The following results may or may not be used in the final exam.

1. For multivariate linear regression model: The LSE is $\hat{\beta} = (Z'Z)^{-1}Z'Y$. The estimator of Σ is $S = 1/(n-r-1)\hat{\epsilon}'\hat{\epsilon}$, where $\hat{\epsilon} = Y - Z\hat{\beta}$ is $n \times m$ matrix.

$$\hat{\beta}_{(j)} - \beta_{(j)} \sim MN(0,\ \sigma_{jj}(Z'Z)^{-1})$$

where σ_{kl} is the (k,l)-th element of the $m \times m$ matrix Σ, which is the variance of the error.

$$(\hat{\beta}_{(j)} - \beta_{(j)})'(Z'Z)(\hat{\beta}_{(j)} - \beta_{(j)})/s_{jj} \sim (r+1)F_{r+1,\,n-r-1}$$

where s_{kl} is the (k,l)-th element of S.

For prediction,

$$Y_0 - \hat{\beta}'z_0 \sim MN\Big(0,\ \Sigma(1 + z_0'(Z'Z)^{-1}z_0)\Big).$$

$$T^2 \equiv \frac{[(\hat{\beta}-\beta)'z_0]'}{\sqrt{z_0'(Z'Z)^{-1}z_0}} S^{-1} \frac{[(\hat{\beta}-\beta)'z_0]}{\sqrt{z_0'(Z'Z)^{-1}z_0}} \sim \frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}$$

Confidence region for $\beta' z_0 = E(Y_0)$ at confidence level $1-\alpha$:

$$\Big\{\beta' z_0 \in R^m : (\hat{\beta}'z_0 - \beta'z_0)'S^{-1}(\hat{\beta}'z_0 - \beta'z_0) \le z_0'(Z'Z)^{-1}z_0 \frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}(\alpha)\Big\}.$$

2. Simultaneous confidence intervals for $\beta'_{(k)}z_0 = E(Y_{0k}), k = 1, ..., m$ at (nominal) confidence level $1-\alpha$:

$$\text{for } \beta'_{(k)}z_0 : \quad \hat{\beta}'_{(k)}z_0 \pm \Big(\frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}(\alpha) z_0'(Z'Z)^{-1}z_0 s_{kk}\Big)^{1/2};\ k = 1, ..., m.$$

3. Suppose $X \sim MN(\mu, \Sigma)$. Then, the k-th (population) principal component is a linear combination of X with the coefficients being $\mathbf{e}_k$ and variance being λ_k, where $(\lambda_1, \mathbf{e}_1), ..., (\lambda_p, \mathbf{e}_p)$ are eigenvalue-eigenvector pairs of Σ such that $\lambda_1 \ge ... \ge \lambda_p$.

$$\sqrt{n}\Big[\begin{pmatrix} \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_p \end{pmatrix} - \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_p \end{pmatrix}\Big] \to MN(0, 2\Lambda^2),$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix}.$$

4. The population canonical variates (U_i, V_i) are

$$U_i = \mathbf{e}_i'\Sigma_{11}^{-1/2}\mathbf{X}^{(1)}, \qquad V_i = \mathbf{f}_i'\Sigma_{22}^{-1/2}\mathbf{X}^{(2)},$$

$$\rho_i = corr(U_i, V_i),$$

where

$$(\rho_1^2, \mathbf{e}_1), ..., (\rho_p^2, \mathbf{e}_p)$$

are the p eigenvalue-eigenvector pairs of the $p \times p$ matrix $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$ such that $\rho_1^2 \ge \cdots \ge \rho_p^2 > 0$. Moreover, let $\mathbf{f}_k = 1/\rho_k \Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\mathbf{e}_k, k = 1, ..., p$. Then,

$$(\rho_1^2, \mathbf{f}_1), \cdots, (\rho_p^2, \mathbf{f}_p), (0, \mathbf{f}_{p+1}), ..., (0, \mathbf{f}_q)$$

are q eigenvalue-eigenvector pairs of the $q \times q$ matrix $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}$.

5. The matrices of errors are

$$S_{11} - \sum_{i=1}^{r} \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{a}}^{(i)})' = \sum_{i=r+1}^{p} \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{a}}^{(i)})'$$

$$S_{22} - \sum_{i=1}^{r} \hat{\mathbf{b}}^{(i)}(\hat{\mathbf{b}}^{(i)})' = \sum_{i=r+1}^{q} \hat{\mathbf{b}}^{(i)}(\hat{\mathbf{b}}^{(i)})'$$

$$S_{12} - \sum_{i=1}^{r} \hat{\rho}_i \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{b}}^{(i)})' = \sum_{i=r+1}^{p} \hat{\rho}_i \hat{\mathbf{a}}^{(i)}(\hat{\mathbf{b}}^{(i)})'$$

where

$$\hat{\mathbf{A}}^{-1} = S_{11}^{1/2}(\hat{\mathbf{e}}_1 \vdots \cdots \vdots \hat{\mathbf{e}}_p) = (\hat{\mathbf{a}}^{(1)} \vdots \cdots \vdots \hat{\mathbf{a}}^{(p)})$$

$$\hat{\mathbf{B}}^{-1} = S_{22}^{1/2}(\hat{\mathbf{f}}_1 \vdots \cdots \vdots \hat{\mathbf{f}}_q) = (\hat{\mathbf{b}}^{(1)} \vdots \cdots \vdots \hat{\mathbf{b}}^{(q)}).$$

6. The expected cost of classification is

$$ECM = c(2|1)p(2|1)p_1 + c(1|2)p(1|2)p_2$$

and the optimal classification rule which minimizing the ECM has

$$R_1 = \left\{ x : \frac{f_1(x)}{f_2(x)} \geq \frac{c(1|2)p_2}{c(2|1)p_1} \right\}$$

and $R_2 = R_1^c$. If f_1 is the density of $MN(\mu_1, \Sigma)$ and f_2 is the density of $MN(\mu_2, \Sigma)$. Then, the optimal classification rule

$$R_1 = \left\{ x : (\mu_1 - \mu_2)'\Sigma^{-1}x - \frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) \geq \log\left[\frac{c(1|2)p_2}{c(2|1)p_1}\right] \right\}$$

and $R_2 = R_1^c$.