UNIVERSITI SAINS MALAYSIA

Master of Business Administration
(On Line Mode)

Second Semester Examination
Academic Session 2007/2008

April 2008

ADW615 – Advanced Business Statistics

Duration: 3 hours

Please check that this examination paper consists of **TWENTY** pages of printed material before you begin the examination.

Answer **FIVE** questions. Section A is **COMPULSORY** and answer any **ONE (1)** question from Section B.

Section A: (COMPULSORY)

Question 1

(a) Ahmad always researches the share market before making an investment. Now he is evaluating whether he should invest in Acer Power Co. By looking at past performances he has divide the performance to 5 categories. The returns are for an investment of RM 150.

ROI	0.00	10.00	15.00	25.00	50.00
Probability	0.20	0.25	0.30	0.15	0.10

If he will invest only if he gets an expected return of more than 10%, will he invest in Acer Power Co.?

[5 marks]

(b) Aquarius Health Club has been advertising a rigorous program for body conditioning. The club claims that after 1 month in the program, the average participant should be able to do eight more push-ups in 2 minutes than he or she could do at the start. Does the random sample of 10 program participants given below support the club's claim? Use the 0.05 level of significance.

Participant	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Before	38	11	34	25	17	38	12	27	32	29
After	45	24	41	39	30	44	30	39	40	41

[5 marks]

(c) Because refunds are paid more quickly on tax returns that are filed electronically, the Commissioner of the Internal Revenue Service was wondering whether refunds due on returns filed by mail were smaller than those due on returns filed electronically. Looking only at returns claiming refunds, a sample of 17 filed by mail had an average refund of $563, and a standard deviation of $378. The average refund on a sample of 13 electronically filed returns was $958, and the sample standard deviation was $619. At $\alpha = 0.01$, do these data support the commissioner's speculation?

[5 marks]

Question 2

The following table shows sample retail prices for three brands of shoes. Use the Kruskal-Wallis test to determine whether there are any differences among the retail prices of the brands throughout the country. Use the 001 level of significance.

Brand A	$89	90	92	81	76	88	85	95	97	86	100
Brand B	$78	93	81	87	89	71	90	96	82	85	
Brand C	$80	88	86	85	79	80	84	85	90	92	

[15 marks]

Question 3

Use the Wilcoxon rank sum test on the following data to determine whether the two population locations differ. (Use alpha 10%)

Sample 1	Sample 2
15	8
7	27
22	17
20	25
32	20
18	16
26	21
17	17
23	10
30	18

[10 marks]

Question 4

a) An advertising firm is trying to determine the demographics for a new product. They have randomly selected 75 people in each of 5 different age groups and introduced the product to them. The results of the survey are given below:

	Age Group				
Future Activity	18-29	30-39	40-49	50-59	60-69
Purchase frequently	12	18	17	22	32
Seldom purchase	18	25	29	24	30
Never purchase	45	32	29	29	13

At the 0.10 significance level, does the future activity differ according to the age group of the consumer?

[10 marks]

b) Cost accountants often estimate overhead based on the level of production. At the Standard Knitting Co., they have collected information on overhead expenses and units produced at different plants, and want to estimate a regression equation to predict future overhead.

Overhead	191	170	272	155	280	173	234	116	153	178
Units	40	42	53	35	56	39	48	30	37	40

(i) Develop the regression equation for the cost accountants.
(ii) Predict overhead when 50 units are produced.

[10 marks]

Section B: Answer ONE (1) question.

Question 5

(a) The output below shows the results of a factor analysis done on three independent variables, perceived ease of use (PEU1 – PEU5), perceived usefulness (PU1 – PU3) and perceived enjoyment (PE1 – PE3). Summarize the output into a table and explain how good is the factor analysis by looking at the assumptions and also other criteria's that are used to assess the goodness of this analysis.

[15 marks]

Factor Analysis

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.879
Bartlett's Test of Sphericity	Approx. Chi-Square	9872.077
	df	55
	Sig.	.000

Communalities

	Initial	Extraction
PU1	1.000	.762
PU2	1.000	.791
PU3	1.000	.805
PEU1	1.000	.689
PEU2	1.000	.753
PEU3	1.000	.722
PEU4	1.000	.780
PEU5	1.000	.790
PE1	1.000	.569
PE2	1.000	.682
PE3	1.000	.680

Extraction Method: Principal Component Analysis.

