8
Referências bibliográficas

ABELLO, R. and PHILLIPS, B. Statistical matching of the HES and NHS: an exploration of issues in the use of unconstrained and constrained approaches in creating a basefile for a microsimulation model of the pharmaceutical benefits segment. Technical report, Australian Bureau of Statistics. Methodology Advisory Committee Paper, June, 2004.

ABRAMOVITZ, M. and STEGUN, I. A. Handbok of mathematical functions. Dover publications, 1970.

ADAMS, L. M., and DARWIN, G. Solving the quandary between questionnarie length and response rate in Higher Education. Research in Higher Education, 17, p. 231-240, 1982.

AGRESTI, A., Categorical data analysis. J. Wiley & Sons, New York, 1990.

ALBIERI, S. e BIANCHINI, Z. M. Uma revisão dos principais aspectos dos planos amostrais das pesquisas domiciliares realizadas pelo IBGE, Rio de Janeiro: IBGE, 21p. (Texto para Discussão, n.º 91), 1998.

ALLISON, P. Multiple imputation for missing data: a cautionary tale. Sociological Methods and Research, 28 (3), 301-309, 2000.

ALTER, H. E. Creation of a synthetic data set by linking records of the Canadian survey of consumer finances with the family expenditure survey 1970. Annals of Economic and Social Measurement, 2, 373-394, 1974

ALUJA-BANET, T. e THIO, S. Survey data fusion. Bulletin of Sociological Methodology, 72, 20-36, 2001.

ANDERSON, T.W. An introduction to multivariate analysis. Second Edition New York: John Wiley & Sons, 1984.

ARMSTRONG, J. An evaluation of statistical matching methods. Working Paper no. BSMD 90-003E, Methodology Branch, Statistics Canada, Ottawa, 1989.

ATROSTIC, B. K. A multiple imputation approach to microsimulation. Proceedings of the Survey Research Methods Section, American Statistical Association, 529-534. Washington, DC: American Statistical Association, 1994.

BAKER, K., HARRIS, P. and O'BRIEN, J. Data fusion: an appraisal and experimental evaluation. Journal of the Market Research Society, 31(2), 153-212, 1989.

BAKER, K. HARRIS, P. and O'BRIEN, J. Data fusion: can perfection be reached? Journal of the Market Research Society, 32(3), 473-474, 1990.

BAKKER, B.F.M. and WINKELS, J.W. Why integration of household surveys? – Why POLS? Netherlands Official Statistics, 13, 5.7, 1998.

BARR, R.S. and TURNER, J.S. Quality issues and evidence in statistical file merging. In Data Quality Control: Theory and Pragmatics (G.E. Liepins and V.R.R. Uppuluri (eds)), 245-313. New York: Marcel Dekker, 1990.

BARR, R. S., STEWART, W. H., e TURNER, J. S. An empirical evaluation of statistical matching methodologies. Technical Report. Edwin L. Cox School of Business. Southern Methodist University, Dallas, TX, 1982.

BARR, R. S. e TURNER, J. S. A new, linear programming approach to micro data file merging. In 1978 Compendium of Tax Research. Washington, DC: Office of Tax Analysis, U.S. Department of the Treasury, 131-149, 1978.

BARRY, J.T. An investigation of statistical matching. Journal of Applied Statistics, 15, 275-283, 1988.

BERTSEKAS, D.P. Linear network optimization: algorithms and codes. Cambridge, Massachusetts: MIT Press, 1991.

BOCK, T. and JEPHCOTT, J. The application and validation of data fusion. International Journal of Market Research, 40(3), 1998.

BOURGEOIS, R. The World Bank participation sourcebook. World Bank, Washington, D.C., USA. Available online at: http://www.worldbank.org/wbi/ sourcebook/sbhome.htm, World Bank, 1996. Acesso: nov/2008

BOX, G.E.P., and TIAO, G.C. Bayesian inference in statistical analysis, New York: Wiley, 1992.

BREIMAN, L. Statistical modeling: The Two Cultures (with comments and a rejoinder by the author) Statist. Sci. 16 199-231, 2001.

BURKARD, R. E. and DERIGS, U. Assignment and matching problems: solution methods with FORTRAN-Programs. Berlin: Springer-Verlag, 1980.

BUDD, E. C. The creation of a micro data file for estimating the distribution of income. Review of Income and Wealth, 17, 317-333, 1971.

CARPENTER, R. and WILCOX, S. Data fusion in the British national readership survey – an experiment. Worldwide Readership Research Symposium VII, Berlin, 1995.

