

SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 2004

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 number of cigarettes smoked per day by a random sample of 196 male graduates aged 30-39 living in British households was found to be 3.1, with a sample standard deviation of 7.0.
- (i) Calculate a 95% confidence interval for the mean number of cigarettes smoked per day by male graduates aged 30-39 living in British households.
 - (ii) The mean number of cigarettes smoked per day by all men aged 30-39 living in British households is known to be 6.4 cigarettes. Calculate a z-statistic and use it to test whether this is a plausible mean for the number of cigarettes smoked per day by male graduates aged 30-39 living in British households.
 - (iii) Suppose that the population standard deviation for the number of cigarettes smoked per day by female graduates aged 30-39 living in British households is assumed to be 4.5. How big a sample would be needed to produce a sample mean that one could be 95% confident fell within 1 cigarette of the population mean number of cigarettes smoked per day by female graduates aged 30-39 living in British households? 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 cross-tabulation shows the relationship between religious denomination and trade union membership for a random sample of 4,870 employees in Britain.

	<u>Not a member of a trade union</u>	<u>Member of a trade union</u>	<u>Total</u>
<u>Religious denomination</u>			
None	1,250	750	2,000
Church of England	915	660	1,575
Roman Catholic	285	240	525
Other Christian	387	258	645
Other non-Christian	85	40	125
TOTAL	2,922 (60%)	1,948 (40%)	4,870

- (i) Calculate the chi-square statistic for the above cross-tabulation and use it to test the hypothesis that there is no relationship between religious denomination and trade union membership for employees 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 similarly shaped cross-tabulation, based on the same random sample of 4,870 employees in Britain, and showing the relationship between ethnic group and trade union membership, gave rise to a chi-square statistic of 7.81. 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 also have been used for this purpose.
- (iii) Does the above cross-tabulation suggest that, in Britain, trade union membership varies significantly between employees who are members of the Church of England and employees who are members of the Roman Catholic church? (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 1 degree of freedom is 3.84).

Continued.....

- 3 The following table shows the mean length of marriage (in months), according to the (main) reason why the marriage ended, for a random sample of 1,367 previously married lone mothers in Britain.

<u>Reason marriage ended</u>	<u>Mean</u>	<u>s</u>	<u>n</u>
Adultery	114.7	66.4	388
Spouses did not 'get on'	88.6	64.1	225
Violence	78.6	60.4	180
Alcohol and/or drugs	100.4	66.0	102
Lack of communication	100.9	71.6	105
Money problems	86.2	58.9	88
Other reason	89.5	63.7	279
		TOTAL	1,367

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

- (i) Test the hypothesis that, in the population, the mean length of marriage corresponding to each different reason for the marriage ending is the same. Discuss your findings with reference to the sample means.

(Note: the critical value of F at the 5% level corresponding to 6 degrees of freedom and 1,360 degrees of freedom is 2.10; the between-groups and within-groups sums of squares are 226,800 and 5,712,000 respectively).

- (ii) Test the hypothesis that, in the population, the mean lengths of marriage for lone mothers whose marriages ended because they did not get on with their spouses and for lone mothers whose marriages ended because of violence are the same.

(Note: the critical value of t at the 5% level corresponding to 403 degrees of freedom is 1.96; the pooled sample standard deviation for lone mothers whose marriages ended because they did not get on with their spouses and lone mothers whose marriages ended because of violence is 62.5).

Continued.....

4 In a random sample of 866 mothers in paid employment in Britain, the (Pearson) correlation between monthly earnings and number of children was found to be -0.20.

- (i) Test the hypothesis that there is no relationship between monthly earnings and number of children among mothers in paid employment in Britain.

(Note: the critical value of F at the 5% level corresponding to 1 degree of freedom and 864 degrees of freedom is 3.85).

The regression equation corresponding to the dependence of monthly earnings on number of children among mothers in paid employment in Britain is

$$y = 750.6 - 118.4x_1$$

where y is monthly earnings, and x_1 is number of children.

