Multivariate Linear Regression Models

- Regression analysis is used to predict the value of one or more responses from a set of predictors.
- It can also be used to estimate the linear association between the predictors and reponses.
- Predictors can be continuous or categorical or a mixture of both.
- We first revisit the multiple linear regression model for one dependent variable and then move on to the case where more than one response is measured on each sample unit.

Multiple Regression Analysis

- Let $z_1, z_2, ..., z_r$ be a set of r predictors believed to be related to a response variable Y.

- The linear regression model for the jth sample unit has the form

$$Y_j = \beta_0 + \beta_1 z_{j1} + \beta_2 z_{j2} + ... + \beta_r z_{jr} + \epsilon_j,$$

where ϵ is a random error and the $\beta_i, i = 0, 1, ..., r$ are unknown (and fixed) regression coefficients.

- β_0 is the intercept and sometimes we write $\beta_0 z_{j0}$, where $z_{j0} = 1$ for all j.

- We assume that

$$E(\epsilon_j) = 0, \quad \text{Var}(\epsilon_j) = \sigma^2, \quad \text{Cov}(\epsilon_j, \epsilon_k) = 0 \;\; \forall j \neq j.$$

Multiple Regression Analysis

- With n independent observations, we can write one model for each sample unit or we can organize everything into vectors and matrices so that the model is now

$$Y = Z\beta + \epsilon$$

where $\mathbf{Y}$ is $n \times 1$, Z is $n \times (r+1)$, $\boldsymbol{\beta}$ is $(r+1) \times 1$ and $\boldsymbol{\epsilon}$ is $n \times 1$.

- $\text{Cov}(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2 I$ is an $n \times n$ variance-covariance matrix for the random errors and for $\mathbf{Y}$.

- Then,

$$E(\mathbf{Y}) = Z\boldsymbol{\beta}, \quad \text{Cov}(\mathbf{Y}) = \sigma^2 I.$$

So far we have made no other assumptions about the distribution of $\boldsymbol{\epsilon}$ or $\mathbf{Y}$.

Least Squares Estimation

- One approach to estimating the vector β is to choose the value of $\boldsymbol{\beta}$ that minimizes the sum of squared residuals

$$(\mathbf{Y} - Z\boldsymbol{\beta})'(\mathbf{Y} - Z\boldsymbol{\beta})$$

- We use $\widehat{\boldsymbol{\beta}}$ to denote the least squares estimate of $\boldsymbol{\beta}$. The formula is

$$\widehat{\boldsymbol{\beta}} = (Z'Z)^{-1}Z'\mathbf{Y}.$$

- Predicted values are $\hat{Y} = Z\widehat{\beta} = H\mathbf{Y}$ where $H = Z(Z'Z)^{-1}Z'$ is called the 'hat' matrix.
- The matrix H is idempotent, meaning that $H'H = HH' = I$.

Residuals

- Residuals are computed as

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - Z\hat{\beta} = \mathbf{Y} - Z(Z'Z)^{-1}Z'\mathbf{Y} = (I - H)\mathbf{Y}.$$

- The residual sums of squares (or error sums of squares) is

$$\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = \mathbf{Y}'(I - H)'(I - H)\mathbf{Y} = \mathbf{Y}'(I - H)\mathbf{Y}$$

Sums of squares

- We can partition variability in y into variability due to changes in predictors and variability due to random noise (effects other than the predictors). The sum of squares decomposition is:

$$\underbrace{\sum_{j=1}^{n} (y_j - \bar{y})^2}_{\text{Total SS}} = \underbrace{\sum_{j} (\hat{y} - \bar{y})^2}_{SSReg.} + \underbrace{\sum_{j} \hat{\epsilon}^2}_{SSError} \quad .$$

- The coefficient of multiple determination is

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

- R^2 indicates the proportion of the variability in the observed responses that can be attributed to changes in the predictor variables.

Properties of Estimators and Residuals

- Under the general regression model described earlier and for $\widehat{\boldsymbol{\beta}} = (Z'Z)^{-1}Z'\mathbf{Y}$ we have:

$$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \text{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(Z'Z)^{-1}.$$

- For residuals:

$$E(\widehat{\boldsymbol{\epsilon}}) = 0, \quad \text{Cov}(\widehat{\boldsymbol{\epsilon}}) = \sigma^2(I - H), \quad E(\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}) = (n - r - 1)\sigma^2.$$

- An unbiased estimate of σ^2 is

$$s^2 = \frac{\widehat{\epsilon}'\widehat{\epsilon}}{n - (r + 1)} = \frac{\mathbf{Y}'(I - H)\mathbf{Y}}{n - r - 1} = \frac{SSE}{n - r - 1}$$