CASELLA, G. and BERGER, R. L. Statistical inference. Duxbury Press, 2002.

CASSEL, C.M. Statistical matching – statistical prediction. What is the difference? An evaluation of statistical matching and a special type of prediction using data from a survey on living conditions. Statistisk Tidskrift, 5, 55-63, 1983.

CHATFIELD, C. Model uncertainty, data mining and statistical inference (with discussion). Journal of the Royal Statistical Society, Series A, 158, 419-466, 1995.

CHEN, J. and SHAO, J. Nearest neighbor imputation for survey data. Journal of Official Statistics, 16(2), 113-131, 2000.

CHEN, J. and SHAO, J. Jackknife variance estimation for nearest-neighbor imputation. Journal of the American Statistical Association, 96, 260-269, 2001.

CHEUNG, S. and SEKO, C. A study of the effects of imputation groups in the nearest neighborhood imputation method for the national farm survey. Survey Methodology, 12(1), 99-106, 1986.

CHIPPERFIELD, J. O. and STEEL, D. G. Design and estimation for split questionnaire surveys. Journal of the Official Statistics, 25, 227-244, 2009.

CICCHITELI, G., HERZEL, A. e MONTANARI, G. E. Il campionamento statistico, Il Mulino, 1992.

CITONI, G., DI NICOLA, F., LUGARESI, S. and PROTO, G. Statistical matching for tax-benefit micro simulation modeling: a project for Italy. Workshop on Statistical Matching, ISPE, 26 September 1991, Rome, Italy, 1991.

CITRO, C.F. and HANUSHEK, E. A. Improving information for social policy decisions: the uses of micro simulation modeling, Volume II. Washington, DC: National Academy Press, 1991.

COCHRAN, W.G. and RUBIN, D. B. Controlling bias in observational studies: a review. Sankhya, Series A, 35, 417-446, 1973.

COCHRAN, W.G. Sampling techniques (3rd ed.). New York: John Wiley, 1977.

COHEN, M.L. Statistical matching and microsimulation models. In Improving information for social policy decisions: the uses of microsimulation modeling, Volume II, Technical Papers, Citro, C. F. and Hanushek E. A. (eds). Washington, D.C.: National Academy Press, 62-88, 1991a.

COHEN, M.L. Variance estimation of microsimulation models through sample reuse. In Improving information for social policy decisions: the uses of microsimulation modeling, Volume II, Technical Papers, Citro, C. F. and Hanushek E. A. (eds). Washington, D.C.: National Academy Press, 237-254, 1991b.

COHEN, M.L. Improving Information for social policy decisions: the uses of microsimulation modeling: Volume II, Technical Papers, 1991.

COHEN, M.P. A new approach to imputation. Proceedings of the Section on Survey Research Methods, American Statistical Association, 293-298. Washington, DC: American Statistical Association, 1996.

COLI, A. and TARTAMELLA, F. The link between national accounts and household's micro data. Paper presented at the Meeting of the Siena Group on Social Statistics, Maastricht (Netherlands), 22-25, May 2000.

COLI, A. and TARTAMELLA, F. A pilot social accounting matrix for Italy with a focus on households. Paper presented to the 26th General Conference of the International Association for Research in Income and Wealth, Cracow (Poland), 27 August – 2 September, 2000.

COX, D.R., and WERMUTH, N. Multivariate Dependencies, London: Chapman and Hall, 1996.

DE CASTRO, L. I. “Affiliation and the Revenue Ranking of Auctions” preliminary version, 2006.

DE GROOT, M. H., FELDER, P. I. e Goel, P. K. Matchmaking. The Annals of Math. Statistics 42, p. 578-593, 1971.

DE GROOT, M. H. and GOEL, P. The matching problem for multivariate normal data. Sankyā, 38 (Series B, Part 1), 14-28, 1976.

DE GROOT, M. H. Record linkage and matching systems. p. 649-654 in Encyclopedia of Statistical Sciences, Vol. 7. S. Kotz and N. L. Johnson , eds. New York : Wiley Interscience, 1987.

DENK, M. and HACKL, P. Data integration and record matching: an Austrian contribution to research in official statistics. Austrian Journal of Statistics, 32, 305-321, 2003.

DILLMAN, D. A., SINCLAIR, M. D. and CLARK, J. R. Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed Census mail surveys. Public Opinion Quarterly, 57, p. 289-304, 1993.

D’ORAZIO, M., DI ZIO, M. and SCANU, M. Statistical Matching: Theory and Practice, Wiley, New York, 2006.

