Solutions to Methods in Spring 2006

1. ADL0 describes the ADL level before the trial. ADL describes the ADL level during the trail. Since the repeated measurement took almost in equal time period, we can treat $ADL - ADL0$ as the difference of decline. In the model, ADL0 must be an offset term. We should add random effects in the model on both the intercept term and the slope of the time. The rate of decline is the slope of time. The interaction effect between the time and dual use will tell us whether dual use will enhance the rate or not.

   (a) Let d_u be how many days taking urinary incontinence medicine (if any) and $d_u = 0$ if not dual used. Let d_s be how many days since the first evaluation. We can fit the model as

   $$\begin{aligned} ADL =& offset(ADL0) + \beta_0 + Age\beta_1 + Sex\beta_2 + \beta_3 d_u + \beta_4 d_s + \beta_5(ds : du) \\ &+ \gamma_0 + \gamma_1 d_u + \gamma_2 d_s + \epsilon. \end{aligned}$$

   where $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ are fixed effects and $\gamma_0, \gamma_1, \gamma_2$ are random effect. We assume $\gamma_j \sin N(0, \sigma_j^2)$ and $\epsilon \sim N(0, \sigma^2)$ independently. The model can describe the rate changes with respect to Age, Sex, days of dual used, and days of evaluations. The random effect part tells us the individual difference for staring rate, dual used days and days of evaluations.

   (b) The model assumptions are list in part (a). When data is given, we can get the predicted values of the response and the random effect part. Look at plots of the predicted value of the random effect part can tell us the assessment of the model assumptions.

2. (a) Since the residual deviance is 29.7723 based on 36 degrees of freedom, the model fits the data.

   (b) The odds of the occurrence of transient vasoconstriction will be about $e^{3.882} = 48.52$ times when volume increases 1 units if the rate keeps the same.

   (c) The overdispersion problem is not serious here and so it is not worth considering allowing the dispersion parameter to vary.

   (d) We need to solve a value for volume so that $\pi = 0.5$ which means $logit(\pi) = 0$. It can be solved from

   $$-9.529259 + 3.882007V + 2.649036R = 0.$$

   Plug the mean rate 1.69 in. We have

   $$-9.5293 + 3.882V + 2.649(1.69) = 0 \Rightarrow V = 1.3015.$$

(e) The new factor variables has 2 degrees of freedom. The difference of residual deviance is $29.77 - 26.16 = 3.16$. The p-value is about 0.2 indicating that the new factor variable is not significant. Thus, we can exclude from the model and conclude the first model is preferable.

3. First, we need to compute the MLE of p. The total number of observations is 1000. Thus, we can write the log-likelihood based on multinomial distribution as

$$\begin{aligned}\ell(p) =& \log \frac{1000!}{442!514!38!6!} + 442\log(\frac{p}{2}) + 514\log(\frac{p^2}{2} + pq) + 38\log(\frac{q}{2}) + 6\log(\frac{q^2}{2}) \\ =& C + 442\log(p) + 514\log(2p - p^2) + 50\log(1-p).\end{aligned}$$

Then, we have

$$\ell'(p) = \frac{442}{p} + \frac{514(2-2p)}{2p-p^2} - \frac{50}{1-p}.$$

Let $\ell'(p) = 0$. We have

$$442(1-p)(2-p) + 1028(1-p)^2 - 50p(2-p) = 0 \Rightarrow 1520p^2 - 3482p + 1912 = 0.$$

Then, we have $\hat{p} = 0.9129$ or $\hat{p} = 1.3778$. The second one does not make sense and we delete it. Thus, the MLE is $\hat{p} = 0.9129$. Based on the MLE, we have the following predicted $\hat{n}_{11} = 456.48$, $\hat{n}_{12} = 496.21$, $\hat{n}_{21} = 43.53$ and $\hat{n}_{22} = 3.79$.

