SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 1998

SURVEYS, SECONDARY ANALYSIS AND SOCIAL STATISTICS

Candidates should answer THREE questions, including at least ONE from Section A and at least ONE from Section B. In Section A candidates are required to provide commentaries on their answers.

Time allowed: 2 hours

Read carefully the instructions on your answerbook and make sure that the particulars required are entered on each answerbook.

Approved calculators may be used

SECTION A

1   The mean duration of pre-marital cohabitation in a random sample of 289 married couples in London (who cohabited before marriage) was found to be 25.6 months, with a sample standard deviation of 25.5 months.

   (i)   Calculate a 95% confidence interval for the mean duration of pre-marital cohabitation among married couples in London.

   (ii)   The mean duration of pre-marital cohabitation for all British married couples is known to be 21.1 months. Calculate a z-statistic and use it to test whether this is a plausible mean duration of pre-marital cohabitation for married couples in London.

   (iii) If the population standard deviation for the mean duration of pre-marital cohabitation for married couples in Scotland is assumed to be 18.0 months, how big a sample would be needed to produce a sample mean that one could be 95% confident fell within 3 months of the population mean duration of pre-marital cohabitation for married couples in Scotland? Comment on your answer in relation to your answer to part (i) of this question.

   Explain how and why your answer to (i) enables you to answer part (ii) without calculating a z-statistic.

2   The following table shows the relationship between cohabitation before marriage and (stated) religion for a random sample of 150 married people aged less than 40 years.

Cohabited before marriage

	Yes	No	Total
Religion			
Church of England	22	18	40
Other	15	35	50
None	38	22	60
TOTAL	75 (50%)	75 (50%)	150

(i)   Calculate the chi-square statistic for the above table and use it to test the hypothesis that there is no relationship between cohabitation before marriage and religion for married people aged less than 40 years.

(Note: the critical value at the 5% level of a chi-square statistic with 2 degrees of freedom is 5.99).

(ii)   A similar table based on a random sample of 250 married people aged 40 years or more, once again showing the relationship between cohabitation before marriage and religion, gave rise to a chi-square statistic of 4.1. Use Cramer's V to compare the strengths of the relationships in the two tables, and explain why the two chi-square statistics could not have been used for this purpose.

(iii)   Does the above table suggest that the probability of having cohabited before marriage differs between married people aged less than 40 years whose stated religion is Church of England and married people aged less than 40 years who state that they do not have a religion?

(Note: the critical value at the 5% level of a chi-square statistic with 1 degree of freedom is 3.84).

3   The following table shows the mean number of dependent children in a woman's household according to her (own) occupational class for a random sample of 254 British women aged 30 years.

Occupational class	Mean	s	n
I and II	0.79	1.00	54
IIINM	1.24	1.12	108
IIIM and IV	1.54	1.08	75
V	2.23	1.48	17
		TOTAL	254

(Occupational class is defined as the Registrar General's Social Class of the woman's own current or last occupation).

(s is sample standard deviation; n is sample size).

(i)   Test the hypothesis that, in the population, 30 year-old women of all occupational classes have the same mean number of dependent children in their households. Discuss your findings with reference to the sample means.

(Note: the critical value of F at the 5% level corresponding to 3 degrees of freedom and 250 degrees of freedom is 2.64; the between-groups and within-groups sums of squares are 35.4 and 295.0 respectively).

(ii)   Test the hypothesis that, in the population, 30 year-old women in occupational classes I and II and 30 year-old women in occupational class IIINM have the same mean number of dependent children in their households.

(Note: the critical value of t at the 5% level corresponding to 160 degrees of freedom is 1.98; the pooled sample standard deviation for women in occupational classes I and II and women in occupational class IIINM is 1.08).

4   In a random sample of 19 higher education institutions in 1994, the (Pearson) correlation between the total number of students accepted onto degree courses and the percentage of students accepted onto degree courses who were aged 30 years or older was found to be -0.57.

    (i)   Test the hypothesis that there is no relationship between the total number of students and the percentage of students aged 30 years or older.

          (Note: $(0.57)^2 = 0.32$; the critical value of F at the 5% level corresponding to 1 degree of freedom and 17 degrees of freedom is 4.45).

The regression equation corresponding to the dependence of the percentage of students aged 30 years or older on the total number of students is

$$y = 14.5 - 3.1x_1$$

where y is the percentage of students aged 30 years or older, and x_1 is the total number of students (in thousands).

