STA244 **2/5/2003**

Homework 4

Due 2/12/2003

Consider the usual linear model $Y \sim N(X\beta, \sigma^2 I)$. and X a $n \times p$ rank $r \leq n$ matrix.

1. Show that the MSE, $\hat{\sigma}^2 \equiv Y'(I - P_X)Y/(n-r)$, where P_X is a rank r orthogonal projection on the $S(X)$, is a minimum variance unbiased estimator for σ^2.

2. For $c'\beta$ estimable, show that

$$\frac{c'\hat{\beta} - c'\beta}{\sqrt{\hat{\sigma}^2 c'(X'X)^- c}}$$

has a Student t distribution with $n - r$ degrees of freedom. Find the form for a confidence interval for $c'\beta$ and the form of a test of $H_o : c'\beta = 0$ versus $H_a : c'\beta \neq 0$. (See Appendix E of Christensen).

3. Suppose that we want to predict the value of a future observation, Y_f, at a vector x'_f $(1 \times p)$, where $Y_f \sim N(x'_f\beta, \sigma^2)$ and is independent of Y.

   (a) Find the distribution of

$$\frac{Y_f - x'_f\hat{\beta}}{\sqrt{\hat{\sigma}^2[1 + x'_f(X'X)^- x_f]}}$$

   (b) Let $\eta \in (0, 0.5]$. The 100ηth percentile of the distribution of Y_f is, say, $\gamma(\eta) = x'_f\beta + z(\eta)\sigma$ where $z(\eta)$ is the 100ηth percentile of a standard normal; note that $z(\eta)$ is negative. This lower confidence bound is referred to as a lower η tolerance point with confidence coefficient (1 - α) 100%. For example, $\eta = 0.1$, $\alpha = 0.05$, and Y_f is the octane value of a batch of gasoline manufactured under conditions x'_f, then we are 95% confident that no more than 10% of all batches produced at x'_f will have an octane value below the tolerance point. Find a (1-α)100% lower confidence interval for $\gamma(\eta)$. *Hint: use a non-central t distribution based on* $x'_f\hat{\beta} - \gamma(\eta)$ *i.e. location is not 0.*

4. Suppose we have a $n \times p$ matrix Q and a $p \times p$ upper triangular matrix R such that $Q'Q = I_p$ and $QR = X$. *assume that* X *is of rank* p

   (a) Show that $R'R = X'X$.

   (b) Show that $\hat{\beta} = R^{-1}Q'Y$. *Thus to efficiently compute* $\hat{\beta}$, *one computes* $z = Q'Y$ *then solves the system of equations* $R\hat{\beta} = z$ *by back substitution without explicit inversion of* $X'X$.

   (c) Show that QQ' is an orthogonal projection of rank p onto the $S(X)$ and that $\hat{Y} = QQ'Y$.

   (d) Find e (the residuals) and residual sum of squares in terms of Y and Q.

(e) Show that the variance of a linear combination $c'\hat{\beta}$ can be written as $\sigma^2 d'd$ where $d = R^{-T}c$, where $R^{-T} = (R')^{-1}$. To find d one can use back substitution in the system of equations $R'd = c$.

(f) Suppose that

$$R = \begin{bmatrix} 2 & 4 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 8 \end{bmatrix}$$

Assuming that $\sigma^2 = 1$, find the variance of $\hat{\beta}_0$, the variance of $\hat{\beta}_1$ and the covariance between $\hat{\beta}_0$ and $\hat{\beta}_1$ using back substitution. The latter requires a slight extension of the above results.