...6/-

Anti-image Matrices

		PU1	PU2	PU3	PEU1	PEU2	PEU3	PEU4	PEU5	PE1	PE2	PE3
Anti-image Covariance	PU1	.471	-.141	-.161	-.015	-.009	-.027	.015	-.010	-.016	.003	.000
	PU2	-.141	.443	-.191	-.008	.019	-.007	-.010	-.004	-.011	-.010	-.038
	PU3	-.161	-.191	.399	-.014	-.026	.013	-.012	-.009	-.016	.009	.002
	PEU1	-.015	-.008	-.014	.395	-.156	-.048	-.039	-.024	-.012	-.022	-.038
	PEU2	-.009	.019	-.026	-.156	.337	-.080	-.023	-.055	-.048	-.014	-.013
	PEU3	-.027	-.007	.013	-.048	-.080	.406	-.074	-.063	-.036	.010	.011
	PEU4	.015	-.010	-.012	-.039	-.023	-.074	.295	-.155	-.039	.012	-.001
	PEU5	-.010	-.004	-.009	-.024	-.055	-.063	-.155	.287	-.021	.027	.016
	PE1	-.016	-.011	-.016	-.012	-.048	-.036	-.039	-.021	.694	-.071	-.143
	PE2	.003	-.010	.009	-.022	-.014	.010	.012	.027	-.071	.818	-.286
	PE3	.000	-.038	.002	-.038	-.013	.011	-.001	.016	-.143	-.286	.757
Anti-image Correlation	PU1	.869[a]	-.309	-.371	-.035	-.021	-.063	.040	-.027	-.028	.004	.000
	PU2	-.309	.835[a]	-.454	-.018	.048	-.015	-.029	-.013	-.020	-.017	-.066
	PU3	-.371	-.454	.837[a]	-.035	-.070	.032	-.035	-.026	-.030	.016	.004
	PEU1	-.035	-.018	-.035	.914[a]	-.426	-.120	-.115	-.070	-.023	-.038	-.070
	PEU2	-.021	.048	-.070	-.426	.901[a]	-.216	-.072	-.177	-.100	-.027	-.025
	PEU3	-.063	-.015	.032	-.120	-.216	.940[a]	-.214	-.183	-.069	.017	.020
	PEU4	.040	-.029	-.035	-.115	-.072	-.214	.877[a]	-.534	-.085	.024	-.002
	PEU5	-.027	-.013	-.026	-.070	-.177	-.183	-.534	.878[a]	-.046	.055	.034
	PE1	-.028	-.020	-.030	-.023	-.100	-.069	-.085	-.046	.945[a]	-.095	-.197
	PE2	.004	-.017	.016	-.038	-.027	.017	.024	.055	-.095	.624[a]	-.364
	PE3	.000	-.066	.004	-.070	-.025	.020	-.002	.034	-.197	-.364	.736[a]

a. Measures of Sampling Adequacy(MSA)

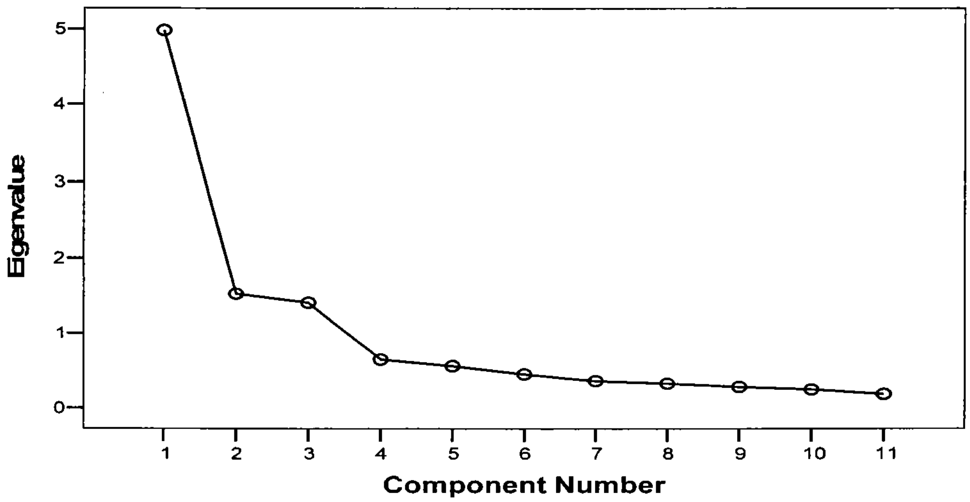
...7/-

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.984	45.307	45.307	4.984	45.307	45.307	3.940	35.817	35.817
2	1.529	13.899	59.205	1.529	13.899	59.205	2.418	21.984	57.801
3	1.410	12.821	72.027	1.410	12.821	72.027	1.565	14.226	72.027
4	.645	5.867	77.894						
5	.560	5.094	82.988						
6	.453	4.121	87.109						
7	.364	3.309	90.419						
8	.331	3.008	93.427						
9	.286	2.597	96.024						
10	.250	2.272	98.296						
11	.187	1.704	100.000						

Extraction Method: Principal Component Analysis.

