SO 2010

UNIVERSITY OF WARWICK

September Examinations 1999

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 age at first heterosexual intercourse in a random sample of 100 male graduates born in the early 1960s was found to be 18.53 years, with a sample standard deviation of 2.50 years.

(i) Calculate a 95% confidence interval for the mean age at first heterosexual intercourse among male graduates born in the early 1960s.

(ii) The mean age at first heterosexual intercourse for all men born in the early 1960s is known to be 17.08 years. Calculate a z-statistic and use it to test whether this is a plausible mean age at first heterosexual intercourse for male graduates born in the early 1960s.

(iii) If the population standard deviation for the age at first heterosexual intercourse of female graduates born in the early 1960s is assumed to be 2.15 years, how big a sample would be needed to produce a sample mean that one could be 95% confident fell within 0.5 years of the population mean age at first heterosexual intercourse of female graduates born in the early 1960s? 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.

Continued.....

2 The following table shows the relationship between the importance of religion/religious beliefs to respondents and self-reported health for a random sample of 960 adult women in Britain.

	Self-reported health			
	Good	Average	Poor	Total
Religion/religious beliefs				
Very important	120	26	14	160
Fairly important	238	66	16	320
Not important	362	100	18	480
TOTAL	720 (75%)	192 (20%)	48 (5%)	960

(i) Calculate the chi-square statistic for the above table and use it to test the hypothesis that there is no relationship between the importance of religion/religious beliefs and self-reported health for adult women in Britain.

(Note: the critical value at the 5% level of a chi-square statistic with 4 degrees of freedom is 9.49).

(ii) A similar table based on a random sample of 750 adult men in Britain, once again showing the relationship between the importance of religion/religious beliefs and self-reported health, gave rise to a chi-square statistic of 3.4. 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 likelihood of 'poor' self-reported health (as compared to 'good' or 'average' self-reported health) varies according to the importance of religion/religious beliefs to adult women in Britain?

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

Continued.....

3 The following table shows the mean occupational status scores according to marital status of a random sample of 3,205 British men aged 40-49.

Marital status	Mean	s	n
Single	27.4	18.0	330
Married	31.5	18.0	2,250
Remarried	30.9	16.8	353
Divorced	29.4	18.1	250
Widowed	25.7	15.0	22
		TOTAL	3,205

(A high score indicates high occupational status).

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

(i) Test the hypothesis that, in the population, men of each marital status have the same mean occupational status score. 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 3,200 degrees of freedom is 2.22; the between-groups and within-groups sums of squares are 5,200 and 1,040,000 respectively).

(ii) Test the hypothesis that, in the population, married men and divorced men have the same mean occupational status score. (Note: 'married' here refers to the Married category only, not to the Remarried category).

(Note: the critical value of t at the 5% level corresponding to 2,498 degrees of freedom is 1.96; the pooled sample standard deviation for divorced men and married men is 18.0).

Continued.....

4 In a random sample of 9,993 adults in Britain, the (Pearson) correlation between (average) number of cigarettes smoked per day and age (in years) was found to be 0.03.

(i) Test the hypothesis that there is no relationship between number of cigarettes smoked per day and age (in years).

(Note: $(0.03)^2 = 0.0009$; the critical value of F at the 5% level corresponding to 1 degree of freedom and 9,991 degrees of freedom is 3.84).

The regression equation corresponding to the dependence of number of cigarettes smoked per day on age (in years) is

$$y = 4.80 + 0.02x_1$$

where y is the number of cigarettes smoked per day, and x_1 is age (in years).

(ii) Use the above equation to predict the number of cigarettes smoked per day by someone of 20 years of age and by someone of 50 years of age. Comment on the difference between the two predicted values. Use the above equation to predict the number of cigarettes smoked per day by a child of 10 years of age. Is this third predicted value useful? Why is the above linear regression equation likely to be an inappropriate model of the relationship between the two variables once children are brought into consideration?

The addition to the regression analysis of a second independent variable, x_2, which is the (average) number of units of alcohol consumed per week, leads to the following equation

$$y = 3.50 + 0.01x_1 + 0.20x_2$$

(iii) Explain why the coefficient of x_1, age (in years), changes between the two equations. Use the second regression equation to predict the number of cigarettes smoked per day by someone of 20 years of age who consumes 0 (zero) units of alcohol per week and by someone of 20 years of age who consumes 20 units of alcohol per week. Comment on the predicted values obtained.

Continued.....

5 The following crosstabulation is of class of own occupation [C], coded as non-manual or manual, by limiting long-term illness [I] by age [A] for a random sample of 2,517 working women in Britain. ('Yes' indicates that a woman has a limiting long-term illness; 'No' indicates that she does not).

AGE = 16-34 years	Yes	No	TOTAL
Non-manual	12	749	761
Manual	11	394	405
TOTAL	23	1143	1166
AGE = 35-49 years	Yes	No	TOTAL
Non-manual	24	479	503
Manual	26	264	290
TOTAL	50	743	793
AGE = 50-64 years	Yes	No	TOTAL
Non-manual	32	269	301
Manual	30	227	257
TOTAL	62	496	558

(i) Use odds ratios to summarise the way in which the relationship between class and limiting long-term illness varies according to age. The chi-square statistics for the three sub-tables are 1.8, 5.5 and 0.2. Using these chi-square statistics, test the relationship in each sub-table for significance.

(ii) Use odds ratios to summarise the relationships between:
(a) class and age; (b) limiting long-term illness 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**	**Comp -ared to model**
1	[C] [I] [A]	93.6	7	0.000				
2	[CI] [A]	84.6	6	0.000	9.0	1	0.003	1
3	[IA] [C]	28.7	5	0.000	64.9	2	0.000	1
4	[CA] [I]	72.8	5	0.000	20.8	2	0.000	1
5	[CI] [IA]	19.8	4	0.001	8.9	1	0.003	3
6	[CI] [CA]	63.8	4	0.000	9.0	1	0.003	4
7	[CA] [IA]	8.0	3	0.047	20.7	2	0.000	3
8	[CI] [CA] [IA]	2.5	2	0.286	5.5	1	0.020	7
9	[CIA]	0.0	0		2.5	2	0.286	8

Continued.....

SECTION B

6 'The value of a secondary analysis depends upon the nature and quality of the data available'. Discuss, with particular reference to ONE existing social survey.

7 Discuss, with reference to ONE hypothetical survey of your choice, the key roles played by the operationalization of concepts and questionnaire design in effective and conceptually adequate research.

8 What are the main shared features of, and key differences between, qualitative interviews and social survey interviews?

9 Under what circumstances is it appropriate and useful to use both quantitative and qualitative methods 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].

HIGHEST QUALIFICATION by POLITICAL KNOWLEDGE

Political Knowledge (scale)

	Score = Low		Score = Medium		Score = High	
Highest qual.		%		%		%
None	507	(31.3)	722	(44.6)	390	(24.1)
'O' level#	167	(24.8)	279	(41.5)	227	(33.7)
'A' level#	119	(13.1)	331	(36.6)	455	(50.3)
Degree	3	(1.2)	41	(16.1)	211	(82.7)
Other	10	(18.9)	28	(52.8)	15	(28.3)

[Notes: Low scores on the 'Political Knowledge' scale correspond to limited political knowledge; high scores on the scale correspond to extensive political knowledge; # = 'Or equivalent'].

END