

Statistical Data Editing References william.e.winkler@census.gov 2005Jun03

- Bankier, M. (1991), "Alternative Method of Doing Quantitative Variable Imputation," Statistics Canada Memorandum.
- Bankier, M. (2000), "2001 Canadian Census Minimum Change Donor Imputation Methodology," U.N. Economic Commission for Europe Work Session on Statistical Data Editing, Cardiff, UK, October 2000 (also available at <http://www.unece.org/stats/documents/2000.10.sde.htm>).
- Bankier, M., Houle, A.-M., Luc, M., and Newcombe, P. (1997), "1996 Canadian Census Demographic Variables Imputation," *American Statistical Association, Proceedings of the 1997 Section on Survey Research Methods*, 389-394.
- Bankier, M., Mason, P., and Poirier, P. (2002), "Imputation of Demographic Variables in the 2001 Canadian Census of Population," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Barahona, F., and Anbil, R. (1998), "The Volume Algorithm: Producing Primal Solutions with a Subgradient Method," IBM Research Report RC 21103(94395)(19FEB98).
- Barcaroli, G., and Venturi, M. (1993), "An Integrated System for Edit and Imputation of Data: an Application To the Italian Labour Force Survey," *Proceedings of the 49th Session of the International Statistical Institute*, Florence, Italy, September 1993.
- Barcaroli, G., and Venturi, M. (1997), "DAISY (Design, Analysis and Imputation System): Structure, Methodology, and First Applications," in (J. Kovar and L. Granquist, eds.) *Statistical Data Editing, Volume II*, U.N. Economic Commission for Europe, 40-51.
- Berthelot, J.-M., and Latouche, M. (1993), "Improving the Efficiency of Data Collection: A Generic Respondent Follow-up Strategy for Economic Surveys" *Journal of Business and Economic Statistics*, 11 (4), 417-424.
- Boskovitz, A., Goré, R., and Hegland, M. (2003), A Logic Formalisation of the Fellegi-Holt Method of Data Cleaning, in *Advances in Intelligent Data Analysis*, Springer LNCS 2810, 554-565 (also <http://arp.anu.edu.au/ftp/techreports/2003/TR-ARP-02-03.pdf>).
- Bruni, R. (2003), "Solving Error Correction for Large Data Sets by Means of a SAT Solver," in (E. Giunchiglia and A. Tachella, eds.) *Lecture Notes in Computer Science 2919*, Springer-Verlag, 229-241.
- Bruni, R. (2004), "Discrete Models for Data Imputation," *Discrete Applied Mathematics*, 144, 59-69.
- Bruni, R. (2005), "Error Correction for Massive Datasets," *Optimization Methods and Software*, to appear.
- Bruni, R., Reale, A., and Torelli, R. (2001), "Optimization Techniques for Edit Validation and Data Imputation," Statistics Canada Symposium 2001, Ottawa, Ontario, Canada, October 2001.
- Bruni, R., and Sassano, A. (2001) "Logic and Optimization Techniques for an Error Free Data Collecting," Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza."
- Charlton, J. C. (2003), "Evaluating New Methods for Data Editing and Imputation - Results from the Euredit Project (UK)," UNECE Statistical Data Editing Worksession, Madrid, Spain, <http://www.unece.org/stats/documents/2003/10/sde/wp.25.e.pdf>.
- Chen, B.-C. (1998), "Set Covering Algorithms in Edit Generation," *American Statistical Association, Proceedings of the Section on Statistical Computing*, 91-96 (also available as Statistical Research Division Report rr98/06 at <http://www.census.gov/srd/www/byyear.html>).
- Chen, B.-C., Thibaudeau, Y., and Winkler, W. (2002), "A Comparison Study of ACS If-Then-Else, NIM, and DISCRETE Edit and Imputation Systems Using ACS Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, to CD-ROM.
- Chen, B.-C., Thibaudeau, Y., and Winkler, W. (2003), "A Comparison Study of ACS If-the-else, NIM and DISCRETE Edit Systems using ACS Data," UNECE Statistical Data Editing Worksession, Madrid, Spain, <http://www.unece.org/stats/documents/2003/10/sde/wp.7.e.pdf>.
- Chen, B.-C., and Winkler, W. (2002), "An Efficient Formulation of Age Comparison in the DISCRETE Edit System," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM, also Research Report SRC 2002/02 at <http://www.census.gov/srd/www/byyear.html>.
- Chen, B.-C., and Winkler, W. (2003), "Pre-order and Set Covering in the DISCRETE Edit System," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM, also research Report SRC 2003/01 at <http://www.census.gov/srd/www/byyear.html>.
