SO2010

UNIVERSITY OF WARWICK

Summer Examinations 2013/14

SURVEYS, SECONDARY ANALYSIS AND SOCIAL STATISTICS

Candidates should choose THREE of the four sections and answer ONE question from each of these THREE sections.

Time allowed: 2 hours plus 15 minutes reading time during which notes may be made (on the question paper only) but NO ANSWER MAY BE BEGUN.

Read the instructions on the answerbook carefully and make sure that the particulars required are entered on each answerbook.

SECTION A

---

1. Discuss, with reference to a substantive topic of your choice, the main issues that a researcher should consider when carrying out a secondary analysis of existing survey data.

---

2. Outline how **AT LEAST TWO** measures operationalise concepts, with reference to a substantive topic of your choice. (The measures should include **AT LEAST ONE** scale constructed from a number of items and **AT LEAST ONE** categorical measure.)

SECTION B

---

3. Assess the benefits of applying survival analysis techniques within a specific substantive field, with reference to **AT LEAST ONE** published article.

---

4. What are the central issues that arise in the construction and use of measures of inequality? Discuss, with reference to a particular substantive area and using **AT LEAST TWO** published articles relating to it.

---

5. Illustrate the value of the insights into a particular substantive topic that can be provided by the application of clustering and/or scaling techniques, with reference to **AT LEAST ONE** published article.

Continued...

SECTION C

6. Table 6.1 shows the relationship in a random sample of parents of school-age children in Britain between their social (occupational) class and how likely they think it is that their children will go to university. Table 6.2 disaggregates Table 6.1 according to the parent's educational level. Chi-square and Cramér's V values are given for each cross-tabulation.

TABLE 6.1

Social class	Very/ Fairly likely	Not very/ Not at all likely	Total
Service/Intermediate	106 (88.3%)	14 (11.7%)	120
Working	52 (68.4%)	24 (31.6%)	76
TOTAL	158 (80.6%)	38 (19.4%)	196

Chi-square = 11.8 (1 d.f.; $p = 0.001$); Cramér's V = 0.245

TABLE 6.2

Educational level = Degree			
Social class	Very/ Fairly likely	Not very/ Not at all likely	Total
Service/Intermediate	34 (100.0%)	0 (0.0%)	34
Working	4 (100.0%)	0 (0.0%)	4
TOTAL	38 (100.0%)	0 (0.0%)	38
Educational level = 'A' level or above, but below degree			
Social class	Very/ Fairly likely	Not very/ Not at all likely	Total
Service/Intermediate	35 (92.1%)	3 (7.9%)	38
Working	14 (66.7%)	7 (33.3%)	21
TOTAL	49 (83.1%)	10 (16.9%)	59
Educational level = Below 'A' level			
Social class	Very/ Fairly likely	Not very/ Not at all likely	Total
Service/Intermediate	37 (77.1%)	11 (22.9%)	48
Working	34 (66.7%)	17 (33.3%)	51
TOTAL	71 (71.7%)	28 (28.3%)	99

Degree: Chi-square = Undefined; Cramér's V = Undefined
A' level or above: Chi-square = 6.2 (1 d.f.; $p = 0.013$); Cramér's V = 0.325
Below 'A' level: Chi-square = 1.3 (1 d.f.; $p = 0.250$); Cramér's V = 0.116

Continued...

(i) Calculate the (three) odds ratios for the relationship between social class and how likely parents think that it is that their children will go to university corresponding to Table 6.1 and to two of the three layers within Table 6.2. Discuss what can be learned from this multivariate cross-tabulation analysis about the relationships between the three variables.

(ii) Use odds ratios to summarise the relationships between educational level and: (a) social class; (b) how likely parents think that it is that their children will go to university.

(iii) Use the following results corresponding to the goodness-of-fit of various log-linear models to determine the most appropriate model of the three-way cross-tabulation. Justify your choice, and explain how the model that you have selected relates to your findings from parts (i) and (ii).