Properties of Estimators and Residuals

- If we assume that the $n \times 1$ vector $\epsilon \sim N_n(0, \sigma^2 I)$, then it follows that

$$\begin{aligned} Y &\sim N_n(Z\beta, \sigma^2 I) \\ \hat{\boldsymbol{\beta}} &\sim N_{r+1}(\boldsymbol{\beta}, \sigma^2 (Z'Z)^{-1}). \end{aligned}$$

- $\hat{\beta}$ is distributed independent of $\hat{\epsilon}$ and furthermore

$$\hat{\boldsymbol{\epsilon}} = (I - H)\mathbf{y} \sim N(0, \sigma^2 (I - H))$$

$$(n - r - 1)s^2 = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} \sim \sigma^2 \chi^2_{n-r-1}.$$

Confidence Intervals

- Estimating $\text{Cov}(\boldsymbol{\beta})$ as $\widehat{\text{Cov}}(\widehat{\boldsymbol{\beta}}) = s^2(Z'Z)^{-1}$, a $100(1-\alpha)\%$ confidence region for β is the set of values of β that satisfy:

$$\frac{1}{s^2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})'Z'Z(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \leq (r+1)F_{r+1,n-r-1}(\alpha),$$

  where $r+1$ is the rank of Z.

- Simultaneous confidence intervals for any number of linear combinations of the regression coefficients are obtained as:

$$c'\widehat{\boldsymbol{\beta}} \pm \sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}\sqrt{s^2c'(Z'Z)^{-1}c},$$

  These are known as Scheffe' confidence intervals.

Inferences about the regression function at z_0

- When $z = z_0 = [1,\ \ z_{01},\ \ ..., z_{0r}]'$ the response has conditional mean $E(Y_0|z_0) = z_0'\beta$
- An unbiased estimate is $\hat{\mathbf{Y}}_0 = z_0'\hat{\boldsymbol{\beta}}$ with variance $z_0'(Z'Z)^{-1}z_0\sigma^2$.
- We might be interested in a confidence interval for the mean response at $z = z_0$ or in a prediction interval for a new observation Y_0 at $z = z_0$.

Inferences about the regression function

- A $100(1-\alpha)\%$ confidence interval for $E(Y_0|z_0) = z_0'\beta$, the expected response at $z = z_0$, is given by

$$z_0'\widehat{\beta} \pm t_{n-r-1}(\alpha/2)\sqrt{z_0'(Z'Z)^{-1}z_0 s^2}.$$

- A $100(1-\alpha)\%$ confidence region for $E(Y_0|z_0) = z_0'\beta$, for all z_0 in some region is obtained form the Scheffe' method as

$$z_0'\widehat{\beta} \pm \sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}\sqrt{z_0'(Z'Z)^{-1}z_0 s^2}.$$

Inferences about the regression function at z_0

- If we wish to predict the value of a *future observation* Y_0, we need the variance of the prediction error $Y_0 - z_0'\widehat{\beta}$:

  $$\text{Var}(Y_0 - z_0'\widehat{\beta}) = \sigma^2 + \sigma^2 z_0'(Z'Z)^{-1}z_0 = \sigma^2(1 + z_0'(Z'Z)^{-1}z_0).$$

  Note that the uncertainty is higher when predicting a future observation than when predicting the mean response at $z = z_0$.

- Then, a $100(1-\alpha)\%$ prediction interval for a future observation at $z = z_0$ is given by

  $$z_0'\widehat{\beta} \pm t_{n-r-1}(\alpha/2)\sqrt{(1 + z_0'(Z'Z)^{-1}z_0)s^2}.$$

Multivariate Multiple Regression

- We now extend the regression model to the situation where we have measured m responses $Y_1, Y_2, ..., Y_p$ and the same set of r predictors $z_1, z_2, ..., z_r$ on each sample unit.

- Each response follows its own regression model:

$$\begin{aligned} Y_1 &= \beta_{01} + \beta_{11} z_1 + ... + \beta_{r1} z_r + \epsilon_1 \\ Y_2 &= \beta_{02} + \beta_{12} z_1 + ... + \beta_{r2} z_r + \epsilon_2 \\ \vdots & \quad \vdots \\ Y_p &= \beta_{0p} + \beta_{1p} z_1 + ... + \beta_{rp} z_r + \epsilon_p \end{aligned}$$

- $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_p)'$ has expectation $\mathbf{0}$ and variance matrix $\Sigma_{p \times p}$. The errors associated with different responses on the same sample unit may have different variances and may be correlated.

Multivariate Multiple Regression

- Suppose we have a sample of size n. As before, the design matrix Z has dimension $n \times (r+1)$. But now:

$$Y_{n\times m} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1p} \\ Y_{21} & Y_{22} & \cdots & Y_{2p} \\ \vdots & \vdots & \vdots\vdots & \\ Y_{n1} & Y_{n2} & \cdots & Y_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{(1)} & \mathbf{Y}_{(2)} & \cdots & \mathbf{Y}_{(p)} \end{bmatrix},$$

where $Y_{(i)}$ is the vector of n measurements of the ith variable. Also,

$$\beta_{(r+1)\times m} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rm} \end{bmatrix} = \begin{bmatrix} \beta_{(1)} & \beta_{(2)} & \cdots & \beta_{(m)} \end{bmatrix},$$

where $\beta_{(i)}$ are the $(r+1)$ regression coefficients in the model for the ith variable.

