SO 2010

UNIVERSITY OF WARWICK

September Examinations 2001

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 income (per adult; in thousands of pounds) in the households of a random sample of 169 adult women in Britain who attended private (fee-paying) schools was found to be 14.95, with a sample standard deviation of 9.75.

   (i)   Calculate a 95% confidence interval for the mean income (per adult) in the households of adult women in Britain who attended private (fee-paying) schools.

   (ii)   The mean income (per adult) in the households of adult <u>men</u> in Britain who attended private (fee-paying) schools is known to be 16.30. Calculate a z-statistic and use it to test whether this is a plausible mean for the income (per adult) in the households of adult women in Britain who attended private (fee-paying) schools

   (iii)   Suppose that the population standard deviation for the income (per adult) in the households of <u>all</u> adult women in Britain is assumed to be 8.25. How big a sample would be needed to produce a sample mean that one could be 95% confident fell within 1.50 of the population mean income (per adult) in the households of <u>all</u> adult women in Britain? 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 cross-tabulation shows the relationship between ethnic group and marital status for a random sample of 5,580 adult women in Britain.

Ethnic group	Never married	Ever married	Total
'White'	870	4530	5400
'Black'	33	39	72
'Asian'	15	57	72
'Other'	12	24	36
TOTAL	930 (16.67%)	4650 (83.33%)	5580

'Ever married' includes women who are currently cohabiting; ethnic group 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 cross-tabulation and use it to test the hypothesis that there is no relationship between ethnic group and marital status for adult women in Britain.

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

(ii)   A similar cross-tabulation, based on a random sample of 4,800 adult <u>men</u> in Britain, and again showing the relationship between ethnic group and marital status, gave rise to a chi-square statistic of 27.0. Use Cramer's V to compare the strengths of the relationships in the two cross-tabulations, and explain why the two chi-square statistics could not have been used for this purpose.

(iii)   Does the above cross-tabulation suggest that, in Britain, the proportion of adult women who have never married varies significantly between the 'Black' and 'Asian' ethnic groups?

(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 population density (in persons per hectare), according to highest educational qualification, of the local areas of residence of a random sample of 2,954 adults in Britain.

Highest qualification	Mean	s	n
Degree	32.35	37.56	278
'A' level (or equivalent)	22.85	24.30	716
'O' level (or equivalent)	23.40	24.13	560
None of the above	26.15	25.34	1,400
		TOTAL	2,954

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

(i)   Test the hypothesis that, in the population, the local areas of residence of adults at each educational level have the same mean population density. 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 2,950 degrees of freedom is 2.60; the between-groups and within-groups sums of squares are 21,606 and 2,042,875 respectively).

(ii)   Test the hypothesis that, in the population, adults with none of the qualifications listed and adults with 'O' levels (or some equivalent) live in areas with the same mean population density.

(Note: the critical value of t at the 5% level corresponding to 1,958 degrees of freedom is 1.96; the pooled sample standard deviation for adults with none of the qualifications listed and adults with 'O' levels, or some equivalent, is 25.00).

4   In a random sample of 589 women aged 18-59 in Britain, the (Pearson) correlation between their usual hours of paid employment per week and the number of children in their household was found to be -0.24.

(i)   Test the hypothesis that there is no relationship between usual hours of paid employment per week and the number of children in the household.

(Note: the critical value of F at the 5% level corresponding to 1 degree of freedom and 587 degrees of freedom is 3.86).

The regression equation corresponding to the dependence of usual hours of paid employment per week on the number of children in the household is

$$y = 22.00 - 4.63x_1$$

where y is usual hours of paid employment per week, and x_1 is the number of children in the household.

(ii)   Use the above equation to predict the usual hours of paid employment per week for a woman with two children in her household and the usual hours of paid employment per week for a woman with six children in her household. Is the second predicted value useful? Why might the above <u>linear</u> regression equation be an inappropriate model of the relationship between the two variables?

The addition to the regression analysis of a second independent variable, x_2, relating to the presence in the household of a child of under 5 years of age (with 1 = Yes and 0 = No), leads to the following equation

$$y = 22.43 - 3.24x_1 - 8.51x_2$$

(iii)   Explain why the coefficient of x_1, the number of children in the household, changes between the two equations. Use the second regression equation to predict the usual hours of paid employment per week of a woman with two children in her household, neither of whom is under 5, and to predict the usual hours of employment per week of a woman with two children in her household, at least one of whom is under 5. Comment on the results.

5   The following cross-tabulation is of an individual's age in years [A] by whether they voted in the last local election [V] by region [R] for a random sample of 1,574 adults in Britain.

REGION = South	Did not vote	Voted	TOTAL
Aged under 30	87	59	146
Aged 30 or more	265	406	671
TOTAL	352	465	817
REGION = North	Did not vote	Voted	TOTAL
Aged under 30	61	29	90
Aged 30 or more	150	292	442
TOTAL	211	321	532
REGION = London	Did not vote	Voted	TOTAL
Aged under 30	18	17	35
Aged 30 or more	67	123	190
TOTAL	85	140	225

(i)   Use odds ratios to summarise the way in which the relationship between whether an individual voted in the last local election and their age varies according to region. The chi-square statistics for the three sub-tables are 19.7, 35.8 and 3.3. Using these chi-square statistics, test the relationship in each sub-table for significance.

(ii)   Use odds ratios to summarise the relationships between:
(a) age and region; (b) whether voted in the last local election and region.

(iii)   Use the following results corresponding to the goodness-of-fit of various log-linear models to determine the most appropriate model of the cross-tabulation 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	[R] [A] [V]	61.4	7	0.000				
2	[RA] [V]	60.7	5	0.000	0.7	2	0.697	1
3	[RV] [A]	58.6	5	0.000	2.8	2	0.245	1
4	[AV] [R]	7.7	6	0.256	53.7	1	0.000	1
5	[RA] [RV]	57.9	3	0.000	0.7	2	0.697	3
6	[AV] [RA]	7.0	4	0.134	0.7	2	0.697	4
7	[AV] [RV]	4.9	4	0.293	2.8	2	0.245	4
8	[RA][RV][AV]	4.6	2	0.099	0.3	2	0.851	7
9	[AVR]	0.0	0		4.6	2	0.099	8

SECTION B

6 What are the constraints and benefits of carrying out secondary analyses using an existing data source? Discuss, with particular reference to ONE social survey of your choice.

7 To what extent does the success of a survey depend on processes of concept operationalization and questionnaire design? Discuss, with reference to ONE hypothetical survey of your choice.

8 'Qualitative interviews and survey interviews use similar processes to generate different forms of data'. Discuss.

9 Under what circumstances should researchers combine the use of quantitative and qualitative methods within a single project?

10 Critically discuss the following cross-tabulation. 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 cross-tabulation is statistically significant; you should specify any more focused statistical tests that you would ideally like to carry out].

CLASS by DEGREE OF PESSIMISM ABOUT THE FUTURE
Degree of pessimism about the future (scale)

Class	Score = Low	%	Score = Medium	%	Score = High	%
Salariat	63	(21.8)	189	(65.4)	37	(12.8)
Routine non-manual	37	(16.6)	153	(68.6)	33	(14.8)
Petty bourgeoisie	8	(11.3)	43	(60.6)	20	(28.2)
Higher manual	7	(12.7)	39	(70.9)	9	(16.4)
Working class	32	(12.5)	167	(65.0)	58	(22.6)

[Note: Low scores on the 'Degree of pessimism about the future' scale correspond to a more optimistic view about what life in Britain will be like in the future; high scores on the scale correspond to a more pessimistic view about what life in Britain will be like in the future].

END