Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2019/2020

COURSE NAME	:	APPLIED MULTIVARIATE ANALYSIS
COURSE CODE	:	BWB 32103
PROGRAMME CODE	:	BWQ
EXAMINATION DATE	:	DECEMBER 2019 / JANUARY 2020
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWER **ALL** QUESTIONS

THIS QUESTION PAPER CONSISTS OF **SIX (6)** PAGES

Q1 For each of the following statements indicate whether it is **TRUE** or **FALSE.**

(i) Factor analysis does not classify variables as dependent or independent.

(ii) A factor is an underlying dimension that explains the correlations among a set of variables.

(iii) The factors identified in factor analysis are overtly observed in the population.

(iv) Factors can be estimated so that their factor scores are not correlated and the first factor accounts for the highest variance in the data, the second factor the second highest and so on.

(v) The variables to be included in the factor analysis should be specified based on past research, theory, and the judgement of the researcher.

(vi) Discriminant functions are linear combinations of the predictor or independent variables, which will best discriminate between the categories of the criterion or dependent variable (groups).

(vii) The direct method is an approach to discriminant analysis that involves estimating the discriminant function so that all the predictors are included simultaneously.

(viii) With the leave-one-out cross-validation option in SPSS, the discriminant model is re-estimated as many times as there are respondents in the sample.

(ix) The hit ratio is the number of variables found to be significant by the discriminant analysis.

(x) The linear combinations of independent variables developed by discriminant analysis that will best discriminate between the categories of the dependent variable are characteristic profiles.

(10 marks)

Q2 The following data matrix contains data on test scores, with x_1 = score on first test, x_2 = score on second test, and x_3= total score on the two tests.

$$X=\begin{bmatrix}12 & 17 & 29\\ 18 & 20 & 38\\ 14 & 16 & 30\\ 20 & 18 & 38\\ 16 & 19 & 35\end{bmatrix}$$

(a) Calculate the mean vector and variance-covariance matrix.

(8 marks)

(b) Prove that the generalized variance is zero.

(5 marks)

Q3 (a) List **TWO (2)** assumptions that your data must meet in order for a one-way MANOVA to give you a valid result.

(2 marks)

(b) A researcher wants to test the effectiveness of different types of learning intervention on exam performance. Exam performance was assessed as the scores achieved in a humanities exam and a science exam. These two example scores are the two dependent variables the researcher wants to measure. The different types of learning intervention were the current learning method (called the "Regular" programme), a rote learning programme (called the "Rote" programme) and a learning programme based on reasoning skills (called the "Reasoning" programme). These three types of learning intervention are the three categories of the independent variable.

(i) List all the variables under study.

(4 marks)

(ii) A one-way MANOVA was used to assess the effect of these different learning interventions on student performance in the humanities and science exams. The Minitab output for the one-way MANOVA is shown in **Table Q3(b)**. Conduct an appropriate MANOVA test and derive your conclusion. Use 0.05 significant level.

Table Q3(b)

MANOVA for Intervention				
s=2 m = -0.5 n = 27.0				
Criterion	Test Statistic	F	Num	P
Wilks'	0.70765	5.285	4	0.001
Lawley-Hotelling	0.40227	5.531	4	0.000
Pillai's	0.30004	5.030	4	0.001
Roy's	0.37317			

(6 marks)

Q4 (a) Given a covariance matrix

$$\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

(i) Determine the population principal component Y_1 and Y_2.

(8 marks)

(ii) Calculate the proportion of the total population variance explained by the first principal component.

