

Introduction to Survey Data Analysis: Bibliography

Basic Data Analysis Techniques

- Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences*. Upper Saddle River, NJ: Prentice-Hall.
- Moser, C. A. & Kalton, G. (1972). *Survey methods in social investigation* (2nd ed.). New York: Basic Books. (see Chapter 17).
- Rosenberg, M. (1968). *The logic of survey analysis*. New York: Basic Books.
- Weisberg, H. F., Krosnick, J. A. & Bowen, B. D. (1996). *An introduction to survey research, polling, and data analysis*. Thousand Oaks, CA: Sage.

Dealing with Missing Data

- Afifi, A. A., & Clark, V. (1996). *Computer-aided multivariate analysis* (3rd ed.). London: Chapman & Hall, Chapter 9, pp. 197-202.
- Allison, P. D. (1987). Estimation of linear models with incomplete data. Pp. 71-103 in C. Clogg (ed.), *Sociological Methodology 1987*. Washington, DC: American Sociological Association.
- Allison, P. D. (2000). Multiple imputation for missing data. *Sociological Methods & Research*, 28, 301–309.
- Anderson, A. B., Basilevsky, A., & Hum, D. P. J. (1983). Missing data. Pp. 415-94 in P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research*. New York: Academic Press.
- Arbuckle, J. (1996a). *Amos users' guide: Version 3.6*. Chicago: SmallWaters Corp.
- Arbuckle, J. (1996b). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (eds.), *Advanced structural equation modeling: Issues and techniques*. Mahwah NJ: LEA.
- Bentler, P. (1990). Fit indexes, Lagrange multipliers, constraint changes, and incomplete data in structural equation models. *Multivariate Behavioral Research*, 25, 163–72.
- Bentler, P. (1996). *EQS: Structural equations program manual*. Los Angeles: BMDP Statistical Software.
- Berk, R.A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review*, 48, 386–98.
- Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling* 1:287-316.
- Cohen, J., & Cohen, P. (1985). *Applied multiple regression and correlation analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B: Methodological* 39:1-38.
- Duncan, T. E., Oman, R., & Ducan, S. C. (1994). Modeling incomplete data in exercise behavior research using structural equation methodology. *Journal of Sport and Exercise Psychology* 16:187-205.
- Fay, R. E. (1996). Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association* 91: 490-98.
- Glasser, M. (1964). Linear regression analysis with missing observations among the independent variables. *Journal of the American Statistical Association* 59: 834-844.
- Glynn, R. (1985). *Regression estimates when nonresponse depends on the outcome variable*. Unpublished D.Sc. dissertation. Harvard University School of Public Health.

- Gourieroux, C., & Monfort, A. (1981). On the problem of missing data in linear models. *Review of Economic Studies* 48: 579-586.
- Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology* 78:119-128.
- Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing values patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research* 31:197-218.
- Graham, J.W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. In L. M. Collins and L. Seitz (eds.), *Advances in data analysis for prevention intervention research*. NIDA Research Monograph. Series (#142), Washington DC: National Institute on Drug Abuse.
- Heckman, J. J. (1976). The common structure of statistical models of truncated, sample selection and limited dependent variables, and a simple estimator of such models. *Annals of Economic and Social Measurement* 5: 475-492.
- Hedeker, D. & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal surveys. *Psychological Methods* 2: 64-78.
- Hill, M., & Dixon, W. (1981). Missing data: Search for patterns. In *Proceedings of the Statistical Computing Section*, 57-60. American Statistical Association.
- Joreskog, K., & Sorbom, D. (1994). *PRELIS 2: User's Reference Guide*. Chicago: Scientific Software.
- Kalton, G. & Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology* 12: 1-16.
- Kim, J. O. & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods & Analysis* 6: 215-240.
- Little, R. J. A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association* 77: 237-250.
- Little, R.J.A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association* 87: 1227-37.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 88: 125-134.
- Little, R. J. A. (1994) "A Class of Pattern-Mixture Models for Normal Incomplete Data." *Biometrika* 81: 471-483.
- Little, R., & Rubin, D. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Little, R., & Rubin, D. (1989). The analysis of social science data with missing values. *Sociological Methods and Research* 18:292-326.
- Little, R. & Schenker, N. (1995). Missing data. In G. Arminger, C. C. Clogg, & M. E. Sobel, (eds.) *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum.
- Little, R. J. A., & Smith, P. J. (1987). Editing and imputation for quantitative survey data. *Journal of the American Statistical Association* 82: 58-68.
- Muthen, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika* 52:431-462, 1987.
- Rovine, M. J. (1994). Latent variables models with missing data analysis. In A. von Eye & C. C. Clogg (eds.), *Latent variable analysis*. Thousand Oaks, CA: Sage.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63: 581-592.
- Rubin, D. (1987). *Multiple imputations for nonresponse in surveys*. New York: Wiley.
- Rubin, D. (1991). EM and beyond. *Psychometrika* 56:241-254.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association* 91: 473-489.