D’ORAZIO, M., DI ZIO, M. and SCANU, Some experiments on Statistical Matching in the R environment,http://www.r-project.org/user-2006/Slides/DOrazioEtAl.pdf , 2006b. Acesso: nov/2008

D'ORAZIO, M., DI ZIO, M., and SCANU, M. "Statistical matching and the likelihood principle: uncertainty and logical constraints", ISTAT Technical Report 1/2004, 2004.

D’ORAZIO, M., DI ZIO, M. and SCANU, M. Statistical matching and official statistics. Rivista di Statistica Ufficiale, 2002/1, 5-24, 2002.

D’ORAZIO, DI ZIO, M. and SCANU, M. Statistical matching: a tool for integrating data in National Statistical Institutes. Second International Seminar of Exchange of Technology and Know-How/Fourth New Techniques and Technologies for Statistics Seminar. Crete, 2001.

DRAPER, D. et al,. Combining Information Statistical Issues and Opportunities for Research, American Statistical Association, Washington, D.C. Contemporary Statistics Number 1, p. 123-134, 1992.

ELBERS, C., LANJOUW, J. O., LANJOUW P. and LEITE, P. G. Poverty and inequality in Brazil new estimates from combined PPV-PNAD Data, DECRG, the World Bank, 2004.

FELLEGI, I. P., and SUNTER, A. B. A theory for record linkage. Journal of the American Statistical Association, 64, p.1183-1210, 1969.

FILIPPELLO, R., GUARNERA, U. and JONA LASINIO, G. Use of auxiliary information in statistical matching. In Proceedings of the XLII Conference of the Italian Statistical Society, pp. 37 – 40. Bari (Italy), 9 – 11 June 2004. Padua: CLEUP, 2004.

FISHER, N., DERQUENNE, C. and SAPORTA, S. A new method to match data set applied to the electric market. Fourth New Techniques and Technologies for Statistics Seminar- Fourth Annual Exchange of Technology and Knowhow Seminar. June 18-21, Crete, 2001.

FUKUNAGA, K. and NARENDRA, P.M. A branch and bound algorithm for computing k-nearest neighbor. IEEE Transactions on Computers, 24, 750-753, 1975.

GOEL, P. K. and RAMALINGAM, T. Some properties of the maximum likelihood strategy for repairing broken random sample. J. Statistics Planning Inference. 16: 237-248, 1987.

GOEL, P. K. and RAMALINGAM, T. The matching methodology: some statistical properties. Lecture Notes in Statistics Series, No. 52. . New York : Springer-Verlag, 1989.

GRAHAM, P., YOUNG, J., PENNY, R. Multiply Imputed Synthetic Data: Evaluation of Hierarchical Bayesian Imputation Models vol. 25, no 2, pp 245, 2009.

GRONE, R., JOHNSON, C. R., SÁ, E. M., e Wolkowicz, H. Positive definite completions of partial hermit an matrices, Linear Algebra and its Applications, 58, pp. 109-124, 1984

HERZOG, A. R., and BACHMAN, J. G. Effects of questionnaire length on response quality. Public Opinion Quarterly, 45, p. 549-559, 1980.

HOAGLIN, D. C., MOSTELLER, F. e TUKEY, J. W. Análise exploratória de dados. Técnicas Robustas. Um Guia. Lisboa: Edições Salamandra, 1992.

INGRAM, D., O'HARE, J., SCHEUREN, F. and TUREK, J. Statistical matching: a new validation case study. Proceedings of the Survey Research Methods Section, American Statistical Association, 2000.

INGRAM, D., TUREK, J., SCHEUREN, F. and O'HARE, J. Statistical matching: continued validation. Proceedings of the Social Sciences Section, American Statistical Association, 2001.

ISTAT, I consume delle famiglie, anno 2000. Annuario Settore Famiglia e Società, no. 7 (in Italian), 2002.

JEPHCOTT, J. and BOCK, T. The application and validation of data fusion. Journal of the Market Research Society, 40(3), 185-205, 1998.

JOHNSON, R. e WICHERN, D. Applied Multivariate Statistical Analysis, Prentice Hall, 2002.

JUDGE, G.G., GRIFFITHS, W.E., HILL, R.C. and LEE, T.C. The theory and econometrics. New York: Wiley, 1980.

KADANE, J. B. Some statistical problems in merging data files. Journal of Official Statistics, 17, pp. 423-433, 1978.