Scree Plot



Component Matrix[a]

	Component		
	1	2	3
PU1	.629	.554	-.244
PU2	.613	.609	-.210
PU3	.666	.549	-.246
PEU1	.798	-.225	.050
PEU2	.824	-.271	.032
PEU3	.789	-.311	-.049
PEU4	.818	-.327	-.062
PEU5	.823	-.323	-.094
PE1	.600	-.032	.329
PE2	.154	.244	.774
PE3	.319	.241	.722

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Rotated Component Matrix [a]

	Component		
	1	2	3
PU1	.225	.842	.052
PU2	.181	.865	.096
PU3	.259	.857	.056
PEU1	.789	.213	.145
PEU2	.837	.197	.120
PEU3	.831	.175	.027
PEU4	.864	.182	.015
PEU5	.866	.199	-.013
PE1	.513	.168	.421
PE2	-.018	.005	.826
PE3	.122	.105	.809

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

(b) The Inland Revenue Board (IRB) of Malaysia has spent millions of dollars developing an online tax return system called e-Filing. Based on the Technology Acceptance Model (TAM) the Inland Revenue Board would like to model the factors that would influence the intention to use the e-Filing system. Ali identified Perceived Usefulness (PU), Perceived Ease of Use (PEU) and Perceived Enjoyment (PE) three main characteristics of the system will influence the intention to use the system. He collects data from 422 tax payers to test the model that he has proposed collecting data about **Perceived Usefulness** (PU), **Perceived Ease of Use** (PEU) and **Perceived Enjoyment** (PE) and **Intention** (BI). The data was analysed using SPSS and the output is given below:

<u>**Explanation of the Variables**</u>

PU	Perceived usefulness of the e-Filing system
PEU	Perceived ease of use of the e-Filing system
PE	Perceived enjoyment of using the e-Filing system
BI	Intention to use the e-Filing system

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	PE, PU, PEU[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: BI

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.670[a]	.449	.445	.97646	1.871

a. Predictors: (Constant), PE, PU, PEU

b. Dependent Variable: BI

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	324.140	3	108.047	113.319	.000[a]
	Residual	398.550	418	.953		
	Total	722.690	421			

a. Predictors: (Constant), PE, PU, PEU

b. Dependent Variable: BI

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.201	.263		.764	.446		
	PU	.039	.040	.041	.989	.323	.757	1.321
	PEU	.605	.052	.506	11.555	.000	.689	1.451
	PE	.335	.046	.278	7.226	.000	.894	1.118

a. Dependent Variable: BI

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	PU	PEU	PE
1	1	3.889	1.000	.00	.00	.00	.00
	2	.061	7.998	.02	.53	.00	.36
	3	.027	12.007	.18	.46	.44	.55
	4	.023	12.956	.80	.01	.55	.09

a. Dependent Variable: BI

...10/-

Casewise Diagnostics(a)

Case Number	Std. Residual	BI	Predicted Value	Residual
181	-3.645	1.00	4.5595	-3.55950
193	-3.246	1.00	4.1693	-3.16925
345	3.131	7.00	3.9426	3.05740