Model No.	**Model**	**Deviance**	**d. f.**	**P**	**Change in deviance**	**d. f.**	**P**	**Comp -ared to model**
1	[S] [U] [E]	50.7	7	0.000				
2	[SU] [E]	39.1	6	0.000	11.6	1	0.001	1
3	[SE] [U]	28.5	5	0.000	22.2	2	0.000	1
4	[UE] [S]	29.5	5	0.000	21.2	2	0.000	1
5	[SU] [SE]	16.9	4	0.002	11.6	1	0.001	3
6	[SU] [UE]	18.0	4	0.001	11.5	1	0.001	4
7	[SE] [UE]	7.3	3	0.063	21.2	2	0.000	3
8	[SU][SE][UE]	2.1	2	0.355	5.2	1	0.022	7
9	[SUE]	0.0	0		2.1	2	0.355	8

[S] = Social class; [U] = Children will go to university?; [E] = Educational level.

Continued…

SECTION D

7. The following are extracts from an article focusing on social distance in relation to people with mental illness and incorporating a linear regression analysis:

"Social distance is commonly considered as a paradigm of public attitudes toward persons with mental illness and defined as the relative willingness of one person to participate in relationships of varying degrees of intimacy with a person who has a stigmatized identity (Bowman, 1987). … little is yet known about the psychological and sociological factors that influence social distance toward people with mental illness. We therefore conducted a survey on public attitudes toward mental illness and psychiatric treatment … [drawing] a representative sample of the residential Swiss population aged 16 to 76 years living in private households (n = 1737). Out of all Swiss phone numbers (hit rate 89.7%), a random sample of households was drawn. A target person in each household was randomly selected … The response rate was 63%. Main reasons for refusal were 'no interest' (39%), 'disapprove of opinion polls' (20%), and 'no time' (15%)." (p266).

"The public attitudes were assessed using computer-assisted telephone interviewing (CATI). CATI should particularly reduce potential measurement error associated with item wording and order, interviewers' verbal behavior, and data processing (Lavrakas, 1993). All interviewers were thoroughly trained and supervised during the survey." (p267).

"According to our understanding, attitudes depend on different levels. Thus, we included demographic, psychological, and sociological concepts in our study. The questionnaire comprises three consecutive parts … (1) General questions; e.g. to the perceived discrimination (Cronbach's α-coefficient of reliability: 0.72), the attitudes to restrictions toward the mentally ill (Cronbach's α: 0.48), to community psychiatry (Cronbach's α: 0.75), to lay helping in psychiatry (Cronbach's α: 0.65), and to psychotropics (Cronbach's α: 0.67). (2) A vignette depicting a case of either major depression or schizophrenia. … Half of the interviewees (n = 893) were randomly given the vignette along with the respective psychiatric diagnosis. The remaining 844 interviewees (that were a representative subsample of the complete sample) were not informed of the diagnosis but were asked to indicate whether the person described was either suffering from a 'mental illness' or experiencing a 'crisis.' ... Questions to assess emotional reactions (Cronbach's α: 0.68) and social distance (Cronbachs α = .86) toward the case depicted followed immediately. Social distance was assessed by the Social Distance Scale SDS (Link et al., 1987; German translation: Angermeyer et al., 1987). The SDS includes seven questions assessing the willingness to interact with a person with mental illness in various social situations (see Table 1). Each item is rated on a 5-point Likert scale anchored by 1 'definitely unwilling' and 5 'definitely willing.' (3) Treatment recommendations in favor of the person depicted, acceptance of psychotropic side effects (Cronbach's α: 0.84), contact with mentally ill people (Cronbach's α: 0.49), and demographic factors were assessed … In order to estimate the effect of different factors on social distance, we carried out a multiple linear regression analysis. All independent variables were entered simultaneously." (p267).

"Table 1 shows the respective percentage of the seven items of the Social Distance Scale. As the level of social distance *increases* the more the situation described implies '*social closeness.*' The greatest social distance was therefore indicated when the respondent was asked whether the person depicted should *look after children* or *marry* into their family." (pp268-269).

