

ICPSR BULLETIN

Volume XXXI, No. 1

Fall 2010



Introducing Survey Documentation and Analysis Version 3.5

By JoAnne McFarland
Director, SAMHDA

ICPSR is pleased to announce the release of version 3.5 of the Survey Documentation and Analysis software. SDA 3.5 provides online data analysis of datasets with complex sample designs with the same level of statistical rigor as Stata, SAS, and SUDAAN.

The new version is a collaboration between the SDA creator, the University of California, Berkeley, and SAMHDA (funded by the federal Center for Behavioral Health

Statistics and Quality, formerly the Office of Applied Studies, Substance Abuse and Mental Health Services Administration [SAMHSA]), with contributions from the Survey Research Center at the University of Michigan's Institute for Social Research.

New features of SDA 3.5 include:

- Corrections to standard errors (SE) in the regression programs (multiple, logit and probit) for surveys based on complex sample designs. Design-based corrections to SE in SDA were required in order to obtain improved significance testing and valid statistical inference. For a discussion of complex samples, see Heeringa, West, and Berglund, (2010). *Applied Survey Data Analysis*. Boca Raton, FL: Chapman and Hall, <http://www.isr.umich.edu/src/smp/asda>.
- Precision-based (PB) specifications for suppression of cell percentage in the crosstabulation program and the cell mean in the means



JoAnne McFarland is director of the Substance Abuse and Mental Health Data Archive, which was recently renewed for its

fourth five-year funding cycle. She has been with ICPSR since 1997, first as a senior research associate and, since 2000, as an archivist. McFarland's research specialties include disclosure risk analysis, and she is leading ICPSR's development of a new Virtual Data Enclave that will allow remote access to restricted-use data. She has a Master of social work degree from the University of Michigan, and graduated magna cum laude from Western Michigan University with a major in psychology.

program. This tool is designed to alert researchers to estimates that have low precision (e.g., the analysis is based on too few cases, with a large standard error).

See SDA, p. 2

In this issue

Introducing SDA ver. 3.5	1
ICPSR offers guidance on data management plans	3
News and Notes	4
Data Fair attracts nearly 300 participants	5

SDA (cont. from page 2)

- Improvements that move SDA to a Web services model and allow SDA output for crosstabs and frequencies to be output in XML. This function will improve ICPSR's ability to display data in various ways and help eliminate the need to recode variables in order to achieve a desired display (e.g., to change a scale from (1) A-E to 1-5 or (2) from High-Low to Low-High for Web display).

The new features in version 3.5 continue the development of SDA. Previous improvements have included the ability to set disclosure protections for restricted-use data in SDA version 3.3, known as Secure SDA or SSDA, and corrected standard errors in better significance testing for crosstabs and

means in version 3.4 (which ICPSR is currently running).

The implications of the new capabilities of SDA version 3.5 are numerous:

- Users have expanded ability to analyze data online
- ICPSR is better able to meet funders' requirements to make data — especially data with complex sample designs — available to a broad user base
- Instructors can more easily use ICPSR data and SDA for teaching
- The ability to access restricted-use data in combination with corrected SEs and disclosure protection settings gives analysts all the tools

needed to conduct publishable research solely using SDA

There are also several improvements to the user interface with the new version of SDA. For example, XML will provide ICPSR the ability to customize the SDA crosstabulation interface (more than 90 percent of SDA analyses are crosstabs). Plans for the crosstab interface include separating the current screen selections so that users first select variables and initial stats for the table, next run the table, and then make other selections such as graphical display options (pie chart vs. bar chart, etc.).

Also, returning to previous screens, changing variables, and changing options will be easier and faster. ■

Chief contributors to the development of SDA versions 3.4 and 3.5 were:

Standard errors and significance tests:

- Survey Research Center/Institute for Social Research: Steve Heeringa, Pat Berglund, Brady West
- ICPSR/ISR: Mohammad Rahman, Sarah Rush
- SDA/Berkeley: Tom Piazza, Yuteh Cheng, and Charlie Thomas
- SAMHSA: Jonaki Bose

XML:

- ICPSR/ISR: Matthew Richardson, Peggy Overcashier
- SDA/Berkeley: Charlie Thomas, Tom Piazza

Precision-based suppression:

- Research Triangle Institute (RTI)
- Center for Behavior Health Statistics and Quality (CBHSQ, formerly the Office of Applied Studies), SAMHSA

CBHSQ, SAMHSA supported the development of SDA 3.4 and 3.5 through the SAMHDA contract.

Chief contributors to the development of SDA v.3.3 were:

- ICPSR/ISR: JoAnne McFarland and Chris Maxwell (now of Michigan State University)
- SDA/Berkeley: Tom Piazza

SDA v.3.3 was supported by the Office of Juvenile Justice and Delinquency Prevention (OJJDP) through its National Archive of Criminal Justice Data (NACJD) Inter-agency agreement.

