Advanced Methods in Biostatistics 2 (140.752)
Winter 2005

Homework Assignment #6
(Due Monday, December 5, 2005)

Please hand in a hard copy of your R code, and send an electronic version of it to Kenny (kshum@jhsph.edu).

1. The following data give the mean and standard deviation on an examination (ILEA score) for a random sample of 7 schools from the population of all high schools in England. There were 50 students examined in each school.

School	ILEA Mean	ILEA s. d.
1	28.3	13.4
2	21.1	14.2
3	14.3	11.8
4	16.3	14.9
5	26.5	13.4
6	20.3	12.1
7	16.8	13.9

   (a) Explain why a one-way random effects model is appropriate for this data set.
   (b) Obtain the analysis of variance table.
   (c) Obtain estimates for the two variance components and a confidence interval for the ratio of the two variance components (i. e. for σ_1^2/σ^2). Comment on the results.

2. There are three frequently occurring test statistics, the likelihood ratio test statistic, the Wald test, and the score test. If $\mathbf{Y}$ has the probability density function $f(\mathbf{y};\boldsymbol{\beta})$ at $\mathbf{Y}=\mathbf{y}$, where $\boldsymbol{\beta}$ is $p\times 1$, then hypothesis of interest are often of the form $H_0 : \mathbf{L}'\boldsymbol{\beta} = \boldsymbol{\xi}$ versus $H_1 : \mathbf{L}'\boldsymbol{\beta} \neq \boldsymbol{\xi}$, where $\mathbf{L}'$ is $s\times p$ of rank $s\le p$. Let

   - $\hat{\boldsymbol{\beta}}$ denote the maximum likelihood estimate of $\boldsymbol{\beta}$ under the full model,
   - $\tilde{\boldsymbol{\beta}}$ denote the maximum likelihood estimate of $\boldsymbol{\beta}$ under the model assuming the null hypothesis is true,
   - $\ell(\boldsymbol{\beta}) = \log[f(\mathbf{y};\boldsymbol{\beta})]$ denote the log likelihood function,
   - $\mathbf{s}(\boldsymbol{\beta})$ be the vector of scores with j^{th} component
   $$s_j(\boldsymbol{\beta}) = \frac{\delta\ell(\boldsymbol{\beta})}{\delta\beta_j},$$
   - $\Im(\boldsymbol{\beta})$ be Fisher's information matrix which has j,k element equal to
   $$-E\left[\frac{\delta^2\ell(\boldsymbol{\beta})}{\delta\beta_j\delta\beta_k}\right].$$

   The three test statistics in this case are the likelihood ratio test statistic, given by
   $$-2[\ell(\tilde{\boldsymbol{\beta}}) - \ell(\hat{\boldsymbol{\beta}})],$$
   the Wald test statistic, given by
   $$(\mathbf{L}'\hat{\boldsymbol{\beta}} - \boldsymbol{\xi})'[\mathbf{L}'\Im(\hat{\boldsymbol{\beta}})^{-1}\mathbf{L}]^{-1}(\mathbf{L}'\hat{\boldsymbol{\beta}} - \boldsymbol{\xi}),$$
   and the score test, given by
   $$\mathbf{s}'(\tilde{\boldsymbol{\beta}})\Im(\tilde{\boldsymbol{\beta}})^{-1}\mathbf{s}(\tilde{\boldsymbol{\beta}}).$$

   For the linear model $\mathbf{Y}\sim \text{MVN}(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \sigma^2\mathbf{I})$, where $\mathbf{X}_1$ is $n\times q$ of rank q, $\mathbf{X}_2$ is $n\times(p-q)$ of rank $p-q$, $\mathbf{X}=[\mathbf{X}_1,\mathbf{X}_2]$ is $n\times p$ of rank p, and σ^2 is known, derive the three test statistics for testing $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$. Comment.