Exam 2, 22S:161, Fall 2009

NAME: SOLUTIONS

This exam has 5 questions; most have several parts. You may use your "formula sheet," a table of percentage points of the t-distribution, a table of upper 5% points of the F distribution, and a calculator. All work should be done on the examination paper. Use backs of facing pages if you need extra room. Show your work as clearly and neatly as possible to receive partial credit. You have 90 minutes for this exam. Good luck!

Points awarded:

1. 18/18
2. 15/15
3. 25/25
4. 22/22
5. 20/20

100/100

1. (18 pts.) Consider an experimental situation in which 4 variables were measured on each of 40 individuals. The ~~38~~ 40 individuals each received one of 5 treatments, as follows:

   8 individuals received Treatment 1
   10 individuals received Treatment 2
   9 individuals received Treatment 3
   7 individuals received Treatment 4
   6 individuals received Treatment 5

   $p = 4$
   $g = 5$

   A one-way MANOVA was carried out for the data. Let Λ^* represent Wilks' lambda statistic for testing the null hypothesis H_0 of no treatment effects (against the alternative hypothesis H_1 that at least one of the treatments has an effect on at least one of the variables).

   (a) Suppose $\Lambda^* = 0.30$. Carry out an approximate size-.05 test of H_0 versus H_1.

   $$-\left(n-1-\frac{p+g}{2}\right)\ln \Lambda^* = -\left(40-1-\frac{4+5}{2}\right)\ln 0.30$$
   $$= 41.54$$

   Compare to $\chi^2_{.05,\, p(g-1)} = \chi^2_{.05,16} = 26.30$

   Since $41.54 > 26.30$, reject H_0.

   (b) Suppose that Treatment 1 is a "control" treatment and the other 4 treatments are newly developed "experimental" treatments. Suppose that one important objective of this experiment was to determine if the effects of the four experimental treatments were different than the effect of the control treatment. Thus, simultaneous confidence intervals for differences between the control treatment effect and the 4 experimental treatment effects would be of interest. That is, we would be interested in obtaining a set of intervals for the differences

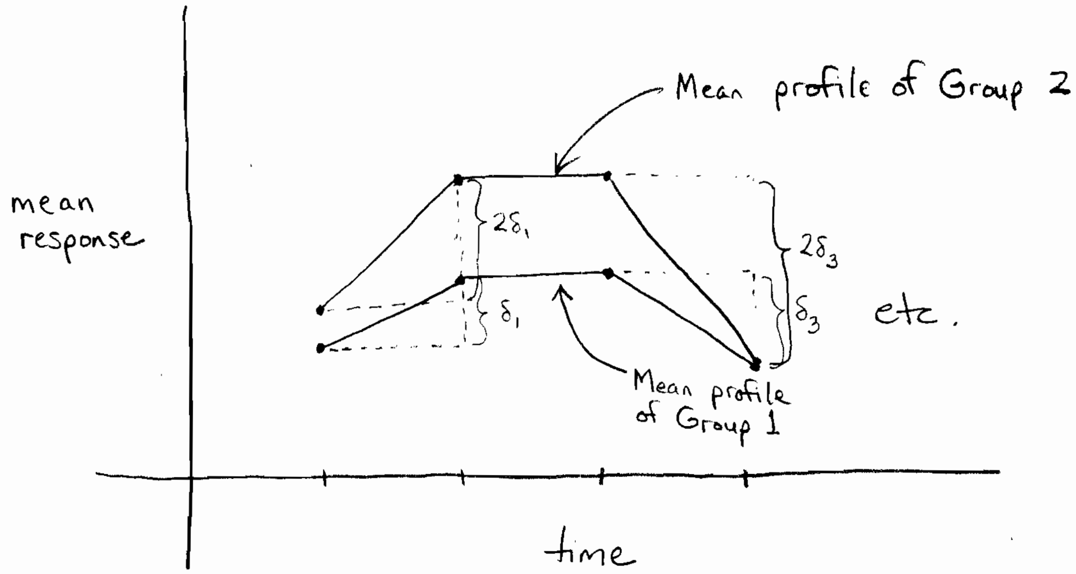
   $$\tau_{1i} - \tau_{\ell i} \quad (\ell = 2, 3, 4, 5;\ i = 1, 2, 3, 4)$$

   whose simultaneous coverage probability is at least $1-\alpha$. Give a mathematical expression for such intervals, in as simplified a form as possible.