Multivariate Multiple Regression

- Finally, the p $n-$dimensional vectors of errors $\epsilon_{(i)}, i = 1, ..., p$ are also arranged in an $n \times p$ matrix

$$\epsilon = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1p} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{np} \end{bmatrix} = \begin{bmatrix} \epsilon_{(1)} & \epsilon_{(2)} & \cdots & \epsilon_{(p)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\epsilon}_1' \\ \boldsymbol{\epsilon}_2' \\ \vdots \\ \boldsymbol{\epsilon}_n' \end{bmatrix},$$

where the $p-$dimensional row vector ϵ_j' includes the residuals for each of the p response variables for the j-th subject or unit.

Multivariate Multiple Regression

- We can now formulate the multivariate multiple regression model:

$$Y_{n\times p} = Z_{n\times(r+1)}\beta_{(r+1)\times p} + \epsilon_{n\times p},$$

$$E(\epsilon_{(i)}) = 0, \quad \mathrm{Cov}(\epsilon_{(i)}, \epsilon_{(k)}) = \sigma_{ik} I, \quad i, k = 1, 2, ..., p.$$

- The m measurements on the jth sample unit have covariance matrix Σ but the n sample units are assumed to respond independently.

- Unknown parameters in the model are $\beta_{(r+1)\times p}$ and the elements of Σ.

- The design matrix Z has jth row $\begin{bmatrix} z_{j0} & z_{j1} & \cdots & z_{jr} \end{bmatrix}$, where typically $z_{j0} = 1$.

Multivariate Multiple Regression

- We estimate the regression coefficients associated with the ith response using only the measurements taken from the n sample units for the ith variable. Using Least Squares and with Z of full column rank:

$$\hat{\boldsymbol{\beta}}_{(i)} = (Z'Z)^{-1}Z'\mathbf{Y}_{(i)}.$$

- Collecting all univariate estimates into a matrix:

$$\hat{\beta} = \left[\begin{array}{cccc} \hat{\boldsymbol{\beta}}_{(1)} & \hat{\boldsymbol{\beta}}_{(2)} & \cdots & \hat{\boldsymbol{\beta}}_{(p)} \end{array} \right] = (Z'Z)^{-1}Z' \left[\begin{array}{cccc} \mathbf{Y}_{(1)} & \mathbf{Y}_{(2)} & \cdots & \mathbf{Y}_{(p)} \end{array} \right],$$

or equivalently $\hat{\beta}_{(r+1)\times p} = (Z'Z)^{-1}Z'Y$.

Least Squares Estimation

- The least squares estimator for β minimizes the sums of squares elements on the diagonal of the residual sum of squares and crossproducts matrix $(Y - Z\widehat{\beta})'(Y - Z\widehat{\beta}) =$

$$\begin{bmatrix} (Y_{(1)} - Z\widehat{\beta}_{(1)})'(Y_{(1)} - Z\widehat{\beta}_{(1)}) & \cdots & (Y_{(1)} - Z\widehat{\beta}_{(1)})'(Y_{(p)} - Z\widehat{\beta}_{(p)}) \\ (Y_{(2)} - Z\widehat{\beta}_{(2)})'(Y_{(1)} - Z\widehat{\beta}_{(1)}) & \cdots & (Y_{(2)} - Z\widehat{\beta}_{(2)})'(Y_{(p)} - Z\widehat{\beta}_{(p)}) \\ \vdots & \cdots & \vdots \\ (Y_{(p)} - Z\widehat{\beta}_{(p)})'(Y_{(1)} - Z\widehat{\beta}_{(1)}) & \cdots & (Y_{(p)} - Z\widehat{\beta}_{(p)})'(Y_{(p)} - Z\widehat{\beta}_{(p)}) \end{bmatrix},$$

  Consequently the matrix $(Y - Z\widehat{\beta})'(Y - Z\widehat{\beta})$ has smallest possible trace.

