

Semester: Jan – Mar 24		
Maximum Marks: 50 Examination: ETE Exam Date: 28-03-2024 Duration: 3 Hours		
Programme code: 01 Programme: MBA	Class: SY	Semester/Trimester: VI
College: K. J. Somaiya Institute of Management	Name of the department/Section/Center: Business Analytics	
Course Code: 217P01M622	Name of the Course: Machine Learning approach for Multivariate Data Analysis	
Instructions: 1. You must attempt 5 questions in all. All questions carry equal marks. 2. Question 1 is compulsory. 3. All subparts to a question must be answered. 4. Calculator is allowed.		

Question No.		Max. Marks				
1	<p>A researcher examined the purchasing behaviour in an online shop. The aim is to determine the influencing factors that lead a person to buy any product from the online shop after visiting the website. The online shop provides the data collected for this purpose. The researcher used logistic regression analysis to find the influencing factors. The variables used for the analysis are mentioned below.</p> <p>Dependent variable: Purchasing behaviour (Nominal scale) 0-Not purchasing any product 1-Purchasing any product</p> <p>Independent variables: Gender (Nominal scale) Age (Ratio scale) Marital status (Nominal scale) Income (Ratio scale) Time spent in the online shop (Ratio scale) Preferred mode of payment (Nominal scale)</p> <p>The output is given below:</p>	10				
Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Gender						
[®] Male						
Female	-2.869	1.113	6.651	1	.010	
Age	-.127	.047	7.313	1	.007	
Marital Status						
[®] Single						

Married	-.782	.854	.839	1	.360	
Income						
® 30000-49999						
50000-64999	.842	.985	.730	1	.393	
65000 & above	-.330	1.083	.093	1	.761	
Time spent in online shopping	-.025	.020	1.636	1	.201	
Preferred mode of payment						
® UPI						
Credit card	1.788	.901	3.942	1	.047	
Constant	5.951	1.916	9.646	1	.002	384.079

a. Variable(s) entered on step 1: Gender, Age, Marital status, Income, Time spent in online shopping, Preferred mode of payment.

Note: ® refers to reference category

- a) Interpret the odds ratios.
- b) Construct the null and alternative hypothesis and conclude the result at 5% level of significance.

2

The researcher wants to test the effect of shoe brands and runner age group on race finishing times (in minutes) in a marathon. He collected the data for the study and ran a Two-way ANOVA to fulfil the objective. Variable description is given below:

10

Age group (Nominal scale)

1. 18-30
2. 31-45

Shoe brands (Nominal scale)

1. Nike
2. Adidas
3. Hoka

Race finishing time (Ratio scale)

The output table is given below:

Tests of Between-Subjects Effects					
Dependent Variable: race_finish_time					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2566.200 ^a	5	513.240	1.445	.223
Intercept	83477.400	1	83477.400	234.971	.000
age_group	317.400	1	317.400	.893	.349
shoe_brands	1016.400	2	508.200	1.430	.248
age_group * shoe_brands	1232.400	2	616.200	1.734	.186
Error	19184.400	54	355.267		
Total	105228.000	60			
Corrected Total	21750.600	59			

- a) How the degrees of freedom for error is 54?
- b) How the degrees of freedom for total is 60?
- c) Construct the null and alternative hypothesis.
- d) Conclude the result at 5% level of significance.

3

The researcher wants to determine the impact of Social skills, Intellectual skills and motivation on Job Performance. All the variables are composites of multiple items. The explanation is given below:

