

SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 1999

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 distance travelled in a random sample of 729 graduates migrating internally within Britain was found to be 53.2 kilometres, with a sample standard deviation of 67.5 kilometres.
- (i) Calculate a 95% confidence interval for the mean distance travelled by graduate internal migrants.
 - (ii) The mean distance travelled by people in general migrating internally within Britain is known to be 28.2 kilometres. Calculate a z-statistic and use it to test whether this is a plausible mean distance travelled for graduate internal migrants.
 - (iii) Suppose that the population standard deviation for the mean distance travelled by retired internal migrants is assumed to be 55.0 kilometres. How big a sample would then be needed to produce a sample mean that one could be 95% confident fell within 5 kilometres of the population mean distance travelled by retired internal migrants? 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 table shows the relationship between feelings of loneliness and sex for a random sample of 200 married people in Britain.

	<u>Men</u>	<u>Women</u>	<u>Total</u>
<u>Feel lonely</u>			
Never	77	63	140
Sometimes	20	28	48
Often/Always	3	9	12
TOTAL	100 (50%)	100 (50%)	200

- (i) Calculate the chi-square statistic for the above table and use it to test the hypothesis that there is no relationship between feelings of loneliness and sex for married people in Britain.

(Note: the critical value at the 5% level of a chi-square statistic with 2 degrees of freedom is 5.99).

- (ii) A similar table, once again showing the relationship between feelings of loneliness and sex, was generated from a random sample of 1,000 single people in Britain. This second table gave rise to a chi-square statistic of 7.0. Use Cramer's V to compare the strengths of the relationships in the two tables, and explain why the two chi-square statistics could not have been used for this purpose.
- (iii) Does the above table suggest that the likelihood of ever feeling lonely (i.e. 'Sometimes' or 'Often/Always' as opposed to 'Never') differs between married men and married women in Britain?

(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 weekly alcohol consumption of male adults according to marital status for a random sample of 4,204 British men.

<u>Marital status</u>	<u>Mean</u>	<u>s</u>	<u>n</u>
Widowed	9.9	16.3	181
Divorced/Separated	19.2	23.6	210
Married	12.7	17.6	2,940
Single	18.3	21.0	873
		TOTAL	4,204

(Alcohol consumption is measured in units of alcohol).

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

- (i) Test the hypothesis that, in the population, men of each marital status have the same mean weekly alcohol consumption. 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 4,200 degrees of freedom is 2.60; the between-groups and within-groups sums of squares are 28,080 and 1,512,000 respectively).

- (ii) Test the hypothesis that, in the population, divorced/separated men and married men have the same mean weekly alcohol consumption.

(Note: the critical value of t at the 5% level corresponding to 3,148 degrees of freedom is 1.96; the pooled sample standard deviation for divorced/separated men and married men is 18.2).

Continued.....

- 4 In a random sample of 38 local authorities in England and Wales, the (Pearson) correlation between the teenage pregnancy rate (conceptions per year per 1,000 women aged 15-19) and the percentage of 'low status' households (in which the head of household has a semi-skilled or unskilled manual occupation) was found to be 0.68.

- (i) Test the hypothesis that there is no relationship between the teenage pregnancy rate and the percentage of low status households.

(Note: $(0.68)^2 = 0.46$; the critical value of F at the 5% level corresponding to 1 degree of freedom and 36 degrees of freedom is 4.11).

The regression equation corresponding to the dependence of the teenage pregnancy rate on the percentage of low status households is

$$y = 5.7x_1 - 10.7$$

where y is the teenage pregnancy rate, and x_1 is the percentage of low status households.

- (ii) Use the above equation to predict the teenage pregnancy rates for a local authority with an average percentage of low status households (11%) and for one with no low status households (0%). Is the second predicted value useful? Why might the above linear regression equation be an inappropriate model of the relationship between the two variables? In Richmond-upon-Thames 5% of households are low status. Use the above equation to predict the teenage pregnancy rate in Richmond-upon-Thames. In fact, the actual rate for Richmond-upon-Thames is 31.3. Suggest an explanation for the difference between the predicted and actual rates.