KALTON, G. e KASPRIZYK, D. The treatment of missing data. Survey Methodology, 12, p. 1-16, 1986.

KAMAKURA, W. A. and WEDEL, M. Statistical data fusion for cross-tabulation. Journal of Marketing Research, 34, 485-498, 1997.

KELLEY, R. P. A preliminary study of the error structure of statistical matching. Pp. 206-208 in Proceedings of the Section on Social Statistics. VA: American Statistical Association, 1983.

KISH, L. Survey Sampling. New York: Wiley, 1965.

KLEVMARKEN, N.A. Comment on paass In G.H. Orcutt, J. Merz and H. Quinke (eds). Microanalystic simulation models to support social and financial policy, pp. 421 – 422. Amsterdam: Elsevier Science, 1986.

KNUTH, D. E. The art of computer programming, Addison-Wesley, 1973.

KOVACEVIC, M. S. and LIU, T. P. Statistical matching of survey data files. Proceedings of the Survey Research Methods Section, American Statistical Association. Washington, DC: American Statistical Association.

KROESE, B., RENSSEN, R.H. and TRIJSSENAAR, M. Weighting or imputation: constructing a consistent set of estimates based on data from different sources. Netherlands Official Statistics, 15, 23-31, 2000.

LEHMANN, E. L. Some concepts of dependence. Annals of Mathematical Statistics, 37, 1137-1153, 1966.

LITLLE, R. J. A. Missing-Data Adjustments in Large Surveys, Journal of Business & Economic Statistics, 6, 3, p. 287-297, 1988.

LITLLE, R. J. A. and Rubin D. B. Statistical analysis with missing data. John Wiley and Sons, New York, 1987.

LIU, T.P. and KOVACEVIC, M.S. An empirical study on categorically constrained matching. Proceedings of the Survey Methods Section, Statistical Society of Canada, 167-178, 1997.

LIU, J. S. and WU, Y.N. Parameter expansion scheme for data augmentation. Journal of the American Statistical Association, 94, 1264-1274, 1999.

MARDIA, K.V., KENT, J.T. and BIBBY, J.M. Multivariate analysis. London: Academic Press, 1979.

MEIMAND, M. Statistical matching – a first experience. New Zealand Statistics Review, 1995 (March), 9-19, 1995.

MILGROM, P. R AND WEBER, R. J. “A Theory of Auctions and Competitive Bidding,” Econometrica, 50, 1089-1122, 1982a.

MORIARITY, C. and SCHEUREN, F. Statistical matching: a paradigm for assessing the uncertainty in the procedure. Journal of Official Statistics, 17, 407-422, 2001.

MORIARITY, C., and SCHEUREN, F. A Note on Rubin's statistical matching using file concatenation with adjusted weights and multiple imputations, Journal of Business and Educational Studies, 21, pp. 65-73, 2003.

MORIARITY and SCHEUREN. Regression-based statistical matching: recent developments, Proceedings of the Joint Statistical Meetings, American Statistical Association, 4050-4057, 2004.

MORIARITY and SCHEUREN. Statistical matching with assessment of uncertainty in the procedure*:* new findings, Proceedings of the Joint Statistical Meetings, American Statistical Association, 2904-2909, 2003.

MORRISON, D.F. Multivariate statistical methods, McGraw-Hill, 1967.

MUNGER, G. F. and LLOYD, B. H. The use of multiple matrix sampling for survey research, The Journal of Experimental Education, 56, p. 187-191, 1988.

NAVARRO, A. and GRIFFIN, (1993). Matrix Sampling Design for the Year 2000 Census, Proceedings of the Section of Survey Research Methods, American Statistical Association, p. 448-485

OKAUCHI, S. The Japanese fusion experience (validation of data fusion by means of ACR data splitting). ARF Week of Workships. October 10, 2002, New York City, 2002.

OKNER, B. Constructing a new data base from existing micro data sets: The 1966 merge file. Annals of Economic and Social Measurement 1:325–362, 1972.

OKNER, B.A. Data matching and merging: an overview. Annals of Economic and Social Measurement 3 (2), 347 – 352, 1974.

PAASS, G. Statistical record linkage methodology, state of the art and future prospects. In Bulletin of the International Statistical Institute , Proceedings of the 45 the Session, volume LI, Book 2, 1985.

PAASS, G. Statistical match: evaluation of existing procedures and improvements by using additional information. In G.H. Orcutt, J. Merz and H. Quinke (eds) Microanalytic Simulation Models to Support Social and Financial Policy, pp. 401 – 422. Amsterdam: Elsevier Science, 1986.

