UNIVERSITI SAINS MALAYSIA

Master of Business Administration

Second Semester Examination
Academic Session 2007/2008

April 2008

AGW615 - Advanced Business Statistics

Duration: 3 hours

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

Answer any **FIVE** questions. Each question carries 20 marks.

Note:

- Scientific calculator is allowed.
- Statistical table and graph sheets should be obtained only from the invigilator.

Question 1

(a) The descriptive statistics table for Engine Displacement in cubic inches of 406 cars are given below:

Descriptive Statistics

Engine Displacement (cu. inches)

N	406
Mean	194.04
Std. Error of Mean	5.221
Median	148.50
Mode	97
Std. Deviation	105.207
Variance	11068.589
Skewness	.692
Std. Error of Skewness	.121
Kurtosis	-.791
Std. Error of Kurtosis	.242
Range	451
Minimum	4
Maximum	455

Answer the following questions from the above output:

(i) Do the engine displacement measurements follow a normal distribution? Justify your answer.
(ii) What is the standard error of mean? How do you get 5.221?
(iii) Write the formula to compute the measure of skewness and interpret the value.
(iv) From the Kurtosis value, describe the frequency curve of the data.
(v) Do the 3-sigma limits cover the minimum and maximum range of the data?

[10 marks]

(b) Explain in detail the following terms with suitable examples:

(i) Experiments and Random experiments
(ii) Mutually exclusive events and independent events

[4 marks]

(c) A Professor wants to crack exactly 3 jokes per year throughout his service period of 35 years. He made a policy that out of 3 jokes, 2 jokes may be overlapping but one joke must be distinct. How many jokes does he require for 35 years? Justify your answer.

[3 marks]

(d) From the following discrete probability distribution, compute mean and variance:

Demand per day	: 0	1	2	3	4	5
Probability	: 0.02	0.08	0.25	0.35	0.21	0.09

[3 marks]

Question 2

(a) An inspector of the Alaska Pipeline has the task of comparing the reliability of two pumping stations. Each station is susceptible to two kinds of failure: pump failure and leakage. When either (or both) occurs, the station must be shut down. The data at hand indicate that the following probabilities prevail:

Station	**P(Pump Failure)**	**P(Leakage)**	**P(Both)**
1.	**0.06**	**0.12**	**0.02**
2.	**0.09**	**0.07**	**0.04**

Which station has the higher probability of being shut down?

[4 marks]

(b) Draw the frequency curve for Poisson distribution and explain briefly the applications of it.

[3 marks]

(c) Glenn Howell, VP of personnel for the Standard Insurance Company, has developed a new training program that is entirely self-paced. New employees work various stages at their own pace; completion occurs when the material is learned. Howell's program has been especially effective in speeding up the training process, as an employee's salary during training is only 67 percent of that earned upon completion of the program. In the last several years, average completion time of the program was 42 days, and the standard deviation was 10 days.

(i) Using the normal distribution, find the probability that an employee will finish the program in 31 to 44 days.

(ii) What is the probability of finishing the program below 32 days?

[6 marks]

(d) What are the primary goals of Six Sigma Methodologies? Explain in detail the benefits of Six Sigma implementation in an industry.

[7 marks]

Question 3

Realtors are often interested in seeing how the appraised value of a home varies according to the size of the home. Some data on area (in thousands of square feet) and appraised value (in thousands of dollars) for a sample of 11 homes follow:

Area (X):	1.1	1.5	1.6	1.6	1.4	1.3	1.1	1.7	1.9	1.5	1.3
Value (Y):	75	95	110	102	95	87	82	115	122	98	90

(a) Draw the scatter diagram and comment on your findings.

[2 marks]

(b) Obtain the correlation coefficient between X and Y and the coefficient of determination.

[4 marks]

(c) What do you infer from the coefficient of determination?

[1 mark]

(d) Fit a regression line of Y on X.

[7 marks]

(e) Estimate Y when X = 2.1

[1 mark]

(f) Obtain the trend values and draw the original and trend values on the graph.

[2 marks]

(g) Find the standard error of the regression estimate.

