

SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 2003

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 dependent children in a random sample of 289 lone-father families containing dependent children (in Britain), was found to be 1.69, with a sample standard deviation of 0.85.
- (i) Calculate a 95% confidence interval for the mean number of dependent children in lone-father families containing dependent children.
 - (ii) The mean number of dependent children in two-parent families containing dependent children is known to be 2.09 children. Calculate a z-statistic and use it to test whether this is a plausible mean for the number of dependent children in lone-father families containing dependent children.
 - (iii) Suppose that the population standard deviation for the number of dependent children in lone-mother families containing dependent children is assumed to be 0.95. How big a sample would be needed to produce a sample mean that one could be 95% confident fell within 0.1 children of the population mean number of children corresponding to lone-mother families containing dependent children? 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 educational level (based on qualifications) and full-time paid employment for a random sample of 1,600 married women in Britain aged under 60 and living in households containing dependent children.

	<u>In full-time paid employment</u>	<u>Not in full-time paid employment</u>	<u>Total</u>
<u>Educational level</u>			
Degree	85	115	200
'A' level or equivalent	108	292	400
GCSE or equivalent	119	441	560
None of the above	88	352	440
TOTAL	400 (25%)	1200 (75%)	1600

- (i) Calculate the chi-square statistic for the above cross-tabulation and use it to test the hypothesis that there is no relationship between educational level and full-time paid employment for married women in Britain who are aged under 60 and living in households containing dependent children.

(Note: the critical value at the 5% level of a chi-square statistic with 3 degrees of freedom is 7.81).

- (ii) A similar cross-tabulation, based on a random sample of 1,400 married women in Britain aged under 60 and living in households containing no dependent children, gave rise to a chi-square statistic of 121.8. 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 extent of involvement in full-time paid employment varies significantly between married women (aged under 60 and living in households containing dependent children) in the last three educational level categories, i.e. those who have 'A' levels, those who have GCSEs, and those who have none of the qualifications mentioned? (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).

Continued.....

27
21 2
20

- 3 The following table shows the mean number of rooms, according to current marital status and sex, in the houses (or flats) of a random sample of 15,874 adults in Britain.

<u>Marital status/sex</u>	<u>Mean</u>	<u>s</u>	<u>n</u>
Single men	5.22	1.51	2,093
Single women	5.12	1.46	1,833
Married men/women	5.65	1.49	9,239
Divorced/separated men	4.79	1.36	540
Divorced/separated women	4.96	1.49	810
Widowed men	4.84	1.39	325
Widowed women	4.70	1.51	1,034
		TOTAL	15,874

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

- (i) Test the hypothesis that, in the population, the mean number of rooms corresponding to each combination of marital status and sex 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 15,867 degrees of freedom is 2.10; the between-groups and within-groups sums of squares are 1,755.6 and 34,907.4 respectively).

- (ii) Test the hypothesis that, in the population, the houses (or flats) of divorced/separated women and the houses (or flats) of divorced/separated men have the same mean number of rooms.

(Note: the critical value of t at the 5% level corresponding to 1,348 degrees of freedom is 1.96; the pooled sample standard deviation for divorced/separated men and divorced/separated women is 1.44).

Continued.....

-
- 4 In a random sample of 25 administrative districts in England and Wales, the (Pearson) correlation between the percentage of employed people who travel to work by bus and the percentage of heads of households who are in Registrar General's Social Classes I or II was found to be -0.56.

- (i) Test the hypothesis that there is no relationship between the percentage of employed people who travel to work by bus and the percentage of heads of households who are in Registrar General's Social Classes I or II. (You may assume that $(-0.56)^2 = 0.31$).

(Note: the critical value of F at the 5% level corresponding to 1 degree of freedom and 23 degrees of freedom is 4.28).

The regression equation corresponding to the dependence of the percentage of employed people who travel to work by bus on the percentage of heads of households who are in Registrar General's Social Classes I or II is

$$y = 21.00 - 0.50x_1$$

where y is the percentage of employed people who travel to work by bus, and x_1 is the percentage of heads of households who are in Registrar General's Social Classes I or II.