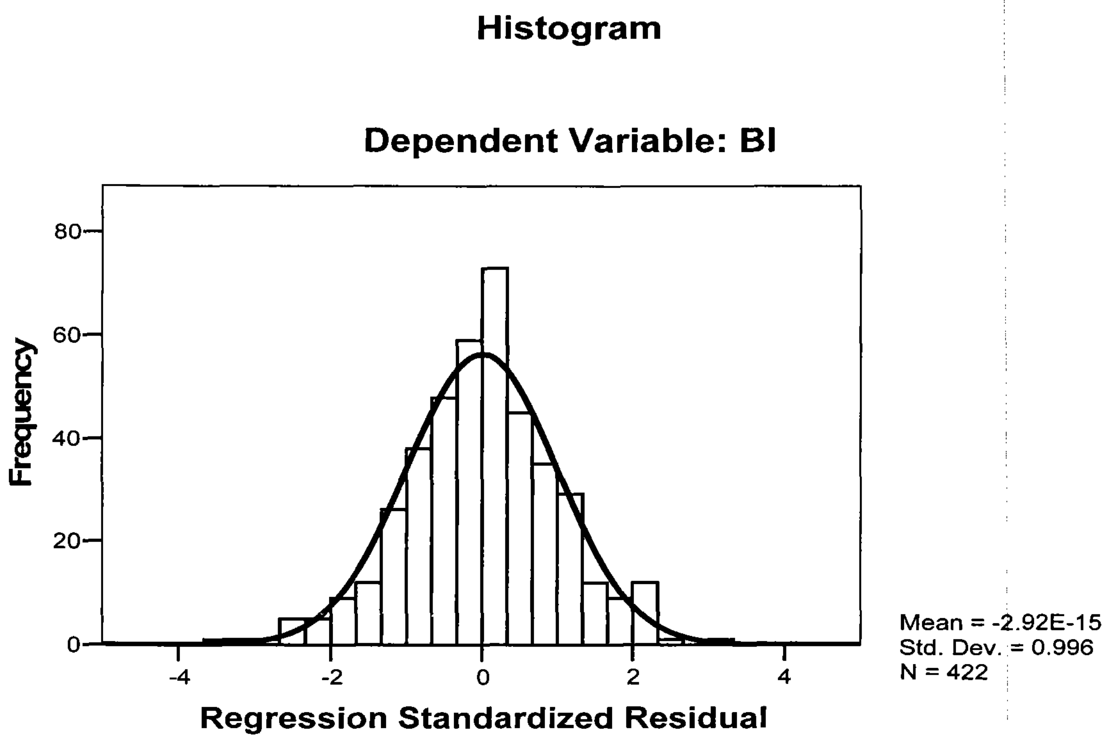
a Dependent Variable: BI

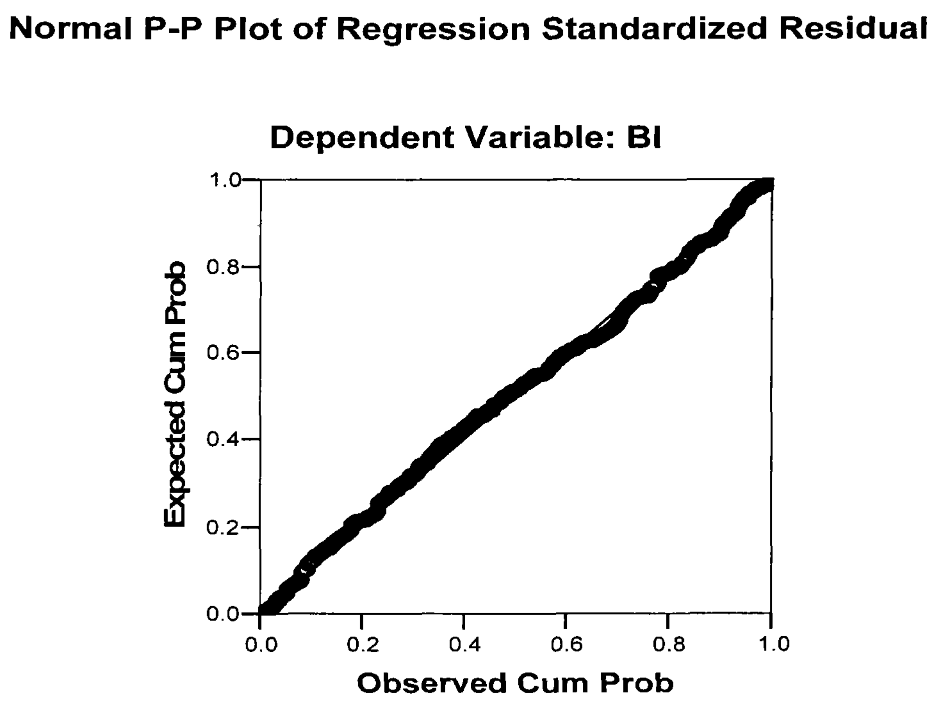
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.6151	7.0520	4.7903	.87746	422
Std. Predicted Value	-3.619	2.578	.000	1.000	422
Standard Error of Predicted Value	.048	.234	.090	.031	422
Adjusted Predicted Value	1.6132	7.0824	4.7902	.87732	422
Residual	-3.55950	3.05740	.00000	.97297	422
Std. Residual	-3.645	3.131	.000	.996	422
Stud. Residual	-3.660	3.143	.000	1.002	422
Deleted Residual	-3.58786	3.08158	.00010	.98300	422
Stud. Deleted Residual	-3.715	3.177	.000	1.005	422
Mahal. Distance	.016	23.109	2.993	3.094	422
Cook's Distance	.000	.029	.003	.005	422
Centered Leverage Value	.000	.055	.007	.007	422

