Human group specific component (Gc) is a plasma transport protein for vitamin D. Polymorphic electrophoretic variants of Gc are found in all human populations. An interesting question is whether the genotypes at the Gc locus influence quantitative differences in plasma concentrations of the Gc protein. Daiger et al. (1986) collected relevant data on 31 identical twin pairs (with same genotypes), 13 fraternal twin pairs, and 45 unrelated controls. For each individual, the concentrations of Gc was available along with additional information about gender, age, and Gc genotype of the individual. The three genotypes 11, 12, and 22 at the Gc locus are distinguishable.

Assume the mean Gc concentration, \( \mathbb{E}(Y) \), is linearly determined by Gc genotypes (denoted as \( X_{ij} \) for genotype \( ij \)), gender (denoted as \( X_2 \)), and age (denoted as \( X_3 \)), i.e.,

\[
\mathbb{E}(Y) = \beta_{11}X_{11} + \beta_{12}X_{12} + \beta_{22}X_{22} + \beta_2X_2 + \beta_3X_3.
\]

Assume \( Y \) follows a normal distribution with variance \( \sigma^2 \). In addition we assume the Gc concentration correlation between two identical twins is \( \rho_1 \), and between two fraternal twins is \( \rho_2 \). Answer the following questions.

1. Obtain the maximum likelihood estimates of the regression parameters, \( \beta_{ij} \), \( \beta_k \), and \( \rho_j \), and their asymptotic covariances.

2. Test \( \beta_{11} = \beta_{12} = \beta_{22} \): using previously derived asymptotic covariance matrix to construct a Wald test type statistic, or doing a likelihood ratio test and estimating significance using the chi-square distribution with 2 degrees of freedom.

3. (Bonus question) Investigate the accuracy of the asymptotic normal/chi-square approximation in question 2) (e.g., compared to a parametric Bootstrap approach to computing significance).
Hint: The twin pairs are dependent, and their joint probability is a bivariate normal distribution with the covariance matrix determined by $\sigma^2$ and the correlation parameters.

References:

You can obtain a copy of the paper from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1684475/. The data to be analyzed is in Table 1.A,B,C. I also put a copy of the data at http://umn.edu/~baolin/teaching/linmods/GcGeno1.xls.