Indian Statistical Institute, Delhi Centre

<u>Linear Models and GLM</u>

<u>Spring 2008</u>

<u>Answer to Quiz # 2</u>

Date: April 4, 2008 (Friday) **Total Points:** $2 \times 5 = 10$

1. Consider the following model:
$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{C}\gamma + \varepsilon \,,$$
where co-ordinates of ε are i.i.d. Normal$\,(0,1)$.
Indicate if the following statements are **True** or **False**:
   (a) If $\mathbf{C}^T\mathbf{C}$ invertible and $\mathbf{P_X}C = \mathbf{0}$ then γ is estimable. <u>`True`</u>
   (b) If γ is estimable with LSE $\hat{\gamma}$ then $\hat{\gamma}$ has a multivariate normal distribution which is singular. <u>`False`</u>
   (c) Suppose γ is estimable, put $\mathbf{Z} = \mathbf{Y} - \mathbf{C}\hat{\gamma}$ then $\mathbf{E}\left[\mathbf{Z}\right] = \mathbf{X}\beta$. <u>`True`</u>
   (d) Two-way classification model with no interaction and one observation per cell can be written as a special case of the above model. <u>`True`</u>

2. Consider the following model:
$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} \;\; 1 \leq k \leq K, \; 1 \leq j \leq J, \; 1 \leq i \leq I \,,$$
where ε_{ijk}'s are i.i.d. Normal$\left(0, \sigma^2\right)$.
Fill in the blanks:
   (a) It is called the <u>`Three`</u>-way classification model with no interaction.
   (b) The degrees of freedom for the *residual sum of square* is $\underline{IJK - I - J - K + 2}$.
   (c) The *maximum likelihood estimate* of σ^2 is given by $\underline{\frac{1}{IJK} \sum_{i,j,k} \left(y_{ijk} - y_{i\cdot\cdot} - y_{\cdot j\cdot} - y_{\cdot\cdot k} + 2y_{\cdot\cdot\cdot}\right)}$.
   (d) The linear parametric functions $\alpha_i - \alpha_{i'}$ for $1 \leq i \neq i' \leq I$ are <u>`estimable`</u>.

3. Consider the following model:
$$y_{ij} = \mu + \alpha_i + \gamma_{ij} + \varepsilon_{ij} \;\; 1 \leq j \leq k_i, \; 1 \leq i \leq I \,,$$
where ε_{ij}'s are i.i.d. Normal$\,(0,1)$.
Indicate if the following statements are **True** or **False**:

(a) $\alpha_1 - \alpha_2$ is estimable. False

(b) If we fix an i then the observations indexed by j form an *one-way classification model.* True

(c) Suppose $k_1 = 10$ then we can do multiple comparison using *Tukey's Honest Significant Difference* to test for $\gamma_{11} - \gamma_{12} = 0$ and $\gamma_{12} - \gamma_{13} = 0$. True

(d) For the multiple comparison in (c) above if we use *Bonferroni's method* then we should do the one-degrees of freedom testing at a level 0.025 to achieve an experimental error rate of 5%. True

4. Indicate if the following statements are **True** or **False**:

   (a) A *log-linear model* is a linear model. False

   (b) A two-way classification data represented as a $I \times J$ table can be modeled by a *log-linear model.* False

   (c) The estimates obtained in *logistic regression* are MLEs under appropriate model. True

   (d) The following *log-linear model* is a *saturated model* False

$$\log m_{ijk} = u_0 + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)} + u_{31(ki)} \,.$$

5. Fill in the blanks:

   (a) For a linear model the residuals are always uncorrelated (independent) of the LSEs.

   (b) One-way classification model is a sub-model of two-way classification model.

   (c) The degrees of freedom for the residual sum of square from a *four-way classification model* with no interaction and one observation per cell is $K^4 - 4K - 3$ where each classification has K categories.

   (d) *Tukey's one degrees of freedom test* is a test of non-additivity.