Solutions to Methods in Fall 2005

1. (a) The total number of plants were $59 + 1 = 60$.

   (b) Yes, they are equal. It is very clear from model 5 to model 7. We can see that the estimated values of x_1 and x_2 are all the same. It means the design matrix of model 5 are orthogonal which exactly indicates equal numbers for pear combination.

   (c) For testing $H_0 : \beta_1 = 0$, we need the t test. It gives t-value $0.28394/0.24405 = 1.16$ indicating H_0 is not rejected. Similarly, the t-value for $H_0 : \beta_3 = 0$ is -0.8947 also indicating H_0 is not reject. To test $H_0 : \beta_1 = \beta_2 = 0$, we need the F-test. Compare the model SS between model 2 and model 3, we get $427.38 - 393.07 = 34.31$. In this case,
   $$\hat{\sigma}^2 = \frac{503.92 - 427.38}{59 - 4} = 1.39.$$
   Thus, the F-value is $34.31/(2\times1.39) = 12.34$ which is greater than 3.2. It indicates significant. In the joint test, at least one of β_1 or β_3 is significant different from 0.

   (d) I suggest model 4. It is clear that X_1^2 and X_2^2 term are not significant from model 1, 2, 3. They should not be included in the model. The interaction term are significant in model 1 and model 4. Even though X_1 is not significant in model 4, since the interaction term is significant, it must be included. The model is
   $$Height = 6.47027 - 0.02953X_1 - 0.87742X_2 + 0.02402X_1X_2.$$
   It tells us that when X_2 is fixed, the height increase about $0.02953 + 0.0240X_2$ when X_1 increase 1 day. Similarly, when X_1 is fixed, the height increase about $0.87742 + 0.0240X_1$ when X_2 increase 1 degree.

   (e) When $X2 = 2$ and $X_1 = 80^o$, then the predicted height is
   $$6.47027 - 0.02953(80) - 0.87742(2) + 0.02402(80)(2) = 6.19623.$$

   (f) Yes, we can. Since the data is balanced, we can write down the design matrix. Based on the design matrix, we can get the correlation between those βs. It needs some complicated computations. Let X be the design matrix. Then, we have $Cov(\hat{\beta}, \hat{\beta}) = \hat{\sigma}^2(X^tX)^{-1}$. It gives the correlation as $\rho_{12} = -0.9926492$, $\rho_{13} = -0.8528029$ and $\rho_{14} = 0.8465341$, $\rho_{23} = 0.009491525$, $\rho_{24} = -1.205273e-04$ and $\rho_{34} = 3.013183e-05$. Thus, the variance of the predicted value is $0.01990930 = 0.1411$. The 95% confidence interval for the predicted value is
   $$[6.19623 - t_{0.025,56}0.1411, 6.19623 - t_{0.025,56}0.1411] = [5.9136, 6.4789].$$

Comments: Since this method is impossible to be computed by calculator, you had better say that you can not and you need the matrix.

2. (a) Here $\hat{\eta} = 2.404 - 0.3235(10) = -0.831$. Thus, the estimated probability is

$$\hat{p} = \frac{e^{-0.831}}{1 + e^{-0.831}} = 0.3034.$$

(b) Here

$$V(\hat{\eta}) = (1.1918)^2 + 10^2(0.1140)^2 + 2(10)(-0.96)(1.1918)(0.1140) = 0.1114.$$

The 95% for η is

$$[-0.831 - 1.96\sqrt{0.1114}, -0.831 + 1.96\sqrt{0.1114}] = [-1.4852, -0.1768].$$

Thus, the 95% confidence interval for p is

$$[\frac{e^{-1.4852}}{1 + e^{-1.4852}}, \frac{e^{-0.1768}}{1 + e^{-0.1768}}] = [0.1846, 0.4559].$$

(c) If $\beta = 0$, then it is the iid case. There are 14 of 1 and 40 of 0. Thus, $\hat{p} = \frac{14}{14+40}$ and

$$\hat{\alpha} = \log(\frac{14}{40}) = -1.0498.$$

The deviance is

$$-2[14\log(\hat{p}) + 40\log(1 - \hat{p})] = 61.806.$$

(d) The difference is $61.806 - 51.017 = 10.789$. It is larger than $\chi^2_{0.05,1} = 3.84$ indicating β is singficant different from 0.

3. We need to approximation for the integral part. It could be computed by

$$\int_{t_{i-1}}^{t_i} I_0 e^{-\frac{2\tau_D}{x}} dx = I_0(t_i - t_{i-1})e^{-\frac{2\tau_D}{(t_i+t_{i-1})/2}} - \frac{8\tau_D}{(a+b)^2} I_0(t_i - t_{i-1})^3 e^{-\frac{2\tau_D}{(t_i+t_{i-1})/2}}$$

or we may need higher-order approximation. We also need to compute its derivative.