(a) We may use the likelihood ratio test statistic and Pearson χ^2 test statistic. The underline assumption for both of them are that the probability table given are correct. There is only a unknown p to be estimated.

(b) The G^2 statistic is

$$G^2 = 2\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}\log\frac{n_{ij}}{\hat{n}_{ij}} = 5.84.$$

and the Pearson statistic is

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = 3.08.$$

Based on 2 degrees of freedom, both are not significant ($\chi^2_{2,0.05} = 5.99$)

4. (a) The model added a random effect on subject. Suppose the effects of subject follows $N(0, \sigma_s^2)$. The output says that $\hat{\sigma}_s = 0.22128$ and $\hat{\sigma} = 0.9352$. For any two subjects, the difference is caused by the error and the random effect. The variance is

$$\hat{\sigma}_s^2 + \hat{\sigma}^2 = 0.9610^2 = 0.9236.$$

This describes the likely difference between the performance of two subjects.

(b) The interaction effect is included in the model and we must consider it. Simple consider the effects, we have the values as 0 for Female and Box, 0.26834 for Male and Box, 0.62026 for Female and Circle and $0.26834+0.62026-0.69902 = 0.18958$ for Male and Circle. Thus, the Female and Circle Combination has the longest contact time.

(c) Each subject has 4 observations and so the total number of observations are $4(108) = 432$. Since Sex and Path are exactly the same for each subject, there are totally only 108 degrees of freedom to estimate their combinations. Note replicate are not the same for each subject, there are 5 parameters left and so the degrees of freedom is $108 - 6 + 1 = 103$.

(d) In the AR(1) structure for subject, it assumes the random effect for the same subject follows $\alpha_i = \phi\alpha_{i-1} + \epsilon_i$, where ϵ_i is iid, where $0 \le \phi < 1$. It gives the correlation $\phi^{|j-k|}$ between (j,k)-th element. We need to compare the difference of the loglikelihood between the two models. It is $-603.59 - (-604.38) = 0.79$. Based on χ_1^2 distribution, it is very small and so we conclude that the correlation effect is not significant.

(e) It is not surprising since in AR(1) model. The correlations between observations are positive. Even when total variation in random effect part is not changed, putting a positive correlation between observation will reduce the estimate of the marginal variances.

5. We consider the MLE estimation in this problem. Suppose $k = 1, \cdots, K_{ij}$, $j = 1, \cdots, J_i$ and $i = 1 \cdots, I$. Then in Scheme 1, $K_{ij} = 2$, $J_i = 2$ and $I = 2$. In Scheme, $k_{11} = 2$, $k_{12} = 1$, $k_{34} = 2$, $k_{35} = 1$ and $k_{46} = 1$. Let $Y_{ij\cdot} = \sum_{k=1}^{K_{ij}} / K_{ij}$, $Y_{i\cdot\cdot} = \sum_{i,k} Y_{ij} / \sum_j K_{ij}$ and $Y_{\cdots} = \sum_{i,jk,} Y_{ijk} / \sum_{i,j} K_{ij}$.

(a) Note that the model is nested. We consider the ANOVA estimation. Note that $Y_{ij1} - Y_{ij2} \sim N(0, 2\sigma_3^2)$ independently. We have $E[(Y_{ij1} - Y_{ij2})^2] = 2\sigma^2$ and so

$$\hat{\sigma}_3^2 = \frac{1}{8}\sum_{i=1}^{2}\sum_{j=1}^{2}(Y_{ij1} - Y_{ij2})^2.$$

Note that $Y_{i1\cdot} - Y_{i2\cdot} \sim N(0, 2\sigma_2^2 + \sigma_3^2)$. Thus, $E[(Y_{i1\cdot} - Y_{i2\cdot})^2] = 2\sigma_2^2 + \sigma_3^2$ and so

$$\hat{\sigma}_2^2 = \frac{1}{4}\sum_{i=1}^{2}(Y_{i1\cdot} - Y_{i2\cdot})^2 - \frac{1}{2}\hat{\sigma}_3^3.$$