- The generalized variance $|(Y - Z\widehat{\beta})'(Y - Z\widehat{\beta})|$ is also minimized by the least squares estimator

Least Squares Estimation

- Using the least squares estimator for β we can obtain predicted values and compute residuals:

$$\begin{aligned}
\widehat{Y} &= Z\widehat{\beta} = Z(Z'Z)^{-1}Z'Y \\
\widehat{\epsilon} &= Y - \widehat{Y} = Y - Z(Z'Z)^{-1}Z'Y = [I - Z(Z'Z)^{-1}Z']Y.
\end{aligned}$$

- The usual decomposition into sums of squares and cross-products can be shown to be:

$$\underbrace{Y'Y}_{TotSSCP} = \underbrace{\widehat{Y}'\widehat{Y}}_{RegSSCP} + \underbrace{\widehat{\epsilon}'\widehat{\epsilon}}_{ErrorSSCP},$$

and the error sums of squares and cross-products can be written as

$$\widehat{\epsilon}'\widehat{\epsilon} = Y'Y - \widehat{Y}'\widehat{Y} = Y'[I - Z(Z'Z)^{-1}Z']Y$$

Propeties of Estimators

- For the multivariate regression model and with Z of full rank $r + 1 < n$:

$$E(\widehat{\beta}) = \beta, \quad \text{Cov}(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\boldsymbol{\beta}}_{(k)}) = \sigma_{ik}(Z'Z)^{-1}, \quad i, k = 1, ..., p.$$

- Estimated residuals $\widehat{\epsilon}_{(i)}$ satisfy $E(\widehat{\epsilon}_{(i)}) = 0$ and

$$E(\widehat{\boldsymbol{\epsilon}}'_{(i)}\widehat{\boldsymbol{\epsilon}}_{(k)}) = (n - r - 1)\sigma_{ik},$$

and therefore

$$E(\widehat{\epsilon}) = 0, \quad E(\widehat{\epsilon}'\widehat{\epsilon}) = (n - r - 1)\Sigma.$$

- $\widehat{\beta}$ and the residuals $\widehat{\epsilon}$ are uncorrelated.

Prediction

- For a given set of predictors $z_0' = \begin{bmatrix} 1 & z_{01} & \cdots & z_{0r} \end{bmatrix}$ we can simultaneously estimate the mean responses $z_0'\beta$ for all p response variables as $z_0'\hat{\beta}$.

- The least squares estimator for the mean responses is unbiased: $E(z_0'\hat{\beta}) = z_0'\beta$.

- The estimation errors $z_0'\hat{\boldsymbol{\beta}}_{(i)} - z_0'\boldsymbol{\beta}_{(i)}$ and $z_0'\hat{\boldsymbol{\beta}}_{(k)} - z_0'\boldsymbol{\beta}_{(k)}$ for the ith and kth response variables have covariances

$$\begin{aligned} E[z_0'(\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)})(\hat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)})'z_0] &= z_0'[E(\hat{\beta}_{(i)} - \beta_{(i)})(\hat{\beta}_{(k)} - \beta_{(k)})']z_0 \\ &= \sigma_{ik} z_0'(Z'Z)^{-1} z_0. \end{aligned}$$

Prediction

- A single observation at $z = z_0$ can also be estimated unbiasedly by $z_0'\widehat{\beta}$ but the forecast errors $(Y_{0i} - z_0'\widehat{\boldsymbol{\beta}}_{(i)})$ and $(Y_{0k} - z_0'\widehat{\boldsymbol{\beta}}_{(k)})$ may be correlated

$$\begin{aligned}
& E(Y_{0i} - z_0'\widehat{\boldsymbol{\beta}}_{(i)})(Y_{0k} - z_0'\widehat{\boldsymbol{\beta}}_{(k)}) \\
& \qquad = E(\epsilon_{(0i)} - z_0'(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)}))(\epsilon_{(0k)} - z_0'(\widehat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)})) \\
& \qquad = E(\epsilon_{(0i)}\epsilon_{(0k)}) + z_0' E(\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)})(\widehat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)}) z_0 \\
& \qquad = \sigma_{ik}(1 + z_0'(Z'Z)^{-1} z_0).
\end{aligned}$$

Likelihood Ratio Tests

- If in the multivariate regression model we assume that $\boldsymbol{\epsilon} \sim N_p(0, \Sigma)$ and if rank$(Z) = r + 1$ and $n \geq (r + 1) + p$, then the least squares estimator is the MLE of β and has a normal distribution with

$$E(\widehat{\beta}) = \beta, \quad \text{Cov}(\widehat{\beta}_{(i)}, \widehat{\beta}_{(k)}) = \sigma_{ik}(Z'Z)^{-1}.$$

- The MLE of Σ is

$$\widehat{\Sigma} = \frac{1}{n}\widehat{\epsilon}'\widehat{\epsilon} = \frac{1}{n}(Y - Z\widehat{\beta})'(Y - Z\widehat{\beta}).$$

- The sampling distribution of $n\widehat{\Sigma}$ is $W_{p,n-r-1}(\Sigma)$.

Likelihood Ratio Tests

- Computations to obtain the least squares estimator (MLE) of the regression coefficients in the multivariate multiple regression model are no more difficult than for the univariate multiple regression model, since the $\widehat{\boldsymbol{\beta}}_{(i)}$ are obtained one at a time.

- The same set of predictors must be used for all p response variables.

- Goodness of fit of the model and model diagnostics are usually carried out for one regression model at a time.

