

SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 2009

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 answer book and make sure that the particulars required are entered on each answer book.

Approved calculators may be used.

SECTION A

1. The mean annual income (in £1000s) earned by a random sample of 144 cohabiting women (in paid employment) in Britain was found to be 20.2, with a sample standard deviation of 15.0.
 - (i) Calculate a 95% confidence interval for the mean annual income of cohabiting women (in paid employment) in Britain.
 - (ii) The mean annual income of married women (in paid employment) in Britain is 17.6. Calculate a z-statistic and use it to test whether this is a plausible mean annual income for cohabiting women (in paid employment) in Britain.
 - (iii) Suppose that the population standard deviation for the annual income of cohabiting men (in paid employment) in Britain is assumed to be 17.5. How big a sample would be needed to produce a sample mean that one could be 95% confident fell within £2,500 of the population mean annual income of cohabiting men (in paid employment) 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.

Continued

-
2. The following cross-tabulation shows the relationship in a random sample of 1,620 women in Britain between occupational class (based on own occupation) and whether a woman owns (or has access to) a bicycle.

	<u>Bicycle: Yes</u>	<u>Bicycle: No</u>	<u>Total</u>
<u>Occupational class</u>			
Managerial/Professional	228	312	540
Intermediate	84	204	288
Small employer/Own account	30	78	108
Lower supervisory/Technical	48	96	144
Semi-routine/Routine	150	390	540
TOTAL	540 (33.33%)	1080 (66.67%)	1620

- (i) Calculate the chi-square statistic for the above cross-tabulation and use it to test the hypothesis that there is no relationship for women in Britain between occupational class and owning a bicycle.

(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 1,490 men in Britain, and again showing the relationship between occupational class and owning a bicycle, gave rise to a chi-square statistic of 51.07. 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, in Britain, the likelihood of a woman owning a bicycle varies significantly between women with occupations in the managerial/professional class and women with occupations in all the other occupational classes combined? 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 number of close friends, according to marital status, for a random sample of 4,906 adults in Britain.

<u>Marital status</u>	<u>Mean</u>	<u>s</u>	<u>n</u>
Married	3.72	3.02	2,352
Cohabiting	3.90	2.90	327
Single	3.96	2.96	1,176
Widowed	3.49	3.25	448
Divorced	3.53	3.10	388
Separated	3.16	2.95	215
		TOTAL	4,906

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

- (i) Test the hypothesis that, in the population, the mean number of close friends corresponding to each marital status 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 4,900 degrees of freedom is 2.21; the between-groups and within-groups sums of squares are 183.0 and 44,835.0 respectively.)

- (ii) Test the hypothesis that, in the population, the mean numbers of close friends for married people and for single people are the same.

(Note: The critical value of t at the 5% level corresponding to 3,526 degrees of freedom is 1.96; the pooled sample standard deviation for married people and single people is 3.00.)

Continued

4. In a random sample of 3,522 UK households, the (Pearson) correlation between household size, i.e. the number of people in the household, and per capita income (in £1000s), i.e. the average income per person, was found to be -0.346 .

- (i) Test the hypothesis that there is no relationship between household size and per capita income (in £1000s) for UK households.

(Note: You may assume that $(-0.346)^2$ is equal to 0.120; the critical value of F at the 5% level corresponding to 1 degree of freedom and 3,520 degrees of freedom is 3.84.)

The regression equation corresponding to the dependence of per capita income on household size for UK households is

$$y = 19.96 - 2.60x_1$$

where y is per capita income (in £1000s), and x_1 is household size.

- (ii) Use the above equation to predict the per capita incomes for three households containing one, two and eight people 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 , the number of children (people aged under 18 years) in the household, leads to the following equation

$$y = 19.15 - 2.11x_1 - 0.85x_2$$

- (iii) Explain why the coefficient of x_1 , household size, changes between the two equations. Use the second regression equation to predict the per capita incomes for the three households mentioned before in part (ii), assuming first that all three households only contain adults, and then that the third household contains two adults and six children. Comment on these predictions, both with reference to each other and also with reference to the predictions made for part (ii).

Continued

5. The following cross-tabulation is of occupational class [C] by whether or not an individual sees herself or himself as underpaid (i.e. paid too little) [U] according to region [R] for a random sample of 1,320 employees in England.

REGION = North	<u>Underpaid: Yes</u>	<u>Underpaid: No</u>	<u>TOTAL</u>
Middle class	79	281	360
Working class	<u>117</u>	<u>175</u>	<u>292</u>
TOTAL	196	456	652
REGION = South	<u>Underpaid: Yes</u>	<u>Underpaid: No</u>	<u>TOTAL</u>
Middle class	74	240	314
Working class	<u>77</u>	<u>114</u>	<u>191</u>
TOTAL	151	354	505
REGION = London	<u>Underpaid: Yes</u>	<u>Underpaid: No</u>	<u>TOTAL</u>
Middle class	28	82	110
Working class	<u>18</u>	<u>35</u>	<u>53</u>
TOTAL	46	117	163

- (i) Use odds ratios to summarise the way in which the relationship between class and viewing oneself as underpaid varies according to region. The chi-square statistics for the three sub-tables are 25.19, 15.89 and 1.28. Using these three values, test the relationship in each sub-table for significance.
- (ii) Use odds ratios to summarise the relationships between region and: (a) class; (b) viewing oneself as underpaid.
- (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] [C] [U]	53.1	7	0.000				
2	[RU] [C]	52.9	5	0.000	0.2	2	0.908	1
3	[CU] [R]	12.7	6	0.049	40.4	1	0.000	1
4	[RC] [U]	42.0	5	0.000	11.0	2	0.004	1
5	[RC] [CU]	1.6	4	0.803	11.0	2	0.004	3
6	[RU] [CU]	12.5	4	0.014	0.2	2	0.908	3
7	[RC] [RU]	41.9	3	0.000	0.2	2	0.908	4
8	[RC][RU][CU]	1.5	2	0.466	0.1	2	0.948	5
9	[RCU]	0.0	0		1.5	2	0.466	8

Continued

SECTION B

6. To what extent does the value of an existing survey to a secondary analyst depend on the nature and quality of the data that it generated? Discuss, with particular reference to ONE social survey of your choice.
-
7. Assess, with reference to ONE real or hypothetical survey of your choice, the importance of processes of concept operationalization and questionnaire design to the conceptual adequacy and effectiveness of a research instrument.
-
8. Is the presence of an interviewer the only important similarity between qualitative interviews and social survey interviews?
-
9. Is using both qualitative methods and quantitative methods within a single research project more fashionable than it is beneficial?
-
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.]

RELIGIOUS IDENTITY by DEPRIVATION INDEX FOR LOCALITY

Deprivation index (scale)

	Score = Low		Score = Medium		Score = High	
<u>Religious identity</u>	%		%		%	
Christian	1502	21.9	4140	60.5	1201	17.6
Hindu	15	15.5	65	67.0	17	17.5
Jewish	13	26.0	33	66.0	4	8.0
Muslim	14	5.7	119	48.8	111	45.5
Other	39	13.6	184	64.1	64	22.3
None	267	21.0	745	58.6	259	20.4

[Note: A low score on the deprivation index corresponds to a low level of deprivation in the locality in which an individual lives; a high score on the index corresponds to a high level of deprivation in the locality in which an individual lives.]

END