STATISTICS 575 SAMPLE FINAL EXAM

Instructions: Answer any four of these six equally weighted questions. If you answer more than four then the best four will count..

Time: 3 hours.

Some possibly useful formulae:

1. Under the usual normality assumptions,

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

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

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

4. $vec\,(\mathbf{ABC}) = \left(\mathbf{A} \otimes \mathbf{C}^T\right) vec\mathbf{B}$.

5. In the usual notation, the likelihood ratio procedure for MANOVA and multivariate regression leads to the use of

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|}.$$

The large-sample approximation to the null distribution is that $-n\log\Lambda^*$ is chi-squared, on the appropriate degrees of freedom.

...over

Questions:

1. Consider the multivariate regression model

$$\mathbf{Y}_{n\times m} = \mathbf{Z}_{n\times(r+1)}\mathbf{B}_{(r+1)\times m} + \mathbf{E}_{n\times m},$$

in which the rows of $\mathbf{E}$ are i.i.d. $N_m(\mathbf{0}, \mathbf{\Sigma})$ vectors and $\mathbf{Z}$ has full column rank. The mle of $\mathbf{B}$ is $\hat{\mathbf{B}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{Y}$.

   (a) Define $\mathbf{C} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{1/2}\left(\hat{\mathbf{B}} - \mathbf{B}\right)$; determine the distribution of $\mathbf{C}^T\mathbf{C}$.

   (b) Describe how you would obtain a $100(1-\alpha)\%$ prediction region for a *new* random vector which will be observed with the regressors set at a particular level $\mathbf{z}_*$. Be very explicit as to what would be computed, degrees of freedom, etc.

2. Suppose that you are given two random vectors $\mathbf{x}_{p\times 1}, \mathbf{y}_{q\times 1}$ $(p \leq q)$ with covariance

$$\mathrm{COV}\left[\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}\right] = \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}.$$

   (a) State the Singular Value Decomposition Theorem.

   (b) Let $\tilde{\boldsymbol{x}} = \mathbf{\Sigma}_{11}^{-1/2}\mathbf{x}$ and $\tilde{\boldsymbol{y}} = \mathbf{\Sigma}_{22}^{-1/2}$ be the standardized versions of $\mathbf{x}$ and $\mathbf{y}$, and consider the problem of finding linear combinations $\mathbf{a}^T\tilde{\boldsymbol{x}}, \mathbf{b}^T\tilde{\boldsymbol{y}}$, with $\|\mathbf{a}\| = \|\mathbf{b}\| = 1$, with maximum correlation. Derive the forms of $\mathbf{a}$ and $\mathbf{b}$. How can the maximum (absolute) correlation be computed, without a knowledge of $\mathbf{a}$ and $\mathbf{b}$?

   (c) Suppose now that $p = 1$. Let R denote the maximum correlation. Show that $1 - R^2$ can be expressed as a conditional variance divided by an unconditional variance.

3. Consider the orthogonal factor model with m common factors:

$$\begin{aligned} \mathbf{x}_{p\times 1} &= \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + \boldsymbol{\varepsilon}, \\ \mathrm{COV}[\mathbf{x}] &= \mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}. \end{aligned}$$

   (a) Suppose that you are given sample values $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with average $\bar{\mathbf{x}}$ and covariance $\mathbf{S}$. Describe how you would estimate the matrix of factor loadings, the communalities and the specific variances. Use the Principal Component method applied to the *covariance* matrix.

   (b) Suppose that the factor scores $\mathbf{f}_j$, $j = 1, \ldots, n$ are to be estimated by the Principal Component (i.e. Ordinary Least Squares) method. Derive the relationship between the $\hat{\mathbf{f}}_j$ and the sample principal components.

...over

4. A p-dimensional random vector $\mathbf{x}$ is known to be distributed as $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ with probability q_i, $i = 1, 2, 3$. $(\sum q_i = 1)$.

   (a) Consider a linear combination $\mathbf{c}^T\mathbf{x}$, and let α_i and σ^2 denote the mean and variance of $\mathbf{c}^T\mathbf{x}$ when the distribution of $\mathbf{x}$ is $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. Let us say that the choice $\mathbf{c}$ which best discriminates between the populations is the one which maximizes $Q = \sum_{i=1}^{3} q_i (\alpha_i - \bar{\alpha})^2 / \sigma^2$, where $\bar{\alpha} = \sum_{i=1}^{3} q_i \alpha_i$. Show that the best choice of $\mathbf{c}$ is the eigenvector, corresponding to the maximum eigenvalue, of a matrix of the form $\boldsymbol{\Sigma}^{-1}\mathbf{G}$. Exhibit the matrix $\mathbf{G}$.

   (b) Describe in detail how you would implement the procedure described in (a), given samples from each of the three populations. Your method should address the fact that the $\{q_i\}$ are unknown, as are the other parameters.

   (c) Given a new observation $\mathbf{x}$ from one of the three populations, explain how you would classify it using your procedure in (b). Explain briefly why your method is sensible.

5. Describe two common algorithms which can be employed to carry out <u>agglomerative</u> clustering. One method should yield clusters with no particular structure; the other should yield clusters with an elliptical structure.

6. Let $\mathbf{x} = (X_1, X_2, X_3)^T$ represent three physiological measurements obtained from a randomly sampled subject. Suppose that $\mathbf{x}$ has covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

   (a) Write out the principal components of $\mathbf{x}$ as linear combinations of the three measurements.

   (b) Evaluate the variance of each principal component in (a).

   (c) Do you think the first principal component provides an adequate approximation of the joint variation among the three measurements? Explain.

   (d) Exhibit two <u>uncorrelated</u> linear functions of <u>all three</u> elements of $\mathbf{x}$, each of which has a variance of 1. By 'all three' it is meant that all three coefficients must be non-zero.