OLLSCOIL na hÉIREANN, GAILLIMH
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SUMMER EXAMINATIONS, 2001
M.B.S. EXAMINATION

Research Methodology MG 525
Part A: Statistics

Prof. Paul Willman
Prof. Roy Green
Ms. Mary Quinn

Time allowed: THREE hours
Answer FOUR questions.
All questions carry equal marks

Notes: (i) Appendix of Statistical formulae supplied as part of this Paper
(ii) Mathematics Tables and New Cambridge Statistical Tables are available to all Candidates.

Question 1.

(i) Moneymaker Corporation has several subsidiaries. Moneymaker's strategic planner believes that expenditure on advertising can be a useful predictor of total sales. As an aid to long term planning, she gathers the following sales and advertising information from several of Moneymaker's companies for the year 2000 (in hundred thousands of pounds).

Company	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Advertising (x) (£ 00,000)	12.5	3.7	21.6	60.0	37.6	6.1	16.8	41.2	52.4
Sales (y) (£ 00,000)	148	55	338	994	541	89	126	379	781

Given $\bar{x} = 27.99$ $\quad\quad$ $\bar{y} = 383.44$

$s_x = 20.5$ $\quad\quad$ $s_y = 330.5$

$\Sigma(x^2) = 10412.91$ $\quad\quad$ $r = 0.958$

$\Sigma(y_i - \hat{y}_i)^2 = 71556.5$ $\quad\quad$ $b_0 = \bar{y} - b_1\bar{x}$

Given that the assumptions underlying linear regression analysis are fulfilled, and given also the partial printout from the regression analysis below:

Predictor	**Coeff**	**Stand. Dev.**	**t-ratio**	**p-value**
Constant	-48.937	59.307	-.825	0.437
Advertising	15.448	1.744	8.860	0.000

Analysis of Variance

Source	**d.f.**	**s.s.**	**m.s.**	**F-ratio**	**p-value**
Regression		802465.695		78.501	0.000
Error					
Total					

(a) What is the regression equation?
(b) What does the regression equation tell you?
(c) Determine if there is a significant linear relationship between advertising and sales.
(d) Complete the ANOVA table. What does the ANOVA table tell you?
(e) What is the value of the coefficient of determination? How do you interpret this?
(f) What are the assumptions underlying regression referred to above? (You may use an illustration to explain if you wish).
(g) What is the predicted value of sales
    (i) when the total advertising expenditure is £480,000?
    (ii) when the total advertising expenditure is £7,200,000?
    Comment on your results.

(80 Marks)

Cont'd....

.... Cont'd

(ii) In what circumstances would a researcher typically consider carrying out a regression analysis and what would the typical objective(s) be in carrying out a regression analysis in the circumstances you have described?

(20 marks)
[Total 100 Marks]

Question 2.

(i) Suppose the average cost of a particular surgical procedure is £30,000 per patient. The costs are normally distributed, with a standard deviation of £9,000.

(a) What proportion of the procedures cost between £15,000 and £40,000?
(b) What proportion of the procedures cost in excess of £50,000?
(c) Suppose the standard deviation was unknown, but it was known that 90.82% of the costs were more than £7,000. What would be the value of the standard deviation?
(d) Suppose the mean cost was unknown but it was known that the standard deviation was £9,000. What would the average cost be if 79.95% of the costs were less than £33,000?

(50 Marks)
Cont'd ....

... Cont'd

(ii) Suppose an insurance company wants to investigate whether there is a difference in the average estimates of value (for insurance purposes) being given by three different jewelry firms. They select five different valuable pieces of jewelry and ask the three firms to give estimates of the value of each of the five pieces. The table below gives the estimates obtained (in hundreds of pounds)

		Appraisers (Jewelry Firms)		
		A	**B**	**C**
Individual	**1**	78	82	79
Pieces of	**2**	102	102	99
Jewelry	**3**	68	74	70
	4	83	88	86
	5	95	99	92

An incomplete computer printout of the analysis of variance for this experiment is given below.

Analysis of Variance

Source	**ss**	**df**	**ms**	**F-ratio**
Jewelry Firms	48.13			
Individual Piece				
Error	22.53			
Total	1829.73			

(a) Complete the ANOVA table.
(b) Identify which is the blocking variable and which is the treatment variable.
(c) Explain what the terms "treatment variable" and "blocking variable"mean.
(d) Develop the appropriate null hypotheses and alternative hypotheses, primary and secondary.
(e) Test your null hypotheses (primary and secondary) against the appropriate alternative hypotheses at $\alpha = 0.05$ and explain what these results mean.
(f) What conclusions do you think the management of the insurance company can draw from these results and what further investigation or action might they undertake?

(50 Marks)
[Total 100 marks]

Question 3.

(i) A researcher into social behaviour patterns has conducted a survey which includes information on a range of lifestyle behaviours and choices. One of his interests is in possible links between respondents' attitude to smoking and their health status. Health status is evaluated on a three-point scale depending on past and current health history. Attitude to smoking is also measured on a three-point scale. The table below gives data on the results of the cross-tabulation of these two items for the 350 survey respondents.

	Health Status		
Attitude to Smoking	**Very Healthy**	**Moderately Healthy**	**Unhealthy**
Completely Intolerant	**30**	**110**	**30**
Somewhat Indifferent	**50**	**60**	**20**
Completely Permissive	**20**	**20**	**10**

On the basis of these data, what, if anything, can be concluded regarding attitude to smoking and health status? Use a significance level of 5% for your any test you do.