Continued…

TABLE 1

Social Distance to a Person with Mental Illness in Different Social Contexts (N = 844; missing data <2,7%)

Would you . . .	*Definitely Unwilling*	*Unwilling*	*Neutral*	*Willing*	*Definitely Willing*
. . . be willing to start work with a person like B.*?	2,6%	6,4%	24,5%	38,0%	27,5%
. . . like to move next door to a person like B.*?	1,9%	10,1%	26,5%	36,1%	25,1%
. . . make friends with a person like B.*?	9,8%	17,5%	31,0%	26,8%	14,6%
. . . rent a room to a person like B.*?	11,5%	22,9%	34,7%	22,0%	8,6%
. . . recommend a person like B.* for a job?	9,4%	24,2%	35,4%	22,0%	8,3%
. . . like your child to marry a person like B.*?	13,3%	25,6%	32,8%	17,8%	7,9%
. . . trust a person like B.* to take care of your child?	25,1%	30,2%	25,8%	14,2%	4,5%

*The person depicted in the vignette was called Beat.

[Table presented on p268.]

"Linear regression analysis could be performed with the data of 594 respondents. The difference to the 844 interviewees is due to missing values. Linear regression analysis (Table 2) showed the following factors to *increase* social distance: the vignette depicting a case of schizophrenia (β = .231; $p < 0.000$), negative emotions (β = .159; $p < 0.000$), acceptance of negative sanctions (β = .150; $p < 0.000$), age (β = .150; $p < 0.000$), female gender (β = .137; $p < 0.000$), living in the Italian-speaking part of Switzerland (β = .120; $p < 0.000$), acceptance of psychotropic side effects (β = .114; $p = 0.001$), favoring medical treatment (β = .106; $p = 0.001$), correct recognition of the vignette as illness (β = .104; $p < 0.01$), and perceived discrimination (β = .080; $p < 0.05$). Social distance is *decreased* by the following variables: A positive attitude to lay helping (β = −.211; $p < 0.000$) and to community psychiatry (β = −.167; $p < 0.000$) as well as an interest in psychiatric topics in the mass media (β = −.124; $p < 0.000$), and contacts to persons with mental illness (β = −.074; $p < 0.05$). The *explained variance* of the model is 44.8% (R^2 adj.)." (p269).

Continued…

TABLE 2

Results of the Multiple Linear Regression Analysis for Social Distance Toward People with Mental Illness in Switzerland (N = 594; R^2 adj. = .448)

	β	*p*
Schizophrenia vignette	.231	<0.000
Positive attitude to lay helping	–.211	<0.000
Positive attitude to community psychiatry	–.167	<0.000
Negative emotions	.159	<0.000
Acceptance of negative sanctions	.150	<0.000
Age	.150	<0.000
Sex (female)	.137	<0.000
Interest in psychiatric topics in the mass media	–.124	<0.000
Living in the Italian-speaking part of Switzerland	.120	<0.000
Acceptance of psychotropic side effects	.114	0.001
Favoring medical treatment	.106	0.001
Correct recognition of the vignette as an illness	.104	<0.01
Perceived discrimination	.080	<0.05
Contact to person with mental illness	–.074	<0.05

[Table presented on p269.]

"[Discussion] ... we examined how demographic, psychological, and sociological factors predict social distance. Four groups of predictor variables were found: The *illness* depicted (i.e. schizophrenia), attitudes to *general aspects* of mental health (lay helping, community psychiatry), *emotions* toward those affected, and attitude toward *consequences* of mental illness (medical treatment, medication side effects, negative sanctions, e.g. withdrawal of the driver license)." (p270).

Extracts from: Lauber, C., Nordt, C., Falcato, L. and Rössler, W. 2004. 'Factors Influencing Social Distance Toward People with Mental Illness', Community Mental Health Journal 40 3: 264-274.

(Note that the material inserted between square brackets within the above extracts has been added to the text for the purposes of this exam question).

***

(i) What are the strengths and limitations of the above linear regression analysis, including the way in which it is reported?

(ii) The article provides very little further information about the linear regression analysis, or about the variables used in it. What additional material could the authors have included which would have helped you assess the merits and weaknesses of their analysis?

Continued...

8. The following are extracts from an article focusing on poverty, delinquency, and educational attainment, and incorporating a number of logistic regression analyses:

"The present study addresses [a] gap in the literature by examining whether high socioeconomic status protects youth from the negative impact of delinquency on their educational future." (p575).