   $$(\bar{X}_{1i} - \bar{X}_{\ell i}) \pm t_{\alpha/32,\, 35}\sqrt{\frac{W_{ii}}{35}\left(\frac{1}{6} + \frac{1}{n_\ell}\right)}$$

   $$(\ell = 2, 3, 4, 5;\ i = 1, 2, 3, 4)$$

2. (15 pts.) Consider a profile analysis involving individuals from two groups that are repeatedly measured over time with respect to some characteristic. Suppose we wish to test, as a null hypothesis, whether the change in Group 2's mean profile over successive occasions is always twice as large as the change in Group 1's mean profile over successive occasions. The following picture conveys the idea:



Give an expression for a Hotelling's T^2-statistic that would be appropriate for testing for this null hypothesis against an unrestricted alternative hypothesis. (Hint: First write out the null hypothesis in terms of the elements of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$.)

$$\left.\begin{aligned} H_0:\ & 2(\mu_{12}-\mu_{11}) = \mu_{22}-\mu_{21} \\ & 2(\mu_{13}-\mu_{12}) = \mu_{23}-\mu_{22} \\ & \vdots \\ & 2(\mu_{1p}-\mu_{1,p-1}) = \mu_{2p}-\mu_{2,p-1} \end{aligned}\right\} \text{ i.e. } H_0: 2\underset{\sim}{C}\underset{\sim}{\mu}_1 = \underset{\sim}{C}\underset{\sim}{\mu}_2$$

$$\text{where } C = \begin{pmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

Equivalently, $H_0: \underset{\sim}{C}(2\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) = \underset{\sim}{0}$.

So,

$$T^2 = (2\bar{\underset{\sim}{X}}_1 - \bar{\underset{\sim}{X}}_2)'\underset{\sim}{C}'\left[\left(\frac{4}{n_1}+\frac{1}{n_2}\right)\underset{\sim}{C}\underset{\sim}{S}_{pooled}\underset{\sim}{C}'\right]^{-1}\underset{\sim}{C}(2\bar{\underset{\sim}{X}}_1 - \bar{\underset{\sim}{X}}_2)$$

$$\left(\text{since } \operatorname{var}\left[\underset{\sim}{C}(2\bar{\underset{\sim}{X}}_1 - \bar{\underset{\sim}{X}}_2)\right] = \underset{\sim}{C}\left(\frac{4}{n_1}\underset{\sim}{\Sigma} + \frac{1}{n_2}\underset{\sim}{\Sigma}\right)\underset{\sim}{C}' = \left(\frac{4}{n_1}+\frac{1}{n_2}\right)\underset{\sim}{C}\underset{\sim}{\Sigma}\underset{\sim}{C}'\right)$$

3. (25 pts.) Suppose $\mathbf{X}$ is a 4×1 random vector with correlation matrix

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & -\rho \\ 0 & 0 & -\rho & 1 \end{bmatrix}, \quad \text{where } 0 < \rho < 1.$$

The eigenvalues of $\boldsymbol{\rho}$ are the solutions to the characteristic equation

$$[(1-\lambda)^2 - \rho^2]^2 = 0.$$

Four orthonormal eigenvectors of $\boldsymbol{\rho}$ are, in no particular order,

$$\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}.$$

(a) Determine the first two principal components (based on the correlation matrix).

$$[(1-\lambda)^2-\rho^2]^2=0 \Leftrightarrow (1-\lambda)^2-\rho^2=0$$
$$\Leftrightarrow 1-\lambda-\rho=0 \quad \text{and} \quad 1-\lambda+\rho=0$$
$$\Leftrightarrow \lambda=1-\rho \quad \text{and} \quad \lambda=1+\rho.$$

Since $0<\rho<1$, the eigenvalues in order (largest to smallest) are $1+\rho$, $1+\rho$, $1-\rho$, $1-\rho$.

$$\begin{pmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & -\rho \\ 0 & 0 & -\rho & 1 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \\ 0 \end{pmatrix} = (1+\rho)\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & -\rho \\ 0 & 0 & -\rho & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} = (1+\rho)\begin{pmatrix} 0 \\ 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}.$$

So we may take the first 2 principal components to be

$$\frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{2}} Z_2$$

and

$$\frac{1}{\sqrt{2}} Z_3 - \frac{1}{\sqrt{2}} Z_4$$

(where the Z_i's are the standardized variables)

(b) Explain, in 15 words or less, why no intuitive meaning should be ascribed to the first two principal components in this situation.

