

SO 2010

UNIVERSITY OF WARWICK

Summer Examinations 2005

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 age at first cohabitation or marriage for a random sample of 1,024 women aged 40-49 living in Britain (in 1990) who attended single-sex secondary schools was found to be 22.25 years, with a sample standard deviation of 4.00.
- (i) Calculate a 95% confidence interval for the mean age at first cohabitation or marriage of women aged 40-49 living in Britain who attended single-sex secondary schools.
  - (ii) The mean age at first cohabitation or marriage of all women aged 40-49 living in Britain is known to be 21.85 years. Calculate a z-statistic and use it to test whether this is a plausible mean age at first cohabitation or marriage of women aged 40-49 living in Britain who attended single-sex secondary schools.
  - (iii) Suppose that the population standard deviation for the age at first cohabitation or marriage of men aged 40-49 living in Britain who attended single-sex secondary schools is assumed to be 4.50. How big a sample would be needed to produce a sample mean that one could be 95% confident fell within 3 months (0.25 years) of the population mean age at first cohabitation or marriage of men aged 40-49 living in Britain who attended single-sex secondary schools? 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 highest educational qualification and past cohabitation for a random sample of 450 single (never married) people in Britain who were 23-25 years old.

	<u>Ever cohabited</u>	<u>Never cohabited</u>	<u>Total</u>
<u>Highest qualification</u>			
Degree	15	135	150
'A' level or above	42	138	180
GCSE or 'O' level	22	58	80
None of above	11	29	40
TOTAL	90 (20%)	360 (80%)	450

- (i) Calculate the chi-square statistic for the above cross-tabulation and use it to test the hypothesis that there is no relationship between highest educational qualification and past cohabitation for single (never married) people in Britain aged 23-25 years.

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

- (ii) A similarly shaped cross-tabulation, based on a random sample of 700 single (never married) people in Britain who were 26-30 years old, and again showing the relationship between highest educational qualification and past cohabitation, gave rise to a chi-square statistic of 2.94. 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, in Britain, past cohabitation varies significantly between single (never married) people aged 23-25 years with degrees and single (never married) people aged 23-25 years without degrees? (Calculate a chi-square statistic to answer this part of the question, and comment on the result 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 1 degree of freedom is 3.84).

Continued.....

- 3 The following table shows the mean number of hours per week spent reading (as a leisure activity), according to employment status, for a random sample of 1,347 adult women in Britain.

<u>Employment status</u>	<u>Mean</u>	<u>s</u>	<u>n</u>
Full-time paid employment	4.11	5.77	360
Part-time paid employment	3.42	5.12	240
Unemployed	3.56	5.92	55
Looking after home	4.64	7.17	246
Retired	6.48	7.85	340
Full-time student	3.70	4.03	26
Other	6.30	8.71	80
		TOTAL	1,347

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

- (i) Test the hypothesis that, in the population, the mean number of hours per week spent reading corresponding to each different employment status is the same. Discuss your findings with reference to the sample means.

(Note: the critical value of F at the 5% level corresponding to 6 degrees of freedom and 1,340 degrees of freedom is 2.10; the between-groups and within-groups sums of squares are 1,890.0 and 60,300.0 respectively).

- (ii) Test the hypothesis that, in the population, the mean numbers of hours per week spent reading for women in full-time paid employment and for women in part-time paid employment are the same.

(Note: the critical value of t at the 5% level corresponding to 598 degrees of freedom is 1.96; the pooled sample standard deviation for women in full-time paid employment and women in part-time paid employment is 5.52).

Continued.....

- 
- 4 In a random sample of 38 Local Authorities in England and Wales, the (Pearson) correlation between the annual out-migration rate (expressed as a percentage) and the percentage of the population who were 45-59 years old was found to be  $-0.44$ .

- (i) Test the hypothesis that there is no relationship between out-migration rate and the percentage of the population aged 45-59 years for Local Authorities in England and Wales.

(Note: You may assume that  $(-0.44)^2$  is equal to 0.19; 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 out-migration rate on the percentage of the population aged 45-59 years for Local Authorities in England and Wales is

$$y = 11.80 - 0.36x_1$$

where  $y$  is the out-migration rate, and  $x_1$  is the percentage of the population aged 45-59 years.