a. Dependent Variable: BI

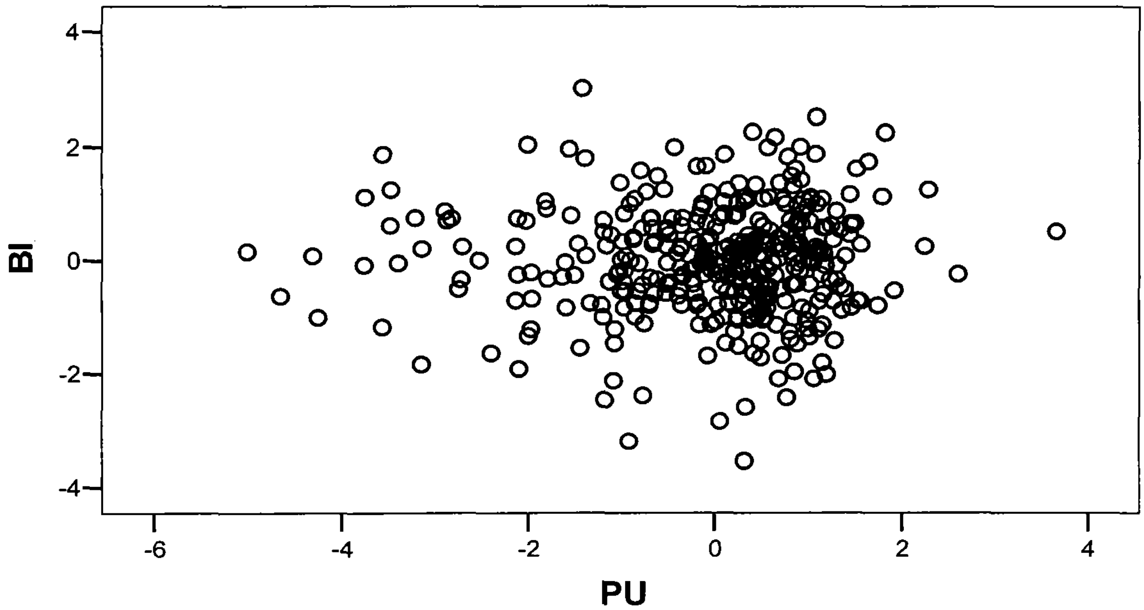
Charts


Histogram
Dependent Variable: BI
Frequency
80
60
40
20
0
-4
-2
0
2
4
Regression Standardized Residual
Mean = -2.92E-15
Std. Dev. = 0.996
N = 422


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: BI
Expected Cum Prob
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Observed Cum Prob

