THE UNIVERSITY OF CHICAGO
Graduate School of Business
Business 41912, Spring Quarter 2006, Mr. Ruey S. Tsay

Solutions to Midterm

1. Suppose that $\boldsymbol{X} = (X_1, X_2, X_3)'$ follows a 3-dimensional normal distribution with mean $\boldsymbol{\mu} = (1, 2, 3)'$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Note that

$$\boldsymbol{\Sigma} = 0.308\boldsymbol{e}_1\boldsymbol{e}_1' + 0.6431\boldsymbol{e}_2\boldsymbol{e}_2' + 5.0489\boldsymbol{e}_3\boldsymbol{e}_3',$$

where

$$[\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3] = \begin{bmatrix} 0.591 & 0.737 & 0.??? \\ -0.737 & 0.328 & 0.591 \\ 0.328 & -0.591 & 0.??? \end{bmatrix}.$$

Answer the following equations:

(a) What is the distribution of $\boldsymbol{Z} = (X_1 - X_2, X_2 - X_3)'$?
A: Note that $\boldsymbol{Z} = \boldsymbol{C}\boldsymbol{X}$, where

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}.$$

Using properties of multivariate normal distribution,

$$\boldsymbol{Z} \sim N_2\left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}, \boldsymbol{I}_2\right).$$

(b) What is the distribution of X_2 given $X_3 = 2$?
A: Focus on the joint distribution of $(X_2, X_3)'$. It is easy to see that $X_2|X_3 = 2$ is univariate normal with mean $\frac{4}{3}$ and variance $\frac{2}{3}$.

(c) Find a linear combination of $\boldsymbol{X}$ with length 1 that has the maximum variance among all possible linear combinations.
A: Make use of the eigen-value and eigen-vector decomposition. First, the third eigenvector is easily seen to be $(0.328, .591, .737)'$. Second, from the eigenvalues, the linear combination is $Y = 0.329X_1 + 0.591X_2 + 0.737X_3$ and the associated variance is 5.0489.

(d) Find the coefficient $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ of the least squares regression

$$X_3 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

A: First, $E(X_3) = \beta_0 + \beta_1 E(X_1) + \beta_2 E(X_2)$. Second, $(\beta_1, \beta_2)'$ can be obtained by

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = [\text{Cov}((X_1, X_2)')]^{-1}\text{Cov}((X_1, X_2)', X_3) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Consequently, $\boldsymbol{\beta} = (1, 0, 1)'$.

2. Amitriptyline is prescribed by some physicians as an antidepressant. However, there are also conjectured side effects that seem to be related to the use of the drug: irregular heartbeat, abnormal bllod pressures, and irregular waves on the electrocardiogram, among other things. Data gathered on 17 patients who were admitted to the hospital after an amitriptyline overdose are collected. The two response variables are (a) Y_1 = total TCAD plasma level and (b) Y_2 = amount of amitriptyline present in TCAD plasma level. The five predictor variables are

- Z_1 = Gender; 1 if female, 0 if male.
- Z_2 = Amount of antidepressants taken at time of overdose.
- Z_3 = PR wave measurement.
- Z_4 = Diastolic blood pressure.
- Z_5 = QRS wave measurement.

We perform a multivariate multiple linear regression

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

to analyze the data, where the first column of $\boldsymbol{Z}$ is a vector of 1. R output is attached. Answer the following questions.

(a) Write down the fitted model.

A: From the output, we obtain

$$\boldsymbol{\beta} = \begin{bmatrix} -2879.5 & -2728.7 \\ 675.65 & 763.03 \\ 0.284 & 0.306 \\ 10.27 & 8.90 \\ 7.25 & 7.21 \\ 7.60 & 4.99 \end{bmatrix}.$$

(b) What are the standard error and t-ratio of the least squares estimate of $\beta_{2,1}$, the $(2,1)$th element of $\boldsymbol{\beta}$?

A: $\sqrt{26262.03} = 162.06$. The square-root of the (2,2)th element of the Kronecker product. The t-ratio $= \frac{675.65}{162.06} = 4.17$.

(c) Construct a 95% simultaneous prediction intervals for the two individual responses Y_{0i} at $\boldsymbol{z}_0 = (1, 1, 1200, 140, 70, 85)'$.

A: Based on the formula (7.49) of the textbook, the 95% C.I. are

$$\boldsymbol{z}_0'\boldsymbol{\beta} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m}(0.05)} \sqrt{(1 + \boldsymbol{z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_0)\left(\frac{n\hat{\sigma}_{ii}}{n-r-1}\right)}.$$

Therefore, for 1st component

$$729.52 \pm \sqrt{\frac{2(17-5-1)}{17-5-2} 4.103} \sqrt{1.236 \frac{17 \times 79091.66}{11}},$$

$$i.e., \quad 729.53 \pm 1167.79 \Rightarrow (-438.27, 1897.31)$$

For the 2nd component

$$575.73 \pm \sqrt{\frac{2(17-5-1)}{17-5-2} 4.103} \sqrt{1.236 \frac{17 \times 85518.99}{11}},$$

$$i.e., \quad 575.73 \pm 1214.31 \Rightarrow (-638.58, 1790.04)$$

(d) Perform a likelihood ratio test for the null hypothesis that Z_5 can be dropped from the multivariate linear regression at the 5% level. Draw your conclusion.

