

The Polytechnic of North London
Faculty of Environmental and Social Studies

Post Qualifying Scheme

Level: Postgraduate (15 points at CNAA Master level)

Module Number: SR501

Module Title: Survey Analysis Workshop

Location: Policy Studies and Social Research

Module Convenor: John Hall (Director, Survey Research Unit)

Study Requirement:

6-9 hours per week of which 3 hours will involve timetabled classes (normally 1 hour instruction followed by 2 hour workshop/discussion). 3-6 hours should be used for private study and/or keyboard experience and follow-up exercises.

Module Objectives:

By the end of the module you will:

- a) acquire practical and intellectual skills in data management and statistical analysis of single variables (univariate), two variables (bivariate) and many variables (multivariate)
- b) be familiar with the language and logic of data analysis (with an emphasis on explanation as well as description) and the interface between theory and data
- c) be able critically to assess published reports which include analysis of survey and similar data
- d) become sufficiently confident and proficient to tackle your own research projects in college, on placement and in employment, or as a basis for more advanced methods
- e) understand how to code data from questionnaire surveys to a standard data layout and how to enter them into a file
- f) understand how to define data and associated dictionary information for entry into SPSS-X and save this in a system file for future use
- g) understand how to prepare and use supporting documentation
- h) acquire a working knowledge of the Vax control language, VMS, and the screen editor, EDT
- j) enjoy a distinct advantage in the employment market
- k) discover that survey analysis is fun and you **can** do it!

Module Assessment:

The course will be assessed by three components:

Component 1: Data Capture and Documentation (20%)

Component 2: Analysis and Report (60%)

Component 3: Descriptive and Inferential Statistics (20%)

The first assignment will be to select from a British Social Attitudes Survey a topic of interest to yourself, to select questions relevant to your topic, and to use SPSS to read the relevant survey data and construct a "system file" with missing value specifications, labelling, and a frequency count, together with appropriate user-documentation. (20%)

The second will be to conduct an analysis of your chosen topic and to write a short report on your findings. (60%)

The third will consist of a set of exercises involving data management and descriptive and inferential statistics, to be designed, conducted and interpreted within a limited time. (20%)

All work for assessment must be submitted (preferably typed) double spaced and single sided on A4 size paper including SPSS output which must be burst before stapling and clearly marked with your correct assessment number.

For components one and two, you should prepare an outline proposal identifying your research topic and listing the variables (and related questions/items) you propose to use and your initial ideas for the line of enquiry you intend to pursue. This should be submitted on the official proposal form not later than 4pm on Friday 13th March 1992.

Assessment date(s):

Component one must be submitted not later than 4pm on Friday 27th March 1992

Components two and three must be submitted not later than 4pm on Friday 19 June 1992

All three components must have been submitted before any marks can be considered by the Examination Board.

There is no provision for extensions. Work submitted late must be accompanied by a statement of the reason(s) for lateness and, if appropriate, copies of supporting evidence.

Study Programme:

This course is heavily skill-based, but with an emphasis throughout on logic and professional standards. Statistics as such are not taught, although the procedures for producing them will be used and their rationale and results explained (in non-mathematical language!)

Teaching programme

Block I From questionnaire to SPSS-X system file (Norušis 1990 Ch 1-6)

- 1 Data matrix. CASES, VARIABLES, VALUES. Coding of questionnaire data. Levels of measurement. The use of computers in survey research. Intro to Vax computer. Use of computer terminals and printer. Simple VMS commands. Special keys. Files on the Vax. Demonstration of SPSS-X. Creating and editing files with the screen editor EDT. Entering questionnaire responses into a data file.
- 2 Intro to SPSS-X. Basic structure of SPSS-X language; commands, sub-commands and specifications. Using SPSS-X to read an external data file. Records, fields, formats. Naming variables. Dictionary, active file. Displaying contents of dictionary and active files.
- 3 Extending a dictionary. Labelling variables and values. Missing values. Saving an external system file.

Block II One Variable (Norušis 1990 Ch 7-8,10)

- 4 Describing data. Univariate distributions. Graphical representations. Retrieving an external system file. Selecting variables for analysis. Frequencies for nominal and ordinal variables. General and integer mode; treatment of missing values; absolute, relative, adjusted frequencies. Barcharts. Utilities for printing
- 5 Frequencies for interval variables. Cumulative percentages. Univariate statistics. Measures of central tendency and dispersion. Histograms, percentiles. Condensed format for variables with many values
- 6 Data transformations. Changing the coding scheme. Derived variables. Selecting cases for analysis. Conditional frequency distributions.

