AUSTRALIAN DATA SETS AVAILABLE FROM ACSPRI

The following Australian data sets have become available since the last edition of the Newsletter

*   *   *

N.S.W. Anti-Discrimination Board

A Survey of New South Wales Government Employees, 1978

Self-completion questionnaires were sent to a sample of women office workers employed in 42 Departments and 10 Statutory Authorities. Men in similar occupations were sampled as a control group. The sample was stratified by sex, occupation and departmental/ organisational size and the sample sizes chosen to equalise the error rates for the major samples. An overall response rate of 65% was achieved, giving 2199 respondents, of whom 1677 are women.

The questionnaire was designed to provide the Board with information on the experiences, attitudes and perceptions of women in the Administration. The data is in card-image format with six cards/case Copies of the questionnaire and codebook, which contains a description of the survey methodology, are available from local ACSPRI representatives.

*   *   *

Troy, Patrick N, Urban Research Unit, Research School of Social Sciences, Australian National University.

Environmental Quality in four Melbourne Suburbs

Personal interviews were carried out ·in 1118 households spread over the four Melbourne suburbs of Clayton, North Melbourne, Strathmore and St Kilda between February and October 1970. Interviewers were conducted by trained social workers who were asked to try to interview an equal number of heads and
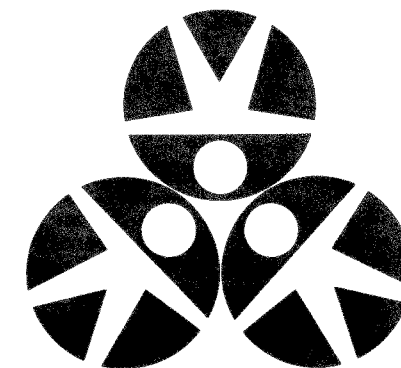
spouses or males and females in families and households with adults of both sexes present. One interview was conducted in each responding household. Interviews were conducted mainly in the evening during the week and during the day at weekends. An 'out' was generally recorded only after a minimum of six calls. Simple random samples were selected in each area using electricity meter readers' lists to obtain sample sizes of 350 dwellings. Response rates were 86%, 83%, 85% and 65% in Clayton, North Melbourne, Strathmore and St Kilda respectively. (Full details are provided in the URU Monograph 'Environmental Quality in Four Melbourne Suburbs', Troy, P. N., 1972).

The data for each suburb is organised into two files, a Household file and a Person file. The Person file contains one record (one card-image) for each person in responding dwellings and includes demographic and household description variables. The Household files have 4 card-images per case and contain attitudes to the dwelling, the convenience of the location, the environment and the people in the area as well as background data for all members of the household.

*   *   *

ACSPRI newsletter

Australian Consortium for Social and Political Research Incorporated

CONTENTS

EXTRACTS FROM THE ANNUAL REPORT

ACSPRI is now entering its fifth year of operation and continues to grow with Melbourne State College joining the Consortium on 1 July 1980 bringing the membership to nineteen institutions, an increase of three over last year. The official representative for the College is Dr Peter Rogers, Computer Facility, Melbourne State College.

Five ACSPRI nominees attended the 1980 ICPSR Training Program in Ann

Arbor. A total travel subsidy of US$1200 was provided by ICPSR.

A total of 81 data sets have been requested during the year. 49 of the data sets were obtained from ICPSR, the remainder being held locally (through earlier requests). 5 of the requests were for Australian data archived by ACSPRI.

A request from the SSRC Survey Archive in Essex for six 'Age Poll' data sets has been processed and the ANU has received materials from Essex. Two data sets have been ordered from the Data and Program Library Service at the University of Wisconsin, Madison.