ICPSR Offers Guidance on Data Management Plans

By Dan Meisler

ICPSR Editor

As one of the leaders in data archiving for nearly 50 years, ICPSR is in a unique position to assist researchers with the data management plans increasingly being required by federal funding agencies such as the National Institutes of Health and the National Science Foundation.

Therefore, we are pleased to offer a [new set of guidelines](#) for preparing data management plans. We hope these resources can assist researchers in all disciplines meet these new requirements, and better understand how to preserve and disseminate their research data.

Our guidelines provide a detailed list of the elements of a data management plan, developed through a gap analysis of existing recommendations for such plans. We also list existing examples of data management plans, as well as other resources available on the Web.

Our page includes links to the data management and sharing policies of several federal agencies.

Depositing data with ICPSR can be an integral part of a data management plan. In fact, the NSF's Directorate for Social, Behavioral & Economic Sciences [recently named ICPSR](#) as an archive that would help grant applicants meet its requirement for data storage. Our data management plan guidelines also provide information on how to deposit data with ICPSR.

The NSF plans to have its requirements for data management plans in place in

[January 18, 2011](#), and is implementing additional guidance for engineering funding proposals.

As stated in NSF's [guide](#) to grant applications, data management plans are to be no longer than two pages, and may include:

- The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project
- The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies)
- Policies for access and sharing including provisions for

appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements

- Policies and provisions for re-use, re-distribution, and the production of derivatives
- Plans for archiving data, samples, and other research products, and for preservation of access to them

And, according to the agency's ["Award and Administration Guide,"](#) grant recipients are "expected to encourage and facilitate" the sharing of primary data with other researchers.

Our guidelines can help researchers reach that goal. We plan to update them regularly, and welcome any suggestions via email at web-support@icpsr.umich.edu. ■

MyData Login/Create Account | Search for Data | Search Web site

Home | Find & Analyze Data | About ICPSR | Deposit Data & Findings | Digital Curation | Membership | Partners & Projects | Teaching & Learning | Help

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Guidelines for Effective Data Management Plans

Many federal funding agencies, including NIH and most recently NSF, are requiring that grant applications contain data management plans for projects involving data collection. To support researchers in meeting this requirement, ICPSR is providing guidance on creating such plans.

Elements of a Data Management Plan

This [list of elements](#) is informed by a gap analysis that ICPSR conducted of existing recommendations for data management plans and other forms of guidance made available for researchers generating data. The result of the gap analysis was a [comparison of existing forms of guidance](#). Elements that are highly recommended for inclusion in effective data management plans are noted.

Data Management Plan Resources and Examples

These sites provide guidance and recommendations on creating data management plans.

[List of resources](#)

Depositing Data with ICPSR for Long-term Data Management

ICPSR has been providing access to data since 1962 and has a proven track record for managing and preserving data. Depositing your data with ICPSR is a good way to ensure that your data will be well-managed and accessible in the future.

Planning Ahead

Good data management starts long before data arrive at ICPSR, however. Ideally, you should plan for eventual archiving and dissemination of your project's data before the data even come into existence. The [ICPSR Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle](#) provides a wealth of information on how to plan ahead for data sharing and how to manage data across all stages of the life cycle.

ICPSR Services

ICPSR provides a variety of services to safeguard data for the future and to add value to data resources deposited in the ICPSR Archive. ICPSR's [data curation services](#) include:

- Review of data for [confidentiality issues](#)
- Dissemination of [restricted-use data](#) when appropriate
- Generation of multiple data formats for dissemination and preservation
- Creation of a searchable [metadata record](#) with a bibliographic citation and a [Digital Object Identifier \(DOI\)](#)
- [Protection of documentation consistent with the NIH specification](#)

Federal Agency Policies on Data Management and Sharing

- Centers for Disease Control and Prevention (CDC) - [Policy on Releasing and Sharing Data](#)
- Council on Governmental Relations - [Access to and Retention of Research Data: Rights and Responsibilities, March 2006](#)
- National Institute of Justice - [Data Resources Program](#)
- National Institutes of Health - [Data Sharing Policy](#)
- National Science Foundation - [Proposal and Award Policies and Procedures Guide](#)
- National Science Foundation, Social, Behavioral and Economic Sciences (SBE) - [Data Archiving Policy](#)
- National Science Foundation (NSF) - [Dissemination and](#)

News and Notes ...

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

George Alter, Acting Director
Mary Vardigan, Assistant Director

COUNCIL MEMBERS

Ann Wolpert, Chair
Massachusetts Institute of Technology

Francine Berman
Rensselaer Polytechnic Institute

G. Sayeed Choudhury
Johns Hopkins University

Paul N. Courant
University of Michigan

Catherine A. Fitch
University of Minnesota

Thomas LaVeist
Johns Hopkins University

Jeffrey Moon
Queen's University

Gregory N. Price
Morehouse College

Rogelio Saenz
Texas A&M University

Barbara Schneider
Michigan State University

Lori M. Weber
California State University, Chico

Christopher Zorn
Pennsylvania State University

Aletha C. Huston, Past Chair
University of Texas, Austin

BULLETIN STAFF

Editor: Dan Meisler
dmeisler@umich.edu
(734) 615-7904

ICPSR is a center within the Institute
for Social Research at the University of
Michigan.