- Chen, B.-C., and Winkler, W. (2004), "A Cutting Plane Algorithm for the Error Localization Problem," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, to appear, CD-ROM, also research at <http://www.census.gov/srd/www/byyear.html>

- Chen, B.-C., Winkler, W. E., and Hemmig, R. J. (2000), "Using the DISCRETE Edit System for ACS Surveys," Statistical Research Division Report rr00/03 at <http://www.census.gov/srd/www/byyear.html> .
- Chernikova, N.V. (1964), "Algorithm for Finding a General Formula for the Non-negative Solutions of System of Linear Equations," *USSR Computational Mathematics and Mathematical Physics*, 4, 151-158.
- Chernikova, N.V. (1965), "Algorithm for Finding a General Formula for the Non-negative Solutions of System of Linear Inequalities," *USSR Computational Mathematics and Mathematical Physics*, 5, 228-233.
- Coppola, L., Di Zio, M., Luzi, O., Ponti, A., and Scanu, M. (2002), "Bayesian Networks for Imputation: A Case Study," University of Jyväskylä, Finland, *Proceedings of the DataClean Conference*.
- DesJardins, D. (1998), "A New Graphical Techniques for the Analysis of Census Data", *Statistics Canada Conference Proceedings*.
- De Kleer, J. (1992), "An Improved Incremental Algorithm for Generating Prime Implicates," AAAI '92, 780-785.
- De Waal, T. (1996), "CherryPi: A Computer Program for Automatic Edit Error Localization," Paper presented at the UN Work Session on Statistical Data Editing, 4-7 November 1996, Voorburg, the Netherlands.
- De Waal, T. (1997), "A Recipe for Applying CherryPi to the Edit Process," Paper presented at the UN Work Session on Statistical Data Editing, 14-17 October 1997, Prague, Czech Republic (also available at <http://www.unece.org/stats/documents/1997.10.sde.htm>).
- De Waal, T. (2000), "New Developments in Automatic Edit and Imputation at Statistics Netherlands," U.N. Economic Commission for Europe Work Session on Statistical Data Editing, Cardiff, UK, October 2000 (also available at <http://www.unece.org/stats/documents/2000.10.sde.htm>).
- De Waal, T. (2003a), "Solving the Error Localization Problem by Means of Vertex Generation," *Survey Methodology*, 29 (1), 71-79.
- De Waal, T. (2003b), "A Fast and Simple Algorithm for Automatic Editing of Mixed Data," *Journal of Official Statistics*, 19 (4), 383-402.
- De Waal, T. (2003c), "Computational Results with Various Error Localization Algorithms," UNECE Statistical Data Editing Worksession, Madrid, Spain, <http://www.unece.org/stats/documents/2003/10/sde/wp.22.e.pdf>.
- De Waal, T. (2003d), *Processing of Erroneous and Unsafe Data*, ERIM Research in Management: Rotterdam.
- De Waal, T., Van De Pol, F., Renssen, R. (2000), "Graphical Macro Editing: Possibilities and Pitfalls," in *The Second International Conference on Establishment Surveys: Survey Methods for Businesses, Farms, and Institutions*, American Statistical Association, Alexandria, Virginia, pp. 579-588.
- Di Zio, M., Luzi, O., and Manzari, A. (2002), "Evaluating Editing and Imputation Processes: The Italian Experience," UNECE Statistical Data Editing Worksession, Helsinki, Finland, <http://www.unece.org/stats/documents/2002/05/sde/12.e.pdf> .
- Di Zio, M., Sacco, G., Scanu, M., and Viccard, P. (2004), "Multivariate Techniques for Imputation Based on Bayesian Networks," *Compstat 2004 Symposium*, to appear.
- Di Zio, M., Scanu, M., Coppola, L., Luzi, O., and Ponti, A. (2004), "Bayesian Networks for Imputation," *Journal of the Royal Statistical Society, A*, 167 (2), 309-322.
- Draper, L., and Winkler, W.E. (1997), "Balancing and Ratio Editing with the new SPEER system," *American Statistical Association, Proceedings of the 1997 Section on Survey Research Methods*, 570-575 (also available as Statistical Research Division Report rr97/05 at <http://www.census.gov/srd/www/byyear.html>).
- DuMouchel, W., Volinsky, C., Johnson, T., Cortes, C., and Pregibon, D. (1999), "Squashing Flat Files Flatter," *Proceedings of the ACM Knowledge Discovery and Data Mining Conference*, 6-15.
