5.70999 > f_{0.95,3,38} = 2.851$ with P-value=0.0025. We will reject the null hypothesis, so drug A and drug B do not have equal means.

v To test the parallel profiles, the hypotheses are

$H_0 : \mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} = \mu_{13} - \mu_{23}$ v.s. $H_A :$ at least two difference not equal

or we can rewrite the null hypothesis as

$$H_0 : \boldsymbol{C_2}(\boldsymbol{y} - \boldsymbol{z}) = 0$$

where $\boldsymbol{C_2} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$.

The statistics are

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\boldsymbol{C}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{z}}))'(\boldsymbol{CSC}')^{-1}(\boldsymbol{C}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{z}})) = 14.2658$$
$$F = \frac{n_1 + n_2 - c - 1}{(n_1 + n_2 - 2)c} T^2 = 6.7849 \sim f_{c,n_1+n_2-c-1}$$

Because $F = 6.7849 > f_{0.95,2,39} = 3.238$ with P-value=0.0030. We will reject the null hypothesis, so there is a significant interaction between drug and time.

Appendices

A SAS

B Output