STA244 **2/26/2004**

Due March 4 2004

1. Consider the linear model $Y = X\beta + e$, $e \sim N(0, I_n/\phi)$ with p dimensional vector of regression coefficients β, and precision ϕ (1/variance). Zellner's g-prior is

$$p(\phi) \propto \phi^{-1} \quad (1)$$

$$\beta|\phi \sim N(0, \frac{g}{\phi}(X'X)^{-1}). \quad (2)$$

   (a) Find the posterior distribution of β.

   (b) Refer to the data in Exerise 2.10.4. Find the posterior mean for β_1 and β_2 with $g = 15$. Construct 98% probability intervals for β_2 and $\beta_2 - \beta_1$.

2. A test of lack of fit of the linear model can be developed when there are replicate observations. Let Y_{ij} denote the ith observation in group j, where

$$Y_{ij} = \mu_j + e_{ij}$$

   for $i = 1, \ldots, n_j$, $j = 1, \ldots, J$, $\sum_j n_j = n$. This is the standard One-Way AOV model.

   Under the linear regression model, the means are $\mu_j = \beta_0 + \beta_1 x_j$. All observations in group j are assumed to be identically distributed, as before, but now a linear restriction has been placed on the means.

   (a) For the AOV model, give a basis for Ω (the mean space under H_1). What is the dimension of Ω?

   (b) For the null hypothesis (H_0) that the means follow the linear regression on x, give a basis for Ω_0. What is the dimension of Ω_0?

   (c) Show that Ω_0 is contained in Ω? i.e show that the linear regression places a restriction on the group means in the AOV model.

   (d) Construct an Analysis of Variance table for testing the linear regression model against the one-way AOV model. (indicate sum of squares and mean squares using the appropriate projections, i.e. P_Ω, Q_Ω, P_{Ω_0}.

   (e) Under H_1 (and the usual iid $N(0, \sigma^2)$ errors), what is the distribution of the F-statistic in the above table?

3. The following problem is based on data from Leger et al. 1979, "Factors associated with Cardiac Mortality in Developed Countries with Particular Reference to the Consumption of Wine", *Lancet*. The data are on the course calendar and are based on the average wine consumption rates (in liter/person) and number of ischemic heart disease deaths (per 1000 men aged 55 to 64 years old) for 18 industrialized countries. Do these data suggest that the heart disease rate is associated with average wine consumption? If so, how can that relationship best be described? (do you need to transform any of the variables to achieve a linear relationship)? Is there any evidence that a linear model is inappropriate? (carry out the lack of fit test developed in part 2, as well as examine residual plots). Is the Lack of Fit test valid if you do not transform the variables? Explain.

4. Turn in a one page (typed) summary of your analysis, written in an popular science article style, i.e. suitable for the *New York Times* Science Times section.