Statistics 232A: Lab #4

Fall 2013 R. Beran

DUE in Discussion on 5 November 2013. Use class techniques and results to do the project. Visit `www.stat.ucdavis.edu/~beran/s232a/monkey.dat` for the data. Attach your computer code to your report.

The monkey data, printed on p. 189 of Scheffé's text and reformatted in `monkey.dat`, reports responses to a certain stimulus that were measured for 5 different monkey-pairs (the subjects) in 5 different periods under 5 different conditions. Factor 1 is the monkey-pair; Factor 2 is the period; and Factor 3 is the conditions. The 25 factor level triples (i, j, k) at which we have a single observation form a structured subset B of all 125 factor level triples $\{(i, j, k): 1 \leq i, j, k \leq 5\}$. The data structure in this experiment is called a **Latin Square design**. We will treat the data as an incomplete three-way layout with one observation per observed triple of factor levels. The **general** Gaussian model states: the observed responses $y_{ijk} = m_{ijk} + e_{ijk}$, where $(i, j, k) \in B$ and the errors $\{e_{ijk}: (i, j, k) \in B\}$ are independent, identically distributed, $N(0, \sigma^2)$ random variables.

Consider the associated (mostly unobserved) complete layout of means $\{m_{ijk}: 1 \leq i, j, k \leq 5\}$. Let $m = \{\{\{m_{ijk}: 1 \leq i \leq 5\}, 1 \leq j \leq 5\}, 1 \leq k \leq 5\}$ denote the 125×1 mirror-dictionary vectorization of these means. Let D be the matrix such that the vector Dm records the 25 means $\{m_{ijk}: (i, j, k) \in B\}$ for which we have an observation. Let y denote the data $\{y_{ijk}: (i, j, k) \in B\}$ vectorized so that $\eta = \mathrm{E}(y) = Dm$. The **general** Gaussian model for the incomplete layout described above can restated: $y = Dm + e$, where e has the $N(0, \sigma^2 I_{25})$ distribution.

The classical treatment of a Latin Square assumes the more restrictive **additive** model in which the means satisfy

$$m_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k, \qquad 1 \leq i, j, k \leq 5$$

subject to the constraints $\alpha_+ = \beta_+ = \gamma_+ = 0$.

a) Show that this additive model for the means is equivalent to assuming that $m = P\beta$ for some $\beta \in R^{125}$, with

$$P = P_0 + P_1 + P_2 + P_3.$$

Here $P_0 = J_3 \otimes J_2 \otimes J_1$, $P_1 = J_3 \otimes J_2 \otimes H_1$, $P_2 = J_3 \otimes H_2 \otimes J_1$, and $P_3 = H_3 \otimes J_2 \otimes J_1$.

Consequently, the Gaussian additive model for the Latin Square layout asserts: $y = DP\beta + e$, where $\beta \in R^{125}$ and e has the $N(0, \sigma^2 I_{25})$ distribution.

b) Assuming the additive model, use results from class to show that the least squares estimator of $\eta = \mathrm{E}(y) = DP\beta$ is $\hat{\eta} = D(DP)^+ y$. Describe the associated estimator $\hat{\sigma}^2$ of σ^2 and give its distribution under the additive model. A key step is to calculate the rank of the design matrix, either numerically or theoretically. Compute the numerical values of $\hat{\eta}$ and $\hat{\sigma}^2$ for the monkey data.

c) Compute the residual vector $\hat{e} = y - \hat{\eta}$. Plot the ordered residuals against the corresponding quantiles of the standard normal distribution. Add to the plot the straight line that passes through the plotted points associated with the first and third quartiles of the residuals. Does this plot of the residuals cast serious doubt on appropriateness of the Gaussian additive model?

d) For $1 \leq k \leq 3$, verify numerically that $P_k = P_k(DP)^+DP$. Deduce from this that $P_k m = P_k(DP)^+\eta$ under the additive model. Hence show that $P_k m$ is linearly estimable under the additive model. (This is the point of the Latin square design).

e) Given that $P_k m$ is linearly estimable under the additive model, show that its least squares estimator under that model is $P_k(DP)^+y$. Thereby compute and report the least squares estimates, under the additive model, of the parameters μ, $\{\alpha_i\colon 1 \leq i \leq 5\}$, $\{\beta_j\colon 1 \leq j \leq 5\}$, and $\{\gamma_k\colon 1 \leq k \leq 5\}$ in the additive model.