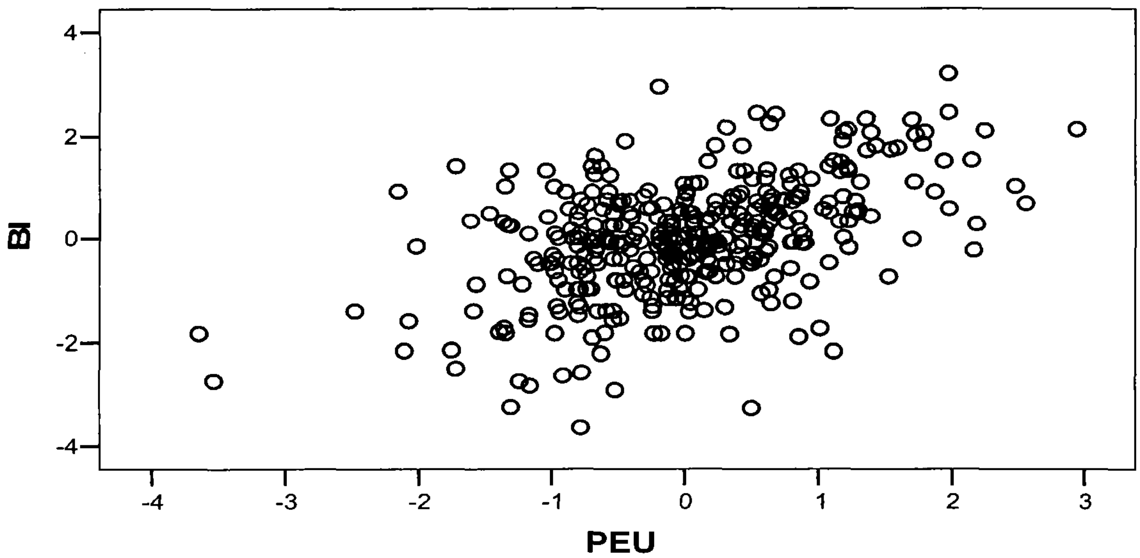
Partial Regression Plot

Dependent Variable: BI



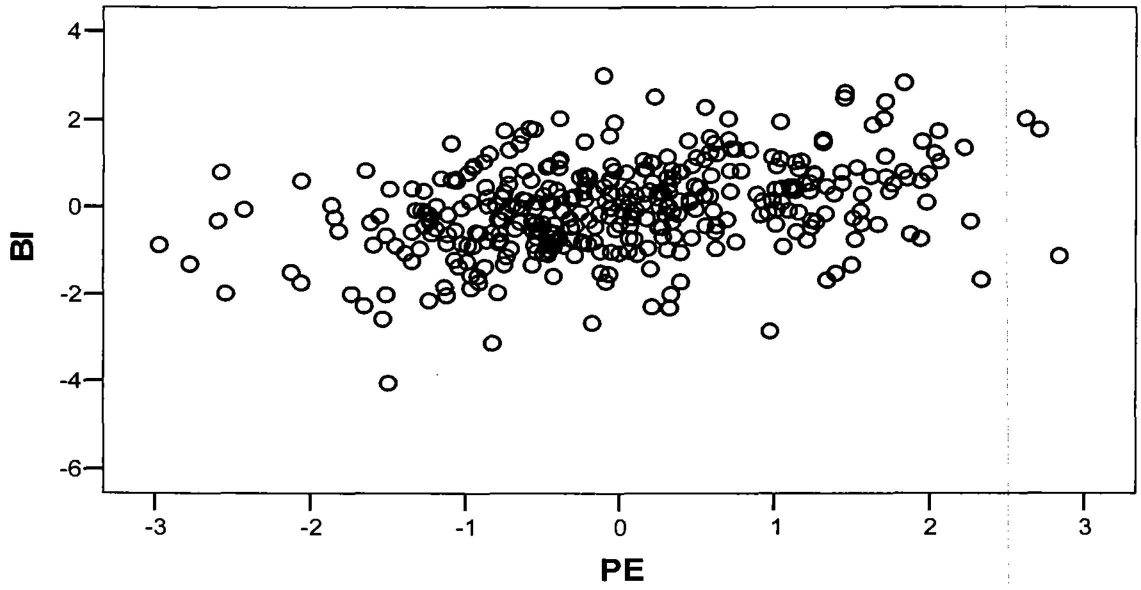
Partial Regression Plot

Dependent Variable: BI



Partial Regression Plot

Dependent Variable: BI



Based on the output answer the questions below:

(a) Can we develop a regression equation?

[3 marks]

(b) What can you conclude from the coefficient of determination? Explain.

[3 marks]

(c) Develop the hypotheses and test each of the following:

(i) The higher the perceived usefulness the higher the intention to use
(ii) The higher the perceived ease of use the higher the intention to use
(iii) There will be a relationship between perceived enjoyment and intention to use

[12 marks]

(d) Is there a problem of multicollinearity? Support your answer.

[2 marks]

(e) Has all the assumptions of the regression fulfilled? If not, explain how they should be tested?

