SO 2010

UNIVERSITY OF WARWICK

September Examinations 1996

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 heterosexual partners in the last year in a random sample of 529 16-24 year-old men was found to be 1.42 partners, with a sample standard deviation of 2.30 partners.

(i) Calculate a 95% confidence interval for the mean number of heterosexual partners in the last year among 16-24 year-old men.

(ii) The mean number of heterosexual partners in the last year for all men is known to be 1.20 partners. Calculate a z-statistic and use it to test whether this is a plausible mean number of heterosexual partners in the last year for 16-24 year-old men.

(iii) If the population standard deviation for the number of heterosexual partners in the last year among 16-24 year-old women is assumed to be 1.40 partners, how big a sample would be needed to produce a sample mean that one could be 95% confident fell within 0.20 partners of the population mean number of heterosexual partners in the last year among 16-24 year-old women? 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 men's views on whether extra-marital sex (adultery) is always morally wrong according to their marital status for a random sample of 600 men.

Marital status	Yes	No	TOTAL
Married	205	195	400
Cohabiting	12	18	30
Sep.\Div.	20	10	30
Widowed	12	8	20
Not married	51	69	120
TOTAL	300 (50%)	300 (50%)	600

Sep.\Div. = Separated or Divorced.

(i) Calculate the chi-square statistic for the above table and use it to test the hypothesis that there is no relationship between men's views on extra-marital sex and their marital status.

(Note: the critical value at the 5% level of a chi-square statistic with 4 degrees of freedom is 9.49).

(ii) A similar table based on a random sample of 600 women, once again showing the relationship between views on extra-marital sex and marital status, gave rise to a chi-square statistic of 16.2. Use Cramer's V to compare the strengths of the relationships in the two tables, and explain why the two chi-square statistics could also have been used for this purpose.

(iii) Does the above table suggest that views on extra-marital sex differ between the following two groups: men who may not have been married (cohabiting men; unmarried men), and men who definitely are or have been married (married men; divorced\separated men; widowed men)?

(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 age (in years) at the time of becoming part of a couple of a random sample of 856 members of couples, according to the way they first met their partners.

Met partner via	Mean	s	n
School\college	20.0	7.0	56
Work	25.4	7.6	147
Friends & outings	24.3	7.9	400
Ads. & agencies	33.3	9.6	5
Clubs & activities	26.9	11.2	81
Other	23.9	7.0	167
		TOTAL	856

Ads. & agencies = Advertisements and dating agencies.

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

(i) Test the hypothesis that, in the population, the mean age at the time of becoming part of a couple is the same for each way of first meeting one's partner. 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 850 degrees of freedom is 2.22; the between-groups and within-groups sums of squares are 2170.0 and 52700.0 respectively).

(ii) Test the hypothesis that, in the population, the mean age at the time of becoming part of a couple is the same for people who first met their partners via friends and outings as for people who first met their partners via advertisements and agencies.

(Note: the critical value of t at the 5% level corresponding to 403 degrees of freedom is 1.96; the pooled sample standard deviation for people who first met their partners via friends and outings and people who first met their partners via advertisements and agencies is 8.0)

/continued.....

4 For a random sample of 62 British adults, the (Pearson) correlation between length of residence at current address (in years) and respondent's age (in years) was found to be 0.50.

(i) Test the hypothesis that there is no relationship between length of residence at current address and respondent's age.

(Note: the critical value of F at the 5% level corresponding to 1 degree of freedom and 60 degrees of freedom is 4.00).

The regression equation corresponding to the dependence of length of residence at current address on respondent's age is

$$y = 0.30x_1 - 1.50$$

where y is length of residence at current address (in years), and x_1 is respondent's age (in years).

(ii) Use the above equation to predict the lengths of residence at current address of someone aged 16 years and someone aged 21 years. Does the difference between the two predicted values make sense in substantive terms? Why might the above linear regression equation be an inappropriate model of the relationship between length of residence at current address and respondent's age? Use the above equation to predict the length of residence at current address of a 5 year-old child. Is the predicted value useful?