Likelihood Ratio Tests

- As in the case of the univariate model, we can construct a likelihood ratio test to decide whether a set of $r-q$ predictors $z_{q+1}, z_{q+2}, ..., z_r$ is associated with the m responses.

- The appropriate hypothesis is

$$H_0 : \beta_{(2)} = 0, \quad \text{where } \beta = \begin{bmatrix} \beta_{(1),(q+1)\times p} \\ \beta_{(2),(r-q)\times p} \end{bmatrix}.$$

- If we set $Z = \begin{bmatrix} Z_{(1),n\times(q+1)} & Z_{(2),n\times(r-q)} \end{bmatrix}$, then

$$E(Y) = Z_{(1)}\beta_{(1)} + Z_{(2)}\beta_{(2)}.$$

Likelihood Ratio Tests

- Under H_0, $Y = Z_{(1)}\beta_{(1)} + \epsilon$. The likelihood ratio test consists in rejecting H_0 if Λ is small where

$$\Lambda = \frac{\max_{\beta_{(1)},\Sigma} L(\beta_{(1)},\Sigma)}{\max_{\beta,\Sigma} L(\beta,\Sigma)} = \frac{L(\hat{\beta}_{(1)},\hat{\Sigma}_1)}{L(\hat{\beta},\hat{\Sigma})} = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_1|}\right)^{n/2}$$

- Equivalently, we reject H_0 for large values of

$$-2\ln\Lambda = -n\ln\left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_1|}\right) = -n\ln\frac{|n\hat{\Sigma}|}{|n\hat{\Sigma} + n(\hat{\Sigma}_1 - \hat{\Sigma})|}.$$

Likelihood Ratio Tests

- For large n:

$$-[n-r-1-\frac{1}{2}(p-r+q+1)] \ln\left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_1|}\right) \sim \chi^2_{p(r-q)}, \text{ approximately.}$$

- As always, the degrees of freedom are equal to the difference in the number of "free" parameters under H_0 and under H_1.

Other Tests

- Most software (R/SAS) report values on test statistics such as:

  1. Wilk's Lambda: $\Lambda^* = \frac{|\hat{\mathbf{\Sigma}}|}{|\hat{\mathbf{\Sigma}}_0|}$

  2. Pillai's trace criterion: $trace[(\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0)\hat{\mathbf{\Sigma}}^{-1}]$

  3. Lawley-Hotelling's trace: $trace[(\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0)\hat{\mathbf{\Sigma}}^{-1}]$

  4. Roy's Maximum Root test: largest eigenvalue of $\hat{\mathbf{\Sigma}}_0\hat{\mathbf{\Sigma}}^{-1}$

- Note that Wilks' Lambda is directly related to the Likelihood Ratio test. Also, $\Lambda^* = \prod_{i=1}^{p}(1 + l_i)$, where l_i are the roots of $|\hat{\mathbf{\Sigma}}_0 - l(\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0)| = 0$.

Distribution of Wilks' Lambda

- Let n_d be the degrees of freedom of $\hat{\Sigma}$, *i.e.* $n_d = n - r - 1$.

- Let $j = r - q + 1$, $r = \frac{pj}{2} - 1$, and $s = \sqrt{\frac{p^2 j^2 - 4}{p^2 + j^2 - 5}}$, and $k = n_d - \frac{1}{2}(p - j + 1)$.

- Then
$$\frac{1 - \Lambda^{*1/s}}{\Lambda^{*1/s}} \frac{ks - r}{pj} \sim F_{pj, ks-r}$$

- The distribution is exact for $p = 1$ or $p = 2$ or for $m = 1$ or $p = 2$. In other cases, it is approximate.

Likelihood Ratio Tests

- Note that

  $$(Y-Z_{(1)}\widehat{\boldsymbol{\beta}}_{(1)})'(Y-Z_{(1)}\widehat{\boldsymbol{\beta}}_{(1)})-(Y-Z\widehat{\beta})'(Y-Z\widehat{\beta}) = n(\hat{\Sigma}_1-\hat{\Sigma}),$$

  and therefore, the likelihood ratio test is also a test of *extra sums of squares and cross-products*.

- If $\hat{\Sigma}_1 \approx \hat{\Sigma}$, then the extra predictors do not contribute to reducing the size of the error sums of squares and crossproducts, this translates into a small value for $-2\ln\Lambda$ and we fail to reject H_0.

Example-Exercise 7.25

- Amitriptyline is an antidepressant suspected to have serious side effects.

- Data on Y_1 = total TCAD plasma level and Y_2 = amount of the drug present in total plasma level were measured on 17 patients who overdosed on the drug. [We divided both responses by 1000.]

- Potential predictors are:

  z_1 = gender (1 = female, 0 = male)

  z_2 = amount of antidepressant taken at time of overdose

  z_3 = PR wave measurements

  z_4 = diastolic blood pressure

  z_5 = QRS wave measurement.