[5 marks]

Question 6

The Malaysian Airlines is planning to use online ticketing on a full scale in the future. In order to understand what factors will impact whether a person will or will not use the online ticketing system they carried out a survey. Based on prior literature they would like to test whether the Theory of Planned Behavior (TPB) proposed by Ajzen (1991). This theory proposes that three variables influence the intention and the actual behavior of a person which are attitude (ATT), subjective norm (SN) and perceived behavioral control (PBC). The descriptions of the variables are as follows:

Attitude (ATT)	Attitude towards using the online ticketing system
Subjective Norms (SN)	The pressure from those around them to use the online ticketing system (Colleagues, friends, relatives, etc)
Perceived Behavioral Control (PBC)	Beliefs about the presence of factors that may facilitate or impede the use of the online ticketing system

Descriptive analysis as well as a two group discriminant analysis was conducted and the results of this analysis are tabulated below. Using figures from these results to support your statements and conclusions, answer the following questions.

(a) Using the descriptive statistics given in the output, what are the likely variables that can differentiate between the online ticketing user from a non user of the online ticketing system?

[5 marks]

(b) In one or two lines describe what are structure matrix and the standardized canonical discriminant function? How good is the discriminant model?

[4 marks]

(c) Which variable has the highest discriminatory power? Explain.

[6 marks]

(d) Write the discriminant function?

[4 marks]

(e) Describe briefly the predictive power of the discriminant function – compare it to the standard benchmarks?

[15 marks]

(f) Calculate the cutting score. Why is this cutting score needed?

[6 marks]

Discriminant

Analysis Case Processing Summary

Unweighted Cases		N	Percent
Valid		98	65.3
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Unselected	52	34.7
	Total	52	34.7
Total		150	100.0

Group Statistics

USAGE		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
Non User	ATT	3.7778	.95950	36	36.000
	SN	3.8218	.84425	36	36.000
	PBC	4.1111	1.37898	36	36.000
User	ATT	5.3710	.78951	62	62.000
	SN	5.3051	.64142	62	62.000
	PBC	5.4758	1.05748	62	62.000
Total	ATT	4.7857	1.14898	98	98.000
	SN	4.7602	1.01614	98	98.000
	PBC	4.9745	1.35154	98	98.000

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
ATT	.549	79.005	1	96	.000
SN	.500	96.135	1	96	.000
PBC	.761	30.215	1	96	.000

...16/-

Analysis 1

Box's Test of Equality of Covariance Matrices

Log Determinants

USAGE	Rank	Log Determinant
Non User	3	-.441
User	3	-1.993
Pooled within-groups	3	-1.257

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Test Results

Box's M		16.284
F	Approx.	2.612
	df1	6
	df2	35271.71
	Sig.	.016

Tests null hypothesis of equal population covariance matrices.

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.225[a]	100.0	100.0	.742

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

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.449	75.586	3	.000

Standardized Canonical Discriminant Function Coefficients

	Function
	1
ATT	.504
SN	.667
PBC	-.031

Structure Matrix

	Function 1
SN	.904
ATT	.820
PBC	.507

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

Canonical Discriminant Function Coefficients

	Function 1
ATT	.589
SN	.924
PBC	-.026
(Constant)	-7.085

Unstandardized coefficients

Functions at Group Centroids

USAGE	Function 1
Non User	-1.438
User	.835

Unstandardized canonical discriminant functions evaluated at group means

Classification Statistics

Classification Processing Summary

Processed		150
Excluded	Missing or out-of-range group codes	0
	At least one missing discriminating variable	0
Used in Output		150

Prior Probabilities for Groups

		Cases Used in Analysis	
USAGE	Prior	Unweighted	Weighted
Non User	.367	36	36.000
User	.633	62	62.000
Total	1.000	98	98.000

ADW615: ADVANCED BUSINESS STATISTICS FORMULAE

Mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Variance: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum x^2 - n\bar{x}^2}{n-1}$

Sampling Distribution: $\mu_{\bar{x}} = \mu \quad \mu_{\hat{p}} = p$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \sigma_{\hat{p}} = \sqrt{\frac{p(1\text{-}p)}{n}}$

Confidence Interval: $\bar{x} \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$ or $\bar{x} \pm t_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$ or $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p}}{n}}$

Sample Sizes: $n = \left(\frac{z_{\alpha/2}\sigma}{B}\right)^2$ **or** $n = \left(\frac{t_{\alpha/2,n-1}s}{B}\right)^2$ **or** $n = \left(\frac{z_{\alpha/2}\sqrt{p(1-p)}}{B}\right)^2$