The addition to the regression analysis of a second independent variable, x_2, which is a dummy variable taking the value 1 if the respondent is aged between 16 and 18 years inclusive and the value 0 otherwise, results in the following equation

$$y = 0.35x_1 + 7.50x_2 - 3.00$$

(iii) Comment on the coefficient of x_2, and explain why the coefficient of x_1, respondent's age, changes between the two equations. Use the second equation to again predict the lengths of residence at current address of someone aged 16 years and someone aged 21 years, and comment on the results.

/continued.....

5 The following crosstabulation is of female partner's religious denomination [F] by male partner's religious denomination [M] by age group [A] for a random sample of 465 British heterosexual Christian couples

AGE GROUP = 18-44	Roman Catholic	Other	TOTAL
Roman Catholic	11	17	28
Other	15	118	133
TOTAL	26	135	161

AGE GROUP = 45-64	Roman Catholic	Other	TOTAL
Roman Catholic	18	15	33
Other	8	146	154
TOTAL	26	161	187

AGE GROUP = 65 plus	Roman Catholic	Other	TOTAL
Roman Catholic	8	8	16
Other	0	101	101
TOTAL	8	109	117

(i) Use odds ratios to summarise the way in which the relationship between female partner's religious denomination and male partner's religious denomination varies according to age group. The chi-square statistics for the three sub-tables are 13.4, 55.3 and 54.2. Using these chi-square statistics, test the relationship in each sub-table for significance.

(ii) Use odds ratios to summarise the relationships between female partner's religious denomination and age group, and between male partner's religious denomination and age group.

(iii) Use the following results corresponding to the 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 part (i).

(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.	[F][M][A]	96.8	7	0.000				
2.	[FM][A]	20.4	6	0.002	76.4	1	0.000	1
3.	[MA][F]	90.7	5	0.000	6.1	2	0.048	1
4.	[FA][M]	95.8	5	0.000	1.0	2	0.613	1
5.	[FM][MA]	14.3	4	0.006	6.1	2	0.048	2
6.	[FM][FA]	19.4	4	0.000	1.0	2	0.613	2
7.	[FA][MA]	89.7	3	0.000	1.0	2	0.613	3
8.	[FM][RA][MA]	14.1	2	0.001	0.2	2	0.919	5
9.	[FMA]	0.0	0	-----	14.1	2	0.008	8

/continued.....

SECTION B

6 `Many research questions can be answered adequately using existing data and without the collection of any new data whatsoever'. Discuss, with reference to the potential as a resource for secondary analysis of ONE social survey of your choice.

---

7 `The construction of a high quality questionnaire is as much a conceptual exercise as it is a technical exercise'. Discuss, with reference to ONE hypothetical survey-based research project of your choice.

---

8 `Social survey interviews and ethnographic interviews are very different; as a consequence the role of the interviewer also differs markedly between these two types of interview'. Discuss.

---

9 Why do a greater proportion of sociological research projects than in the past use both quantitative and qualitative approaches within the same project?

---

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 focussed statistical tests that you would ideally like to carry out].

REGION OF BRITAIN by LEVEL OF ENVIRONMENTALLY-FRIENDLY BEHAVIOUR

	Level of environmentally-friendly behaviour					
	Score = Low		Score= Medium		Score = High	
Region		%		%		%
Scotland	55	(47.4)	42	(36.2)	19	(16.4)
North	137	(36.7)	136	(36.5)	100	(26.8)
Midlands	82	(38.1)	62	(28.8)	71	(33.0)
Wales	40	(53.3)	23	(30.7)	12	(16.0)
South	122	(29.0)	136	(32.3)	163	(38.7)
London	42	(32.8)	52	(40.6)	34	(26.6)

[Note: The variable `Level of environmentally-friendly behaviour' is a scale. Low scores on the scale supposedly correspond to a limited or negligible extent of `green' behaviour; high scores on the scale supposedly correspond to extensive `green' behaviour].