(a) The method should be based an approximation to the integral part. It does not have analytic form, but it can be computed approximately. Let $w_i = 1/(t_i - t_{i-1})$. Let

$$g_i(\tau_D) = \int_{t_{i-1}}^{t_i} e^{-\frac{2\tau_D}{x}} dx.$$

Then

$$g_i'(\tau_D) = -\int_{t_{i-1}}^{t_i} \frac{2}{x} e^{-\frac{2\tau_D}{x}} dx$$

Then, we need to find I_0 and τ_D to minimize

$$L(I_0, \tau_D) = \sum_{i=1}^{n} w_i(y_i - I_0 g_i(\tau_D))^2.$$

Note that

$$\begin{aligned} \frac{\partial L}{\partial I_0} &= -\sum_{i=1}^{n} w_i(y_i - I_0 g(\tau_D))g_i(\tau_D) \\ \frac{\partial L}{\partial \tau_D} &= -I_0 \sum_{i=1}^{n} w_i(y_i - I_0(g(\tau_D))g_i'(\tau_D). \end{aligned}$$

The above equations can be solved numerically such as by Newton-Raphson method.

(b) Let $h = t_i - t_{i-1}$. In this case, the error is iid. Suppose the predicted values $\hat{y}_i$ is obtained. Note that if the model is true, y_i is about $I_0 e^{-\frac{4\tau_D}{h}}$ if h is small. Thus, if we plot $\log(y_i)$ versus $2/(t_i + t_{i-1})$, we should see an approximately linear function if the f is correct. The assumptions of the error term can be evaluated by look at model residuals by a lot of plots (omitted).

(c) The standard deviation of $\hat{\omega}$ in design 1 is less than that in design 2. However, for $\hat{\tau}$, it is opposite. However, since he wants both the estimate of ω and τ_D in just one design, it may suggest that neither is better than the other. Dependent on which one is more interested in, we can choose the designs.

4. (a) The additive model with all the main effects suggest a not significant residual deviance ($\chi^2_{0.05,4} = 9.488$). However, since the interaction effect between sizep and age reduces 6.63 residual deviance. It is better to keep this in the model. Thus, the best model includes all the main effects and the interaction effect between sizep and age.

(b) We can use the loglikelihood test statistic based on the assumption of χ^2 distribution.

(c) The table is $2 \times 2 \times 2$.

(d) Since the sizep:age interaction effect is included, the p-value of main effect does not really reflected the significance. This result is inconsistent with the previous one but we trust the previous one.

(e) The model is

$$\hat{\eta} = 0.90 - 3.46 Sizep_S - 0.36 I + 5.43 Sisef_S - 3.66(Sizep_L)(I).$$

When pirate is S and Feeder is L, the model becomes

$$\hat{\eta} = -2.56 - 4.02 I.$$

Thus, the relative odds is

$$e^{-4.02} = 0.0179.$$

(f) It is

$$\hat{p} = \frac{e^{0.90}}{1 + e^{0.90}} = 0.7109.$$

5. Let $h(t)$ be the hazard function, $f(t)$ be the density function and $F(t)$ be the CDF. The formulae are

$$h(t) = \frac{f(t)}{1 - F(t)}$$

and

$$F(t) = 1 - e^{-\int_0^t h(x)dx}.$$

Thus, we have

$$f(t) = h(t)e^{-\int_0^t h(x)dx}.$$

For this particular model, we have

$$f(t) = \begin{cases} \lambda_1 e^{-\lambda_1 t}, & \text{when } 0 \leq t < \pi_1 \\ \lambda_2 e^{-\lambda_1\pi_1 - \lambda_2(t-\pi_1)} & \text{when } \pi_1 \leq t < \pi_2 \\ \lambda_3 e^{-\lambda_1\pi_1 - \lambda_2(\pi_2-\pi_1) - \lambda_3(t-\pi_2)} & \text{when } t \geq \pi_2 \end{cases}$$

and

$$S(t) = \begin{cases} e^{-\lambda_1 t}, & \text{when } 0 \leq t < \pi_1 \\ e^{-\lambda_1\pi_1 - \lambda_2(t-\pi_1)} & \text{when } \pi_1 \leq t < \pi_2 \\ e^{-\lambda_1\pi_1 - \lambda_2(\pi_2-\pi_1) - \lambda_3(t-\pi_2)} & \text{when } t \geq \pi_2 \end{cases}$$

