SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 2008

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 gifts (of money) given to charity per year by a random sample of 1,764 adults in Britain belonging to Christian religious denominations was found to be 13.6, with a sample standard deviation of 21.0.

(i) Calculate a 95% confidence interval for the mean number of gifts given to charity per year by adults in Britain belonging to Christian religious denominations.

(ii) The mean number of gifts given to charity per year for <u>all</u> adults in Britain is 12.4. Calculate a z-statistic and use it to test whether this is a plausible mean number of gifts given to charity per year by adults in Britain belonging to Christian religious denominations.

(iii) Suppose that the population standard deviation for the number of gifts given to charity per year by adults in Britain belonging to non-Christian religious denominations is assumed to be 25.5. How big a sample would be needed to produce a sample mean that one could be 95% confident fell within 1 gift of the population mean number of gifts given to charity per year by adults in Britain belonging to non-Christian religious denominations? 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 1,000 single adults in Britain between region of residence and whether an individual has access to the internet within their home.

	Access: Yes	Access: No	Total
Region			
North (of England)	192	208	400
South (of England)	162	138	300
London	102	78	180
Wales	8	32	40
Scotland	36	44	80
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 for single adults in Britain between region and whether an individual has access to the internet within their home.

(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 1,987 married adults in Britain, and again showing the relationship between region and whether an individual has access to the internet within their home, gave rise to a chi-square statistic of 12.90. 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 the likelihood of a single adult having access to the internet within their home varies significantly between regions if Scotland is combined with the North and London is combined with the South? Does the above cross-tabulation suggest that the likelihood of a single adult having access to the internet within their home varies significantly between Wales and the rest of Britain? Calculate two chi-square statistics to answer this part of the question, and comment on the results 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; 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 age difference between spouses (husband's age in years minus wife's age in years), according to the highest educational qualification of the wife, for a random sample of 1,253 married couples in England and Wales.

Highest qualification	Mean	s	n
Degree	2.00	5.03	324
'O' or 'A' level, or equivalent	2.75	4.98	576
None of the above	2.66	5.33	353
		TOTAL	1,253

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

(i) Test the hypothesis that, in the population, the mean age difference between spouses corresponding to each level of highest qualification is the same. Discuss your findings with reference to the sample means.

(Note: the critical value of F at the 5% level corresponding to 2 degrees of freedom and 1,250 degrees of freedom is 3.01; the between-groups and within-groups sums of squares are 124.8 and 32,500.00 respectively.)

(ii) Test the hypothesis that, in the population, the mean age differences between spouses for couples where the wife has a degree and for couples where the wife has 'O' or 'A' level qualifications are the same.

(Note: the critical value of t at the 5% level corresponding to 898 degrees of freedom is 1.96; the pooled sample standard deviation for couples where the wife has a degree and couples where the wife has 'O' or 'A' level qualifications is 5.00.)

Continued.....

4 In a random sample of 981 adults in England and Wales, the (Pearson) correlation between income (in £1000s) and the number of years for which an individual had lived in their current neighbourhood was found to be –0.144.

(i) Test the hypothesis that there is no relationship between income (in £1000s) and number of years lived in current neighbourhood for adults in England and Wales.

(Note: you may assume that $(-0.144)^2$ is equal to 0.021; the critical value of F at the 5% level corresponding to 1 degree of freedom and 979 degrees of freedom is 3.85.)

The regression equation corresponding to the dependence of number of years lived in current neighbourhood on income for adults in England and Wales is

$$y = 19.03 - 0.12x_1$$

where y is the number of years lived in current neighbourhood, and x_1 is income.

(ii) Use the above equation to predict the numbers of years lived in current neighbourhood for three individuals whose incomes are £5,000, £50,000 and £200,000. 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 takes the value 1 if an individual is in employment and 0 otherwise, leads to the following equation

$$y = 23.12 - 0.05x_1 - 8.65x_2$$

(iii) Explain why the coefficient of x_1, income, changes between the two equations. Use the second regression equation to predict the numbers of years lived in current neighbourhood for the three individuals from part (ii), assuming that (a) they are in employment, and (b) they are not in employment. Comment on these predictions with reference to the predictions made for part (ii).

Continued.....

5 The following cross-tabulation is of possession of a degree [D] by self-identification as a feminist [F] according to party political identification [P] for a random sample of 1,130 adult women in Britain.

PARTY = Conservative	Feminist: Yes	Feminist: No	TOTAL
Has a degree	2	38	40
Does not have a degree	9	223	232
TOTAL	11	261	272
PARTY = Labour	Feminist: Yes	Feminist: No	TOTAL
Has a degree	17	46	63
Does not have a degree	26	341	367
TOTAL	43	387	430
PARTY = Other or None	Feminist: Yes	Feminist: No	TOTAL
Has a degree	18	58	76
Does not have a degree	22	330	352
TOTAL	40	388	428

(i) Use odds ratios to summarise the way in which the relationship between having a degree and feminist self-identification varies according to party. The chi-square statistics for the three sub-tables are 0.11, 23.66 and 22.42. Using these three values, test the relationship in each sub-table for significance.

(ii) Use odds ratios to summarise the relationships between party identification and: (a) possession of a degree; (b) self-identification as a feminist.

(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	[P] [D] [F]	48.3	7	0.000				
2	[PF] [D]	38.3	5	0.000	10.0	2	0.007	1
3	[DF] [P]	14.6	6	0.024	33.7	1	0.000	1
4	[PD] [F]	46.4	5	0.000	1.9	2	0.391	1
5	[PD] [DF]	12.7	4	0.013	33.7	1	0.000	4
6	[PF] [DF]	4.6	4	0.330	10.0	2	0.007	3
7	[PD] [PF]	36.4	3	0.000	10.0	2	0.007	4
8	[PD][PF][DF]	2.8	2	0.242	1.8	2	0.412	6
9	[PDF]	0.0	0		2.8	2	0.242	8

Continued.....

SECTION B

6 Do the inherent constraints of using existing survey data to address a research question outweigh any advantages of carrying out a secondary analysis? 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 impact of the operationalization of concepts and questionnaire design on the validity and value of survey data.

---

8 To what extent and in what ways are a successful qualitative interview and a successful survey interview likely to be similar to each other?

---

9 Should quantitative methods and qualitative methods never be used together to address a single research topic? Or should they always be used together?

---

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.]

SOCIAL CLASS by PERCEIVED SAFETY OF LOCAL NEIGHBOURHOOD

Safety of neighbourhood (scale)

	Score = Low		Score = Medium		Score = High	
Social class	%		%		%	
Service class	112	(51.6)	93	(42.9)	12	(5.5)
Intermediate	67	(44.1)	67	(44.1)	18	(11.8)
Working class	40	(24.0)	97	(58.1)	30	(18.0)
Unemployed	5	(21.7)	16	(69.6)	2	(8.7)
Full-time student	4	(25.0)	8	(50.0)	4	(25.0)

[Notes: low scores on the safety of neighbourhood scale correspond to a high degree of perceived safety; high scores on the scale correspond to a lower degree of perceived safety.]

END