- Fellegi, I. P., and Holt, D. (1976), "A Systematic Approach to Automatic Edit and Imputation," *Journal of the American Statistical Association*, 71, 17-35.
- Filion, J.-M., and Schopiu-Kratina, I. (1993), "On the Use of Chernikova's Algorithm for Error Localization," *Statistics Canada Technical Report*.
- Franconi, E., Palma, A. L., Leone, N., Perri, S., and Scarcello, F. (2001), "Census Data Repair: A Challenging Application of Disjunctive Logic Programming," *8th International Conference on Logic for Programming, Artificial Intelligence and Reasoning*.
- Garfinkel, R. S., Kunnathur, A. S., and Liepins, G. E., (1986), "Optimal Imputation of Erroneous Data: Categorical Data, General Edits," *Operations Research*, 34, 744-751.
- Garfinkel, R. S., Kunnathur, A. S., and Liepins, G. E. (1988), "Error Localization for Erroneous Data: Continuous Data, Linear Constraints," *SIAM Journal on Scientific and Statistical Computing*, 9, 922-931.
- Garcia, M. (2002), "Implied Edit Generation for Ratio and Balance Edits," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Garcia, M. (2003), "Implied edit generation and error localization for ratio and balancing edits," UNECE Statistical Data Editing Worksession, Madrid, Spain, <http://www.unece.org/stats/documents/2003/10/sde/wp.27.e.pdf>.

- Garcia, M. (2004), "Implicit Linear Inequality Edits Generation and Error Localization in the SPEER Edit System," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM, to appear.
- Garcia, M., and Thompson, K. J. (2000), "Applying the Generalized Edit/Imputation System AGGIES to the Annual Capital Expenditures Survey," *Proceedings of the International Conference on Establishment Surveys, II*, 777-789.
- Granquist, L. (1995), "Improving the Traditional Editing Process," in Cox, Binder, Chinappa, Christianson, Colledge and Kott (eds.) *Business Survey Methods*, John Wiley and Sons, 385-401.
- Granquist, L. (1997), "Macro-Editing - The Aggregate Method," *Statistical Data Editing, UN Conference of European Statisticians Statistical Standards and Studies, Geneva (Switzerland)*.
- Granquist, L., and Kovar, J. (1997), "Editing of Survey Data: How Much is Enough?" in Lyberg, L., Biemer, P., Collins, M., De Leeuw, E. Dippo, C., Schwartz, N. and Trewin, D. (eds) *Survey Measurement and Process Quality*, John Wiley and Sons, 415-435.
- Greenberg, B. G., and Surdi, R. (1984), "A Flexible and Interactive Edit and Imputation System for Ratio Edits," SRD report RR-84/18, U.S. Bureau of the Census, Washington, D.C., USA.
- Greenberg, B. G., and Petkunas, T. (1990), "Overview of the SPEER System," SRD report RR-90/15, U.S. Bureau of the Census, Washington, D.C., USA.
- Hidioglou, M.A., and Berthelot, J.-M. (1986), "Statistical Editing and Imputation of Periodic Business Surveys," *Survey Methodology*, 12, 73-83.
- Hogan, H. (1995), "How Exploratory Data Analysis is Improving the Way We Collect Business Statistics," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 102-107.
- Kosinski, A.S. (1999), "A Procedure for the Detection of Multivariate Outliers," *Computational Statistics and Data Analysis*, 29, 145-161.
- Kovar, J.G., MacMillan, J.H., and Whitridge, P. (1991), "Overview and Strategy for the Generalized Edit and Imputation System", Statistics Canada, Methodology Branch Working Paper BSMD 88-007E (updated in 1991).
- Kovar, J.G., and Whitridge, P. (1990), "Generalized Edit and Imputation System: Overview and Applications," *Revista Brasileira de Estadística*, 51, 85-100, Rio de Janeiro.
- Kovar, J. G., and Winkler, W. E. (1996), "Editing Economic Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 81-87 (also available as Statistical Research Division Report rr00/04 at <http://www.census.gov/srd/www/byyear.html>).
- Latouche, M., and Berthelot, J.-M (1992), "Use of a Score Function to Prioritize and Limit Recontacts in Business Surveys" *Journal of Official Statistics*, 8 (3), 389-400.
- Little, R. A., and Rubin, D. B., (1987), *Statistical Analysis with Missing Data*, John Wiley: New York.
