Solutions to Methods in Winter 2003

1. (a) The log-likelihood function is

$$\ell = \log \prod_{i,j,k} \frac{\lambda_{ijk}^{n_{ijk}}}{n_{ijk}!} e^{-\lambda_{ijk}} = \sum_{i,j,k} [n_{ijk} \log \lambda_{ijk} - \lambda_{ijk} - \log n_{ijk}!].$$

There is no constraint for the satured model. Thus, we have

$$\frac{\partial \ell}{\partial \lambda_{111}} = \frac{n_{111}}{\lambda_{111}} - 1.$$

Thus, we have $\hat{\lambda}_{111} = n_{111}$.

(b) If i is independent of (j, k). Then,

$$\frac{\lambda_{ijk}}{\lambda_{+++}} = \frac{\lambda_{i++}}{\lambda_{+++}} \frac{\lambda_{+jk}}{\lambda_{+++}} \Rightarrow \lambda_{ijk} = \frac{\lambda_{i++}\lambda_{+jk}}{\lambda_{+++}},$$

where $\lambda_{i++} = \sum_{j,k} \lambda_{ijk}$, $\lambda_{+jk} = \sum_i \lambda_{ijk}$, and $\lambda_{+++} = \sum_{i,j,k} \lambda_{ijk}$. Therefore, we have

$$\log \lambda_{ijk} = \log \lambda_{i++} + \log \lambda_{+jk} - \log \lambda_{+++}.$$

Therefore, the structure of the model is

$$\log \lambda_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk}.$$

Based on the zero-sum constraint, we have

$$\begin{aligned}
(\beta\gamma)_{jk} =& \log \lambda_{+jk} - \frac{1}{J} \sum_j \log \lambda_{+jk} - \frac{1}{K} \sum_k \log \lambda_{+jk} + \frac{1}{JK} \sum_{j,k} \log \lambda_{+jk};\\
\beta_j =& \frac{1}{K} \sum_k \log \lambda_{+jk} - \frac{1}{JK} \sum_{j,k} \log \lambda_{+jk}/JK;\\
\gamma_k =& \frac{1}{J} \sum_j \log \lambda_{+jk} - \frac{1}{JK} \sum_{j,k} \log \lambda_{+jk};\\
\alpha_i =& \log \lambda_{i++} - \frac{1}{I} \sum_i \log \lambda_{i++};\\
\mu =& \frac{1}{I} \sum_i \log \lambda_{i++} + \frac{1}{JK} \sum_{j,k} \log \lambda_{+jk} - \log \lambda_{+++}.
\end{aligned}$$

(c) Based on this conditional, we have

$$\lambda_{ijk} = \frac{\lambda_{i+k}\lambda_{+jk}}{\lambda_{+++}}.$$

We have

$$\log \lambda_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk},$$

Table 1: ANOVA based on x_1 and x_2

Source	DF	Squares	Mean Square	F Value
Model	5	303.2395833	60.6479167	14.52
Error	42	175.4770833	4.1780258	
Total	47	478.7166667		

where

$$
\begin{aligned}
(\beta\gamma)_{jk} =& \log\lambda_{+jk} - \frac{1}{J}\sum_j \log\lambda_{+jk} - \frac{1}{K}\sum_k \log\lambda_{+jk}/K + \frac{1}{JK}\sum_{j,k}\log\lambda_{+jk};\\
(\alpha\gamma)_{ik} =& \log\lambda_{i+k} - \frac{1}{I}\sum_i \log\lambda_{i+k} - \frac{1}{K}\sum_k \log\lambda_{i+k} + \frac{1}{IK}\sum_{i,k}\log\lambda_{i+k};\\
\alpha_i =& \frac{1}{K}\sum_k \log\lambda_{i+k} - \frac{1}{IK}\sum_{i,k}\log\lambda_{i+k};\\
\beta_j =& \frac{1}{K}\sum_k \log\lambda_{+jk} - \frac{1}{JK}\sum_{j,k}\log\lambda_{+jk};\\
\gamma_k =& \frac{1}{I}\sum_i \log\lambda_{i+k} - \frac{1}{IK}\sum_{i,k}\log\lambda_{i+k} + \frac{1}{J}\sum_j \log\lambda_{+jk} - \frac{1}{JK}\sum_{j,k}\log\lambda_{+jk};\\
\mu =& \frac{1}{IK}\sum_{i,k}\log\lambda_{i+k} + \frac{1}{JK}\sum_{j,k}\log\lambda_{+jk} - \log\lambda_{+++}.
\end{aligned}
$$

2. (a) For each treatment group, plot the lines of response versus time for each animal. If we see five parallel curves, then it means given treatment, time and animal interaction effect can be ignored.

   (b) The objective is to study the treatment effect. Thus, we need to look at result of treatment effect from the ANOVA model.

   (c) Animal is nested in treatment group. The animial effect may be put in the error term and not considered in the analysis.

   (d) Do, because the animal is nested. We need to use linear mixed effect model. In the linear mixed effect model, the effect of animal is random variable. Since the interaction effect between animal and time can be ignored. We can predict the value of animal effect witnin each group and then look at the QQ-plot of the estimated values.

3. (a) Tukey additive test; or three-factor interaction plot.

   (b) The ANVOA table is given in table 1.