Block III Two variables (and sometimes three) (Norušis 1990 Ch 9,11,13)

- 7 Joint frequency distributions for two variables. Contingency tables. Dependent and independent variables. Rules for percentaging. Specifying cell contents. Percentage differences (epsilon).
- 8 Introducing a third variable. Conditional contingency tables. Controlling for test variables. Elaboration.
- 9 Handling multiple response questions. Frequencies and contingency tables using multiple response.
- 10 More transformations. Creating simple scales. Comparing averages across different groups of cases.

Block IV Testing hypotheses (Norušis: 1990 Ch 14-16,18-23)

- 11 Testing the differences between means from samples. The t-test. One way analysis of variance.
- 12 Testing for statistical independence of variables in contingency tables; the chi-square test. Observed and expected frequencies; residuals; significance levels.
- 13 Plotting data on scattergrams. Correlation and regression.

Presentation of findings.

- 14 The final session is given over to brief presentations by individuals or groups of their experiences and findings, and to course evaluation.

Essential Reading:

Norušis M J **The SPSS Guide to Data Analysis for Release 4**
(ISBN 0-923967-08-7: SPSS Inc., 1990)

Further Reading:

Norušis M J, **SPSS Introductory Statistics Student Guide**
(ISBN 0-923967-02-8: SPSS Inc., 1990)

Norušis, M J **SPSS Base System User's Guide**
(ISBN 0-918469-63-5: SPSS Inc, 1990)

Technical Report on British Social Attitudes Survey
(Social and Community Planning Research, annually).

British Social Attitudes (Gower, annually)

Survey Analysis Workshop: statistical notes
(Survey Research Unit, PNL, 1988)

Learning Materials:

Facsimile questionnaires and other material from the British Social Attitudes survey.

There is some PNL Computer Services documentation on SPSS-X and the Vax and its operating system, but most of the course relies heavily on extensive documentation by John Hall.

As well as your own exercises on the 1989 British Social Attitudes data, Marija Norušis' book includes some on the 1984 General Social Survey (National Opinion Research Center, Chicago). This data set has been installed on the Vax as a SPSS system file **ASS:GSS84.SYS** and access is open to any Vax user.

Component 1: Data Capture and Documentation (20%)

In the following exercise, think in terms of **variables** for analysis, especially **dependent** and **independent** variables, bearing in mind that the second component will involve using the resultant system files for analysis and writing a report.

Choose a topic from the 1989 British Social Attitudes Survey. Select at least 10, but not more than 20 variables, including attitudes and beliefs, and items which might affect variation in your dependent variables (eg attitudes to private education or NHS facilities could well be affected by experience or usage of such provision). Appropriate demographic variables should be included, as should variables from both interviewer and self-completion sections of the questionnaire. You may interpret "variable" as sometimes comprising more than one data item (eg for attitude scores or for multiple response questions)

1. Using SPSS-X, create a system file containing only those cases necessary to your analysis, plus all your selected variables, together with missing value specifications and appropriate variable and value labels. Include a document.
2. Submit the final version of your SPSS-X command file, together with your user documentation (if any).
3. Display variable labels and document. Produce frequency counts in general mode for all variables in your file.

The data are on file **ASS:BSA89.DAT** and there are 23 records per case.

The coding for open-ended questions and for the letter-coded income questions is not given on the questionnaire. See Brook, Taylor and Prior, **Technical Report**, SCPR, 1990, in Library if you need these.

Some questions are capable of more than one answer (multiple response) and special facilities are available for analysing them. If you wish to use such questions, check first, as the coding schemes vary in complexity (eg 63b, 84f (open-ended) 27b, 33a-c, 41b-d, 67, 907b, 914 (precoded)) and you may need help.

There are no multiple response items in the self-completion questionnaire for Version A. Version B has them in qq 4, 7, 15.

In general, single column fields have 8 (DK) 9 (N/A) as missing, and two column fields 98 and 99. Some variables have values 0 or -1 which need to be treated as missing. Codes 7 and 97 tend to be used for "Other uncodable" and should be treated as missing. Separate documents are supplied giving details of income codes and of data for additional variables entered on record 23.

Printing of SPSS **listing** files is by default on A4 at Ladbroke House, but to print your SPSS **command** file on A4:

```
$ LPRINT _____ .SPS
```

Component 2: Analysis and Report (60%)

Write a report of not less than 2,000 and not more than 3,000 words (excluding figures and tables) to cover the following:

Introduction to the topic chosen and variables selected for your first component, including any preliminary hypotheses or ideas you had about what you expected to find or prove (or disprove) and referring to any relevant literature.

What analyses you performed on the data and why.

What your main findings were.

Methodological comments and insights.

Use the SPSS system file you generated for your first component, but amend any errors or omissions you may have made. Feel free to use any additional variables you think you need (e.g. for multiple response questions). Try to keep your final analysis simple by restricting yourself to a few key variables, if necessary by constructing scales or summary types.