- Rubin, D. B., & Schenker, N. (1991). Multiple imputation in health-care databases: An overview and some applications. *Statistics in Medicine* 10: 585-598.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.
- Walker, A., Acock, A., Bowman, S., & Li, F. (1996). Amount of care given and caregiving satisfaction: a latent growth curve analysis. *Journal of Gerontology: Psychological Sciences* 51:130-142.

Weighting Survey Data

- Fuller, C. H. (1974). Weighting to adjust for survey nonresponse. *Public Opinion Quarterly*, 38, 239-46.
- Holt, D., & Elliot, D. (1991). Methods of weighting for unit non-response. *The Statistician* 40, 333-42.
- Mandell, L. (1974). When to weight: Determining nonresponse bias in survey data. *Public Opinion Quarterly*, 38, 247-51.
- Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61, 317-37.
- Pfeffermann, D. (1996). The use of sampling weights for survey data analysis. *Statistical Methods in Medical Research*, 5, 239-61.

Statistical Analysis of Complex Survey Data

- DuMouchel, W.H. & Duncan, G.J. (1983) Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association* 78: 535-543.
- Korn, E.L. and Graubard, B.I. (1991) Epidemiologic studies utilizing surveys: Accounting for the sampling design. *American Journal of Public Health*, 81, 1166-73.
- LaVange, L. M., Stearns, S. C., Lafata, J. E., Koch, G. G., & Shah, B. V. (1996). Innovative strategies using SUDAAN for analysis of health surveys with complex samples. *Statistical Methods in Medical Research*, 5, 311-29.
- Lee, S. L., Forthofer, R. N., & Lorimor, R. J. (1986). Analysis of complex sample survey data: Problems and strategies. *Sociological Methods & Research*, 15, 69-100.
- Lee, S. L., Forthofer, R. N., & Lorimor, R. J. (1988). *Analyzing complex survey data*. Newbury Park, CA: Sage.
- Pfefferman, D., & Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of the American Statistical Association*, 76, 681-89.
- Pfeffermann, D., & Holmes, D. J. (1985) Robustness considerations in the choice of a method of inference for regression analysis of survey data. *Journal of the Royal Statistical Society, Series A, General*, 148, 168-78.
- Skinner, C. J., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex surveys*. New York: Wiley.
- Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research*, 23, 230-57.

Complex Sample

- Bean, J. (1975). Distribution and properties of variance estimators for complex multistage probability samples: An empirical distribution. *Vital and Health Statistics, Series 2 Number 65*, National Center for Health Statistics.
- Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review* 51, 279-92.
- Brillinger, D. R. (1977). Approximate estimation of the standard errors of complex statistics based on sample surveys. *New Zealand Statistician*, 11(2), 35-41.
- Brogan, D., Flagg, E., Deming, M., & Waldman, R. (1994). Increasing the accuracy of the expanded Programme on Immunization's cluster survey design. *Annals of Epidemiology*, 4(4), 302-11.
- Burt, V. L., & Cohen, S. B. (1984). A comparison of alternative variance estimation strategies for complex survey data. *Proceedings of the American Statistical Association Survey Research Methods Section*.
- Carlson, B. L., Johnson, A. E., & Cohen, S. B. (1993). An evaluation of the use of personal computers for variance estimation with complex survey data. *Journal of Official Statistics*, 9(4), 795-814.
- Cohen, S. B., Xanthopoulos, J. A., & Jones, G. A. (1988). An evaluation of statistical software procedures appropriate for the regression analysis of complex survey data. *Journal of Official Statistics*, 4, 17-34.
- Dean AG, JA Dean, D Coulombier, KA Brendel, DC Smith, AH Burton, RC Dicker, K Sullivan, RF Fagan, & TG Arner (1995). *Epi Info, Version 6: A word processing, database, and statistics program for public health on IBM-compatible microcomputers*. Atlanta: Centers for Disease Control and Prevention.
- Dippo, C. S., Fay, R. E., & Morganstein, D. H. (1984). Computing variances from complex samples with replicate weights. *Proceedings of the American Statistical Association Survey Research Methods Section*.
- Fay, R. E. (1984). Some properties of estimates of variance based on replication methods. *Proceedings of the American Statistical Association Survey Research Methods Section*.
- Fay, R. E. (1990). VPLX: Variance estimates for complex samples. *Proceedings of the American Statistical Association Survey Research Methods Section*.
- Flyer, P., Rust, K., & Morganstein, D. (1989). Complex survey variance estimation and contingency table analysis using replication. *Proceedings of the American Statistical Association Survey Research Methods Section*.
- Frankel, M. R. (1971). *Inference from survey samples*. Ann Arbor, MI: Institute for Social Research, University of Michigan.
- Groves, R. (1989). *Survey errors and survey costs*. New York: Wiley.
- Hansen MH, WN Hurwitz, & WG Madow (1953). *Sample survey methods and theory, Volume I: Methods and applications*. New York: Wiley (Section 10.16).
- Hansen MH, WG Madow, & BJ Tepping (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78(384), 776-93.
- Kaplan B, I Francis, and J Sedransk (1979). A Comparison of Methods and Programs for Computing Variances of Estimators from Complex Sample Surveys. *Proceedings of the American Statistical Association Survey Research Methods Section* (pp. 97-100).
- Kish, L. (1965). *Survey sampling*. New York: Wiley, p. 162.
- Kish L and M Frankel (1968). Balanced Repeated Replications for Analytical Statistics. *Proceedings of the American Statistical Association Social Statistics Section*, pp. 2-10.