- Manzari, A., and Reale, A. (2001), "Towards a new method of edit and imputation of the Italian Census: A Comparison with the Canadian Nearest-Neighbour Methodology," Presented at the International Statistical Institute Meeting in Seoul, Korea (September 2001).
- Manzari, A., and Reale, A. (2002), "Towards a New Method of Edit and Imputation of the Italian Census: A Comparison with the Canadian Nearest-Neighbour Methodology," paper presented at the U.N Economic Commission for Europe worksession, May 27-29, 2002, Helsinki, Finland (available at <http://www.unece.org/stats/documents/2002.05.sde.htm>).
- McKeown, P.G., (1984), "A Mathematical Programming Approach to Editing of Continuous Survey Data," *SIAM Journal on Scientific and Statistical Computing*, 5, 784-797.
- Moore, A. W., and Lee, M. S. (1998), "Cached Sufficient Statistics for Efficient Machine Learning with Large Datasets," *Journal of Artificial Intelligence Research*, 8, 67-91.
- Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression*, Reading, MA: Addison-Wesley.
- Naus, J. (1975), *Data Quality Control and Editing*, Marcel Dekker: New York.
- Nemhauser, G. L., and Wolsey, L. A., (1988), *Integer and Combinatorial Optimization*, John Wiley: New York.
- Nordbotten, S. (1963), *Automatic Editing of Individual Statistical Observations*, Statistical Standards and Studies, Handbook No. 2, United Nations: New York.
- Nordbotten, S. (1995), "Editing Statistical Records by Neural Networks," *Journal of Official Statistics*, 11, 393-414.
- Nordbotten, S. (1996), "Neural Network Imputation Applied to the Norwegian 1990 Population Census Data," *Journal of Official Statistics*, 12, 385-402.
- Owen, A. (2003), "Data Squashing by Empirical Likelihood," *Data Mining and Knowledge Discovery*, 7 (1), 101-113.
- Pritzker, L., Ogus, J., and Hansen, M. H. (1965), "Computer Editing Methods--Some Applications and Results," *Bulletin of the International Statistical Institute, Proceedings of the 35th Session*, Belgrade, 395-417.
- Ragsdale, C.T., and McKeown, P.G. (1996), "On Solving the Continuous Data Editing Problem," *Computers & Operations Research*, 23, 263-273.

- Riera-Ledesma, J., and Salazar-Gonzalez, J.-J. (2003), "New Algorithms for the Editing and Imputation Problem," UNECE Statistical Data Editing Worksession, Madrid, Spain, <http://www.unece.org/stats/documents/2003/10/sde/wp.5.e.pdf>.
- Riera-Ledesma, J., and Salazar-Gonzalez, J.-J. (2004), "A Branch-and-Cut Algorithm for the Error Localization Problem in Data Cleaning," technical report, Universidad de la Laguna, Tenerife, Spain.
- Rocke, D.M., and Woodruff, D.L. (1996), "Identification of Outliers in Multivariate Data," *Journal of the American Statistical Association*, 91, 1047-1061.
- Rubin, D.S. (1975), "Vertex Generation in Cardinality Constrained Linear Programs," *Operations Research*, 23, 555-565.
- Sande, G. (1979), "Numerical Edit and Imputation," *Proceedings of the 42nd Session of the International Statistical Institute*, Manila, Philippines.
- Schaffer, J., (1987), "Procedure for Solving the Data-Editing Problem with Both Continuous and Discrete Data Types," *Naval Research Logistics*, 34, 879-890.
- Schiopu-Kratina, I., and Kovar, J.G. (1989), "Use of Chernikova's Algorithm in the Generalized Edit and Imputation System," Statistics Canada, Methodology Branch Working Paper BSMD 89-001E.
- Sigman, R., and Wagner, D. (1997), "Algorithms for Adjusting Survey Data that Fail Balance Equations," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 576-581
- Simon, L., and del Val, A. (2001), "Efficient Consequence Finding," *Proceedings of the International Joint Conference on Artificial Intelligence*, 359-370.
- Thibaudeau, Y. (1999), "Model Explicit Item Imputation in Demographic Surveys and Censuses," in *Statistical Data Editing*, Rome: ISTAT, pp. 153-167 (also available at <http://www.census.gov/srd/www/byyear.html>).
- Thibaudeau, Y., and Winkler, W. E. (2002), "Bayesian Network Representations, Generalized Imputation, and Synthetic Data Satisfying Analytic Restraints," (Research report RRS 2002/09 <http://www.census.gov/srd/www/byyear.html>).