CONTRIBUTIONS TO THE NEWSLETTER

Contributions from readers are encouraged and reports on the use of data supplied through ACSPRI would be of particular interest. Contributions and enquiries about ACSPRI should be addressed to:

Roger Jones
The ANU Survey Research Centre,
Australian National University,
P.O. Box 4, Canberra 2600
Telephone (062) 49 4400

AN AUSTRALIAN GENERAL SOCIAL SURVEY

Many people have, I am sure, considered the possible benefits to the Australian social science community of a General Social Survey similar to that conducted by the National Opinion Research Center in the United States. The purpose of that survey is 'to generate data trends in social characteristics and opinions' and 'to make fresh, interesting, high-quality data available to social scientists and students' who would otherwise not have access to such data. Following recent discussions within ANU, Dr. Jonathan Kelley, Senior Research Fellow in the Department of Sociology, RSSS, has set down a tentative plan for such a project. Your reactions, comments and any suggestions on this proposal are invited, and anyone wishing to be involved in choosing suitable items for the questionnaire should write to Roger Jones at the ANU Survey Research Centre.

* * *

Topics to be Covered

The idea would be to divide the survey into two parts, the first and largest to cover all the standard demographic and background items together with a wide range of social and political items of interest to a broad academic audience. These "core" items would be repeated in subsequent editions of the survey and so provide measures of change over time. The second and smaller part would be items of topical interest or special interest to the principal investigators taking responsibility for a particular edition of the survey and would presumably not be repeated in subsequent years.

Core Items   In addition to standard demographic and background items (age, ethnicity, occupation, education, income, region, number of children, family background, political party, and the like), which have to be included, I would suggest that we compile a large and diverse collection of items that

might be of interest to researchers in one field or another. Given what we know of the normal range of users for social surveys, I presume that the bulk would tap political and social views, but there is no reason to restrict ourselves. I would, however, strongly suggest that we not write new items — a notoriously tricky business — but restrict ourselves to items that have already been used successfully in national surveys here and abroad. As you know, this is not much of a restriction since a very wide variety of items have been used in various surveys in Britain, the EEC, and the U.S. as well as in Australia; choosing from existing items also has the great advantage that we can compare Australia with other countries. Then we can send xerox copies of all the questions to a wide range of potential users at universities throughout Australia and ask them to indicate which they might be interested in. This would allow us to choose a set of items that are of use to a broad group of researchers and, incidentally, generate some interest and support for the project. Jim Davis followed this procedure when he designed the National Opinion Research Center's ongoing survey in the U.S. and it seems to have worked to the satisfaction of all. We could then set up a working committee of three or four knowledgeable people from different fields to help vet the final choice of items.

Topical Items   A relatively small part of the survey might be devoted to items of interest to the principal investigators. For a project of this size, we might want to think of three principal investigators from different universities and several different disciplines. That would spread the intellectual labour and risk, increase the number of people actively analyzing the data, and improve our chances of funding. Assuming that we plan for a questionnaire something like an hour long, a usual figure for this kind of survey, and that the core items take up something like three-quarters of the time, this would leave roughly five

minutes for each principal investigator's items. That gives enough time for a fair although not enormous number of questions (e.g. a lengthy attitude scale) and, given that a wide variety of information would already be available in the core section of the questionnaire, should suffice.

Data Collection and Coding

It is, I suspect, most practical to have the survey done commercially. For a once every three years project, setting up our own organization does not seem wise or practical, especially not when satisfactory commercial organizations already exist.

We probably would, however, want to do the coding here (or at least the non-routine parts of it). Occupation and industry are particularly difficult and I doubt that commercial organizations are set up to do it with the care needed. Presumably we could get the key punching done here too.

Use of the Data

The basic idea is to make all the data widely and immediately available to researchers throughout Australia. Judging from experience in the U.S. I would expect it to be widely used both in research and in teaching. The National Opinion Research Center's General Social Survey has become the most widely used data set in sociology while the analogous political surveys out of Michigan's Survey Research Center are the most widely used source in political science. Both surveys have been enormous successes and hardly an issue of the top journals is published without articles using them. They are also widely used in teaching; students in methods and data analysis classes find it enormously more interesting to work with high quality data on a representative national sample than on more limited data.

The second main use of the data would be by the principal investigators

working on their special topics. I presume that we can find three reasonable investigators who could be counted on to actually get the work done (if nothing else, it is important for the granting agency to have some assurance that something will come of their money). I would be happy, as you might have guessed, to be one of them.