Example-Exercise 7.25

- We first fit a full multivariate regression model, with all five predictors. We then fit a reduced model, with only z_1, z_2, and performed a Likelihood ratio test to decide if z_3, z_4, z_5 contribute information about the two response variables that is not provided by z_1 and z_2.

- From the output:

$$\hat{\Sigma} = \frac{1}{17}\begin{bmatrix} 0.87 & 0.7657 \\ 0.7657 & 0.9407 \end{bmatrix}, \quad \hat{\Sigma}_1 = \frac{1}{17}\begin{bmatrix} 1.8004 & 1.5462 \\ 1.5462 & 1.6207 \end{bmatrix}.$$

Example-Exercise 7.25

- We wish to test $H_0 : \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5 = 0$ against H_1 : at least one of them is not zero, using a likelihood ratio test with $\alpha = 0.05$.

- The two determinants are $|\hat{\Sigma}| = 0.0008$ and $|\hat{\Sigma}_1| = 0.0018$. Then, for $n = 17, p = 2, r = 5, q = 2$ we have:

$$-(n - r - 1 - \frac{1}{2}(p - r + q + 1)) \;\; \ln\left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_1|}\right) =$$
$$-(17 - 5 - 1 - \frac{1}{2}(2 - 5 + 2 + 1)) \;\; \ln\left(\frac{0.0008}{0.0018}\right) = 8.92.$$

Example-Exercise 7.25

- The critical value is $\chi^2_{2(5-2)}(0.05) = 12.59$. Since $8.92 < 12.59$ we fail to reject H_0. The three last predictors do not provide much information about changes in the means for the two response variables beyond what gender and dose provide.

- Note that we have used the MLE's of Σ under the two hypotheses, meaning, we use n as the divisor for the matrix of sums of squares and cross-products of the errors. We do not use the usual unbiased estimator, obtained by dividing the matrix E (or W) by $n - r - 1$, the error degrees of freedom.

- **What we have not done but should:** Before relying on the results of these analyses, we should carefully inspect the residuals and carry out the usual tests.

Prediction

- If the model $Y = Z\beta + \epsilon$ was fit to the data and found to be adequate, we can use it for prediction.

- Suppose we wish to predict the mean response at some value z_0 of the predictors. We know that

$$z_0'\widehat{\beta} \sim \mathsf{N}_p(z_0'\beta, z_0'(Z'Z)^{-1}z_0\Sigma).$$

- We can then compute a Hotelling T^2 statistic as

$$T^2 = \left(\frac{z_0'\widehat{\beta} - z_0'\beta}{\sqrt{z_0'(Z'Z)^{-1}z_0}}\right)' \left(\frac{n}{n-r-1}\widehat{\Sigma}\right)^{-1} \left(\frac{z_0'\widehat{\beta} - z_0'\beta}{\sqrt{z_0'(Z'Z)^{-1}z_0}}\right).$$

Confidence Ellipsoids and Intervals

- Then, a $100(1-\alpha)\%$ CR for $x_0'\beta$ is given by all $z_0'\beta$ that satisfy

$$(z_0'\hat{\beta} - z_0'\beta)' \left(\frac{n}{n-r-1}\hat{\Sigma}\right)^{-1} (z_0'\hat{\beta} - z_0'\beta)$$
$$\leq z_0'(Z'Z)^{-1}z_0 \left[\left(\frac{p(n-r-1)}{n-r-p}\right) F_{p,n-r-p}(\alpha)\right].$$

- The simultaneous $100(1-\alpha)\%$ confidence intervals for the means of each response $E(Y_i) = z_0'\boldsymbol{\beta}_{(i)}$, i=1,2,...,p, are

$$z_0'\hat{\boldsymbol{\beta}}_{(i)} \pm \sqrt{\left(\frac{p(n-r-1)}{n-r-p}\right) F_{p,n-r-p}(\alpha)} \sqrt{z_0'(Z'Z)^{-1}z_0 \left(\frac{n}{n-r-1}\hat{\sigma}_{ii}\right)}.$$

Confidence Ellipsoids and Intervals

- We might also be interested in predicting (forecasting) a single $p-$dimensional response at $z = z_0$ or $\mathbf{Y}_0 = z_0'\beta + \boldsymbol{\epsilon}_0$.
- The point predictor of $\mathbf{Y}_0$ is still $z_0'\hat{\beta}$.
- The forecast error

  $\mathbf{Y}_0 - z_0'\hat{\beta} = (\beta - z_0'\hat{\beta}) + \boldsymbol{\epsilon}_0$ is distributed as $\mathrm{N}_p(0, (1 + z_0'(Z'Z)^{-1}z_0)\Sigma)$.