GESIS Spring Seminar Approaching

GESIS - Leibniz Institute for the
Social Sciences is offering a [spring
seminar](#) in March 2011 on "Latent
Structures in Multidimensional
Tables — Models and
Visualizations" in Cologne,
Germany.

The seminar is a three-week course
for social scientists interested
in advanced techniques of data
analysis, comprised of lectures,
exercises, and hands-on work. The
course is split into three one-week
modules — a detailed schedule will
be posted on the seminar [Web site](#).
The dates are March 14 through
April 1.

The scheduled lecturers (all of
whom will lecture in English) are:

- Jacques Hagenaars (emeritus)
and Ruud Luijkx, Tilburg
University, The Netherlands,
on Categorical Latent Variable
Models, and Extended Latent
Class Analysis, March 14-18
- Jörg Blasius, University of Bonn,
Germany, on Correspondence
Analysis, Related Methods, and
the Quality of Data, March 21-
25
- Michael Greenacre, University
Pompeu Fabra, Barcelona
Graduate School of Economics,
Spain, and Oleg Nenadic,
University of Göttingen,
Germany, on Biplots in Practice.

NAHDAP Seeking RFPs for Acquisition Subcontracts

The [National Addiction & HIV
Archive Program](#) is accepting
requests for proposals for
subcontracts for acquiring datasets
on HIV/AIDS and drug abuse.

The archive plans to award up
to five contracts for a maximum
of \$50,000 each in direct costs
and limited indirect costs. The
contracts will run from Feb. 1,
2011, through Dec. 15, 2011.

Subcontractors must be willing
and able to deposit datasets on
HIV/AIDS or drug addiction
resulting from research funded by
the National Institute on Drug
Abuse. Details are available in the

[subcontract competition document](#)
(PDF).

Applications are due by Dec. 10,
2010.

NAHDAP is funded through a
contract with NIDA with the
goal of acquiring, preserving, and
disseminating research relevant to
drug addiction and HIV research.

James McNally, director of ICPSR's
National Archive of Computerized
Data on Aging (NACDA),
recently became director of
NAHDAP as well. His work on
HIV dates back to the 1980s, and
he is currently tracking attitudes
toward HIV in a sample population
30 years later, as the group enters
its 50s and 60s.



Participation High at ICPSR's Virtual Data Fair

By Linda Detterman

Director, Marketing & Membership

Nearly 300 people participated in ICPSR's Virtual Data Fair from Nov. 8-11.

Topics covered in the webinars included data management plans, social sciences in North America, international data, and using social science data in teaching.

Slides and videos of all the presentations are available on the [Data Fair Web site](#).

The fair consisted of 14 sessions over four days, and 281 people took part. Each participant viewed an average of 2.4 sessions.

The presentation on data management plans by Katherine McNeill of MIT Libraries was the best attended session, with 134.

A presentation on "America's Most Wanted: Top U.S. Government Data Resources" from Lynda Kellam of the University of North Carolina drew the second-highest number of attendees with 92.

Of the 136 institutions represented by attendees, 124 were ICPSR member institutions.

Seventy-one of the session attendees were ICPSR Official Representatives or Designated Representatives.

The attendance figures compare favorably to the 2009 OR meeting, which consisted of 24 sessions over five days — 227 attendees participated in that event.

Several ICPSR researchers and staff members deserve thanks for their

hard work in putting the Data Fair together:

- Arun Mathur, Membership & User Support
- Mary Vardigan, Assistant Director
- Peter Granda, Director, General Archive
- Lynette Hoelter, Director, Instructional Resources and Development

A special thanks to all presenters, as well as everyone who attended.

We are also asking for feedback — please [fill out a survey](#) on your experiences at the Data Fair, or [email us](#) with any comments or suggestions. See you next time! ■

ICPSR Publishes Access Policy Framework

ICPSR has published an [Access Policy Framework](#) focusing on ICPSR's mission to provide access to data for social science research and instruction. The policy describes ICPSR's mandate in this area, the objectives and scope of the endeavor, the guiding principles, and related risks and challenges. The appendices provide additional information on access rules, roles and responsibilities, and related links.

The policy was developed by a committee chaired by ICPSR Assistant Director Mary Vardigan, with other members coming from various archives, the Computing and Network Services division, and the Web development team. The policy is meant to coordinate with ICPSR's policies on digital preservation, data access, and deposit, and is part of the consortium's response to the Trustworthy Repositories Audit and Certification (TRAC) guidelines.

The objectives of ICPSR's Access Policy Framework are to support the community of data users, ensure uninterrupted access to data, protect the privacy of survey participants, and support the goals of our sponsoring agencies.

In 2007, ICPSR published its Digital Preservation Policy Framework, which makes explicit ICPSR's commitment to preserving the digital assets in its collections through the development and evolution of a comprehensive digital preservation program.