- (ii) Use the above equation to predict the percentages of employed people who travel to work by bus in three administrative districts in which the percentages of heads of households who are in Registrar General's Social Classes I or II are 0%, 25% and 50%. 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 population density (in persons per hectare), leads to the following equation

$$y = 16.95 - 0.42x_1 + 0.15x_2$$

- (iii) Explain why the coefficient of x_1 , the percentage of heads of households who are in Registrar General's Social Classes I or II, changes between the two equations. Use both the first and second regression equations to predict the percentage of employed people who travel to work by bus in Kingston & Richmond, a district in which 40% of heads of households are in Registrar General's Social Classes I or II and for which the population density is 30. Comment on the difference between the two predictions.

Continued.....

- 5 The following cross-tabulation is of educational level [E] by whether or not a person smokes [S] by region [R] for a random sample of 10,378 adults in England. (The 'A' level category includes higher or equivalent qualifications).

REGION = North	<u>Does not smoke</u>	<u>Smokes</u>	<u>TOTAL</u>
'A' levels, etc.	813	268	1081
None	<u>1233</u>	<u>769</u>	<u>2002</u>
TOTAL	2046	1037	3083
REGION = Midlands	<u>Does not smoke</u>	<u>Smokes</u>	<u>TOTAL</u>
'A' levels, etc.	801	284	1085
None	<u>1358</u>	<u>679</u>	<u>2037</u>
TOTAL	2159	963	3122
REGION = South	<u>Does not smoke</u>	<u>Smokes</u>	<u>TOTAL</u>
'A' levels, etc.	1331	446	1777
None	<u>1578</u>	<u>818</u>	<u>2396</u>
TOTAL	2909	1264	4173

- (i) Use odds ratios to summarise the way in which the relationship between educational level and smoking varies according to region. The chi-square statistics for the three sub-tables are 58.3, 17.0 and 39.5. Using these chi-square statistics, test the relationship in each sub-table for significance.
- (ii) Use odds ratios to summarise the relationships between:
(a) educational level and region; (b) smoking 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	Compared to model
1	[R] [E] [S]	189.3	7	0.000				
2	[RE] [S]	127.1	5	0.000	62.2	2	0.000	1
3	[RS] [E]	179.4	5	0.000	9.9	2	0.007	1
4	[ES] [R]	76.1	6	0.000	113.2	1	0.000	1
5	[RE] [RS]	117.2	3	0.000	9.9	2	0.007	2
6	[ES] [RE]	13.9	4	0.008	62.2	2	0.000	4
7	[ES] [RS]	66.2	4	0.000	9.9	2	0.007	4
8	[RE][RS][ES]	6.5	2	0.038	7.4	2	0.025	6
9	[RES]	0.0	0		6.5	2	0.038	8

Continued.....

SECTION B

- 6 To what extent is the value of secondary analysis contingent on the constraints imposed by the use of existing data? Discuss, with particular reference to ONE social survey of your choice.
-
- 7 In what ways do the processes of concept operationalisation and questionnaire design underpin the effectiveness of survey research? Discuss, with reference to ONE real or hypothetical survey of your choice.
-
- 8 Are the shared features of quantitative and qualitative interviews, such as the interviewer, less significant than the differences between them, such as the degree of standardisation?
-
- 9 Is combining quantitative and qualitative methods within a single research project a more attractive idea in theory than it is in practice?
-
- 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].

WORK STATUS by UNHEALTHY EATING BEHAVIOUR

Unhealthy eating behaviour (scale)

<u>Work status</u>	Score = Low		Score = Medium		Score = High	
	%		%		%	
Paid employment	2385	(47.8)	1837	(36.8)	770	(15.4)
Unemployed	127	(28.0)	197	(43.4)	130	(28.6)
Sick or disabled	77	(41.0)	70	(37.2)	41	(21.8)
Retired	1071	(53.6)	746	(37.3)	181	(9.1)
Student	60	(55.6)	39	(36.1)	9	(8.3)
Housework	665	(52.7)	450	(35.7)	147	(11.6)

[Note: Low scores on the 'Unhealthy eating behaviour' scale correspond to relatively healthy eating behaviour; high scores on the scale correspond to relatively unhealthy eating behaviour].