(60 Marks)

(ii) Random samples taken from two shifts at the Hitech Industries factory in Galway are given below. They indicate the mean number of completed units of product from the Gizmo department over several days for each shift. The two shifts employ equal numbers of staff and equipment.
Do the data provide sufficient evidence to indicate that the two shifts are equally productive? Use $\alpha = 0.01$.

Shift 1	$n_1 = 12$	$\overline{x}_1 = 32.1$	$s_1^2 = 144$
Shift 2	$n_2 = 9$	$\overline{x}_2 = 29.6$	$s_2^2 = 133$

(40 Marks)

[Total 100 Marks]

Question 4.

(i) An MBS student is carrying out a research project involving a large-scale survey on attitudes of chief executives in a particular region to trade union recognition. She is aware of the possibility of the survey respondents not being truly representative of the underlying population. She knows it is important that the respondents are representative of the base population in relation to their known features, particularly features which may be linked to their attitude to trade union recognition.

One such feature is the size of the organisations the senior managers surveyed work in, as measured by the number of employees in the organization. Latest data from the Department of Enterprise and Employment indicate that the current distribution of organisations in the region by number of people employed is as follows:

	Number Employed in Organisation					
	0-25	**26-50**	**51-200**	**201-500**	**500-1000**	**1001+**
No. of Orgs in Region	78	52	47	32	11	8

The respondent data is as follows:

	Number Employed in Organisation					
	0-25	**26-50**	**51-200**	**201-500**	**500-1000**	**1001+**
No. of Respondents	18	17	12	7	6	3

What test would you advise her to carry out to ascertain whether the respondents are representative of the underlying population on this criterion?

Carry out the test that you feel is appropriate, indicating clearly what conclusion you would draw from the result. Use $\alpha = 0.05$.

(50 Marks)

(ii) A furniture removals company has found that 9% of its contracts give rise to claims for damage to material or good in transit.

(a) In a random sample of 20 contracts, what is the probability that there will be fewer than 4 claims for damage?
(b) What is the probability that there are exactly 6 claims?
(c) Calculate the mean and variance of this distribution.

(30 Marks)

(iii) When computer packages for the Chi-square test are used for 2x2 contingency analysis, they usually give two values – without continuity correction and with continuity correction.

(a) Why is this?
(b) Which one would you advise using, and why?

(20 Marks)
[Total 100 Marks]

Question 5

(i) A pilot sample (sample size = 25) yields a variance of 85. The variance computed from a later more comprehensive survey, with a sample size of 120, is 125. Do these results indicate that the estimate of the pilot-sample variance may have been too low? Use $\alpha = 0.05$.

(20 Marks)

(ii) A local politician who almost lost his seat in the last election four years ago is undecided whether he should retire from politics altogether. To inform his decision, he wants to discover whether he has lost popularity since the last election, when he got 17.6% of the votes.
He has commissioned a survey of 350 randomly chosen voters. Of these, 74 indicated their intention to vote for him.
What conclusion can he draw from this survey? Use $\alpha = 0.05$.
Would the conclusion be the same if you carried out the test using $\alpha = 0.01$?
What advice would you give him in relation to the decision he is trying to make?

(30 Marks)

(iii) A consumer association wants to calculate a measure of the relationship between the prices of different brands of face make-up and the length of time the different brands last before needing re-application. The price for 20 grams of 12 different brands is given below, together with the ranking of the 12 brands in order of their time to re-application. (1 is shortest, 12 the longest)

Brand	**A**	**B**	**C**	**D**	**E**	**F**
Price (£)	7.90	8.50	12.35	4.50	3.99	7.99
Ranking	6	9	11	2	3	8

Brand	**G**	**H**	**I**	**J**	**K**	**L**
Price (£)	6.25	9.25	6.75	5.50	4.55	5.75
Ranking	5	12	7	10	1	4

(a) Compute an appropriate correlation coefficient and explain what this actually means with respect to the above data.
(b) Indicate why you have chosen this particular correlation coefficient
(c) Test whether the correlation coefficient you have calculated is significant at the 0.01 level and explain what this result means.

(50 marks)
[Total 100 Marks]

Question 6

(i) Explain briefly the multiplicative model in time series analysis. In your explanation, include a brief note on each term in the model.

(20 Marks)

(ii) Given below are the quarterly sales of ice cream in litres in the Happy Times ice cream parlour and sweet shop in High Street, Ballytown.

(a) Using these data, compute the 4-quarter centred moving average.
(b) Which part of the model referred to in (i) above does this moving average describe?
(c) Explain to the proprietor what this 4-quarter centred moving average is telling her.
(d) If you are now told that the seasonal indices are 95%, 103%, 109% and 91.7% for the first, second, third and fourth quarters respectively, explain to the proprietor what these are telling her.
(e) Calculate the seasonally adjusted figures for 2000 and explain to the proprietor what they mean for the business.

Sales in Litres

	1997	**1998**	**1999**	**2000**
Quarters				
1	3107	3168	3227	3287
2	3253	3411	3565	3595
3	3596	3587	3842	3963
4	2986	3084	3178	3224

(60 Marks)

(iii) In hypothesis testing, what is meant by
(a) Type I error
(b) Type II error

(20 Marks)
[Total 100 Marks]