UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

COLLEGE OF INFORMATICS AND ELECTRONICS

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MA4128 **SEMESTER: Spring 2003**

MODULE TITLE: Advanced Data Modelling **DURATION: 5 hours**

LECTURER: Dr. Vicki Livingstone

EXTERNAL EXAMINER: Prof. Philip Boland

INSTRUCTIONS TO CANDIDATES:

- **This is an open book exam.**
- **Answer all questions.**
- **All questions carry equal marks.**
- **The MINITAB worksheets are on the disk provided.**
- **Your answers may be submitted on paper or in the form of a word document or as a combination of both. Word documents should be stored on the disk provided at the start of the exam and should also be emailed to me at victoria.livingstone@ul.ie. Remember to put your name on the disk and on all paper submitted.**
- **The exam is worth 70% of the final grade for this module.**

Q1.

The estimates of the average protein consumption from different food sources for the inhabitants of 25 European countries was published by Weber in 1973.

The data is stored in a worksheet called PROTEIN.MTW. The first column contains the country name. The remaining nine columns contain the food sources.

C1: Country
C2: Red meat
C3: White meat
C4: Eggs
C5: Milk
C6: Fish
C7: Cereals
C8: Starchy foods
C9: Pulses, Nuts and Oilseeds
C10: Fruits and Vegetables

(a)

(i) Perform a Principal Components Analysis of the variables C2 to C10 and store the principal component scores. Interpret the results of the analysis.

(ii) Do you think that the covariance matrix or the correlation matrix should have been used when performing Principal Components Analysis on this data? Justify your answer.

(b) Perform a two factor analysis on the variables C2 to C10 and store the factor scores. Interpret the two factors.

(c) Perform a clustering of the variables C2 to C10 and comment on the outcome.

(d) Use the variables C2 to C10 to cluster the countries in the data set. Use average linkage and squared Euclidean distance. Interpret the results of the analysis.

Q2.

A manager of a psychiatric unit in a hospital has a policy of asking all patients to fill out a questionnaire on their admission to the hospital. The results of the questionnaire for a random sample of 294 patients are contained in columns C1 to C17 of the worksheet DEPRESS.MTW as follows:

C1: Sex (1 = Male, 2 = Female)

C2: Age in years

C3: Spent entire day in bed in the last two months?
1 = Yes and 0 = No.

Respondents were asked to indicate their agreement with each of the following statements using a five-point scale with 1 = strongly disagree to 5 = strongly agree.

C4: I felt that I was as good as other people.
C5: I felt that I could not shake off the blues even with the help of my family or friends.
C6: I felt hopeful about the future.
C7: I was happy.
C8: I felt depressed.
C9: I felt lonely.
C10: I enjoyed life.
C11: I felt sad.
C12: I felt fearful.
C13: I felt that everything was an effort.
C14: I felt that people liked me.
C15: I talked less than usual.
C16: People were unfriendly.

After the patient leaves the psychiatrist diagnoses whether or not the patient is depressed and records the information in C17 of DEPRESS.MTW as follows:

C17: Depressed: 1 = Yes and 0 = No.

The psychiatrist wants to (a) describe the population of patients who are admitted, (b) determine characteristics which distinguish between those diagnosed with depression and those not, and (c) investigate whether the results of the questionnaire may be used to identify, in advance, patients who are more likely to be diagnosed with depression. Carry out whatever analyses you deem appropriate in order to address the psychiatrist's needs.