Usually $\delta_i = 0$ means dropoff. The likelihood function is

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^n [\frac{f(t_i)}{S(t_i)}]^{\delta_i} S(t_i) = \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i)$$

and the loglikelihood function is

$$l = \sum_{i=1}^n \delta_i \log[h(t_i)] + \sum_{i=1}^n \log[S(t_i)].$$

(a) Let $n_1 = \#\{\delta_i : 0 \leq t_i < \pi_1\}$, $n_2 = \#\{\delta_i : \pi_1 \leq t_i < \pi_2\}$ and $n_3 = \#\{\delta_i : t_i \geq \pi_3\}$. Let $t_{1+} = \sum_{i=1}^n t_i I_{[0,\pi_1)}(t_i)$, $t_{2+} = \sum_{i=1}^n t_i I_{[\pi_1,\pi_2)}(t_i)$ and $t_{3+} = \sum_{i=1}^n t_i I_{[\pi_2,\infty)}(t_i)$ For this particular model, we have

$$\begin{aligned} l =& \sum_{i=1}^n \delta_i[\log(\lambda_1) I_{[0,\pi_1)}(t_i) + \log(\lambda_2) I_{[\pi_1,\pi_2)}(t_i) + \log(\lambda_3) I_{[\pi_2,\infty)}(t_i)] \\ & - \lambda_1 t_i I_{[0,\pi_1)}(t_i) - [(\lambda_1 - \lambda_2)\pi_1 + \lambda_2 t_i] I_{[\pi_1,\pi_2)}(t_i) \\ & - [(\lambda_1 - \lambda_2)\pi_1 + (\lambda_2 - \lambda_3)\pi_2 + \lambda_3 t_i] I_{[\pi_2,\infty)}(t_i) \\ =& \sum_{i=1}^n [\log(\lambda_1) n_1 + \log(\lambda_2) n_2 + \log(\lambda_3) n_3] - \lambda_1 t_{1+} - \lambda_2 t_{2+} - \lambda_3 t_{3+} \\ & - 2(\lambda_1 - \lambda_2)\pi_1 - (\lambda_2 - \lambda_3)\pi_2. \end{aligned}$$

(b) Compute the partial derivatives, we have

$$
\begin{aligned}
\frac{\partial \ell}{\partial \lambda_1} =& \frac{n_1}{\lambda_1} - t_{1+} - 2\pi_1 \\
\frac{\partial \ell}{\partial \lambda_2} =& \frac{n_2}{\lambda_2} - t_{2+} + 2\pi_1 - \pi_2 \\
\frac{\partial \ell}{\partial \lambda_3} =& \frac{n_3}{\lambda_3} - t_{3+} + \pi_2
\end{aligned}
$$

We have

$$
\begin{aligned}
\hat{\lambda}_1 =& \frac{n_1}{t_{1+} + 2\pi_1} \\
\hat{\lambda}_2 =& \frac{n_2}{t_{2+} + \pi_2 - 2\pi_1} \\
\hat{\lambda}_3 =& \frac{n_3}{t_{3+} - \pi_2}.
\end{aligned}
$$

(c) We compute the Fisher Information Matrix. Note that the cross-derivative are all 0. We only need

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda_1^2} =& - \frac{n_1}{\lambda_1^2} \\
\frac{\partial \ell^2}{\partial \lambda_2^2} =& - \frac{n_2}{\lambda_2^2} \\
\frac{\partial \ell^2}{\partial \lambda_3^2} =& - \frac{n_3}{\lambda_3^2}.
\end{aligned}
$$

Therefore, the standard error of the MLE can be computed by

$$
\begin{aligned}
V(\hat{\lambda}_1) =& \frac{n_1}{n\lambda_1^2} \\
V(\hat{\lambda}_2) =& \frac{n_1}{n\lambda_2^2} \\
V(\hat{\lambda}_3) =& \frac{n_1}{n\lambda_3^2}
\end{aligned}
$$

by plugging the MLEs.

6. The design is orthogonal. So we can compute the SS of material (SSM) as

$$
SSM = 4 \sum_{i=A}^{D} (\bar{y}_A - \bar{y})^2 = 46.215.
$$

Thus, we have

$$
SSE = 74.440 - 9.865 - 14.685 - 46.215 = 3.675.
$$

(a) The F-value is

$$
F^* = \frac{SSM/3}{SSE/6} = \frac{46.215/3}{3.765/6} = 24.55.
$$

Comparing it with $F_{0.05,3,6} = 4.76$, it is too large. Thus, we conclude that the is a significant difference among the materials.

(b) Since the estimated variances of the differences are exactly the same and they are

$$\frac{1}{2}\hat{\sigma}^2 = \frac{1}{2}\frac{SSE}{6} = 0.31375.$$

It gives the critical value $t_{0.025,6}\sqrt{0.21375} = 1.3706$. Then, the difference between (A,B), (A,C), (A,D) and (B,C) are significant. Only the difference between (B,D) and (C,D) are not.

(c) Based on additive model, we can also compute the differences between treatments of Position and Application and test the significance of the differences.

(d) It means that the levels of application are not significant different from each other. This can be done by an F test. The F-value is

$$F^* = \frac{9.865/3}{3.675/6} = 5.368$$

which is greater than the critical value 4.76. Thus, this claim is rejected at $\alpha = 0.05$.

(e) If we exclude the application from the model, then

$$SSE = 3.675 + 9.865 = 13.54.$$

The F-value becomes

$$F^* = \frac{46.215/3}{13.54/9} = 10.24$$

which is also larger than $F_{0.05,3,9} = 3.86$. Thus, there is still a significant difference.