Note that $Y_{1\cdot\cdot} - Y_{2\cdot\cdot} \sim N(0, 2\sigma_1^2 + \sigma_2^2 + \frac{1}{2}\sigma_3^3)$. Thus, we have

$$\hat{\sigma}_1^2 = \frac{1}{2}(Y_{1\cdot\cdot} - Y_{2\cdot\cdot})^2 - \frac{1}{2}\hat{\sigma}_2^2 - \frac{1}{4}\hat{\sigma}_3^2.$$

(b) This is unbalanced case. We denote $Y_1, \cdots, Y_8$ as the 8 observations. There are a lot unbiased estimators. We have $Y_1 - Y_2 \sim N(0, 2\sigma_3^2)$, $\frac{1}{2}(Y_1 + Y_2) - Y_3 \sim N(0, 2\sigma_2^2 + \frac{2}{3}\sigma_3^2)$, $Y_3 - Y_4 \sim N(0, 2\sigma_1^2 + 2\sigma_2^2 + 2\sigma_3^2)$, $Y_5 - Y_6 \sim N(0, 2\sigma_3^3)$, $\frac{1}{2}(Y_5 + Y_6) - Y_7 \sim N(0, 2\sigma_2^2 + \frac{3}{2}\sigma_3^2)$ and $Y_7 - Y_8 \sim N(0, 2\sigma_1^2 + 2\sigma_2^2 + 2\sigma_3^3)$. Therefore, we have

$$\hat{\sigma}_3^2 = \frac{1}{4}[(Y_1 - Y_2)^2 + (Y_5 - Y_6)^2],$$

$$\hat{\sigma}_2^2 = \frac{1}{4}\{[\frac{1}{2}(Y_1 + Y_2) - Y_3]^2 + [\frac{1}{2}(Y_5 + Y_6) - Y_7]^2\} - \frac{3}{4}\hat{\sigma}_3^2.$$

For σ_1^2, we have the following independent sample as

$$\frac{1}{4}(Y_1 + Y_2) + \frac{Y_3}{2} \sim N(\mu, \sigma_1^2 + \frac{\sigma_2^2}{2} + \frac{3\sigma_3^2}{8})$$

$$Y_4 \sim N(\mu, \sigma_1^2 + \sigma_2^2 + \sigma_3^3)$$

$$\frac{1}{4}(Y_5 + Y_6) + \frac{Y_7}{2} \sim N(\mu, \sigma_1^2 + \frac{\sigma_2^2}{2} + \frac{3\sigma_3^2}{8})$$

and

$$Y_8 \sim N(\mu, \sigma_1^2 + \sigma_2^2 + \sigma_3^3).$$

Thus, we have

$$\frac{1}{4}(Y_1 + Y_2) + \frac{Y_3}{2} - Y_4 \sim N(0, 2\sigma_1^2 + \sigma_2^2 + \frac{3\sigma_3^2}{4})$$

and

$$\frac{1}{4}(Y_5 + Y_6) + \frac{Y_7}{2} - Y_8 \sim N(0, 2\sigma_1^2 + \sigma_2^2 + \frac{3\sigma_3^2}{4}).$$

$$\hat{\sigma}_1^2 = \frac{1}{4}[\frac{1}{4}(Y_1 + Y_2) + \frac{Y_3}{2} - Y_4]^2 + [\frac{1}{4}(Y_5 + Y_6) + \frac{Y_7}{2} - Y_8]^2 - \frac{1}{2}\hat{\sigma}_2^2 - \frac{3}{8}\hat{\sigma}_3^2.$$

(c) It is clear that only $\hat{\sigma}_3^2$ is χ^2-distributed since the other 2 could be negative with positive probability. In Scheme 1, it has 4 degrees of freedom and in Scheme 2 it has 2 degrees of freedom.

(d) Both $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ could be negative.