- (ii) Use the above equation to predict the out-migration rates for three hypothetical Local Authorities in England and Wales, in which 0%, 20% and 40% of the population are aged 45-59 years. 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 corresponds to the percentage of the population aged 20-24 years, leads to the following equation

$$y = 0.27 + 0.03x_1 + 0.70x_2$$

- (iii) Explain why the coefficient of  $x_1$ , the percentage of the population aged 45-59 years, changes between the two equations. Use the second regression equation to predict the out-migration rate for the Warwick Local Authority, in which 20% of the population is aged 45-59 years and 7% of the population is aged 20-24 years. Comment on this prediction with reference to one of the predictions made for part (ii).

---

Continued.....

- 5 The following cross-tabulation is of household social class [C] by housing tenure [T] by family type [F] for a random sample of 2,438 families living in households in Britain.

<b>FAMILY TYPE = Lone parent</b>	<u>Owens</u>	<u>Rents</u>	<u>TOTAL</u>
Middle class	116	86	202
Working class	<u>82</u>	<u>255</u>	<u>337</u>
TOTAL	198	341	539
<b>FAMILY TYPE = Stepfamily</b>	<u>Owens</u>	<u>Rents</u>	<u>TOTAL</u>
Middle class	64	16	80
Working class	<u>34</u>	<u>37</u>	<u>71</u>
TOTAL	98	53	151
<b>FAMILY TYPE = Other family</b>	<u>Owens</u>	<u>Rents</u>	<u>TOTAL</u>
Middle class	1074	98	1172
Working class	<u>380</u>	<u>196</u>	<u>576</u>
TOTAL	1454	294	1748

- (i) Use odds ratios to summarise the way in which the relationship between household class and housing tenure varies according to family type. The chi-square statistics for the three sub-tables are 59.5, 17.0 and 181.8. Using these chi-square statistics, test the relationship in each sub-table for significance.
- (ii) Use odds ratios to summarise the relationships between:  
(a) family type and household class; (b) family type and housing tenure.
- (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	Compared to model
1	[F] [C] [T]	811.7	7	0.000				
2	[CT] [F]	439.1	6	0.000	372.6	1	0.000	1
3	[FT] [C]	398.6	5	0.000	413.1	2	0.000	1
4	[FC] [T]	660.8	5	0.000	150.9	2	0.000	1
5	[CT] [FC]	288.2	4	0.000	150.9	2	0.000	2
6	[FT] [FC]	247.7	3	0.000	150.9	2	0.000	3
7	[FT] [CT]	26.0	4	0.000	372.6	1	0.000	3
8	[FT][CT][FC]	1.8	2	0.411	24.2	2	0.000	7
9	[FCT]	0.0	0		1.8	2	0.411	8

Continued....

---

**SECTION B**


---

- 6 To what extent are the strengths and limitations of secondary analysis-based research specific to its use of existing data? Discuss, with particular reference to ONE social survey of your choice.
- 
- 7 Discuss, with reference to ONE real or hypothetical survey of your choice, the contributions made by the operationalization of concepts and questionnaire design to the validity and effectiveness of the survey research process.
- 
- 8 To what extent should survey interviews and qualitative interviews be viewed as equivalent approaches to collecting data?
- 
- 9 Under what circumstances is there something of value to be gained by using both quantitative and qualitative methods within a single research project?
- 
- 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].

**HIGHEST QUALIFICATION by EXTENT OF 'GREEN' ACTIVITIES**

Extent of 'Green' activities (scale)

<u>Highest qual.</u>	Score = Low		Score = Medium		Score = High	
	%		%		%	
Degree	14	(11.8)	69	(58.0)	36	(30.3)
Other Higher educ.	19	(19.6)	49	(50.5)	29	(29.9)
'A'/'O' level/GCSE	63	(24.0)	154	(58.6)	46	(17.5)
Other qual.	32	(38.1)	41	(48.8)	11	(13.1)
None	27	(18.6)	99	(68.3)	19	(13.1)

[Notes: Low scores on the 'Extent of Green activities' scale correspond to limited engagement in actions aimed at protecting the environment; high scores on the scale correspond to more extensive engagement in such actions].

---

 END