- (ii) Use the above equation to predict the monthly earnings of mothers in paid employment in Britain who have two, four and seven children respectively. Is the third predicted value useful? Why might the above linear 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 of paid work usually worked per week, leads to the following equation

$$y = -34.3 - 19.4x_1 + 25.7x_2$$

- (iii) Explain why the coefficient of x_1 , number of children, changes between the two equations. Use the second regression equation to predict the monthly earnings of a mother with two children, who does 22.5 hours of paid work per week, and the monthly earnings of a mother with four children, who does 15 hours of paid work per week. Comment on these predictions with reference to the predictions made for part (ii).

Continued.....

- 5 The following cross-tabulation is of sex [S] by whether or not a person is active in organisations [O] by social class of current or last job [C] for a random sample of 7,150 adults in households in Britain.

CLASS = Service class	<u>Active</u>	<u>Not active</u>	<u>TOTAL</u>
Male	742	425	1167
Female	<u>575</u>	<u>448</u>	<u>1033</u>
TOTAL	1317	873	2190
CLASS = Intermediate class	<u>Active</u>	<u>Not active</u>	<u>TOTAL</u>
Male	540	502	1042
Female	<u>800</u>	<u>940</u>	<u>1740</u>
TOTAL	1340	1442	2782
CLASS = Working class	<u>Active</u>	<u>Not active</u>	<u>TOTAL</u>
Male	533	686	1219
Female	<u>297</u>	<u>662</u>	<u>959</u>
TOTAL	830	1348	2178

- (i) Use odds ratios to summarise the way in which the relationship between sex and activity within organisations varies according to social class. The chi-square statistics for the three sub-tables are 12.4, 8.9 and 37.0. Using these chi-square statistics, test the relationship in each sub-table for significance.
- (ii) Use odds ratios to summarise the relationships between:
(a) sex and social class; (b) activity in organisations and social class.
- (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	[C] [S] [O]	478.6	7	0.000				
2	[CO] [S]	264.1	5	0.000	214.5	2	0.000	1
3	[SO] [C]	432.6	6	0.000	46.0	1	0.000	1
4	[CS] [O]	273.1	5	0.000	205.5	2	0.000	1
5	[CO] [SO]	218.1	4	0.000	46.0	1	0.000	2
6	[CS] [SO]	227.1	4	0.000	46.0	1	0.000	4
7	[CS] [CO]	58.6	3	0.000	205.5	2	0.000	2
8	[CS][CO][SO]	7.3	2	0.027	51.3	1	0.000	7
9	[CSO]	0.0	0		7.3	2	0.027	8

Continued....

SECTION B

- 6 Critically assess, with particular reference to ONE social survey of your choice, the advantages and disadvantages of secondary analysis-based research.
-
- 7 Discuss, with reference to ONE real or hypothetical survey of your choice, the roles played by processes of concept operationalization and questionnaire design in generating valid data that allow researchers to achieve their theoretical objectives.
-
- 8 Do differences with respect to features such as standardisation and structure mean that survey interviews and qualitative interviews are fundamentally different data collection methods?
-
- 9 Does using both quantitative and qualitative methods within a single research project necessarily bring particular benefits?
-
- 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].

MEN'S SOCIAL CLASS by PARTICIPATION IN HOUSEHOLD TASKS
Participation in household tasks (scale)

	Score = Low		Score = Medium		Score = High	
<u>Social class</u>	%		%		%	
Salariat	260	(28.6)	460	(50.6)	189	(20.8)
Routine non-manual	39	(30.7)	61	(48.0)	27	(21.3)
Petty bourgeoisie	75	(21.1)	189	(53.2)	91	(25.6)
Working class	298	(31.3)	471	(49.4)	184	(19.3)
Unemployed	66	(31.7)	92	(44.2)	50	(24.0)

[Notes: Low scores on the 'Participation in household tasks' scale correspond to limited (or no) participation; high scores on the scale correspond to higher levels of participation. The petty bourgeoisie category consists of self-employed men].

 END