There is a third possibility, although I am somewhat unsure about it. Insofar as we use items from surveys in other countries, it might be instructive to compare Australia with the other countries. There are, for example, interesting questions on confidence in government and other institutions in the Michigan and NORC surveys and interesting items on, for example, women's liberation in the Euro-Barometer series for EEC countries. Since these surveys are all publicly available, it would not be hard to do the comparisons and they might tell us a good deal about Australia. Similar comparisons have proved useful in the U.S. and, indeed, have been done with government support (Social Indicators 1977 put out by the Department of Commerce is the best example). We might think of doing something on that line in Australia. It would be of interest beyond academic circles and offer the granting agency another assured payoff for their money.
Jonathan Kelley
Senior Research Fellow.

* * *

PRODUCING A CODEBOOK WITH 'MANUAL'

No analysis of survey data, and particularly secondary analysis, is possible without the existence of adequate documentation. The most important piece of documentation is the codebook which relates the location of information on the file to the substance of the study. A good codebook contains not only the location of each variable on the tape but a full description of the study itself, including a title, the name of the principal investigator and his affiliation, sponsorship, the fieldwork

agency, dates and method of data collection, sampling information and response rates and a list of publications. In addition the exact wording used in the questionnaire and the way in which the data is coded should be explained.

The Statistical Package for the Social Sciences (SPSS) provides a means of producing a codebook for a survey data file through the DOCUMENT control card, which allows descriptive information to be permanently retained on a SPSS system file. The information and the VALUE LABELS, VARIABLE LABELS and MISSING VALUES can then be recovered, either in printed form with the LIST FILEINFO procedure or in the form of standard SPSS data-definition cards for regeneration of the file at another installation through the WRITE FILEINFO facility.

A shortcoming of the SPSS facility is the limitation placed on the length of variable and value labels – 40 characters and 20 characters respectively. While these are enough to give the user familiar with the data an indication of the variable under consideration, a full description of the question and response categories is often not possible within these limitations. The researcher then has to refer elsewhere to obtain the required details.

The program MANUAL, developed by Professor D Anderson of the School of Education, Flinders University is designed to overcome these shortcomings. With the two programs MANUAL and CODE, the researcher can produce an enhanced version of the SPSS codebook produced by

LIST FILEINFO and create the SPSS data definition runstream necessary to produce a fully labeled system file.

Input to the program MANUAL is a file of study description records. These records may be of seven types, with each type indicated by the appropriate letter in column one.

T   Title (max 50 characters)
H   Headings (max 50 characters)
S   Sub-headings (max 50 chars)
C   Comments (max 50 characters)
V   Variable descriptions including variable name and label, card no., format and missing values
D   Value labels
X   Column skips

The 'Title' and the current 'Heading' appear at the top of each page of the codebook. 'Headings', 'Sub-headings' and 'Comments' can be used anywhere in the file. Variable names, variable labels and value labels are limited to 8, 40 and 20 characters respectively as they are used to label the SPSS system file. In the example given below, the variable number 'VAR 17', location code 'LOCATION: 26' and column numbers 'COLUMN 26-28' are produced automatically. The question is included by placing a comment card (C) between the variable (V) and value label (D) cards.

Thus all the information usually entered with DOCUMENT can be included on comment cards, with additional headings, sub-headings and comments included where necessary to provide a full description of the data file.

The codebook created by the program MANUAL serves as input to the program CODE which creates the control cards necessary for recreating an SPSS system file. The control cards created from the run are

VARIABLE LIST
INPUT MEDIUM
INPUT FORMAT
N OF CASES   UNKNOWN
VAR LABELS
VALUE LABELS
MISSING VALUES
READ INPUT DATA
FINISH

At present, versions of the programs MANUAL and CODE are only available for DEC-10 and Univac-1100 computer systems, but could readily be modified for use on other systems by a competent programmer (most of the source code is written in FORTRAN). Anyone interested in obtaining a copy of the programs should write to the ANU Survey Research Centre.