- The $100(1-\alpha)\%$ prediction ellipsoid consists of all values of $\mathbf{Y}_0$ such that

$$(\mathbf{Y}_0 - z_0'\hat{\beta})' \left(\frac{n}{n-r-1}\hat{\Sigma}\right)^{-1} (\mathbf{Y}_0 - z_0'\hat{\beta}) \leq (1 + z_0'(Z'Z)^{-1}z_0) \left[\left(\frac{p(n-r-1)}{n-r-p}\right) F_{p,n-r-p}(\alpha)\right].$$

Confidence Ellipsoids and Intervals

- The simultaneous prediction intervals for the p response variables are

$$z_0'\widehat{\boldsymbol{\beta}}_{(i)} \pm \sqrt{\left(\frac{p(n-r-1)}{n-r-p}\right) F_{p,n-r-p}(\alpha)} \sqrt{(1 + z_0'(Z'Z)^{-1}z_0)\left(\frac{n}{n-r-1}\widehat{\sigma}_{ii}\right)},$$

  where $\widehat{\boldsymbol{\beta}}_{(i)}$ is the ith column of $\widehat{\beta}$ (estimated regression coefficients corresponding to the ith variable), and $\widehat{\sigma}_{ii}$ is the ith diagonal element of $\widehat{\Sigma}$.

Example - Exercise 7.25 (cont'd)

- We consider the reduced model with $r = 2$ predictors for $p = 2$ responses we fitted earlier.

- We are interested in the 95% confidence ellipsoid for $E(Y_{01}, Y_{02})$ for women ($z_{01} = 1$) who have taken an overdose of the drug equal to 1,000 units ($z_{02} = 1000$).

- From our previous results we know that:

$$\hat{\beta} = \begin{bmatrix} 0.0567 & -0.2413 \\ 0.5071 & 0.6063 \\ 0.00033 & 0.00032 \end{bmatrix}, \quad \hat{\Sigma} = \begin{bmatrix} 0.1059 & 0.0910 \\ 0.0910 & 0.0953 \end{bmatrix}.$$

- SAS IML code to compute the various pieces of the CR is given next.

Example - Exercise 7.25 (cont'd)

```
proc iml ; reset noprint ;
n = 17 ; p = 2 ; r = 2 ;

tmp = j(n,1,1) ;
use one ;
  read all var{z1 z2} into ztemp ;
close one ;

Z = tmp||ztemp ;
z0 = {1, 1, 1000} ;

ZpZinv = inv(Z`*Z) ; z0ZpZz0 = z0`*ZpZinv*z0 ;
```

Example - Exercise 7.25 (cont'd)

```
betahat = {0.0567 -0.2413, 0.5071 0.6063, 0.00033 0.00032} ;
sigmahat = {0.1059 0.0910, 0.0910 0.0953};

betahatz0 = betahat`*z0 ;
scale = n / (n-r-1) ;
varinv = inv(sigmahat)/scale ;

print z0ZpZz0 ;
print betahatz0 ;
print varinv ;

quit ;
```

Example - Exercise 7.25 (cont'd)

- From the output, we get: $z_0'(Z'Z)^{-1}z_0 = 0.100645$,

$$z_0'\widehat{\beta} = \begin{bmatrix} 0.8938 \\ 0.685 \end{bmatrix}, \quad \left(\frac{n}{n-r-1}\widehat{\Sigma}\right)^{-1} = \begin{bmatrix} 43.33 & -41.37 \\ -41.37 & 48.15 \end{bmatrix}.$$

- Further, $p(n-r-1)/(n-r-p) = 2.1538$ and $F_{2,13}(0.05) = 3.81$.

- Therefore, the 95% confidence ellipsoid for $z_0'\beta$ is given by all values of $z_0'\beta$ that satisfy

$$(z_0'\beta - \begin{bmatrix} 0.8938 \\ 0.685 \end{bmatrix})' \begin{bmatrix} 43.33 & -41.37 \\ -41.37 & 48.15 \end{bmatrix} (z_0'\beta - \begin{bmatrix} 0.8938 \\ 0.685 \end{bmatrix}) \leq 0.100645(2.1538 \times 3.81).$$

Example - Exercise 7.25 (cont'd)

- The simultaneous confidence intervals for each of the expected responses are given by:

$$\begin{aligned} 0.8938 &\pm \sqrt{2.1538 \times 3.81}\sqrt{0.100645 \times (17/14) \times 0.1059} \\ &= 0.8938 \pm 0.3259 \\ 0.685 &\pm \sqrt{2.1538 \times 3.81}\sqrt{0.100645 \times (17/14) \times 0.0953} \\ &= 0.685 \pm 0.309, \end{aligned}$$

  for $E(Y_{01})$ and $E(Y_{02})$, respectively.

- Note: If we had been using s_{ii} (computed as the i,ith element of E over $n - r - 1$) instead of $\widehat{\sigma}_{ii}$ as an estimator of σ_{ii}, we would not be multiplying by 17 and dividing by 14 in the expression above.