(2 marks)

(b) A bank requires eight pieces of information from loan applicants: income, education level, age, length of time at current residence, length of time with current employer, savings, debt, and number of credit cards. A bank administrator wants to analyze this data to determine the best way to group and report it. The administrator collects this information for 30 loan applicants. The administrator performs a principal components analysis to reduce the number of variables to make the data easier to analyze. The software output given in **Figure Q2(b).**

Eigenanalysis of the Correlation Matrix

Eigenvalue	3.5476	2.1320	1.0447	0.5315	0.4112	0.1665	0.1254	0.0411
Proportion	0.443	0.266	0.131	0.066	0.051	0.021	0.016	0.005
Cumulative	0.443	0.710	0.841	0.907	0.958	0.979	0.995	1.000

Eigenvectors

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Income	0.314	0.145	-0.676	-0.347	-0.241	0.494	0.018	-0.030
Education	0.237	0.444	-0.401	0.240	0.622	-0.357	0.103	0.057
Age	0.484	-0.135	-0.004	-0.212	-0.175	-0.487	-0.657	-0.052
Residence	0.466	-0.277	0.091	0.116	-0.035	-0.085	0.487	-0.662
Employ	0.459	-0.304	0.122	-0.017	-0.014	-0.023	0.368	0.739
Savings	0.404	0.219	0.366	0.436	0.143	0.568	-0.348	-0.017
Debt	-0.067	-0.585	-0.078	-0.281	0.681	0.245	-0.196	-0.075
Credit cards	-0.123	-0.452	-0.468	0.703	-0.195	-0.022	-0.158	0.058

Figure Q2(b)

(i) The administrator wants enough components to explain 90% of the variation in the data. Determine the suitable number of components that meet the administrator demands. Justify your answer.

(4 marks)

(ii) Interpret the first three principal components. Identify the important variables that can be a key factor for the administrator to make a decision.

(6 marks)

(iii) Construct a scree plot based on all components. Interpret the plot.

(5 marks)

Q5 The correlation matrix for chicken–bone measurement is given in **Table Q5(a)** and **Table Q5(b)** is the estimated factor loading were extracted by the maximum likelihood procedure.

Table Q5(a)

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.000					
X_2	0.505	1.000				
X_3	0.569	0.422	1.000			
X_4	0.602	0.467	0.926	1.000		
X_5	0.621	0.482	0.877	0.874	1.000	
X_6	0.603	0.450	0.878	0.894	0.937	1.000

Table Q5(b)

Variable	Estimated factor loadings		Varimax rotated estimated factor loadings	
	F_1	F_2	F_1^*	F_2^*
X_1	0.602	0.200	0.484	0.411
X_2	0.467	0.154	0.375	0.319
X_3	0.926	0.143	0.603	0.717
X_4	1.000	0.000	0.519	0.855
X_5	0.874	0.476	0.861	0.499
X_6	0.894	0.327	0.744	0.594

Using the unrotated and rotated estimated factor loadings, obtain the maximum likelihood estimate of the following:

(a) The specific variances.

(8 marks)

(b) The communalities.

(8 marks)

(c) The proportion of variance explained by each factor.

(4 marks)

Q6 (a) Let $\boldsymbol{\rho}_{12} = \begin{bmatrix} \rho & \rho \\ \rho & \rho \end{bmatrix}$ and $\boldsymbol{\rho}_{11} = \boldsymbol{\rho}_{22} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, corresponding to the equal correlation structure where $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ each have two components. Determine the canonical variates corresponding to the nonzero canonical correlation.

(8 marks)

(b) A group of researchers studied on the abilities of student at primary school in certain places. They report that number of student involved in their study is 140 and the students received four tests on $X_1^{(1)}$= reading speed, $X_2^{(1)}$= reading power, $X_1^{(2)}$=

arithmetic speed, and $X_2^{(2)}$= arithmetic power. The correlation for performance are given as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} = \begin{bmatrix} 1.0000 & 0.6328 & 0.2412 & 0.0586 \\ 0.6328 & 1.0000 & -0.0553 & 0.0655 \\ 0.2412 & -0.0553 & 1.0000 & 0.4248 \\ 0.0586 & 0.0655 & 0.4248 & 1.0000 \end{bmatrix}$$

Find all the sample canonical correlation and the sample canonical variates.

(12 marks)

– **END OF QUESTIONS** –