*   *   *

AGREEMENTS WITH OVERSEAS ARCHIVES

ACSPRI has recently negotiated 'one-for-one' exchange agreements with the Roper Center Data Archive in the USA and the SSRC Survey Archive in the UK. These agreements expand considerably the amount of data accessible to researchers in Australia. However, in order for these agreements to operate without charge, ACSPRI must have data sets to exchange. Should a substantial imbalance occur, as ACSPRI requests more data than it can offer, then a charge will be imposed to cover the cost to the overseas archive of processing the request.

Roper Center Data Archive

The Roper Center is the world's largest data archive with holdings of about 9000 data sets from many areas of the world, with an emphasis on public opinion poll data. For many years most of the world's Gallup organisations,

including Australia's Morgan Gallup Polls, have deposited their regular opinion poll data with the Roper Center. While the Morgan Gallup Poll no longer deposit their data, the Center does hold data for surveys conducted between 1943 and 1965, a period during which the Morgan Gallup Poll was the only regular opinion poll being conducted in Australia. To date, no data has been acquired from the Roper Center collection, although evidence of the value of this data for secondary analysis is clearly provided by Professor David Kemp's book 'Society and Electoral Behaviour in Australia'.

SSRC Survey Archive

The SSRC Survey Archive, situated at the University of Essex, is the largest national repository of machine-readable social science data in Britain. It was set up in 1967 with a brief to collect and preserve machine-readable data relating to social and economic affairs from academic, commercial and governmental sources, and to make these data available for secondary analysis. The archive currently holds over 1,500 data sets including British Census data, data from major regular Government surveys such as the two-yearly Family Expenditure Survey and the annual General Household Survey, regular opinion poll data from National Opinion Polls (1963 to date), Social Surveys (Gallup Poll) Ltd (1970 to date) and Opinion Research Centre (1971 to date), and many of the major academic surveys conducted in Britain.

The Archive also maintains a detailed and up-to-date record of social surveys conducted in Britain and information about these surveys, whether or not lodged in the Survey Archive, can generally be made available to enquirers.

A full listing of the Archive's holdings is provided in the SSRC Survey Archive Data Catalogue (see your local representative or Library about obtaining a copy at £15.50). Entries

VAR 17	SALEPR	ESTIMATED SALE PRICE – $000–	/
	LOCATION: 26	FORMAT: F3.0 CODE FOR MISSING 998, 999	
	CARD NO.: 1	COLUMN: 26-28	

Q13   If this dwelling were up for sale, what do you think it would sell for?

VALUE LABELS   (998)   DON'T KNOW
              (999)   NOT APPLICABLE

are divided into over 100 subject categories with extensive cross-referencing. An Abbreviated Guide precedes the main text, containing a shorter description of each study to allow users to identify material of interest as rapidly as possible. Five indices are provided as a further aid – by title, by depositor, by survey number, by geographical region and by population. More recent acquisitions are briefly described in the triannual SSRC Survey Archive Bulletin which is circulated to all local ACSPRI representatives.

During a recent visit, Dr. Terry Beed (Sydney) raised the question of on-line satellite contact between the SSRC Archive and Australia, and this is now being investigated by Eric Roughley, Systems Manager of the Archive.

Note: Data sets requested from both the Roper Center and the SSRC Archive will be supplied to ACSPRI free of charge. Requesters will however have.to pay the normal ACSPRI charges, at present $75 per data tape, to cover the costs of the tape and of processing the request at the ANU Survey Research Centre, through which all requests must be made. Copies of codebooks and questionnaires are supplied at the cost of duplication, plus postage.

*   *   *

OTHER FOREIGN DATA SOURCES

DAEDAC

DAEDAC, acronym for Drug Abuse Epidemiology Data Center, is a Federally funded, non-profit research archive operated by the Institute for Behavioural Research at Texas Christian University. DAEDAC maintains the largest existing file of coded raw data from original drug surveys. Computer tapes of the original data of selected surveys, evaluation studies and other large scale drug abuse data sets that have potential for secondary analysis are available to researchers. The

Center's current Original Data File Directory and User's Guide provides information on the 148 studies available, and includes brief descriptions of the major variables included, the population covered and the sample sizes.