Example - Exercise 7.25 (cont'd)

- If we wish to construct simultaneous confidence intervals for a single response at $z = z_0$ we just have to use $(1 + z_0'(Z'Z)^{-1}z_0)$ instead of $z_0'(Z'Z)^{-1}z_0$. From the output, $(1 + z_0'(Z'Z)^{-1}z_0) = 1.100645$ so that the 95% simultaneous confidence intervals for forecasts $(Y_{01}, Y_{02}$ are given by

$$\begin{aligned} 0.8938 &\pm \sqrt{2.1538 \times 3.81}\sqrt{1.100645 \times (17/14) \times 0.1059} \\ &= 0.8938 \pm 1.078 \\ 0.685 &\pm \sqrt{2.1538 \times 3.81}\sqrt{1.100645 \times (17/14) \times 0.0953} \\ &= 0.685 \pm 1.022. \end{aligned}$$

- As anticipated, the confidence intervals for single forecasts are wider than those for the mean responses at a same set of values for the predictors.

Example - Exercise 7.25 (cont'd)

```
ami$gender <- as.factor(ami$gender)
library(car)
fit.lm <- lm(cbind(TCAD, drug) ~ gender + antidepressant + PR + dBP
                 + QRS, data = ami)
fit.manova <- Manova(fit.lm)

Type II MANOVA Tests: Pillai test statistic
               Df test stat approx F num Df den Df   Pr(>F)
gender          1   0.65521   9.5015      2     10 0.004873 **
antidepressant  1   0.69097  11.1795      2     10 0.002819 **
PR              1   0.34649   2.6509      2     10 0.119200
dBP             1   0.32381   2.3944      2     10 0.141361
QRS             1   0.29184   2.0606      2     10 0.178092
---
```

Example - Exercise 7.25 (cont'd)

```
C <- matrix(c(0, 0, 0, 0 , 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1)
newfit <- linearHypothesis(model = fit.lm, hypothesis.matrix = C)

Sum of squares and products for the hypothesis:
          TCAD      drug
TCAD 0.9303481 0.7805177
drug 0.7805177 0.6799484

Sum of squares and products for error:
          TCAD      drug
TCAD 0.8700083 0.7656765
drug 0.7656765 0.9407089
```

```
Multivariate Tests:
                 Df test stat approx F num Df den Df  Pr(>F)
Pillai            3 0.6038599 1.585910      6     22 0.19830
Wilks             3 0.4405021 1.688991      6     20 0.17553
Hotelling-Lawley  3 1.1694286 1.754143      6     18 0.16574
Roy               3 1.0758181 3.944666      3     11 0.03906 *
---
```

Dropping predictors

```
fit1.lm <- update(fit.lm, .~ . - PR - dBP - QRS)

Coefficients:
                TCAD        drug
(Intercept)      0.0567201  -0.2413479
gender1          0.5070731   0.6063097
antidepressant   0.0003290   0.0003243

fit1.manova <- Manova(fit1.lm)

Type II MANOVA Tests: Pillai test statistic
               Df test stat approx F num Df den Df    Pr(>F)
gender          1   0.45366   5.3974      2     13   0.01966 *
antidepressant  1   0.77420  22.2866      2     13 6.298e-05 ***
```

Predict at a new covariate

```
new <- data.frame(gender = levels(ami$gender)[2], antidepressant = 1)

predict(fit1.lm, newdata = new)
       TCAD      drug
1 0.5641221 0.365286
fit2.lm <- update(fit.lm, .~ . - PR - dBP - QRS + gender:antidepressa
fit2.manova <- Manova(fit2.lm)

predict(fit2.lm, newdata = new)
```

Drop predictors, add interaction term

```
fit2.lm <- update(fit.lm, .~ . - PR - dBP - QRS + gender:antidepressa

fit2.manova <- Manova(fit2.lm)


predict(fit2.lm, newdata = new)
       TCAD      drug
1 0.4501314 0.2600822
```

```
anova.mlm(fit.lm, fit1.lm, test = "Wilks")
Analysis of Variance Table

Model 1: cbind(TCAD, drug) ~ gender + antidepressant + PR + dBP + QRS
Model 2: cbind(TCAD, drug) ~ gender + antidepressant
  Res.Df Df Gen.var.  Wilks approx F num Df den Df Pr(>F)
1     11    0.043803
2     14  3 0.051856 0.4405    1.689      6     20 0.1755


anova(fit2.lm, fit1.lm, test = "Wilks")
Analysis of Variance Table

Model 1: cbind(TCAD, drug) ~ gender + antidepressant + gender:antidep
Model 2: cbind(TCAD, drug) ~ gender + antidepressant
```

```
  Res.Df Df Gen.var.   Wilks approx F num Df den Df   Pr(>F)
1     13    0.034850
2     14  1 0.051856 0.38945   9.4065      2     12 0.003489 **
---
```