[3 marks]

Question 4

(a) The gold prices for 25 grams(in US$) from July 1999 to June 2006 have been analyzed using the curve fitting models and the output is given in Table 4.1.

Table 4.1

Independent: Time

Dependent	Mod	Rsq	d.f.	F	Sigf	b0	b1	b2	b3
us$	LOG	.489	82	78.61	.000	114.005	70.5825		
us$	INV	.109	82	10.08	.002	373.249	-242.16		
us$	QUA	.928	81	524.87	.000	284.592	-1.5730	.0589	
us$	CUB	.931	80	359.80	.000	271.144	.2712	.0050	.0004
us$	COM	.874	82	568.71	.000	236.083	1.0092		
us$	POW	.555	82	102.07	.000	178.124	.1934		
us$	GRO	.874	82	568.71	.000	5.4642	.0092		
us$	EXP	.874	82	568.71	.000	236.083	.0092		

Answer the following questions:

(i) Out of the 8 models fitted, which is the best fit for the data? Justify your answer.
(ii) Write the mathematical equation of the best fit.
(iii) Using the best fit, estimate the Gold price in US$ for April 2008.
(iv) What is degree of freedom (d.f.)?
(v) Why d.f. varies with respect to models although the underlying data is the same?

[10 marks]

(b) The multiple regression model has been used to study the linear relationship between Income in Rupees (1RM=Rupees 11.8) and age in years, computer skills (average=1, above average=2), number of years of service, durations of service in the present job in years and number of Business Processing Outsourcing (BPO's) worked. The SPSS output is summarized in Tables 4.2 – 4.4.

Table 4.2

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.781[a]	.610	.596	3649.061	.610	44.961	5	144	.000

a. Predictors: (Constant), NUMBER OF BPO'S WORKED, COMPUTER SKILLS, AGE IN YEARS, DURATION OF SERVICE IN TH PRESENT JOB, NUMBER OF YEARS OF SERVICE

Table 4.3

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-396.666	2538.366		-.156	.876
	AGE IN YEARS	-101.380	97.689	-.061	-1.038	.301
	COMPUTER SKILLS	6774.752	651.126	.571	10.405	.000
	NUMBER OF YEARS OF SERVICE	1257.498	380.850	.249	3.302	.001
	DURATION OF SERVICE IN THE PRESENT JOB	1577.683	495.660	.205	3.183	.002
	NUMBER OF BPO'S WORKED	146.070	447.053	.019	.327	.744

a. Dependent Variable: INCOME IN RUPEES

Table 4.4

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions					
				(Constant)	AGE IN YEARS	COMPUTER SKILLS	NUMBER OF YEARS OF SERVICE	DURATION OF SERVICE IN THE PRESENT JOB	NUMBER OF BPO'S WORKED
1	1	5.423	1.000	.00	.00	.00	.00	.01	.00
	2	.276	4.429	.00	.00	.01	.01	.63	.04
	3	.167	5.694	.01	.01	.07	.02	.00	.65
	4	.072	8.668	.00	.00	.08	.76	.28	.27
	5	.054	10.037	.04	.06	.72	.10	.09	.00
	6	.007	27.016	.95	.92	.12	.12	.00	.03

a. Dependent Variable: INCOME IN RUPEES

Answer the following questions:

(i) What are the underlying assumptions for fitting a multiple regression model?
(ii) What is the value of the unexplained variation from Table 4.2?
(iii) Write the fitted multiple regression model from Table 4.3 and comment on the model.
(iv) Does Table 4.4 reveal co-linearity in the data? Explain.
(v) Identify the significant independent variable(s) influencing the dependent variable.

[10 marks]

Question 5

(a) It is of interest to know whether the average processing time in minutes for the similar type of jobs for an IT project remains the same or not. Towards this, 10 arbitrary processing times of Operator-1 and 11 samples of Operator -2 have been taken for the study in a week. The two-sample independent t-test has been applied under the usual assumptions and the outputs are given in Tables 5.1 and 5.2 as follows:

Table 5.1

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
PROCESSING TIME IN MINUTES	OPERATOR-1	10	26.5000	2.54951	.80623
	OPERATOR-2	11	22.7273	1.90215	.57352

Table 5.2

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
PROCESSING TIME IN MINUTES	Equal variances assumed	.667	.424	3.868	19	.001	3.77273	.97537	1.73125	5.81420
	Equal variances not assumed			3.813	17.000	.001	3.77273	.98941	1.68132	5.86413

Answer the following questions:

(i) Can you conclude from Table 5.1 that Operator-2 is taking less processing time than Operator-1on the average?