- Kish, L., & Frankel, M. R. (1970). Balanced repeated replications for standard errors. *Journal of the American Statistical Association*, 65(331), 1071–94.
- Kish, L. & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society B*(36), 1–37.
- Korn, E., & Graubard, B. (1995). Analysis of large health surveys: Accounting for the sample design. *Journal of the Royal Statistical Society A*(158), 263–295.
- Lepkowski JM, JA Bromberg, and JR Landis (1981). A program for the analysis of multivariate categorical data from complex sample surveys. *Proceedings of the American Statistical Association Statistical Computing Section*.
- McCarthy, P. (1966). Replication: An approach to the analysis of data from complex surveys. *Vital and Health Statistics, Series 2, Number 14*, National Center for Health Statistics.
- McCarthy, P. (1969). Pseudoreplication: Further evaluation and application of the balanced half-sample technique. *Vital and Health Statistics, Series 2, Number 31*, National Center for Health Statistics.
- Rao JNK & CFJ Wu (1984). Bootstrap inference for sample surveys. *Proceedings of the American Statistical Association Survey Research Methods Section*.
- Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics*, 1(4), 381–397.
- SAS Institute, Inc. (1994). *SAS System for Windows, Release 6.10 Edition*. Cary, NC: Author.
- Skinner C. J., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex surveys*. New York: Wiley.
- SPSS, Inc. (1988). *SPSS/PC+ V2.0 Base Manual*. Chicago: Author.
- Tepping, B. J. (1968). Variance estimation in complex surveys. *Proceedings of the American Statistical Association Social Statistics Section*, pp. 11–18.
- Tukey, J. W. (1958). Bias and confidence in not-quite large samples: Abstract. *Annals of Mathematical Statistics*, 29, 614.
- Wolter, K. M. (1985). *Introduction to variance estimation*. New York: Springer-Verlag.
- Woodruff, R. S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association*, 66(334), 411–14.

Bibliography: Software Documentation, References, and Contact Information

PC Carp

- Fuller, W. A. (1975). Regression analysis for sample surveys. *Sankhya C*, 37, pp. 117-132.
- Fuller W. A., Kennedy, W., Schell, D., Sullivan, G., & Park, H. J. (1989). *PC CARP*. Ames, IA: Statistical Laboratory, Iowa State University.
- Hidirouglou, M. A. (1974). *Estimation of regression parameters for finite populations*. Unpublished Ph.D. thesis. Ames, IA: Iowa State University.
- Contact:** *Statistical Laboratory, Institute for Social Research, Iowa State University at (515) 294-5242 to purchase software. Available for use on personal computers running under DOS.*

Stata

- Stata Corporation (1996). *Stata technical bulletin*. STB-31, College Station, TX, pp. 3-42.
- Contact:** *Stata Corporation, College Station, TX at (800) STATAPC (782-8272) or by Internet at Stata@Stata.com to purchase software. Available for use on personal computer (Windows, DOS, Mac) or workstation (DEC, SPARC, and others).*

SUDAAN

Shah, B. V., Folsom R. E., LaVange, L. M., Boyle, K. E., Wheelless, S. C., & Williams, R. L. (1993). *Statistical methods and mathematical algorithms used in SUDAAN*. Research Triangle Park, NC: Research Triangle Institute.

Shah B. V., Barnwell, B. G., & Bieler, G. S. (1996). *SUDAAN user's Manual, Version 6.4* (2nd ed.). Research Triangle Park, NC: Research Triangle Institute.

Contact: Research Triangle Institute at (919) 541-6602 or by Internet at SUDAAN@rti.org or <http://www.rti.org/patents/sudaan.html> to purchase software. Available for use on mainframe (VAX, IBM, DEC), workstation (VAX, SunOS, RISC-6000, DEC), or personal computer (Windows or DOS).

VPLX

Fay, R. E. (1990). VPLX: Variance estimates for complex samples. *Proceedings of the American Statistical Association Survey Research Methods Section*.

Contact: VPLX and its documentation are available free of charge on the U.S. Bureau of the Census Web site at <http://www.census.gov>. It is available in executable format for use on personal computers (DOS), VAX VMS, and UNIX-based workstations. It is portable to other systems as well—a FORTRAN version is available to copy and compile on any system.

WesVarPC

Brick, J. M., Broene, P., James, P., & Severynse, J. (1996). *A user's guide to WesVarPC*. Rockville, MD: Westat, Inc.

S-PLUS <http://www.stat.psu.edu/~jls/misoftwa.html#top>