10

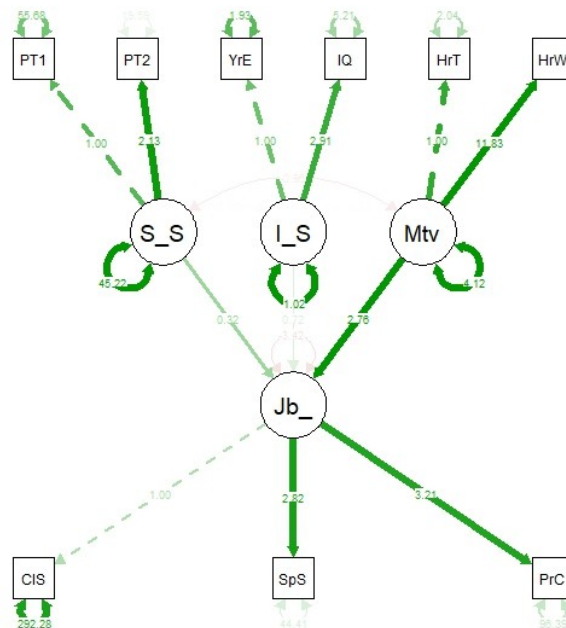
Client satisfaction (ClientSat): A satisfaction rating between 1 and 100 by your main client

Superior satisfaction(SuperSat): A rating on Job Performance between 1 and 100 by your superior
 Project completion (ProjCompl): The percentage of your projects that was successfully delivered
 Psychiatric Test 1 (PsychTest1): a score between 1–100
 Psychiatric Test 2 (PsychTest2):also a score between 1–100
 Tears of education (YrsEdu): Number of years of higher education followed
 IQ: Score on an IQ test
 Hours training (HrsTrain): Number of hours spent on training
 Hours working (HrsWrk): Average number of hours in a workweek
Job performance = ClientSat+SuperSat+ProjCompl
Social Skills = PsychTest1+PsychTest2
Intellectual Skills = YrsEdu+IQ
Motivation = HrsTrn+HrsWrk

The researcher ran a structural equation modelling to fulfil the objective. The output is given below:

Regressions:

	Estimate	Std.Err	z-value	P(> z)	Std.lv	Std.all
Job_performance ~						
Social_Skills	0.325	0.029	11.085	0.000	0.389	0.389
Intlletl_Skils	0.725	0.077	9.361	0.000	0.13	0.13
Motivation	2.758	0.234	11.769	0.000	0.998	0.998



Interpret the output.

4

A firm surveyed its employees to determine the relationship between sales performance and intelligence of the employees. Two collections of variables were measured:

10

Sales Performance:

Sales Growth

Test Scores as a Measure of Intelligence:

Mechanical Reasoning

Sales Profitability	Abstract Reasoning Mathematics
---------------------	-----------------------------------

The researcher then conducted canonical correlation analysis. The output is mentioned below:

Canonical Correlations							
	Correlation	Eigenvalue	Wilks Statistic	F	Num D.F.	Denom D.F.	Sig.
1	.985	32.010	.015	106.514	6.000	90.000	
2	.705	.988	.503	.	.	.	

H0 for Wilks test is that the correlations in the current and following rows are zero

Set 1 Canonical Loadings

Variable	1	2
sales_growth	-.959	-.282
sales_profit	-.995	.101

Set 2 Canonical Loadings

Variable	1	2
mechanica_reasoning	-.758	.042
abstract_reasoning	-.536	-.844
maths_score	-.968	-.062

Proportion of Variance Explained

Canonical Variable	Set 1 by Self	Set 1 by Set 2	Set 2 by Self	Set 2 by Set 1
1	.955	.926	.600	.582
2	.045	.022	.239	.119

- a) Identify the set of dependent and independent variables.
- b) Interpret the output.