(ii) Why do you require Levene's test in Table 5.2? Draw the conclusions from Levene's test.

(iii) When do you apply t-test? Write the inference for t-test.

(iv) Interpret 95% confidence interval for the difference of means.

(v) Can we use paired t-test instead of an un-paired t-test? Justify your answer.

[10 marks]

(b) The pulse rate of 18 dogs were measured after a particular surgery at 5 different time intervals via half an hour after surgery, two hours, four hours, six hours and twenty four hours after surgery. It is of interest to test whether the population pulse rate of dogs is the same for all the time intervals. On the basis of this, the completely randomized design is carried out and the following output has been obtained using SPSS and is provided in Tables 5.3 and 5.4.

Table 5.3

ANOVA

Pulse Rate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3789.667	4	947.417	3.018	.022
Within Groups	26681.889	85	313.905		
Total	30471.556	89			

Table 5.4 : Post Hoc Tests

Multiple Comparisons

Dependent Variable: Pulse Rate

Tukey HSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Pulse rate of dogs Half an hour after Surgery	Two Hours after Surgery	8.500	5.906	.604	-7.96	24.96
	Four Hours after Surgery	11.000	5.906	.345	-5.46	27.46
	Six Hours after Surgery	4.944	5.906	.918	-11.52	21.41
	Twenty Four Hours after Surgery	19.444*	5.906	.012	2.98	35.91
Two Hours after Surgery	Pulse rate of dogs Half an hour after Surgery	-8.500	5.906	.604	-24.96	7.96
	Four Hours after Surgery	2.500	5.906	.993	-13.96	18.96
	Six Hours after Surgery	-3.556	5.906	.974	-20.02	12.91
	Twenty Four Hours after Surgery	10.944	5.906	.351	-5.52	27.41
Four Hours after Surgery	Pulse rate of dogs Half an hour after Surgery	-11.000	5.906	.345	-27.46	5.46
	Two Hours after Surgery	-2.500	5.906	.993	-18.96	13.96
	Six Hours after Surgery	-6.056	5.906	.843	-22.52	10.41
	Twenty Four Hours after Surgery	8.444	5.906	.610	-8.02	24.91
Six Hours after Surgery	Pulse rate of dogs Half an hour after Surgery	-4.944	5.906	.918	-21.41	11.52
	Two Hours after Surgery	3.556	5.906	.974	-12.91	20.02
	Four Hours after Surgery	6.056	5.906	.843	-10.41	22.52
	Twenty Four Hours after Surgery	14.500	5.906	.111	-1.96	30.96
Twenty Four Hours after Surgery	Pulse rate of dogs Half an hour after Surgery	-19.444*	5.906	.012	-35.91	-2.98
	Two Hours after Surgery	-10.944	5.906	.351	-27.41	5.52
	Four Hours after Surgery	-8.444	5.906	.610	-24.91	8.02
	Six Hours after Surgery	-14.500	5.906	.111	-30.96	1.96

*. The mean difference is significant at the .05 level.

Answer the following questions:

(i) Write the null and alternative hypotheses for the problem.
(ii) What is the impact of surgery on pulse rate?
(iii) Why do we use F-distribution in ANOVA? How do you compute F-value as 3.018 in Table 5.3?
(iv) What does the significance value indicates?
(v) What are your conclusions from Post Hoc test by Tukey given in Table 5.4?

[10 marks]

Question 6

(a) The details on the purchasing pattern of consumer durables have been obtained from 400 customers in a research project. The categorical principal component analysis has been applied to study the most influencing factors for purchasing pattern. The significant findings of the analysis are displayed in Tables 6.1-6.2 and in Fig.6.1.

