SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 2007

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 length of current relationship for a random sample of 144 single (never married) British adults with regular but non-resident partners and higher educational qualifications was found to be 61.1 months, with a sample standard deviation of 96.0.

    (i)   Calculate a 95% confidence interval for the mean length of current relationship for single British adults with regular but non-resident partners and higher educational qualifications.

    (ii)   The mean length of current relationship for all single British adults with regular but non-resident partners is known to be 41.9 months. Calculate a z-statistic and use it to test whether this is a plausible mean length of current relationship for single British adults with regular but non-resident partners and higher educational qualifications.

    (iii)   Suppose that the population standard deviation for the mean length of current relationship for <u>formerly married</u> British adults with regular but non-resident partners and higher educational qualifications is assumed to be 136.0. How big a sample would be needed to produce a sample mean that one could be 95% confident fell within 16 months of the population mean length of current relationship for formerly married British adults with regular but non-resident partners and higher educational qualifications? 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 in a random sample of 800 British men between occupational class and whether an individual spent their childhood (up to age 16) living with both their natural parents.

	Yes	No	Total
Occupational class			
Professional/Manager	224	56	280
Intermediate .	33	7	40
Petty Bourgeoisie	95	25	120
Supervisor/Technician	86	34	120
Semi-routine/Routine	162	78	240
TOTAL	600 (75%)	200 (25%)	800

(i)   Calculate the chi-square statistic for the above cross-tabulation and use it to test the hypothesis that there is no relationship for British men between occupational class and whether an individual spent their childhood (up to age 16) living with both their natural parents.

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

(ii)   A similarly shaped cross-tabulation, based on a random sample of 1001 British women, and again showing the relationship between occupational class and whether an individual spent their childhood (up to age 16) living with both their natural parents, gave rise to a chi-square statistic of 4.55. Use Cramér's V to compare the strengths of the relationships in the two cross-tabulations, and explain why the values of the two chi-square statistics could not have been used for this purpose.

(iii)   Does the above cross-tabulation suggest that the likelihood of a British man having spent their childhood (up to age 16) living with both their natural parents varies significantly between men in professional/managerial occupations, men in intermediate occupations and men in the petty bourgeoisie? Calculate a chi-square statistic to answer this part of the question, and comment on the result in relation to the pattern in the cross-tabulation as a whole.

(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 number of people in an individual's workplace, according to Registrar General's Social Class, for a random sample of 3,291 adults in employment in Britain.

Social class	Mean	s	n
I	271.57	307.82	143
II	200.32	271.42	928
III Non-manual	198.34	260.50	860
III Manual	210.25	267.59	684
IV	197.78	259.93	468
V	169.73	243.84	208
		TOTAL	3,291

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

(i)   Test the hypothesis that, in the population, the mean number of people in an individual's workplace corresponding to each social class is the same. Discuss your findings with reference to the sample means.

(Note: the critical value of F at the 5% level corresponding to 5 degrees of freedom and 3,285 degrees of freedom is 2.22; the between-groups and within-groups sums of squares are 976,250 and 233,235,000 respectively).

(ii)   Test the hypothesis that, in the population, the mean numbers of people in an individual's workplace for individuals in Social Class IV and Social Class V are the same.

(Note: the critical value of t at the 5% level corresponding to 674 degrees of freedom is 1.96; the pooled sample standard deviation for individuals in Social Class IV and Social Class V is 255.00).

4   In a random sample of 572 employees in Britain, the (Pearson) correlation between gross earnings (in £1000s) and hours per week spent watching television was found to be –0.224.

(i)   Test the hypothesis that there is no relationship between gross earnings (in £1000s) and hours per week spent watching television among employees in Britain.

(Note: You may assume that $(-0.224)^2$ is equal to 0.050; the critical value of F at the 5% level corresponding to 1 degree of freedom and 570 degrees of freedom is 3.86.)

The regression equation corresponding to the dependence of hours per week spent watching television on gross earnings for employees in Britain is

$$y = 25.02 - 0.26x_1$$

where y is the number of hours per week spent watching television, and x_1 is gross earnings.

(ii)   Use the above equation to predict the numbers of hours per week spent watching television for three individuals whose gross earnings are £5,000, £25,000 and £100,000. Is the third 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, which corresponds to hours usually worked per week, leads to the following equation

$$y = 27.93 - 0.21x_1 - 0.09x_2$$

(iii)   Explain why the coefficient of x_1, gross earnings, changes between the two equations. Use the second regression equation to predict the numbers of hours per week spent watching television for the three individuals from part (ii), assuming that they work for (a) 40 hours, and (b) 20 hours. Comment on these predictions with reference the predictions made for part (ii).

5   The following cross-tabulation is of income [I] by housing tenure [T] by region [R] for a random sample of 1,569 adults in England.

REGION = North	Owns	Rents	TOTAL
Low income	129	71	200
Medium/High income	457	129	586
TOTAL	586	200	786
REGION = South	Owns	Rents	TOTAL
Low income	85	42	127
Medium/High income	367	104	471
TOTAL	452	146	598
REGION = London	Owns	Rents	TOTAL
Low income	21	15	36
Medium/High income	93	56	149
TOTAL	114	71	185

(i)   Use odds ratios to summarise the way in which the relationship between income and housing tenure varies according to region. The chi-square statistics for the three sub-tables are 14.30, 6.55 and 0.20. Using these chi-square statistics, test the relationship in each sub-table for significance.

(ii)   Use odds ratios to summarise the relationships between:
(a) region and income; (b) region and housing tenure.

(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	Compared to model
1	[R] [I] [T]	39.3	7	0.000				
2	[RT] [I]	25.1	5	0.000	14.2	2	0.001	1
3	[IT] [R]	21.8	6	0.001	17.5	1	0.000	1
4	[RI] [T]	34.3	5	0.000	5.0	2	0.082	1
5	[RI] [IT]	16.8	4	0.002	17.5	1	0.000	4
6	[RT] [IT]	7.6	4	0.108	14.2	2	0.001	3
7	[RI] [RT]	20.1	3	0.000	14.2	2	0.001	4
8	[RI][RT][IT]	1.4	2	0.489	6.2	2	0.046	6
9	[RIT]	0.0	0		1.4	2	0.489	8

Continued.....

SECTION B

6   Critically assess, with particular reference to ONE social survey of your choice, the extent to which a secondary analyst's use of existing data can constrain their ability to address research questions thoroughly and validly.

7   Discuss, with reference to ONE real or hypothetical survey of your choice, the conceptual and practical challenges that researchers face when operationalizing concepts and designing research instruments.

8   To what extent do qualitative researchers and survey researchers need to address the same issues to ensure that interviews are effective?

9   What issues does a researcher need to consider when deciding whether or not to use both quantitative methods and qualitative methods within a single research 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].

AGE LEFT EDUCATION SYSTEM by DEGREE OF RACIAL PREJUDICE

Degree of racial prejudice (scale)

Age left educ. sys.	Score = Low		Score = Medium		Score = High	
	%		%		%	
15 or under	271	(35.0)	404	(52.1)	100	(12.9)
16 or 17	373	(48.8)	342	(44.7)	50	(6.5)
18 to 20	136	(52.9)	109	(42.4)	12	(4.7)
21 or over	145	(65.0)	73	(32.7)	5	(2.2)
Other	26	(60.5)	16	(37.2)	1	(2.3)

[Notes: Low scores on the degree of racial prejudice scale correspond to little or no prejudice; high scores on the scale correspond to a more extensive degree of prejudice].

END