The addition to the regression analysis of a second independent variable, x_2 , which is the female unemployment rate (%), leads to the following equation

$$y = 3.2x_1 + 6.5x_2 - 23.0$$

- (iii) Explain why the coefficient of x_1 , the percentage of low status households, changes between the two equations. The average female unemployment rate nationally is 6%; the rate is also 6% in Richmond-upon-Thames. Use the second regression equation to again predict the teenage pregnancy rate in Richmond-upon-Thames. Why is this prediction more accurate than the previous one?

Continued.....

- 5 The following crosstabulation is of sex [S] by attitude to living with a partner [C] by age [A] for a random sample of 1,212 single British people. ('Yes' indicates that the respondent would prefer at this point in time to be living with a partner/spouse; 'No' indicates that they prefer being single):

AGE = 16-19 years	<u>Yes</u>	<u>No</u>	<u>TOTAL</u>
Men	31	193	224
Women	31	190	221
TOTAL	62	383	445
AGE = 20-24 years	<u>Yes</u>	<u>No</u>	<u>TOTAL</u>
Men	41	169	210
Women	51	101	152
TOTAL	92	270	362
AGE = 25+ years	<u>Yes</u>	<u>No</u>	<u>TOTAL</u>
Men	82	154	236
Women	63	106	169
TOTAL	145	260	405

- (i) Use odds ratios to summarise the way in which the relationship between sex and attitude to living with a partner varies according to age. The chi-square statistics for the three sub-tables are 0.0, 9.2 and 0.3. Using these chi-square statistics, test the relationship in each sub-table for significance.
- (ii) Use odds ratios to summarise the relationships between:
(a) sex and age; (b) attitude to living with a partner and age.
- (iii) Use the following results corresponding to the goodness-of-fit of various log-linear models to determine the most appropriate model of the table 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	[S] [C] [A]	72.7	7	0.000				
2	[SC] [A]	70.5	6	0.000	2.2	1	0.138	1
3	[SA] [C]	65.9	5	0.000	6.8	2	0.033	1
4	[CA] [S]	16.1	5	0.007	56.6	2	0.000	1
5	[SC] [SA]	63.7	4	0.000	6.8	2	0.033	2
6	[SC] [CA]	13.9	4	0.008	56.6	2	0.000	2
7	[SA] [CA]	9.3	3	0.026	6.8	2	0.033	4
8	[SC] [SA] [CA]	5.1	2	0.079	4.2	1	0.041	7
9	[SCA]	0.0	0		5.1	2	0.079	8

Continued.....

SECTION B

- 6 Do the inherent constraints of secondary analysis outweigh the benefits of using existing data? Discuss, with particular reference to ONE existing social survey.
-
- 7 'Both the operationalization of concepts and questionnaire design are vital to the production of valid and useful data.' Discuss, with reference to ONE hypothetical survey of your choice.
-
- 8 'Interviews may differ in structure and purpose, but they are always social encounters.' Discuss.
-
- 9 Why do some research projects use quantitative and qualitative approaches in combination when so many other projects do not?
-
- 10 Critically discuss the following table. 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 table is statistically significant; you should specify any more focused statistical tests that you would ideally like to carry out].

HIGHEST QUALIFICATION by ENGAGEMENT IN LEISURE ACTIVITIES
Engagement in Leisure Activities (scale)

	Score = Low		Score = Medium		Score = High	
<u>Highest qual.</u>	%		%		%	
None	490	(39.5)	589	(47.4)	163	(13.1)
'O' level [#]	61	(17.0)	173	(48.2)	125	(34.8)
'A' level [#]	13	(11.7)	47	(42.3)	51	(45.9)
Degree	12	(10.4)	47	(40.9)	56	(48.7)
Other	114	(16.9)	334	(49.6)	225	(33.4)

[Notes: Low scores on the 'Engagement in Leisure Activities' scale correspond to limited participation in leisure activities; high scores on the scale correspond to extensive participation in leisure activities; # = 'Or equivalent'].

 END