SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 2000

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 from first marriage to divorce in a random sample of 400 people in Britain who cohabited with their marriage partner before marriage and whose first marriages ended in divorce was found to be 86.6 months, with a sample standard deviation of 60.0 months.

   (i)   Calculate a 95% confidence interval for the mean duration from first marriage to divorce for people who cohabited before marriage and whose first marriages ended in divorce.

   (ii)   The mean duration from first marriage to divorce for <u>all</u> people whose first marriages ended in divorce is known to be 129.5 months. Calculate a z-statistic and use it to test whether this is a plausible mean duration for people <u>who cohabited before marriage</u> and whose first marriages ended in divorce.

   (iii) Assume that the population standard deviation for the mean duration from first marriage to divorce for <u>all</u> people whose first marriages ended in divorce is 84.0 months. How big a sample would be needed to produce a sample mean that one could be 95% confident fell within 6 months of the population mean duration from first marriage to divorce for <u>all</u> people whose first marriages ended in divorce? 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 ethnicity and housing tenure for a random sample of 384 adults living in London.

	Owner occupier	Other	Total
Ethnic group			
'Black'	14	10	24
'Asian'	28	8	36
'Irish'	14	22	36
'White'	200	88	288
TOTAL	256 (66.67%)	128 (33.33%)	384

Note: 'Owner occupier' included ownership via a mortgage; ethnicity is based on self-identification in relation to a broader range of ethnic categories than those shown above.

(i)   Calculate the chi-square statistic for the above table and use it to test the hypothesis that there is no relationship between ethnicity and housing tenure for adults living in London.

(Note: the critical value at the 5% level of a chi-square statistic with 3 degrees of freedom is 7.82).

(ii)   A similar table based on a random sample of 3,400 adults living in Britain (excluding London), once again showing the relationship between ethnicity and housing tenure, gave rise to a chi-square statistic of 27.2. 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, in London, housing tenure differs between the three minority ethnic groups listed (i.e. between 'Black', 'Asian' and 'Irish' adults)?

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

3   The following table shows the mean age (in years) at leaving full-time education according to degree subject for a random sample of 1,605 graduates in Britain.

Degree subject	Mean	s	n
Sciences	23.04	5.14	574
Social Studies	24.15	5.62	256
Business Studies	21.90	4.99	144
Arts & Humanities	24.03	5.38	284
Other	23.73	5.71	347
		TOTAL	1,605

(The degree subject categories cover all possible degree courses; e.g. a Physics degree would be counted within the Sciences category).

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

(i)   Test the hypothesis that, in the population, graduates in each degree subject category have the same mean age at leaving full-time education. Discuss your findings with reference to the sample means.

(Note: the critical value of F at the 5% level corresponding to 4 degrees of freedom and 1,600 degrees of freedom is 2.37; the between-groups and within-groups sums of squares are 756 and 50,400 respectively).

(ii)   Test the hypothesis that, in the population, graduates with Social Studies degrees and graduates with Business Studies degrees have the same mean age at leaving full-time education

(Note: the critical value of t at the 5% level corresponding to 398 degrees of freedom is 1.97; the pooled sample standard deviation for graduates with Social Studies degrees and graduates with Business Studies degrees is 5.40).

4   In a random sample of 45 urban areas in Britain, the (Pearson) correlation between the long-term illness rate (i.e. the percentage of the population with a long-term illness) and the percentage of women in the population was found to be 0.37.

    (i)   Test the hypothesis that there is no relationship in urban areas between the long-term illness rate and the percentage of women.

        (Note: $(0.37)^2 = 0.14$; the critical value of F at the 5% level corresponding to 1 degree of freedom and 43 degrees of freedom is 4.07).

The regression equation corresponding to the dependence of the long-term illness rate on the percentage of women in the population is

$$y = 1.5x_1 - 62.5$$

where y is the long-term illness rate (%), and x_1 is the percentage of women in the population.

    (ii)   Use the above equation to predict the long-term illness rates for a (hypothetical) urban area containing equal numbers of men and women (50%) and for a (hypothetical) urban area containing markedly less women than men (40%). Is the second predicted value useful? What does the above equation suggest that the long-term illness rate would be in an urban area in which the population was 100% female? The long-term illness rate in urban areas is rarely lower than 8%. What does this suggest about the percentage of women in urban areas?

The addition to the regression analysis of a second independent variable, x_2, which is the percentage of the population aged 75 years or over, leads to the following equation

$$y = 1.1x_1 + 0.5x_2 - 47.0$$

    (iii)   Explain why the coefficient of x_1, the percentage of women in the population, changes between the two equations.

4

5   The following crosstabulation is of marital status [M] by sex [S] by age [A] for a random sample of 1,170 graduates in Britain. (In this crosstabulation marital status focuses on whether or not a graduate has ever been married).

AGE = 30-39 years	Female	Male	TOTAL
Never married	64	87	151
Ever married	151	269	420
TOTAL	215	356	571
AGE = 40-49 years	Female	Male	TOTAL
Never married	16	25	41
Ever married	109	250	359
TOTAL	125	275	400
AGE = 50-59 years	Female	Male	TOTAL
Never married	11	11	22
Ever married	46	131	177
TOTAL	57	142	199

(i)   Use odds ratios to summarise the way in which the relationship between marital status and sex varies according to age. The chi-square statistics for the three sub-tables are 2.0, 1.3 and 5.5. Using these chi-square statistics, test the relationship in each sub-table for significance.

(ii)   Use odds ratios to summarise the relationships between:
(a) marital status and age; (b) sex and age.

(iii)   Use the following results corresponding to the goodness-of-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	[A] [M] [S]	66.4	7	0.000				
2	[MS] [A]	58.0	6	0.000	8.4	1	0.004	1
3	[AS] [M]	59.0	5	0.000	7.4	2	0.025	1
4	[AM] [S]	15.6	5	0.008	50.8	2	0.000	1
5	[AS] [MS]	50.6	4	0.000	8.4	1	0.004	3
6	[AM] [MS]	7.2	4	0.128	8.4	1	0.004	4
7	[AM] [AS]	8.2	3	0.042	7.4	2	0.025	4
8	[AM][AS][MS]	2.4	2	0.301	4.8	2	0.093	6
9	[AMS]	0.0	0		2.4	2	0.301	8

SECTION B

6   Critically assess the merits and limitations of carrying out secondary analyses of existing data, with particular reference to ONE social survey of your choice.

7   'An effective questionnaire or interview schedule contains well-designed questions which act as valid indicators of the full range of concepts of interest to the researcher'. Discuss, with reference to ONE hypothetical survey of your choice.

8   Does the need for standardisation in survey interviews mean that they are fundamentally different to qualitative interviews?

9   Are there always compelling arguments for combining quantitative and qualitative approaches within a single research project?

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].

SOCIAL CLASS by MORAL TRADITIONALISM
Level of moral traditionalism (scale)

	Score = Low		Score = Medium		Score = High	
Social Class	%		%		%	
Salariat	96	(36.0)	130	(48.7)	41	(15.4)
Routine non-manual	67	(24.8)	150	(55.6)	53	(19.6)
Petty bourgeoisie	18	(20.0)	48	(53.3)	24	(26.7)
Higher manual	23	(15.8)	84	(57.5)	39	(26.7)
Working class	47	(19.0)	139	(56.0)	62	(25.0)

[Note: Low scores on the 'Level of moral traditionalism' scale correspond to 'modern' views on moral issues relating to sexual relations, crime and gender roles; high scores on the scale correspond to 'traditional' views on such issues].

END