R Development Core Team 2004 R: A language and environment for statistical computing. Vienna: Foundation for Statistical Computing.

RADNER, D. B. The development of statistical matching in economics. Proceedings of the Social Statistics Section, American Statistical Association, 503-508, 1978.

RADNER, D. B., ALLEN, R., GONZALEZ, M. E., JABINE, T. B. and MULLER H. J. Report exact and statistical matching techniques. Statistical Policy Working Paper 5, Federal Committee on Statistical Methodology,1980.

RAGHUNATHAN, T. E. and GRIZZLE J. E. A split questionnaire survey design, Journal of the American Statistical Association, 90, p. 54-63, 1995.

RAMALINGAM, T. Statistical properties of the file-merging methodology. Doctor of Philosophy Thesis. Lafayette, Indiana: Purdue University, 1985.

RÄSSLER, S., KOLLER, F. and MÄENPÄÄ, C. A Split Questionnaire Survey Design applied to German Media and Consumer Surveys, www.icis.dk/ICIS_papers/D1_2_3.pdf . Acess: 2005, 2001.

RÄSSLER, S. Alternative approaches to statistical matching with an application to media data (alternative verfahren der datenfusion mit einer anwendung auf media-daten), Habilitationsschrift, Nürnberg, 2001.

RÄSSLER, S. Split questionnaire survey sampling. IMPUTE: Symposium on No response, Questionnaire Split and Multiple Imputation, Nuremberg, Germany, September 25, 2002.

RÄSSLER, S. Statistical matching: a frequentist theory, practical applications, and alternative Bayesian approaches. Springer-Verlag New York: New York, 2002.

RÄSSLER, S. A non-iterative Bayesian approach to statistical matching. Staistica Neerlandica, 57(1), 58-74, 2003.

RÄSSLER, S. Data fusion: identification problems, validity, and multiple imputation. Austrian Journal of Statistics, 33(1-2), 153-171, 2004.

RÄSSLER, S. and FLEISHER, K. An evaluation of data fusion techniques. Workshop and Symposium on Combining Data from Difference Sources. May 4-7, 1999, Ottawa, Canada, 1999.

RAO, C. R. On the distance between two populations. Sankhya, 9, 246-248, 1949.

RENSSEN, R. H. "Use of statistical matching techniques in calibration estimation", Survey Methodology, 24, pp. 171-183, Statistics Canada, 1998.

Report 10, ITME (Information Technology and the Market Economy) Project of the Japan Society Promotion of Science.

RIDDER, G. and MOFFITT, R. "The econometrics of data combination", in Heckman, J.J., and E.E. Leamer (eds.), Handbook of Econometrics Volume 6, Amsterdam: North Holland (to appear), 2006.

RODGERS, W. L. An evaluation of statistical matching. J. Bus. Econ. Statistic. 02: 91-102, 1984.

RODGERS, W.L. and DEVOL, E. An evaluation of statistical matching. Proceedings of the Section on Survey Research Methods, American Statistical Association, 128-132. Washington, DC: American Statistical Association, 1982.

RODRIGUES, N. M. C. Amostragem matricial no questionário da amostra – CD 1.02 do Censo Demográfico de 1991. Dissertação de Mestrado em estudos populacionais e pesquisas sociais. Escola Nacional de Ciências Estatísticas. Rio de Janeiro. 132p, 2003.

ROSENBAUM, P. R and RUBIN D. B.. The central role of the propensity score in observational studies for causal effects. Biometrika, 70, p. 41-55, 1983.

ROSENBAUM, P. R and RUBIN D. B.. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician, 39, p. 33-38, 1985.

ROSENBAUM, P. R. Optimal matching for observational studies. J. Amer. Statist. Assoc. 84, p. 1024-1032, 1989.

ROSZKOWSKI, M. J. and BEAN, A. G. Believe it or not! Longer Questionnaires Have Lower Response Rates, Journal of Business and Psychology, - Springer... 4, Summer, 1990.

RUBIN, D. B. Inference and Missing Data, Biometrika, 63, p. 581-592, 1976.

RUBIN, J. N. K. Multiple imputations in sample surveys: a phenomenological bayesian approach to non response , in Proceedings of the Survey Research methods Section, American Statistical Association, ,p.20-34, 1978.

RUBIN, D. B. Statistical matching using file concatenation with adjusted weights and multiple imputations. Journal of Business and Economic Statistics 4, 87-94, 1986.

RUBIN, D. B. Multiple imputation for no response in Surveys. John Wiley and Sons, New York, 1987.