Test of Differences:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}};\cdots if \cdots \sigma_1^2 \neq \sigma_2^2 \qquad v = \frac{\left[(\frac{s_x^2}{n_x}) + (\frac{s_y^2}{n_y})\right]^2}{(\frac{s_x^2}{n_x})^2/(n_x - 1) + (\frac{s_y^2}{n_y})^2/(n_y - 1)}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sigma_1 = \sigma_2 \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \cdots\cdots df = n_1 + n_2 - 2$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \cdots if \cdots D = 0 \qquad z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}} \cdots if \cdots D \neq 0$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \cdots \hat{p}_1 = \frac{x_1}{n_1} \cdots \hat{p}_2 = \frac{x_2}{n_2}$$

Paired Sample t-test:

$$t = \frac{\overline{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \qquad S_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}}$$

Wilcoxon Rank Sum Test: $z = \frac{T - \mu_T}{\sigma_T} \quad \mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \cdots\cdots \sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$

1-way ANOVA

$$F = \frac{MST}{MSE}$$

$$MST = \frac{SST}{k-1} \cdots\cdots MSE = \frac{SSE}{n-k}$$

$$SST = \sum_{j=1}^{k} n_j(\overline{x}_j - \overline{\overline{x}})^2 \cdots\cdots SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2$$

$$\overline{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \cdots\cdots \overline{\overline{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n}$$

Kruskal-Wallis 1-way ANOVA $H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{T_j^2}{n_j}\right] - 3(n+1)$

Randomized Block

$$F = \frac{MST}{MSE}$$

$$MST = \frac{SST}{k-1} \cdots\cdots MSE = \frac{SSE}{n-k-b+1}$$

ANOVA: $SST = \sum_{j=1}^{k} n_j(\overline{x}_{Tj} - \overline{\overline{x}})^2 \cdots\cdots SSE = \sum_{j=1}^{k}\sum_{i=1}^{b}(x_{ij} - \overline{x}_{Tj} - \overline{x}_{Bi} + \overline{\overline{x}})^2$

$$SSB = \sum_{i=1}^{b} m_i(\overline{x}_{bi} - \overline{\overline{x}})^2 \cdots \overline{x}_{Tj} = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \cdots\cdots \overline{x}_{Bi} = \frac{\sum_{j=1}^{m_i} x_{ij}}{m_i} \cdots \overline{\overline{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n}$$

Chi-Squared Tests

$$\chi^2_{k-1} = \sum \frac{(O_i - E_i)^2}{E_i} \text{ where } O_i = \textit{observed frequency};\ E_i = \textit{expected frequency}(=T_i / N)$$

$$\chi^2_{rc-1} = \sum\sum \frac{(O_{ii} - E_{ij})^2}{E_{ij}} \text{ where } O_i = \textit{observed frequency};\ E_i = \textit{expected frequency}(=R_i C_j / N)$$

Simple Regression

$$y = \beta_0 + \beta_1 x \quad \hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$SS_{xy} = \sum (x - \bar{x})(y - \bar{y}) \quad SS_{xx} = \sum (x - \bar{x})^2 = \sum x^2 - n\bar{x}^2$$

$$SST = \sum (y - \bar{y})^2 = \sum y^2 - n\bar{y}^2 \quad SSE = \sum (y - \hat{y})^2 = SS_{yy} - \frac{SS_{xy}}{SS_{xx}}$$

$$SSR = SST - SSE = \sum (\hat{y} - \bar{y})^2 = \frac{SS^2_{xy}}{SS_{xx}}$$

Coefficient of Determination: $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

Standard Error of Estimate: $\hat{\sigma}_e = \sqrt{\frac{SSE}{n-2}}$

Discriminant Analysis

***Maximum Chance Criterion*:** Max($p_1, p_2, p_3, \ldots, p_k$)

***Proportional Chance Criterion*:** $p_1^2 + p_2^2 + p_3^2 + \ldots + p_k^2$

Press Q-Statistic: $Q = \frac{[N - nk]^2}{N(k-1)}$ n = no. of correct classification; N = total number

$$Q \approx \chi^2_1$$

Cutting Score for 2 groups:

$\frac{n_a C_A + n_B C_B}{n_A + n_B}$ n_A = no. in group A; C_A = centroid for group A

Potency Index for i-th Variable

$$P_i = L^2_{i1}\left(\frac{\lambda_1}{\lambda}\right) + L^2_{i2}\left(\frac{\lambda_2}{\lambda}\right) + \ldots + L^2_{ik}\left(\frac{\lambda_k}{\lambda}\right)$$

$\lambda = \sum \lambda_i$ L_{ij} = loading of variable i on discriminant function j

- ooo O ooo -