5 A group of engineers operating in the national construction company in Iraq. The population frame is defined as all engineers working for this company with a minimum of one year experience. The company administrators have been observing that some engineers perform more effectively compared to others. The research is interested in determining which variables (predictors) are useful in predicting job performance (dependent variable). The company has considered four variables that can be recognized as probable discriminators; these include:

- job history to assess the experience (Job_record)
- job test to evaluate knowledge in engineering (Profession_test)
- personality measure that assesses friendliness (Friendliness)
- college GPA to appraise their performance at college (College_GPA)

Dependent variable:
Job performance

1. Better performance
2. Poor performance

The researcher then ran a discriminant analysis to find the predictors which are differentiating between two groups. The output is given

10

below:

		Group Statistics			
				Valid N (listwise)	
Appraisal		Mean	Std. Deviation	Unweighted	Weighted
Better performance	Friendliness	38.76	5.854	25	25.000
	College_GPA	17.16	2.115	25	25.000
	Job_record	18.80	3.416	25	25.000
	Profession_test	18.52	4.727	25	25.000
Poor performance	Friendliness	24.28	9.072	25	25.000
	College_GPA	13.40	4.682	25	25.000
	Job_record	13.92	4.030	25	25.000
	Profession_test	13.84	6.743	25	25.000
Total	Friendliness	31.52	10.516	50	50.000
	College_GPA	15.28	4.066	50	50.000
	Job_record	16.36	4.444	50	50.000
	Profession_test	16.18	6.229	50	50.000

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
Friendliness	.516	44.968	1	48	.000
College_GPA	.782	13.393	1	48	.001
Job_record	.692	21.331	1	48	.000
Profession_test	.856	8.074	1	48	.007

Pooled Within-Groups Matrices					
		Friendliness	College GPA	Job record	Profession test
Correlation	Friendliness	1.000	.446	.241	
	College_GPA	.446	1.000	.025	
	Job_record	.241	.025	1.000	
	Profession_test	.440	.443	.108	1.000

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.163 ^a	100.0	100.0	.733

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.462	35.494	4	.000

Standardized Canonical Discriminant Function

Coefficients

Function	
1	
Friendliness	.743
College_GPA	.180
Job_record	.442
Profession_test	-.074

Interpret the output.

6

An airline company planned to design a product considering four attributes namely Price, Duration, Comfort and Entertainment. Mentioned below the attributes and their levels.

Price	Duration
5000	2 hours
8000	5 hours
10000	
Comfort	Entertainment
Cramped seat	TV screen
Spacious seat	Magazine
	Music system

Using the four attributes and their levels, the company generated some combinations using SPSS orthogonal design. The combinations were considered as questionnaire and circulated among the customers. After getting the response, the company ran a conjoint analysis to finalize the product. The questionnaire and the output tables are given below.

Utilities

		Utility Estimate
Price	5000	.117
	8000	.533
	10000	-.450
Duration	2 hours	.450
	5 hours	-1.183
Comfort	Cramped seat	.350
	Spacious seat	.450
Entertainment	TV screen	.617
	Magazine	.600
	Music system	.465
(Constant)		6.740

Importance value

Price	34
Duration	24

10

Comfort	22			
Entertainment	20			
	Price	Duration	Comfort	Entertainmet
Product 1	10000	5 hours	spacious seat	Magazine
Product 2	5000	2 hours	cramped seat	Magazine
Product 3	10000	2 hours	cramped seat	Music system
Product 4	10000	2 hours	spacious seat	Music system
Product 5	8000	5 hours	spacious seat	Music system
Product 6	5000	2 hours	spacious seat	Magazine
Product 7	8000	2 hours	spacious seat	TV screen
Product 8	10000	2 hours	spacious seat	TV screen
Product 9	8000	2 hours	cramped seat	Magazine
Product 10	10000	5 hours	cramped seat	Music system
Product 11	5000	2 hours	spacious seat	Music system
Product 12	10000	2 hours	cramped seat	Magazine
Product 13	8000	2 hours	cramped seat	Music system
Product 14	10000	2 hours	cramped seat	Magazine
Product 15	8000	2 hours	spacious seat	Magazine
Product 16	8000	2 hours	cramped seat	Music system
Product 17	5000	5 hours	cramped seat	Magazine
Product 18	10000	5 hours	cramped seat	TV screen
Product 19	5000	2 hours	cramped seat	TV screen

The questionnaire does not contain any holdout cases.

- a) Interpret the utilities and importance values tables.
- b) Finalise the product.