(c) If we let $\beta = (\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3)$, and

$$X = \begin{pmatrix} 1_4 & 1_4 & 0 & 1_4 & 0 & 0 \\ 1_4 & 1_4 & 0 & 0 & 1_4 & 0 \\ 1_4 & 1_4 & 0 & 0 & 0 & 1_4 \\ 1_4 & 1_4 & 0 & -1_4 & -1_4 & -1_4 \\ 1_4 & 0 & 1_4 & 1_4 & 0 & 0 \\ 1_4 & 0 & 1_4 & 0 & 1_4 & 0 \\ 1_4 & 0 & 1_4 & 0 & 0 & 1_4 \\ 1_4 & 0 & 1_4 & -1_4 & -1_4 & -1_4 \\ 1_4 & -1_4 & -1_4 & 1_4 & 0 & 0 \\ 1_4 & -1_4 & -1_4 & 0 & 1_4 & 0 \\ 1_4 & -1_4 & -1_4 & 0 & 0 & 1_4 \\ 1_4 & -1_4 & -1_4 & -1_4 & -1_4 & -1_4 \end{pmatrix}$$

where 1_4 is the four-dimensioanl column vector with all element 1.

(d) The estimate of σ^2 is

$$\hat{\sigma}^2 = 4.1780.$$

Thus, the average of each combination is

$$\frac{\hat{\sigma}^2}{4} = \frac{4.178}{4} = 1.0445.$$

4. (a) The three links are logistic, probit and complementary loglog link, as defined as $\log[\pi/(1-\pi)] = \eta$, $\Phi^{-1}(\pi) = \eta$ and $\log[-\log(1-\pi)] = \eta$ respectively, where η is a linear function of predictors.

(b) We can look at G^2 and X^2 from the models by the three links respectively.

(c) (Omited).

5. (a) The estimate is

$$\hat{\theta} = \frac{18 \times (60 - 13)}{13 \times (62 - 18)} = 1.4790.$$

The variance of $\log \hat{\theta}$ is

$$\frac{1}{13} + \frac{1}{44} + \frac{1}{18} + \frac{1}{47} = 0.1765.$$

The 95% confidence interval for θ is

$$[1.4790 \times e^{-1.96 \times \sqrt{0.1765}}, 1.4790 \times e^{1.96 \times \sqrt{0.1765}}] = [0.6492, 3.3696].$$

The confidence interval contains 1. Thus it is not significantly different from one.

(b) The fit is good based on 0.05 level, because the p-value for the deviance is 0.0849 which is greater than 0.05.

(c) The odds ratio when log-dose increase 0.031 is

$$e^{0.031 \times 34.2859} = 2.8946.$$

Its p-value is exactly the p-value of log-dose, which is less than 0.0001. Thus it is significant.

(d) The odds ratio in (a) is not significant in (b) is significant because we only use the part of the table of the data. In logistic model, we use all of the data. Thus, it produces the difference.

6. (a) There are 31 trees used in this data.

(b)

$$\begin{aligned} A =& \frac{3422.55167}{38.65483} = 88.54137; \\ B =& \frac{6845.10334}{7927.4387} = 0.8635; \\ C =& \frac{-52.59692}{13.93248} = -3.7751. \end{aligned}$$

(c) It strongly suggests log-transformation, because the confidence interval for λ does not contain 1 and the estimate of λ is very close to 0.

(d) The fit is good because R^2 is 0.9235 which is close to one. The sign of the two predictor is positive because the sign of the estimate is negative.

(e) The predicted value is

$$\hat{\log Y} = 0.35266 + 12 \times 0.12345 + 75 \times 0.01737 = 3.1368.$$

The variance of the mean of the predicted value is

$$\begin{aligned} &0.1108492879 + 2 \times 12 \times 0.000663884 - 2 \times 75 \times 0.001563494 \\ &+ 144 \times 0.0000807487 - 2 \times 12 \times 75 \times 0.00002264 + 75 \times 75 \times 0.0000244709 \\ =&0.000783. \end{aligned}$$

The variance of the predicted value is

$$0.000783 + 0.14857^2 = 0.022856.$$

Therefore, the 95% confidence interval for log(Y) is

$$[3.1368 - 2.0484 \times \sqrt{0.022856}, 3.1368 + 2.0484 \times \sqrt{0.022856}] = [2.8271, 3.4465].$$

Thus, that for Y is

$$[e^{2.8271}, e^{3.4465}] = [16.8964, 31.3903].$$

(f) Because the correlation between the interaction variable and the two variables are far from 0.

7. We fitted two regression model for U and V as $U = \alpha_U + \beta_u T + \epsilon_U$ and $V = \alpha_V + \beta_v T + \epsilon_V$, where $\epsilon_U \sim N(0, \sigma_U^2)$ and $\epsilon_V \sim N(0, \sigma_V^2)$. We have $\hat{\alpha}_U = 13.8394$, $\beta_U = 43.4194$, $\hat{\alpha}_V = 1362.40$ and $\beta_V = 28.9477$.

(a)
$$T_0 = \frac{1362.40031 - 13.83953}{43.41942 - 28.94772} = 93.19.$$

(b) Let $Y = U - V$. Then, we can fit a model $Y = \alpha + \beta T + \epsilon$. Compute the predicted interval at $T_0 = 90$. If the interval includes 0, then we accept $T_0 = 90$. Otherwise we reject $T_0 = 90$. Let Y_0 be the value at $T_0 = 90$. Then, $\hat{Y}_0 = -46.10775$ and $V(\hat{Y}_0) = 49.53$. The z-value is 6.55. Thus, we reject $T_0 = 90$.

(c) The estimated value is
$$\hat{T}_0 = \frac{\hat{\alpha}_U - \hat{\alpha}_V}{\hat{\beta}_V - \hat{\beta}_U}.$$

To compute the variance of $\hat{T}_0$, we can use δ method. Then, the confidence interval is constructed by Wald method as $\hat{T}_0 \pm t_{0.025,4}\sigma(\hat{T}_0)$.