Table 6.1

Model Summary

Dimension	Cronbach's Alpha	Variance Accounted For	
		Total (Eigenvalue)	% of Variance
1	.669	2.466	27.400
2	.542	1.929	21.428
Total	.869[a]	4.394	48.828

a. Total Cronbach's Alpha is based on the total Eigenvalue.

Table 6.2

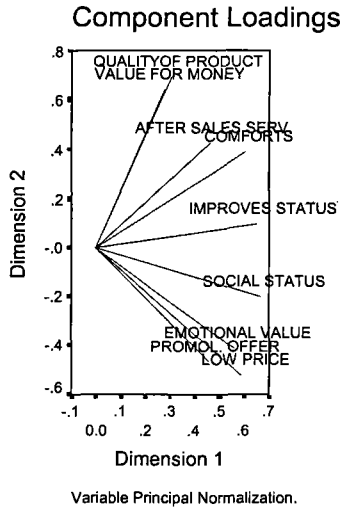
Component Loadings

Component Loadings

	Dimension	
	1	2
PROMOL. OFFER	.455	-.466
LOW PRICE	.588	-.519
EMOTIONAL VALUE	.555	-.412
VALUE FOR MONEY	.275	.642
COMFORTS	.603	.390
AFTER SALES SERV.	.457	.425
SOCIAL STATUS	.666	-.200
IMPROVES STATUS	.649	.096
QUALITYOF PRODUCT	.303	.691

Variable Principal Normalization.

Fig. 6.1



Variable Principal Normalization.

...10/-

Answer the following questions:

(i) Why do we apply Cronbach's Alpha in Table 6.1 and what is the inference you draw from it?
(ii) What is Principal Component? How do you obtain the component loadings in Table 6.2?
(iii) What will happen if we increase the dimensions in Table 6.2?
(iv) From Fig 6.1, what are the significant factor(s) influencing the purchasing pattern of consumer durables?
(v) Summarize the entire results of Categorical Principal Component analysis.

[10 marks]

(b) About 27 patients suffering from depression were administered a standardized questionnaire to measure four types of disabilities namely Occupational, Marital, Personal and Social on a 10-point ordinal rating scale. The higher the scores the patients secure indicates the greater the disability. The non-parametric tests have been applied and the results are summarized in Tables 6.3and 6.4.

Table 6.3 : Mann-Whitney Test

Ranks

	Gender	N	Mean Rank	Sum of Ranks
OCCUPATIONAL DISABILITY	Female	12	6.63	79.50
	Male	15	19.90	298.50
	Total	27		
MARITAL DISABILITY	Female	12	13.88	166.50
	Male	15	14.10	211.50
	Total	27		

Test Statistics [b]

	OCCUPATIONAL DISABILITY	MARITAL DISABILITY
Mann-Whitney U	1.500	88.500
Wilcoxon W	79.500	166.500
Z	-4.382	-.075
Asymp. Sig. (2-tailed)	.000	.940
Exact Sig. [2*(1-tailed Sig.)]	.000 [a]	.943 [a]

a. Not corrected for ties.

b. Grouping Variable: Gender

Table 6.4 : Kruskal-Wallis Test

Ranks

	GROUP	N	Mean Rank
DISABILITY SCORES	PERSONALITY DISABILITY	27	46.41
	SOCIAL DISABILITY	27	43.89
	OCCUPATIONAL DISABILITY	27	60.22
	MARITAL DISABILITY	27	67.48
	Total	108	

Test Statistics[a,b]

	DISABILITY SCORES
Chi-Square	10.680
df	3
Asymp. Sig.	.014

a. Kruskal Wallis Test

b. Grouping Variable: GROUP

Answer the following questions:

(i) When do you apply Mann-Whitney test?

(ii) Can you conclude from Table 6.3 that males and females have the same occupational disability scores?

(iii) Can you conclude from Table 6.3 that males have more marital disability scores than females?

(iv) When do you apply Kruskal Wallis test?

(v) Is there any evidence from Table 6.4 to show that the median scores of all types of disabilities have the same distribution?

[10 marks]

- ooo O ooo -