There is no need to copy tables by hand into your report: just hand in your final selection as SPSS output, making sure that the tables or figures are clearly numbered and titled. You must also clearly indicate in the text which table or figure you are referring to (e.g. See Table 4 or Table 10 here) **Tables do not count towards the 1,500-2,000 words needed in the report.** Do not include more than ten tables.

Component three: Descriptive and Inferential Statistics (20%)

For this component you will have to design, execute and interpret statistical analyses using SPSS-X. The format will be that of an examination paper which you will be required to complete within a limited time. The paper will be distributed on **21 May 1992**.

SPECIMEN ONLY 1992 format, but using 1986 data instead of 1989

Component three: Descriptive and inferential statistics (20%)

You may use abbreviated forms of SPSS-X commands and subcommands. All answers to be on A4 paper, including SPSS-X output, burst, with banner pages attached. No answer to be longer than two A4 sides.

File **ASS:NOPROT.SYS** contains the following variables from the 1986 British Social Attitudes Survey:

SEX REGION PARTY EDQUAL V2018 V2019 V2020 V2021 V2023 AGE

File **ASS:XMAS.SYS** contains details of numbers of injury causing accidents in 41 police authorities in Dec 1986 **INJ86** and in Dec 1987 **INJ87**

Answer ALL questions

Section A (Technical)

Question A1

Using file **ASS:NOPROT.SYS** write a command file in SPSS to perform the following analysis. Construct a score **NOPROT** with a range of 0-20 from items v2018 to v2023 and recode it with four groups (0-3)(4-6)(7-9)(10-20) into **NOPROTGP**. Recode **AGE** into **AGEGROUP** (18-29, 30-44, 45-59, 60+) and **EDQUAL** into **EDGROUP** (GCE O-level and above, CSE2-5 and none) and leaving out foreign qualifications. Write appropriate variable and value labels and take account of missing values.

Produce the following output:

frequency counts (in general mode)

NOPROT with a histogram overlaid by a normal distribution; the mean, standard deviation and standard error; the lower and upper quartiles and the median.

NOPROTGP EDGROUP AGEGROUP

crosstabs (with row percent and chi-square)

Dependent variable:	NOPROTGP (column variable)
Independent variable:	SEX (row variable)
First order test variable:	AGEGROUP
Second order test variable:	EDGROUP

means (in crossbreak format)

Dependent variable:	NOPROT
Independent variable:	SEX
First order test variable:	AGEGROUP

Second order test variable: EDGROUP

t-test

Dependent variable: NOPROT
Independent variable: PARTY (Labour vs SDP/Lib)

oneway (with descriptive statistics and tukey range check)

Dependent variable: NOPROT
Independent variable: REGION

Question A2

Using file ASS:XMAS.SYS write a command file in SPSS to find the mean number of injury causing accidents each year and plot the 1987 figures against the 1986 figures in regression format.

Section B (Interpretation)

Question B1

Write a short account of the effects of sex, age and educational level on anti-protest attitudes using **either** the crosstabs output **or** the means/crossbreak output.

Construct an appropriate summary table using **either** percent "Definitely not allow" **or** mean anti-protest score.

Question B2

Choose **TWO** of the following inferential statistics topics and, from the SPSS output for section A, explain what the test is, what the technical and statistical terms are and why the test was used for these data.

What do the results tell you?

chi-square (but **not** Likelihood ratio or Mantel-Haenszel)

t-test

oneway analysis of variance

linear regression

(Draw an approximate regression line on the plot and comment generally on your results. What would be your best estimate of the number of injury causing accidents in 1987 for an authority which had 300 in 1986?)

Guidelines for preparation of assessed coursework

Component 1: Data capture and documentation (20%)

This component should consist of **two** SPSS command files; the first will be your initial file to include:

data list
missing values
variable labels
value labels
save

The second SPSS command file to include all the commands and specifications (document, recode, compute, select if, if, save) used for documentation, data transformation or construction of derived variables. These can be produced using:

```
$ LPRINT _____ .sps
```

Finally, you should submit one SPSS listing file with:

```
set length 72 width 80 print off.  
title .....(to include your student number)  
get file .....  
display labels.  
display document.  
frequencies ...(in general mode for all variables except those with more than 20 values, e.g. age)
```

Remember, one criterion for this component is to enable someone else to understand what you are doing and to carry on where you leave off, or work on your material when you are not there. You may submit notes in addition to your SPSS output if you wish, but these should not cover more than 2 sides of A4.

Component 2: Analysis and Report (60%)

This should be submitted double-spaced, single-sided (preferably typed) on A4 paper. SPSS output containing tables and figures for your report can be appended to your text. Use:

```
set length 72 width 80 print off.  
title .....(to include your student number)  
subtitle .....(Table/Figure number)
```

Component 3: Descriptive and Inferential Statistics (20%)

To be run as one SPSS job with:

```
set len 72 wid 80  
title (to include your candidate number)  
subtitle (as appropriate)
```

