Mathematical Sciences 807
Applied Multivariate Statistical Analysis
Midterm II

I attest that I have not received nor given any assistance on this examination. I further attest that I have not witnessed any infraction of the University Honor code with respect to this examination. Signed_____

I. Machines producing electrodes are generally calibrated several times a year to reduce the variability of the electrodes. Investigators are interested if there are differences in the electrodes produced from machines located in two different parts of the building. There are five variables measured on each of the 50 electrodes produced by each of the machines. These variables are x_1-Total diameter, x_2-Probe diameter, x_3-Total height, x_4-Basal height, and x_5-Internal diameter.

   (a) An initial analysis found a Wilk's $\Lambda = 0.1190174$ for testing for differences in machines. In a complete sentence, what does this value indicate about the differences in the electrodes produced from the two machines.

   (b) An eigenvalue analysis produced the eigenvalues: $\lambda_1 = 7.402134$, and all other eigenvalues essentially 0. In a complete sentence, what does this mean in terms of the mean vectors of the electrodes produced by the two machines?

II. A researcher is interested in measuring the effects of a drug over an extended period of time. This drug is expected to affect people from different ethnic groups differently, as well as people from different age groups and genders. The researcher plans to measure blood concentrations of the drug for 10 months, taking a blood sample once a month for ten months on each of 10 patients from each ethnic group (G1,G2,G3), each group having 5 males and 5 females, and 2 people from each of 5 age groups.

(a) Describe the design to which the researcher is prescribing by (i) defining the model indicating (ii) what the between factor(s) are, (iii) what the within factor(s) are, and by diagramming how these 300 measurements would be laid out in a data matrix.

(b) What is an appropriate tests for determining differences across ethnic groups, giving the appropriate degrees of freedom for the test chosen? Assume the general form for error E and hypotheses H matrices.

(c) What is an appropriate test for determining differences across gender, giving the appropriate degrees of freedom for the test chosen?

(d) What is an appropriate test for determining if interaction exists, giving the appropriate degrees of freedom for the test chosen?

III. Doctors have discovered that there are two variables that can readily be used in the diagnosis of hemophilia, antihemophilic factor activity (AHF_1) and antihemophilic factor antigen (AHF_2). Data on noncarriers(G_1) and obligatory carriers(G_2) were analyzed for these two variables. 30 individuals were identified as being noncarriers and 45 identified as being obligatory carriers. Linear and quadratic rules were used to classify each of the observations.

(a) From the data $\overline{\boldsymbol{y}}_{G_1} = (-0.1349, -0.0779)$ and $\overline{\boldsymbol{y}}_{G_2} = (-0.3079, -0.0060)$. The linear discriminant function $z = 19.3190(AHF_1)$ - $17.1242(AHF_2)$. Define the rule for classifying individuals to one of the two groups.

(b) Use this rule to classify the following individuals as being noncarriers(G_1) or obligatory carriers(G_2).

Patient	AHF_1	AHF_2	Classification
1	-0.112	-0.279	
2	-0.059	-0.068	
3	0.064	0.012	

(c) Application of the linear classification, assuming proportionality, misclassified 4 of the noncarriers as being obligatory carriers, misclassified 6 obligatory carriers as being noncarriers. What is the *apparent correct classification rate* using the linear classification ?

(d) Similarly a quadratic classification was conducted with the following results. Noncarriers classified as obligatory carriers-4 and obligatory carriers classified as noncarriers - 5. Again compute the *apparent correct classification rate* and comment on the results as compared to the linear classification.