    (ii)   Use the above equation to predict the percentage of students accepted who were aged 30 years or older for a university accepting a total number of 5 thousand students. Is the predicted value useful? Why might the above <u>linear</u> regression equation be an inappropriate model of the relationship between the percentage of students aged 30 years or older and the total number of students? Warwick University accepted a total of 2.3 thousand students in 1994. Use the above equation to predict the percentage of students accepted by Warwick who were aged 30 years or older. The actual figure for Warwick was 3.2%. Why do you think it is that the prediction differs from this actual figure?

The addition to the regression analysis of a second independent variable, x_2, which is a dummy variable taking the value 1 if the institution is an 'old' university and the value 0 otherwise, results in the following equation

$$y = 14.9 - 2.6x_1 - 2.3x_2$$

    (iii)   Explain why the coefficient of x_1, the total number of students accepted, changes between the two equations.

5   The following crosstabulation is of sex [S] by educational level [D] by parental class [C] for a random sample of 3,214 British adults. (The parental class variable is father's Registrar General's Social Class, and the educational level variable distinguishes between graduates and non-graduates):

CLASS = I	Degree	No degree	TOTAL
Male	32	59	91
Female	20	87	107
TOTAL	52	146	198
CLASS = II or IIINM	Degree	No degree	TOTAL
Male	82	416	498
Female	73	469	542
TOTAL	155	885	1040
CLASS = IIIM TO V	Degree	No degree	TOTAL
Male	59	904	963
Female	22	991	1013
TOTAL	81	1895	1976

(i)   Use odds ratios to summarise the way in which the relationship between sex and educational level varies according to parental class. The chi-square statistics for the three sub-tables are 6.9, 1.8 and 19.6. Using these chi-square statistics, test the relationship in each sub-table for significance.

(ii)   Use odds ratios to summarise the relationships between sex and parental class, and between educational level and parental class.

(iii) Use the following results corresponding to the fit of various log-linear models to determine the most appropriate model of the table given above. Justify your choice, and, given the model that you have selected, comment on your findings in parts (i) and (ii)

(Note: the critical value at the 5% level of a chi-square statistic with 2 degrees of freedom is 5.99; the critical value at the 5% level of a chi-square statistic with 1 degree of freedom is 3.84).

Model No.	Model	Deviance	d.f.	P	Change in deviance	d.f.	P	Compared to model
1	[C] [S] [D]	188.7	7	0.000				
2	[SD] [C]	171.0	6	0.000	17.7	1	0.000	1
3	[CD] [S]	29.7	5	0.000	159.0	2	0.000	1
4	[CS] [D]	188.0	5	0.000	0.7	2	0.720	1
5	[CD] [SD]	12.0	4	0.017	17.7	1	0.000	3
6	[CS] [SD]	170.4	4	0.000	0.7	2	0.720	2
7	[CS] [CD]	29.0	3	0.000	0.7	2	0.720	3
8	[CS] [CD] [SD]	8.7	2	0.013	3.4	2	0.190	5
9	[CSD]	0.0	0		8.7	2	0.013	8

SECTION B

6   'The issues faced by secondary analysts are more similar to those faced by other quantitative researchers than they are different from them.' Discuss, with particular reference to the secondary analysis of data from **ONE** existing social survey.

7   'An examination of the processes of concept operationalization and questionnaire design reveals survey research to be an inherently theoretical and essentially social activity.' Discuss, with reference to **ONE** hypothetical survey of your choice.

8   'The most important feature of both social survey interviews and qualitative interviews is a shared one: the interviewer herself or himself.' Discuss.

9   'Combining quantitative and qualitative approaches within a single research project in order to reap the benefits of both styles of research is less sensible and more difficult than it sounds.' Discuss.

10   Critically discuss the following table. Your discussion should include:
     * an account of what you would like to know about the data collection process and the sample;
     * a consideration of the validity of the variables as indicators of underlying concepts;
     * a description of the substantive relationship visible in the table;
     * an outline of how the analysis needs to be extended and/or could be elaborated.
     [Note: You may assume that the overall relationship in the table is statistically significant; you should specify any more focused statistical tests that you would ideally like to carry out].

NEWSPAPER READ BY POSITION ON A LEFT/RIGHT POLITICAL SCALE
Left/Right Political Scale

Newspaper read	Score = Low		Score = Medium		Score = High	
	%		%		%	
None	95	(30.3)	135	(43.0)	84	(26.8)
Daily Mail	10	(17.2)	29	(50.0)	19	(32.8)
Daily Mirror	44	(37.9)	53	(45.7)	19	(16.4)
Sun	19	(17.8)	61	(57.0)	27	(25.2)
Other	63	(28.0)	89	(39.6)	73	(32.4)

[Note: Low scores on the 'Left/Right Political Scale' supposedly correspond to a politically 'left-wing' viewpoint; high scores on the scale supposedly correspond to a politically 'right-wing' viewpoint].

END