RUBIN, D. B. and BELIN, T. R. Recent developments in calibrating error rates for computer matching. Proceedings of the 1991 Annual Research Conference, Bureau of the Census. Washington, D.C. Department of Commerce, 1991.

RUGGLES, N. and RUGGLES, R. A strategy for merging and matching micro data sets. Annals of Economic and Social Measurement, 3, 353-371, 1974.

RUGGLES, N., RUGGLES, R. and WOLFF, E. Merging micro data: rationale, practice and testing. Annals of Economic and Social Measurement, 6, 429-44, 1977.

RUGGLES, N.D. and RUGGLES, R. Macro- and micro-data analyses and their integration. Northampton, MA: Edward Elgar Publishing, 1999.

SAPORTA, G. and CO V. Fusion de fichiers: une nouvelle méthode basée sur l'analyse homogène. In G. Brossier and A. M. Dussaix (eds.) Enquêtes et sondages, 81-93. Dunod. ALONSO, 1999.

SÄRNDAL, C.E., SWENSSON, B. and WRETMAN, J. Model-Assisted survey sampling. New York: Springer-Verlag, 1992.

SCHAFER, J.L. Analysis of incomplete multivariate data. Chapman & Hall, London, 1997.

SCHEUREN, F. The social policy simulation database and model: an example of survey and administrative data integration. (includes a comment). Survey of Current Business, 40-41, 1989.

SCHEUREN, F. Multiple imputation: how it began and continues. The American Statistician, November 315-319, Vol. 59, No. 42005, 2005.

SEBER, G. A. F. Linear regression analysis. New York: Wiley, 1977.

SHOEMAKER, D. M. Principle and Procedures of Multiple Matrix Sampling. Ballinger, Cambridge, M.A, 1973.

SIMS, C. A. Comments on "constructing a new data base from existing micro data sets": "the 1966 merge file", by B. A. Okner. Annals of Economic and Social Measurement 1:343–345, 1972.

SINGH, A. C., MANTEL, H. J., KINACK, M. D. and ROWE, G. Statistical Matching: use of auxiliary information as an alternative to conditional independence assumption. Survey Methodology, 19(1): 59-79, 1993.

SINGH, A. C., MANTEL, H. J., KINACK, M. D. and ROWE, G. *apud* D'ORAZIO, M., DI ZIO, M. and SCANU, M. Statistical Matching: Theory and Practice, Wiley, New York, 2006, p. 39.

SUBCOMMITTEE ON MATCHING TECHNIQUES. Federal committee on Statistical Methodology. Statistical Policy Working Paper no. 5. Office of Federal Statistical Methodology Policy and Standards. Washington, D.C. : U.S. Department of Commerce, 1980.

SUTHERLAND, H., TAYLOR, R. and GOMULKA, J. Combining household income and expenditure data in policy simulations. Technical Report MU0101, Department of Applied Economics, University of Cambridge, 2001.

TEPPING, B. J. A Model for optimum linkage of records. Journal of the American Statistical Association, 63, p. 1321-1332, 1968.

UNITED NATIONS, System of National Accounts. United Nations, New York, 1993.

VAN DER LAAN, P., Integrating administrative registers and household surveys. Netherlands Official Statistics, 15, pp. 7-15, 2000.

VAN DER PUTTEN. P., KOK, J. N., B. and GUPTA, A. Data Fusion through statistical matching, MIT Sloan School of Management, Working Paper 4342-02, 2002.

VANTAGGI, B. The role of coherence for the integration of different sources. In F.G. Cozman, R. Nau and T. Seidenfeld (eds), Proceedings: 4th International

Symposium on Imprecise Probabilities and Their Applications, pp. 269 – 378. Pittsburgh: Brightdocs, 2005.

WASSERMAN, L. All of statistics. Springer, 2003.

WINKELS, J.W. and EVERAERS, P.C.J. Design of an integrated survey in the Netherlands. The case of POLS. Netherlands Official Statistics 13, 8 – 11, 1998.

WINKLER, W.E. Matching and record linkage. In G.P. Cox, D.A. Binder, B.N. Chinappa, A. Christianson, M. Colledge and P.S. Kott (eds), Business Survey Methods, pp. 355 – 384. New York: Wiley, 1995.

WOODBURRY, M. A. Statistical record-matching for files. P. 173-202 in Incomplete Data in Samples Surveys, vol. 3. W. G. Madow, H. Nisselson, and I. Olkin, eds. New York: Academic Press, 1983.