The data files are edited by the DAEDAC staff for completeness, consistency and coding of missing data. In addition, complete documentation is provided in the form of a data dictionary which includes the format of the data, the survey instrument or other source document, descriptive information concerning the study and a listing of all publications generated from it.

An information retrieval system has been developed to enable users to obtain data from a variety of studies as well as from a specified data set, all on the same variable or concept. Retrieval is possible through delimiters or descriptors which give such information as the type of population under study, age range, sex, sample size, types of drugs, racial/ethnic groups and dates of data collection, and/or through the use of locator codes consisting of drug codes, use/effect codes and attribute codes with extensive subcategories.

Researchers requiring further information on DAEDAC holdings should firstly consult the Original Data File Directory and User's Guide which is available from their local ACSPRI representative. If a researcher finds that potentially suitable data are available, a data dictionary may be requested before deciding to order a data tape. The data dictionaries are available free of charge, and data files can be ordered directly from DAEDAC at a cost of about $50.

World Fertility Survey

Dr Terry Beed (Sydney), Vice-Chairman of ACSPRI, recently attended the World Fertility Survey (WFS) Conference in London and reports that the WFS headquarters is now in a position to distribute data sets from participating

countries. Forty-four countries are participating in the WFS programme and at the moment about 12 of these have cleared their data holdings for distribution to research workers. Of particular interest are some special programmes which have been developed by WFS. These include a survey editing system known as CONCOR developed by CELADE, the United Nations Demographic Centre for Latin America, to meet the needs of national institutions processing censuses and sample surveys. A copy of the magnetic tape of CONCOR can be obtained at nominal cost by writing to CELADE, Casilla 91 Santiago, Chile. Another technical publication recently released by WFS is CLUSTERS, a package programme for computation of sampling errors for clustered samples. For information on the WFS programme and survey data sets contact should be made with the Information Office, ISI-WFS, 35-37 Grosvenor Gardens, London, SW1W OBS, U.K.

*   *   *

DIGITISED CENSUS BOUNDARIES

The Australian Division of National Mapping (NATMAP) has generated a set of files in digital form containing many of the boundaries used for the 1976 Census of Population and Housing. The boundaries have been digitised from the maps used by the ABS to assign areas to the census collectors.

The digitised data includes the boundaries of CCDs in and around the major cities and also a number of smaller ones (generally the cities of 25,000 or more population), the Northern Territory and the ACT. For the remainder of the continent the boundaries digitised were those of LGAs or equivalent areas used for statistical purposes. NATMAP has also computed the centroid (mean latitude and longitude) of each CCD or LGA from the boundary data and created a file of centroids for each State.

Boundary data are available in segment or polygon form, as latitudes and longitudes or, in some cases, Universal Transverse Mercator coordinates.

Copies of boundary and centroid files on magnetic tape are available for sale, but certain conditions must be agreed to by the purchaser before the data are released. All requests for the data must be in writing and should include a brief description of the use to which the data will be put. Requests should be sent to the Director, Division of National Mapping, PO Box 548, Queanbeyan, NSW 2620.

NURSING PERSONNEL SURVEY

The Commonwealth Department of Health conducted a full census of nursing personnel in Australia during 1978, sending out almost 190,000 self-completion questionnaires to Professional Nurses, Mothercraft Nurses, Nursing Aides and Nursing Students requesting information on age, sex, marital status, education, nursing qualifications, details of current employment, employment status over the last 5 years, reasons for leaving/returning to nursing. Responses were received from about 136,000 nursing personnel.

The Committee responsible for the study recognises the potential for further research on this data set, and invites applications for access to the data in the form of research proposals. Subject to the quality of applications, a limited number of grants may be provided to support the research.

Enquiries about the data and applications for access should be directed to:
   The Assistant Director-General,
   Nursing Branch,
   Commonwealth Department of Health,
   P.O. Box 100, Woden, ACT. 2606.
The 'Report of the Committee on Nursing Personnel Survey' is available from the AGPS.