**STATISTICS 575 SAMPLE MIDTERM EXAM**

*Instructions:* Answer all four equally weighted questions.
*Time:* 80 minutes.

**Some possibly useful formulae:**

1. Under the usual normality assumptions,

$$T^2 = (\bar{\mathbf{x}}-\boldsymbol{\mu})^T \left(\widehat{cov\,[\bar{\mathbf{x}}]}\right)^{-1} (\bar{\mathbf{x}}-\boldsymbol{\mu}) \sim dfS\frac{df1}{df2}F^{df1}_{df2},$$

where the Wishart distribution associated with the (unbiased) covariance estimate is $W_{df1}(dfS, \cdot)$ and $df1 + df2 = dfS + 1$.

2. Using the usual notation and assumptions, the conditional mean and covariance matrix in the distribution of $\mathbf{x}_1$ given $\mathbf{x}_2$ are $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ and $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

3. The characteristic function of the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution is

$$E[e^{i\mathbf{t}^T\mathbf{x}}] = \exp\{i\mathbf{t}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}\}.$$

4. In the notation introduced in class, the likelihood function for a single $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ sample is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{n}{2}\left[tr\boldsymbol{\Sigma}^{-1}\mathbf{S}_n + (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})\right]\right\}.$$

---

1. (a) Define 'multivariate normal distribution'.

   (b) Let $(X_1, X_2, X_3)$ be multivariate normal. Suppose that $X_1$ and $X_2$ are independent. Does it follow that $X_1$ and $X_2$ are conditionally independent, given $X_3$? Give a brief proof or counterexample.

   (c) Let $\mathbf{y}$ be a $p$-dimensional random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Let $\mathbf{A}$ be a fixed matrix with $p$ columns. State the definition of 'covariance matrix' in the form $E[\cdots]$, and then derive an expression for the covariance matrix of $\mathbf{A}\mathbf{y}$. Is it necessary to assume that $\mathbf{y}$ is multivariate normal?

   (d) Referring to (c), suppose that $\boldsymbol{\mu} = \mathbf{0}$ and that $\boldsymbol{\Sigma}$ has rank $p$. Derive the distribution of $\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}$. Is it necessary to assume that $\mathbf{y}$ is multivariate normal?

.......over

2. Measurements of $X_1$ = petal length (mm) and $X_2$ = petal width (mm) were obtained for samples from three species of a flower. The following summary statistics were calculated:

| | species 1 | species 2 | species 3 |
|---|---|---|---|
| sample size | 10 | 8 | 16 |
| sample mean | $\begin{pmatrix} 41 \\ 9 \end{pmatrix}$ | $\begin{pmatrix} 50 \\ 21 \end{pmatrix}$ | $\begin{pmatrix} 20 \\ 5 \end{pmatrix}$ |
| sample covariance | $\begin{bmatrix} 20 & 7 \\ 7 & 5 \end{bmatrix}$ | $\begin{bmatrix} 30 & 9 \\ 9 & 10 \end{bmatrix}$ | $\begin{bmatrix} 25 & 6 \\ 6 & 11 \end{bmatrix}$ |

(a) Introduce appropriate notation, and then formulate the null hypothesis of no difference between the three species.

(b) Describe in detail how you would carry out the likelihood ratio test of the hypothesis in a) using the given summary statistics . Your answer should include *i)* a description of how any relevant quantities would be calculated (actual calculations not necessary) and *ii)* a description of the manner in which a large-sample approximation to the $p$-value would be calculated.

3. For the same situation and data as in the preceding question, consider the null hypothesis of no difference between the *first two* species. Describe in detail a standard test statistic for this null hypothesis. Your test should incorporate information from the covariance matrix from the third sample. Describe how the $p$-value can be calculated from $F$-tables.

4. Consider the growth curve model, in which the sample data are

$$\{\mathbf{x}_{lj} \mid j = 1, ..., n_l; l = 1, ..., g\},$$

with $\mathbf{x}_{lj} \sim N_p(\mathbf{B}\boldsymbol{\beta}_l, \boldsymbol{\Sigma})$ (all independently distributed). We wish to compare the curves in varying groups ($l = 1, ..., g$). Show that the maximum likelihood estimates of the parameters $\boldsymbol{\beta}_l$ are

$$\hat{\boldsymbol{\beta}}_l = \left(\mathbf{B}^T\hat{\boldsymbol{\Sigma}}\mathbf{B}\right)^{-1}\mathbf{B}^T\hat{\boldsymbol{\Sigma}}\bar{\mathbf{x}}_l,$$

where $\hat{\boldsymbol{\Sigma}}$ is the mle for $\boldsymbol{\Sigma}$ (which you are NOT being asked to derive).