YAHAV, J. A. On matchmaking. P. 497-504 In Statistical Decision Theory and Related Topics III, 2. S. S. Gupta and J. O. Berger, eds. New York : Academic Press, 1982.

9
Apêndices

9.1. Apêndice A

Identificação do ínfimo e supremo

Seguindo uma notação usual:

O supremo e o ínfimo de dois números reais a e b, indicados por $a \vee b$ e $a \wedge b$, onde pode-se verificar que:

$$a \vee b = \frac{1}{2}(a + b + |a - b|)$$

$$a \wedge b = \frac{1}{2}(a + b - |a - b|)$$

Se a e b são números reais, teremos sempre:

$|a + b| \leq |a| + |b|$ e $\big||a| - |b|\big| \leq |a - b|$

Se f e g são funções reais, define-se a função máximo (supremo) de

f e g, anote $f \vee g$; e mínimo (ínfimo) de f e g, anote $f \wedge g$, peças fórmulas:

$\sup x_n = (f \vee g)(x) = f(x) \vee g(x)$

$\inf x_n = (f \wedge g)(x) = f(x) \wedge g(x)$

Caso $g = 0$, escrevemos f^+ no lugar de $f \vee g$, e f^- no lugar de $f \wedge g$. Então:

$$|f| = f^+ + f^-$$

e $$f = f^+ - f^-$$

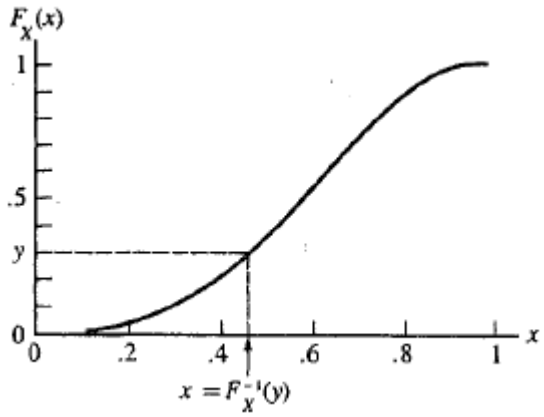
Temos: $$\sup\{f, g\} = \frac{1}{2}(f + g + |f - g|)$$

$$\inf\{f, g\} = \frac{1}{2}(f + g - |f - g|)$$

Sobre uma CDF sabe-se que:

$$\hat{F}_X^{-1}(y) = x \Leftrightarrow F_X(x) = y\}$$

$$\hat{F}_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}$$



9.2.Apêndice B

Três passos para transformações de variáveis continuas

Seja $W = r(X, Y)$ a transformação desejada, os passos para obter f_W são:

1. Para cada w, obter um conjunto A_w = { (x, y) : $r(x, y) \leq w$ }
2. Ache a fdc:

$$F_W(w) = \mathrm{P}(W \leq w) = \mathrm{P}(r(X, Y) \leq w)$$

$$= \mathrm{P}(W \leq w) = \mathrm{P}(\{(x, y); r(x, y) \leq w\}) = \int \int_{A_w} f_{X,Y}(x, y)dxdy$$