"Data from the National Longitudinal Survey of Youth (NLSY) were used … The survey began in 1979 with a national household probability sample of 6,111 youth between the ages of 14 and 21 and two supplementary samples: an oversample of racial and ethnic minorities and low-income youth, and an additional sample of youth in the military. For the current study, I use both the national household probability sample and the oversample of disadvantaged youth. [Endnote states: I use the oversample of disadvantaged youth to help ensure a reliable statistical comparison with middle- and upper-class youth. However, the central findings of this study were also replicated with just the national probability sample.] Because the present study focuses on delinquency's effect on later educational attainment, a subset of individuals between the ages of 14 and 17 enrolled in school in 1979 was selected from the main sample. Youth already outside of school in 1979 were excluded from the sample because their educational attainment is often largely determined at this point and their reported levels of delinquency could be seen as an effect, rather than a cause, of their educational attainment … Sample attrition for the 1990 survey and listwise deletion of missing values reduced the subset by approximately 20 percent. Univariate analyses of missing cases from sample attrition suggested that the final sample was not significantly different from the base sample in regard to the distribution of key variables." (p580).

"Three measures of self-reported misbehavior in 1980 were used in the analyses. Delinquency was first defined by a variety scale of thirteen items related to a wide range of adolescent deviance (α = 0.78). Each item was coded such that an individual scored 1 if a specific delinquent act was reported one or more times and 0 otherwise (scale range: 0–13). … a separate set of analyses included measures of misconduct drawing official response from the criminal justice system or resulting in disciplinary action from the school. Specifically, these measures are the number of times an individual has been arrested or officially charged with criminal activity other than a minor traffic offense and the number of times an individual has been suspended from school (coded 0 for never, 1 for once, 2 for twice, 3 for three times, and 4 for more than three times). … Socioeconomic disadvantage was operationalized using a dummy variable indicating whether a respondent's family income in 1978 was above or below the federal poverty line, taking into account variation in family size. … While most previous studies use measures of social class derived from data on parents' (usually father's) occupational prestige, measures of occupational status appear to be better suited for explaining variation among individuals in the higher classes (Farnworth et al. 1994; Jarjoura 1996). Since the present study is focused on discerning differences between the lower class and the rest of society, an indicator of poverty status seemed most appropriate. … Two measures of educational attainment were examined in the analyses: the raw number of years of school completed and a dummy variable for high school dropout status in 1990." (pp580-581).

"Logistic regression was used to analyze the dropout-status dependent variable. Ordinary least squares regression (OLS) was used to examine variation in the number of years of school completed." (p581).

Continued…

"[T]heoretical predictions … suggest a conditioning effect (or interaction effect) of poverty status on the relationship between delinquency and educational attainment [and] that attainment processes work differently in the lower class than in all other social classes. Therefore, the equations were estimated separately for those in poverty and those not in poverty to compare the strength of the effects of delinquency on educational attainment for both groups. In order to determine whether or not differences in coefficients were statistically significant, a *t*-statistic was calculated …" (p582).

"All models included the standard demographic controls for family structure… , number of siblings, sex …, age, and ethnicity and race. … In addition, statistical controls were added for two frequently used, powerful predictors of educational attainment: educational aspirations and academic ability…" (p582).

Table 2
Logistic Regression Estimates for Dropout Status by Poverty Background

Independent Variables	Nonpoor	Poor	*t*	Nonpoor	Poor	*t*	Nonpoor	Poor	*t*
# of times suspended	.505*	.198*	3.20*	—	—		—	—	
	(.075)	(.060)							
Delinquency scale	—	—		.118*	.046	1.44	—	—	
				(.038)	(.033)				
# of times charged/arrested	—	—		—	—		.533*	.110	2.53*
							(.147)	(.080)	
Educational aspirations	−.224*	−.193*		−.250*	−.211*		−.246*	−.203*	
	(.057)	(.044)		(.058)	(.046)		(.056)	(.044)	
Academic aptitude	−.063*	−.079*		−.068*	−.085*		−.064*	−.080*	
	(.007)	(.008)		(.007)	(.008)		(.007)	(.008)	
Female	−.176	−.151		−.176	−.169		−.332	−.201	
	(.204)	(.158)		(.210)	(.166)		(.199)	(.157)	
Age	−.185	−.217*		−.147	−.189*		−.147	−.202*	
	(.096)	(.076)		(.096)	(.078)		(.094)	(.076)	
African-American	−1.03*	−1.34*		−.731*	−1.23*		−.803*	−1.27*	
	(.327)	(.202)		(.327)	(.207)		(.317)	(.203)	
Hispanic	−.172	−.167		−.105	−.174		−.154	−.181	
	(.380)	(.208)		(.376)	(.214)		(.376)	(.208)	
Two-parent family	−.590*	−.324		−.489*	−.331		−.518*	−.345*	
	(.209)	(.163)		(.213)	(.167)		(.208)	(.163)	
# of siblings	.070	.046		.059	.041		.067	.046	
	(.044)	(.027)		(.045)	(.027)		(.044)	(.027)	
Intercept	5.53	6.41		5.38	6.40		5.51	6.50	
Model χ^2	354	236		275	220		300	228	
Sample size	1,965	1,228		1,898	1,163		1,964	1,230	