Because the eigenvectors associated with $\lambda_1 = \lambda_2 = 1+\rho$ are not unique.

(c) Determine an expression (in terms of ρ) for the percentage of total population variance that is explained by the first two principal components.

$$\frac{2(1+\rho)}{4} \times 100\% = \left(\frac{1+\rho}{2}\right) \times 100\%$$

(d) How, if at all, would your answers to parts (a) and (c) change if the range of values for ρ was $-1 < \rho < 0$ rather than $0 < \rho < 1$?

The eigenvalues in order would be

$1-\rho,\ 1-\rho,\ 1+\rho,\ 1+\rho$.

The first 2 principal components would be

$$\frac{1}{\sqrt{2}} Z_1 - \frac{1}{\sqrt{2}} Z_2$$

and

$$\frac{1}{\sqrt{2}} Z_3 + \frac{1}{\sqrt{2}} Z_4,$$

although these would not be unique either.

The answer to (c) would be $\left(\frac{1-\rho}{2}\right) \times 100\%$.

4. (22 pts.) An investigation was conducted to study two spectral reflectance measures, *nm560* and *nm720*, for three species of trees (*species* = *SS*, *JL*, *LP*) at three different times (*time* = 1,2,3) during the growing season. Four replications of each spectral reflectance measure were taken at each combination of species and time. The data are listed on the first and second pages of the accompanying SAS output.

   On subsequent pages of the output, some results of a two-factor MANOVA (with interaction) are given. (The SAS code generating this output is on the first page.) From the results provided, what conclusions do you draw about the effects of species and time, and their interaction, on the spectral reflectance measures? Be as complete and specific as possible, and justify your conclusions with p-values from appropriate hypothesis tests and comparisons of sample means.

First, we conclude that there is a significant species*time interaction for at least one of the spectral reflectance measures ($\Lambda = 0.087$, $P < .0001$).

From the ANOVA's for nm560 and nm720 separately, we see that this interaction is significant for nm560 ($P < .0001$), but not for nm720 ($P = 0.5741$).

Thus we may test for main effects for nm720, but not for nm560. The main effects of species and time are both significant ($P < .0001$) for nm720.

An examination of the species means suggests that the significant main effect of species for nm720 is due primarily to a much larger mean for species JL than for the other 2 species.

An examination of the time means suggests that the significant main effect of time is due primarily to a much larger mean at time 3 than at the previous times.

5. (20 pts.) Consider the subset of the data described in the previous problem that corresponds to Species SS only. For this subset, a multivariate regression analysis of the spectral reflectance measures on time was carried out. The model upon which this analysis was based is as follows:

$$\begin{aligned} nm560_j &= \beta_{01} + \beta_{11} * time_j + \beta_{21} * time_j^2 + epsilon_{j1} \\ nm720_j &= \beta_{02} + \beta_{12} * time_j + \beta_{22} * time_j^2 + epsilon_{j2} \\ & \quad (j = 1, \ldots, 12). \end{aligned}$$

Note that in contrast to the previous problem, here time is regarded as a quantitative variable — not a classification variable — and the mean of each dependent variable is assumed to be a quadratic function of time. To answer the following questions, and to justify your answers, you may use the SAS code on the first page of the accompanying handout and the output corresponding to it.

(a) Are either or both quadratic terms in these two regressions needed to adequately describe the means of the reflectance measures as functions of time?

"Test2" indicates that at least one of the quadratic terms is needed ($P < .0013$), and the two univariate analyses suggest that the quadratic term for nm560 (β_{21}) is significant ($P = .0314$), but not the quadratic term for nm720 ($P = .2501$).

(b) Are the quadratic terms for the two reflectance measures equal to each other?

"Test5" is the relevant test. Based on it, we do not reject the hypothesis that the two quadratic terms are equal ($P = .4492$).

(c) Based on your answer to Part (a) and any other relevant information in this problem, would it be reasonable to fit a model with cubic terms (in time) to these data on the two reflectance measures? Explain.

Since measurements were taken at only 3 distinct time points, we cannot fit a cubic model (not all the regression coefficients will be estimable).