

SO 2010

UNIVERSITY OF WARWICK

Summer 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 dependent children in a random sample of 225 female-headed lone-parent families was found to be 1.69 children, with a sample standard deviation of 0.75 children.
- (i) Calculate a 95% confidence interval for the mean number of dependent children in female-headed lone-parent families.
- (ii) The mean number of dependent children in two-parent families is known to be 1.86 children. Calculate a z-statistic and use it to test whether this is a plausible mean number of dependent children for female-headed lone-parent families.
- (iii) If the population standard deviation for the mean number of dependent children in male-headed lone-parent families is assumed to be 0.60, 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 dependent children in male-headed lone-parent families? 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 the latter part of (ii) without calculating a z-statistic.

---

/continued.....

- 
- 2 The following table shows whether women smoked cigarettes according to their (Registrar General's) occupational social class for a (stratified) random sample of 750 women.

<u>Class</u>	<u>Yes</u>	<u>No</u>	<u>TOTAL</u>
I or II	57	123	180
III NM	78	192	270
III M	25	50	75
IV or V	90	135	225
TOTAL	250 (33.3%)	500 (66.7%)	750

III NM = III Non-Manual; III M = III Manual.

- (i) Calculate the chi-square statistic for the above table and use it to test the hypothesis that there is no relationship between smoking cigarettes and occupational social class for women.

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

- (ii) A similar table based on a stratified random sample of 750 men, once again showing the relationship between smoking cigarettes and occupational social class, gave rise to a chi-square statistic of 27.8. 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 the likelihood of smoking cigarettes differs between women with non-manual occupations (classes I, II and III NM) and women with manual occupations (classes III M, IV and V)?

(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 household size according to geographical region of a random sample of 864 English households.

<u>Region</u>	<u>Mean</u>	<u>s</u>	<u>n</u>
North	2.75	1.46	259
Midlands	2.72	1.37	173
South	2.56	1.25	324
London	2.41	1.26	108
		TOTAL	864

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

- (i) Test the hypothesis that, in the population, all four regions 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 3 degrees of freedom and 860 degrees of freedom is 2.60; the between-groups and within-groups sums of squares are 21.0 and 1505.0 respectively).

- (ii) Test the hypothesis that, in the population, households in the South and households in London have the same mean household size.

(Note: the critical value of t at the 5% level corresponding to 430 degrees of freedom is 1.96; the pooled sample standard deviation for households in the South and households in London is 1.25)

/continued.....

- 4 In interviews with a random sample of 122 adults, the (Pearson) correlation between interview duration (in minutes) and interviewee's age (in years) was found to be 0.20.

- (i) Test the hypothesis that there is no relationship between interview duration and interviewee's age.

(Note: the critical value of F at the 5% level corresponding to 1 degree of freedom and 120 degrees of freedom is 3.92).

The regression equation corresponding to the dependence of interview duration on interviewee's age is

$$y = 50.0 + 0.20x_1$$

where  $y$  is interview duration (in minutes), and  $x_1$  is interviewee's age (in years).

- (ii) Use the above equation to predict the interview duration of an interviewee aged 20 years. The actual mean duration for 20 year-old interviewees is 52.5 minutes. Why do you think the prediction differs from this figure? Why might the above linear regression equation be an inappropriate model of the relationship between interview duration and interviewee's age? Use the above equation to predict the interview duration for 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 interviewee is in paid employment and the value 0 otherwise, results in the following equation

$$y = 47.5 + 0.25x_1 + 2.10x_2$$

- (iii) Explain why the coefficient of  $x_1$ , interviewee's age, changes between the two equations. (Note that the questionnaire used in the interview contains a number of questions relating to the interviewee's current employment).

---

/continued.....

5 The following crosstabulation is of party identification [P] by religious denomination [R] by age group [A] for a random sample of 2,041 British adults. (The religious denomination variable is limited to a dichotomy, distinguishing between Christianity on the one hand and other denominations, agnosticism and atheism on the other hand):

AGE GROUP = 16-39	Christian	Other	TOTAL
Conservative	181	146	327
Labour	<u>181</u>	<u>247</u>	<u>428</u>
TOTAL	362	393	755

  

AGE GROUP = 40-59	Christian	Other	TOTAL
Conservative	264	111	375
Labour	<u>187</u>	<u>100</u>	<u>287</u>
TOTAL	451	211	662

  

AGE GROUP = 60 plus	Christian	Other	TOTAL
Conservative	293	45	338
Labour	<u>214</u>	<u>72</u>	<u>286</u>
TOTAL	507	117	624

- (i) Use odds ratios to summarise the way in which the relationship between party identification and religious denomination varies according to age group. The chi-square statistics for the three sub-tables are 12.7, 2.1 and 14.3. Using these chi-square statistics, test the relationship in each sub-table for significance.
- (ii) Use odds ratios to summarise the relationships between party identification and age group, and between 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 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.	[P][R][A]	232.6	7	0.000				
2.	[PR][A]	195.8	6	0.000	36.8	1	0.000	1
3.	[RA][P]	57.9	5	0.000	174.7	2	0.000	1
4.	[PA][R]	203.7	5	0.000	28.9	2	0.000	1
5.	[PR][RA]	21.1	4	0.000	36.8	1	0.000	3
6.	[PR][PA]	166.9	4	0.000	28.9	2	0.000	2
7.	[PA][RA]	29.1	3	0.000	28.8	2	0.000	3
8.	[PR][PA][RA]	4.3	2	0.119	16.8	2	0.000	5
9.	[PRA]	0.0	0	-----	4.3	2	0.119	8

/continued.....

SECTION B

6 'The data collected by social surveys form a reservoir of useful information that secondary analysts can exploit'. Discuss, with reference to the potential as a resource for secondary analysis of ONE social survey of your choice.

7 'The operationalisation of concepts and the design of the questionnaire are crucial stages if a survey is to be successful'. Discuss, with reference to ONE hypothetical survey-based research project of your choice.

8 'Social survey interviews generate standardised data in a mechanical fashion, whereas ethnographic interviews are conversations with a purpose'. Discuss.

9 ✗ What are the advantages and difficulties of combining quantitative and qualitative approaches within a single research 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].

**MARITAL STATUS by ATTITUDE TOWARDS MARRIED WOMEN AND PAID WORK**

<u>Marital status</u>	<u>Attitude towards married women and paid work</u>					
	Score = Low		Score = Medium		Score = High	
		%		%		%
Married	168	(29.0)	271	(46.8)	140	(24.2)
Living as married	5	(10.6)	22	(46.8)	20	(42.6)
Separated\Divorced	21	(26.3)	43	(53.8)	16	(20.0)
Widowed	28	(30.4)	54	(58.7)	10	(10.9)
Never married	28	(17.0)	73	(44.2)	64	(38.8)

[Note: The variable 'Attitude towards married women and paid work' is a scale. Low scores on the scale supposedly correspond to a 'traditional' attitude towards married women and paid work, i.e. that in some situations they should not work; high scores on the scale supposedly correspond to a 'modern' attitude towards married women and paid work].