A: Using Result 7.11 of the textbook, the likelihood ratio stattistic is

$$-2\ln(\Lambda) = -17 \ln\left(\frac{803335966}{1134405317}\right) = 5.87.$$

Compared with $\chi^2_2(0.05) = 5.99$. We cannot reject the null hypothesis of zero coefficients at the 5% level. Since the sample size is small, asymptotic result might not hold.

3. Consider Problem 6.26 of the textbook that concerning how consumers in Green Bay, Wisconsin would react to an electrical time-of-use pricing scheme. The measurement is $Y = \log(\text{current consumption}) - \log(\text{baseline consumption})$. The summary statistics of the experiment are given below. Test group: $n_1 = 28$, and $\bar{\boldsymbol{X}}_1 = (.153, -.231, -.322, -.339)'$. Control group: $n_2 = 58$, $\bar{\boldsymbol{X}}_2 = (.151, .180, .256, .257)'$. The pooled covariance matrix is

$$\boldsymbol{S}_{\text{pooled}} = \begin{bmatrix} .804 & .355 & .228 & .232 \\ .355 & .722 & .233 & .199 \\ .228 & .233 & .592 & .239 \\ .232 & .199 & .239 & .479 \end{bmatrix}$$

Perform a profile analsyis of the data. Does time-of-use pricing make any difference in electrical consumption at the 5% level? What is the nature of the difference, if any?

A: For the profile analysis, let

$$\boldsymbol{C} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

The T^2 for testing parallel profiles is

$$T^2 = (\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{C}'[((1/n_1) + (1/n_2))\boldsymbol{C}\boldsymbol{S}_{pooled}C']^{-1}\boldsymbol{C}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) = 9.23,$$

The critical value is

$$c^2 = \frac{(n_1 + n_2 - 2)(p-1)}{n_1 + n_2 - p} F_{p-1, n_1+n_2-p}(\alpha) = 8.36.$$

Thus, reject the null hypothesis. In other words, the profiles are not parallel. Consequently, there is no need to perform further tests.

The time-of-use pricing reduced the usage of the last three measurements. However, the first measurement (at 9 a.m.) was hardly changed. Thus, the first measurement essentially rejects the hypothesis of parallel profiles.

4. Eight mean received a certain drug. The changes in blood sugar and blood pressure (both systolic and diastolic) are collected. The sample mean of the changes are (31.25,−0.75,3.125) and the sample covariance matrix and its inverse are given below

$$\boldsymbol{S} = \begin{bmatrix} 1069.64 & 82.5 & 16.96 \\ 82.5 & 17.36 & 6.39 \\ 16.96 & 6.39 & 4.70 \end{bmatrix}, \quad \boldsymbol{S}^{-1} = \begin{bmatrix} .00167 & -.01143 & .00954 \\ -.01143 & .19403 & -.22277 \\ .00954 & -.22277 & .48170 \end{bmatrix}.$$

Answer the following questions:

(a) Test the null hypothesis $H_o : \boldsymbol{\mu} = \boldsymbol{0}$ versus the alternative hypothesis $H_a : \boldsymbol{\mu} \neq \boldsymbol{0}$ at the 5% level. Draw your conclusion.

A: The Hotelling T^2 statistic is

$$T^2 = n\boldsymbol{\mu}'\boldsymbol{S}^{-1}\boldsymbol{\mu} = 79.1.$$

The critical value is

$$c = \frac{(n-1)p}{n-p} F_{p,n-p}(0.05) = 22.72.$$

Thus, we reject the null hypothesis.

(b) Construct 95% simutaneous confidence intervals for the three changes.

A: The 95% simultaneous C.I.s are

$$\bar{x}_i \pm \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p}(0.05)} \sqrt{\frac{s_{ii}}{n_i}}.$$

That is,

$$31.25 \pm 4.77\sqrt{1069.64/8}; -0.75 \pm 4.77\sqrt{17.36/8}; 3.125 \pm 4.77\sqrt{4.7/8}.$$

$$31.25 \pm 55.16; \quad -0.75 \pm 7.03; \quad 3.125 \pm 3.66.$$

(c) Construct 95% Bonferroni simultaneous confidence intervals for the three changes.

A: The 95% Bonferroni simultaneous C.I.s are

$$\bar{x}_i \pm t_{n-1}(0.05/6)\sqrt{\frac{s_{ii}}{n_i}}.$$

That is,

$$31.25 \pm 4.77\sqrt{1069.64/8}; -0.75 \pm 3.13\sqrt{17.36/8}; 3.125 \pm 3.13\sqrt{4.7/8}.$$

$$31.25 \pm 36.19; \quad -0.75 \pm 4.61; \quad 3.125 \pm 2.40.$$