- Thibaudeau, Y., and Winkler, W. E. (2005), "Smoothing Imputations for Categorical Data in the Linear Regression Paradigm," UNECE Worksession on Statistical Data Editing, Ottawa, Ontario, Canada, May 2005 (<http://www.unece.org/stats/documents/2005.05.sde.htm>).
- Thibaudeau, Y., and Winkler, W. E. (2006), "Graphical Networks, Loglinear Models, and General Methods for Sampling from Contingency Tables Satisfying Analytic Restraints," in progress.
- Thompson, K. J. (1999), "Ratio Edit Tolerance Development Using Variations of Exploratory Data Analysis (EDA) Resistant Fences Methods," *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, Statistical Policy Working Paper 29, 95-104.
- Thompson, K. J. (2002), "Evaluating the Effect of Two Editing Systems on Data Quality: Two Case Studies," UNECE Statistical Data Editing Worksession, Helsinki, Finland, <http://www.unece.org/stats/documents/2002/05/sde/5.e.pdf>
- Thompson, K. J., and Sigman, R. S. (1999), "Statistical Methods for Developing Ratio Edit Tolerances for Economic Data," *Journal of Official Statistics*, 15, 517-535.
- Todaro, T. (1997), "Adapting the SPEER Edit System to Edit Hog Data in the National Agricultural Statistics Service's Quarterly Agricultural Surveys," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 570-575.
- Todaro, T. (1999), "Overview and Evaluation of the AGGIES Automated Edit and Imputation System," paper presented at the U.N Economic Commission for Europe worksession, June 2-4, 1999, Rome, Italy (available at <http://www.unece.org/stats/documents/1999.06.sde.htm>).
- Tukey, J. W. (1988), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.
- Van De Pol, F., and Bethlehem, J. (1997), "Data Editing Perspectives," *Statistical Journal of the United Nations ECE*, 14, 153-171.
- Weng, S. (2002), "Linear Edits Error Localization – An Implementation in C++," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Weng, S. (2004), "Modifications of Chernikova's Algorithm and Error Localization in Linear Editing," technical report, NASS/USDA.
- Williams, T. (1998), "Imputing Person Age for the 2000 Census Short Form: A Model-based Approach," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 680-685 ((available at <http://www.census.gov/srd/www/byyear.html> as research report rr98/07).
- Winkler, W. E. (1995), "Editing Discrete Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 108-113 (also available as Statistical Research Division Report rr97/04 at <http://www.census.gov/srd/www/byyear.html>).

- Winkler, W.E. (1997a), "Set-Covering and Editing Discrete Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 564-569 (also available as Statistical Research Division Report rr98/01 at <http://www.census.gov/srd/www/byyear.html>).
- Winkler, W. E. (1997b), "Problems with Inliers," paper presented at the European Conference of Statisticians, October 14-17, 1997, Prague, Czech Republic, http://www.unece.org/stats/documents/1997/10/data_editing/22.e.pdf
- Winkler, W. E. (1999), "The State of Statistical Data Editing," in *Statistical Data Editing*, Rome: ISTAT, 169-187 (also available at <http://www.census.gov/srd/www/byyear.html>).
- Winkler, W. E. (2003), "A Contingency Table Model for Imputing Data Satisfying Analytic Constraints," *American Statistical Association, Proc. Survey Research Methods Section*, CD-ROM, also research Report SRS 2003/07 at <http://www.census.gov/srd/www/byyear.html>.
- Winker, W. E. (2005a), "Modeling Data and Data Quality," technical report, to appear.
- Winker, W. E. (2005b), "Determining a Set of Edits," technical report, to appear.
- Winker, W. E. (2005c), "Methods and Analyses for Determining Quality," technical report, to appear.
- Winkler, W. E. and Chen, B.-C. (2002), "Extending the Fellegi-Holt Model of Statistical Data Editing," (available at <http://www.census.gov/srd/www/byyear.html>).
- Winkler, W. E., and Draper, L. (1997), "The SPEER Edit System," in (J. Kovar and L. Granquist, eds.) *Statistical Data Editing, Volume II*, U.N. Economic Commission for Europe, 51-55.
- Winkler, W. E., and Garcia, M. M. (2005), "Modeling and Analysis with Data," UNECE Worksession on Statistical Data Editing, Ottawa, Ontario, Canada, May 2005 (<http://www.unece.org/stats/documents/2005.05.sde.htm>).
- Winkler, W. E., and Petkunas, T. (1997), "The DISCRETE Edit System," in (J. Kovar and L. Granquist, eds.) *Statistical Data Editing, Volume II*, U.N. Economic Commission for Europe, 56-62.