Note: Standard errors in parentheses
* $p < 0.05$

[The above table comes from pages 584 and 585]

Continued…

"Table 2 illustrates the effects of the delinquency measures on high-school dropout status separately for youth from poverty backgrounds and youth not from poverty backgrounds. Logistic regression analyses revealed that the effect of the delinquency scale on the log odds of dropping out was statistically significant and positive in the nonpoverty sample (0.118, $p < 0.05$) and practically zero for the poverty group. Likewise, the effect of the number of times charged/arrested on dropout status was positive and significant for the nonpoverty group (0.533, $p < 0.05$) but statistically unrelated for the poverty sample. The effect of the number of times suspended from school was positive and significant for both groups, but was much more pronounced in the nonpoverty sample (0.505, $p < 0.05$) than in the poverty sample (0.198, $p < 0.05$). A *t*-test for equality of regression coefficients showed that the apparent social class differences in the effects of the number of times suspended and number of times charged/arrested are statistically significant ($p < 0.05$). In other words, the differences between the suspension and charge/arrest coefficients for the poverty and nonpoverty groups are, in all likelihood, more than just a product of random variation. While the effect of the delinquency scale on dropout for the nonpoverty group appeared greater than that of the poverty group, the *t*-test suggested that the apparent difference in the effect of the delinquency scale could simply be due to chance." (pp583-588)

Extracts from: Hannon, L. 2003. 'Poverty, Delinquency, and Educational Attainment: Cumulative Disadvantage or Disadvantage Saturation?', Sociological Inquiry 73.4: 575–594.

(Note that the material inserted between square brackets within the above extracts has been added to the text for the purposes of this exam question).

***

(i) What are the strengths and limitations of the above logistic regression analysis, including the way in which it is reported?

(ii) Table 2 contains six different logistic regression analyses with the same dependent variable. What would have been the pros and cons of combining all the analyses into a single logistic regression?

Continued…

9. The following are extracts from an article focusing on the long-term effects of divorce on the self-esteem of young adults, and incorporating analyses of variance (ANOVAs) involving two independent variables:

"... there is a growing body of research suggesting that the well-being of children is affected more by family dynamics than by family structure (Emery, 1982; Glenn & Kramer, 1985; Kulka & Weingarten, 1979; Long, 1986). ... The goal of this research is to quantitatively examine the long-term effects of family structure (intact vs. divorced) and interparental conflict on the general and social self-esteem of young adults (18-24 year old undergraduates). Self-esteem is selected as a dependent variable because it is seen as an indicator of well-being across the life-span (Long, 1986; Wylie, 1974) ... To summarize, there are two theoretical perspectives concerning the relative long-term effects of family structure and interparental conflict on the self-esteem of children. One, both, and/or an interaction of these independent variables may best explain the variance in self-esteem." (pp131-132).

"Data were gathered from 324 male and female college students from 12 ... classes at four New England universities. A total of 306 usable surveys were collected from students 18 through 24 years of age ... The questionnaire consisted of four sections: (1) Demographics; (2) *Schwarz Inter-Personal Conflict Scale* – child form (*IPC*); (3) *Social Self-Concept Scale* (SSCS); (4) *Rosenberg Self-Esteem Scale (RSES)*. Section 1 permitted the assignment of subjects to intact or divorced groups. ... The *IPC* (Schwarz & Zuroff, 1979) is a 37 item 7-point Likert scale (never to at least once a week) developed with an undergraduate sample to assess the recalled frequency of interparental conflict. ... The *SSCS* is a 45 item 4-point Likert scale (strongly disagree to strongly agree) developed with an undergraduate sample to assess social self-esteem. The theoretical rationale for this construct is based in the multi-faceted-hierarchical model of self-concept advanced by Shavelson, Stanton and Huber (1976) ... evidence of the instrument's construct validity is provided by Zorich and Reynolds (1988). ... The *RSES* is a 10 item 4-point Likert scale (strongly disagree to strongly agree) developed to measure general self-esteem. For over 20 years it has been used with a large variety of samples ranging across the life-span from children to older adults." (pp132-133).

