STAT 802 Multivariate Analysis October 7, 2003

Two Formulations of Multivariate Regression:

1. (Section 7.7 pp 383ff)

$Y=Z\beta+\varepsilon$ where there are m response variables, r predictor variables, and n cases.

Here Z is n x (r+1) when there are n cases and r variables (The +1 is for the 1-vector)
Y is n x m when there are m response variables, and
β is (r+1) x m
ε is n x m
Z values are considered fixed values (not random, but set)

Least squares estimate of β. $\hat{\beta} = (Z'Z)^{-1}Z'Y$ and residuals $\hat{\varepsilon} = Y - \hat{Y} = Y - Z\hat{\beta}$
The matrix of error sum of squares and cross-products is
$(Y-Z\hat{\beta})'(Y-Z\hat{\beta})$ and m x m matrix. (If m =1, there is no cross-product of one component of Y with another)

$\hat{\beta}$ is the value of β that minimizes the generalized sample variance
$= \text{trace}\{(Y-Z\hat{\beta})'(Y-Z\hat{\beta})\}$

and the orthogonality of $\hat{\varepsilon}$ and $\hat{Y}$ and the columns of Z holds as in multiple regression.

The analysis of variance looks like :

$Y'Y = \hat{Y}'\hat{Y} + \hat{\varepsilon}'\hat{\varepsilon}$ (like SST=SSR+SSE)

β is (r+1) x m but call the jth column $\beta_{(j)}$. Then $\text{Cov}(\hat{\beta}_{(j)}, \hat{\beta}_{(k)}) = \sigma_{ik}(Z'Z)^{-1}$

2. Suppose $\varepsilon \sim N_m(0, \Sigma)$. Then Y is m-variate normal and can compute mles of β. This model is a conditional regression model. But the least squares estimates of β are the same as the mles of β. See Result 7.10 p 390 and comments in section 7.9 p 409-410. Also, mle of Σ is $n^{-1}\hat{\varepsilon}'\hat{\varepsilon}$. Not much new.

See Comment p 392. The estimation of the regression coefficients can be done one response at a time. But the regression coefficients from one component regression will be correlated with the regression coefficients from another component. This relationship is pretty complicated and not usually of practical interest. An exception might be the very simple case of r=1 and m=2 (one independent variable and two response variables). This is the case illustrated on p 398

Testing for adequacy of submodels (when trying to eliminate predictors) in the Multivariate Regression model. (Top p 393)

Just compute the estimated residual covariance matrix under full and reduced model, $\hat{\Sigma}$ and $\hat{\Sigma}_1$ respectively. Then take the ratio of determinants to form Wilks lambda statistic and use the Chi square approximation to it on p 393. (df = m(r-q) where r+1 is the rank of the full model and q+1 is the rank of the reduced model.)

Example 7.9 p 395

Z matrix is on p 373. This example is adding a response – data not shown.

We can check the rank of the full and reduced Z matrix. How?

Note here r=5, q=3, m=2, n = 18.

Exercise: Use SS matrices given on p 394 for this data to compute the four statistics on p 395. Find sources to evaluate these test statistics in this particular example.

Section 7.10 Multiple Regression with time-dependent errors.

Note that successive errors will be correlated, not independent. Need to build this into model so error can be properly estimated. See p 412.

EX: Compute the error variance for the natural gas data (part of it on p 411) ignoring the correlation over time. Compare with the time series estimate on p 412 (228.89)