SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 2001

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 hours worked per week in a random sample of 81 divorced women who were aged 20-59 and in paid employment was found to be 28.6 hours, with a sample standard deviation of 13.5 hours.

(i) Calculate a 95% confidence interval for the mean number of hours worked per week by divorced women aged 20-59 and in paid employment.

(ii) The mean number of hours worked per week by single women (i.e. never-married women not living with a partner) aged 20-59 and in paid employment is known to be 35.5 hours. Calculate a z-statistic and use it to test whether this is a plausible mean for divorced women aged 20-59 and in paid employment.

(iii) Suppose that the population standard deviation for the mean number of hours worked per week by divorced men aged 20-59 and in paid employment is assumed to be 12.0 hours. How big a sample would then be needed to produce a sample mean that one could be 95% confident fell within 3 hours of the population mean number of hours worked per week by divorced men aged 20-59 and in paid employment? 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 an individual's level of recent contact with family members (outside their own household) and sex for a random sample of 1,000 adults living in Britain.

	Men	Women	Total
Level of contact			
None	40	24	64
Occasional	104	88	192
Regular	260	260	520
Extensive	96	128	224
TOTAL	500 (50%)	500 (50%)	1000

(i) Calculate the chi-square statistic for the above cross-tabulation and use it to test the hypothesis that there is no relationship between level of recent contact with family members and sex for adults living in Britain.

(Note: the critical value at the 5% level of a chi-square statistic with 3 degrees of freedom is 7.82).

(ii) A similar cross-tabulation based on a random sample of 5,000 adults living in Britain, but this time showing the relationship between level of recent contact with friends and sex, gave rise to a chi-square statistic of 6.5. 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 proportion of adults who have had no recent contact with family members varies significantly according to sex?

(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 mean household size (number of persons) according to head of household's (Registrar General's) Social Class for a random sample of 4,686 adults in Britain.

Social Class	Mean	s	n
I	3.10	1.24	320
II	2.93	1.29	1,280
III (Non-manual)	2.54	1.31	621
III (Manual)	2.99	1.26	1,461
IV	2.83	1.37	707
V	2.67	1.53	297
		TOTAL	4,686

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

(i) Test the hypothesis that, in the population, adults in each social class have the same mean household size. 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 4,680 degrees of freedom is 2.21; the between-groups and within-groups sums of squares are 129.5 and 8,190.0 respectively).

(ii) Test the hypothesis that, in the population, adults in Social Class I and adults in Social Class II have the same mean household size.

(Note: the critical value of t at the 5% level corresponding to 1,598 degrees of freedom is 1.96; the pooled sample standard deviation for adults in Social Class I and adults in Social Class II is 1.28).

Continued.....

4 In a random sample of 32 countries, the (Pearson) correlation between the Infant Mortality Rate (per 1,000 infants) and the Total Fertility Rate (which is, in effect, the average number of births per woman) was found to be 0.835.

(i) Test the hypothesis that there is no relationship between the Infant Mortality Rate and the Total Fertility Rate.

(Note: $(0.835)^2 = 0.70$; the critical value of F at the 5% level corresponding to 1 degree of freedom and 30 degrees of freedom is 4.17).

The regression equation corresponding to the dependence of the Infant Mortality Rate on the Total Fertility Rate is

$$y = 21.5x_1 - 26.5$$

where y is the Infant Mortality Rate, and x_1 is the Total Fertility Rate.

(ii) Use the above equation to predict the Infant Mortality Rate for a country with a Total Fertility Rate of 6.0 and the Infant Mortality Rate for a country with a Total Fertility Rate of 1.6. What does the above equation suggest that the Infant Mortality Rate would be for a country with a Total Fertility Rate of 1.2? Is this last 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 is the logarithm of the per capita GNP (Gross National Product) of the country (in US dollars), leads to the following equation

$$y = 14.8x_1 - 14.7x_2 + 107.0$$

(iii) Explain why the coefficient of x_1, the Total Fertility Rate, changes between the two equations.

Continued.....

5 The following cross-tabulation is of social class of origin [C] by possession of a degree [D] by age [A] for a random sample of 512 men aged 30-59 in Britain. (Social class of origin is measured using father's occupation).

AGE = 30-39 years	Degree	No degree	TOTAL
Middle class	12	48	60
Working class	8	81	89
TOTAL	20	129	149
AGE = 40-49 years	Degree	No degree	TOTAL
Middle class	15	43	58
Working class	8	90	98
TOTAL	23	133	156
AGE = 50-59 years	Degree	No degree	TOTAL
Middle class	13	49	62
Working class	5	140	145
TOTAL	18	189	207

(i) Use odds ratios to summarise the way in which the relationship between social class of origin and possession of a degree varies according to age. The chi-square statistics for the three sub-tables are 3.7, 9.1 and 16.8. Using these chi-square statistics, test the relationship in each sub-table for significance.

(ii) Use odds ratios to summarise the relationships between:
(a) social class of origin and age; (b) possession of a degree 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 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	[A] [C] [D]	35.7	7	0.000				
2	[CD] [A]	9.0	6	0.173	26.7	1	0.000	1
3	[AD] [C]	32.0	5	0.000	3.7	2	0.161	1
4	[AC] [D]	31.2	5	0.000	4.5	2	0.107	1
5	[CD] [AC]	4.5	4	0.339	4.5	2	0.107	2
6	[CD] [AD]	5.3	4	0.254	3.7	2	0.161	2
7	[AC] [AD]	27.6	3	0.000	3.7	2	0.161	4
8	[CD][AC][AD]	2.2	2	0.336	2.3	2	0.309	5
9	[CDA]	0.0	0		2.2	2	0.336	8

Continued.....

SECTION B

6 What do secondary analysts gain and lose by using existing data? Discuss, with particular reference to **ONE** social survey of your choice.

7 Discuss, with reference to **ONE** hypothetical survey of your choice, the pivotal contribution to the research process made by concept operationalization and questionnaire design.

8 Do shared features justify the view that qualitative interviews and social survey interviews are variants of the same method?

9 Are the practical and epistemological problems that arise when qualitative and quantitative methods are combined within a single research project necessarily outweighed by the benefits of this joint approach?

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].

REGION by UNHEALTHY BEHAVIOUR

Extent of unhealthy behaviour (scale)

	Score = Low		Score = Medium		Score = High	
Region		%		%		%
Scotland	532	(57.5)	213	(23.0)	180	(19.5)
North	1504	(61.3)	502	(20.5)	446	(18.2)
Midlands	1236	(61.4)	439	(21.8)	337	(16.7)
London	615	(65.1)	198	(21.0)	132	(14.0)
South	1764	(66.1)	528	(19.8)	377	(14.1)

[Note: Low scores on the 'Extent of unhealthy behaviour' scale correspond to a limited level of engagement in activities likely to be damaging to an individual's health; high scores on the scale correspond to a more extensive level of engagement in such activities].

END