"The final sample consisted of 228 (74.3%) females and 78 (25.4%) males. ... a series of t-tests revealed no significant differences between males and females with respect to age, *RSES*, *SSCS* or Grade Point Average (GPA). A series of one-way ANOVAs indicated no statistically significant differences among the 4 universities or 12 classes with respect to *RSES*, *SSCS* or GPA. ... Since the distribution [of interparental conflict values] was slightly skewed, the median was selected to divide the sample into [low and high] conflict groups." (p133).

"Item to total scale correlations computed for the *IPC* ranged from .24 to .67. The internal consistency reliability (α) was found to be .95. Item to total scale correlations computed for the *SSCS* ranged from .22 to .72. The internal consistency reliability (α) was found to be .95. Item to total scale correlations computed for the *RSES* ranged from .59 to .80. The internal consistency reliability (α) was found to be .91. Summarizing, the three measures, *IPC*, *SSCS* and *RSES*, were found to be highly reliable for this study sample and the internal consistency reliabilities were "consistent" with previously reported statistics for each of the three instruments". (p134).

Continued...

"The results were the same for both two-way ANOVA analyses. With respect to general self-esteem and again for social self-esteem, there were no significant main effects for family structure and no significant interactions between family structure and interparental conflict. However, there were significant main effects for interparental conflict with those from the high conflict group demonstrating lower general and social self-esteem than those from the low conflict group, regardless of family structure. The ANOVA source tables are presented in Tables 1 and 2." (p134).

Table 1

Two-Way ANOVA Source Table for General Self-Esteem (RSES)

Source of Variation	Sum of Squares	*df*	Mean Square	*F*	Sig of *F**
IPC	2.759	1	2.759	11.298	.001
Family Structure	.10	1	.103	.423	.516
IPC X Family Structure	.247	1	.247	1.011	.316
Error	66.426	272	.244		
Total	69.432	275	.252		

*Significance of *F*

Table 2

Two-Way ANOVA Source Table for Social-Self Concept (SSCS)

Source of Variation	Sum of Squares	*df*	Mean Square	*F*	Sig of *F**
IPC	1.317	1	1.317	9.420	.002
Family Structure	.000	1	.000	.001	.974
IPC X Family Structure	.000	1	.000	1.001	.971
Error	38.024	272	.140		
Total	39.382	275	.143		

*Significance of *F*

[The above tables are located on pages 135 and 136 of the article].

Continued…

"The results of this study support the psychological-wholeness perspective [since it] was demonstrated that, regardless of family structure, interparental conflict has long-term effects on the general and social self-esteem of young adults. However, the findings do not lend credence to the physical wholeness-perspective as no significant long-term relationship was discovered between family-structure and the general or social self-esteem of young adults.

One should be careful concerning the clinical implications of this study. Since interpersonal conflict only explained 4.5% of the variance in self-esteem, it should not be perceived as the primary factor affecting the well-being of adult children of intact or divorced families. Similarly, one should not assume that divorce is an unimportant factor in the lives of the young adults who frequent our counseling and mental health centers or sit in our college classrooms." (p134).

Extracts from: Garber, R.J. 1992. 'Long-Term Effects of Divorce on the Self-Esteem of Young Adults', Journal of Divorce and Remarriage 17.1-2: 131-137.

(Note that the material inserted between square brackets within the above extracts has been added to the text for the purposes of this exam question).

***

(i) What are the strengths and limitations of the above analyses of variance (ANOVAs), including the way in which they are reported?

(ii) The article provides very little further information about the ANOVAs, including information about the variables/data to which they were applied. What additional material could have been included which would have helped you better assess the analyses' merits and weaknesses?

END