3. Então $f_W(w) = F'_W(w)$ em todos os pontos de x nos quais a F_W seja diferenciável.

9.3. Apêndice C

Código R

```
######
###### Funções para B doador e as adicionais quando A doador
################################################################


################################################
######  Separa A e B de XYZ

 cria_A <-function(x,a){
  {
    a  <- x [seq(1, nrow(x ), by = 2), ]
              ###### registros     Impares ordenados pelos Py

  a  <- data.frame(a[order(a[,4]),])

  }
    return(a)

 }
```

```
 cria_B <-function(x,b){
  {
    b  <- x [seq(2, nrow(x ), by = 2), ]
                    ###### registros Pares ordenados pelos Pz
    b  <- data.frame(b[order(b[,5]),])

  }
    return(b)
 }



    ################################################
    ########## Ordenados os Pz de B,para buscar o Pz mais perto
do PY + residuo_aleatorio (de A).

    Intervalo <- function (a,b)
    {
          nrowA   =    nrow(a)
          nrowB   =    nrow(b)
          b_order  <- data.frame(b[order(b[,3]),])
          Ind <- array(NA,dim=c(nrowA , 4 ))
          Ind[ ,1] <-  a[,4]
          Ind[ ,2] <-  b_order[,3]

    cont<-0
    for(i in 1:nrowA)
    {
      for(j in 1:nrowB)
      {
        if(Ind[i,1]>=Ind[j,2]){
         cont<-cont+1
      }
        Ind[i,3]<- cont
        Ind[i,4]<- cont+1
     }
        cont<-0
    }

    for(i in 1:nrowA){
```

```
      if(Ind[i,1]<=0){
        Ind[i,3] <- 1
        Ind[i,4] <- 1
      } else{
            if (Ind[i,1]>=1){
               Ind[i,3] <- nrowA
               Ind[i,4] <- nrowA
             }
      }

    }
    for(i in 1:nrowA){
      if (a [i , 4] < b_order[ 1 ,3 ]){
          Ind[i,3] <- 1
          Ind[i,4] <- 1
              }
      }
    for(i in 1:nrowA){
      if (Ind[i,3] == nrowA){
            Ind[i,4] <- nrowA
              }
      }
     Ind
     }




    ################################################
    ######  calcula-se  o  Z_estimado  (invertendo  o  cálculo  dos
"percentis" - Pz_estimados).

     Inverte_Interpolacao <- function (a, b, ind) {
       nrowI   =    nrow(ind)
       nrowA   =    nrow(a)
       nrowb   =    nrow(b)
       intervalo <-  array(NA,dim=c(nrowA  ,2))
       intervalo [ ,1] <-  ind[,3]
       intervalo [ ,2] <-  ind[,4]

       x <-  array(NA,dim=c(nrowA  ,2))
```

```
        k = 0
        b_order  <- data.frame(b[order(b[,3]),])

      for(i in 1: nrowA ) {

                      k = intervalo[ i, 1]
                      x [i ,1 ] = k
                      inf <- b_order[intervalo[ i, 1], 2]
                      sup <- b_order[intervalo[ i, 2], 2]
                      ampZ <- sup - inf

                      p_inf  <-  b_order[intervalo[  i,  1],3  ]
### valores do Pz inferior e superior
                      p_sup <- b_order[intervalo[ i, 2],3 ]
                      ampPz  <- p_sup - p_inf

                      w <- a[i , 4]  - p_inf

       {if (ampPz != 0)
                      {
                       x [i ,2 ] <- inf + (( ampZ * w)/ ampPz )
                      }
        else {  x [i ,2 ] <- b_order[intervalo[ i, 1] ,2 ]
          }
      }

     if (( a [i , 4] <= 0) | (a [i , 4] < b_order[ 1 ,3 ])) {  x
[i ,2 ] <- b_order[ 1 ,2 ]}


     }
      x
       }
```

```
     ###########  calcula-se o Pz_estimado
     ############  e verificando quem é o observado mais proximo
no intervalo onde está Pz_estimado).

     Inverte_Intervalar <- function (a,b,ind) {
        nrowI   =    nrow(ind)
        nrowA   =    nrow(a)
        nrowb   =    nrow(b)
        intervalo <-  array(NA,dim=c(nrowA  ,2))
        intervalo [ ,1] <-  ind[,3]
        intervalo [ ,2] <-  ind[,4]
        nrow_IND  =    nrow(intervalo)

        x <-  array(NA,dim=c(nrowA  ,2))
        b_order  <- data.frame(b[order(b[,3]),])

      for(i in 1: nrowA ) {

        anterior = intervalo[ i, 1]
     if (intervalo[ i, 1]<nrowA) { posterior = intervalo [ i, 2] }
           x [i ,1 ] = anterior

         if ( (i < nrowb) )
           {
             if  ( (abs(a  [i  ,  4]  -  b_order[anterior,  3 ])) <=
(abs(a [i , 4] - b_order[posterior,3 ])) )
              { x [i ,2 ] <- b_order[anterior , 2 ] }

            else { x [i ,2 ] <- b_order[posterior, 2 ] }
           }
       else
            {  x [i ,2 ] <- b_order[anterior ,2 ]}
         }
         x
       }
```

```
    ##############################################################
#####
    Soma_residuo_e_percentil <- function (a,xyz,residuo )
    {
       nrowA <- nrow(a)
       ind_res_aleat                                              <-
data.frame(trunc(runif(nrow(a))*(nrow(xyz)- 1)) + 1)
       res_aleat <- data.frame(residuo[ind_res_aleat[ ,1], ])

    C <-  array(NA,dim=c(nrowA ,3 ))
    C[,1] <-  a[,3]                              ### Py
    C[,2] <-  res_aleat[,2]                  ### Residuo


    for(i in 1:nrowA)  {

    Flag <- 0
    if ( (a[i,3]== 0)  )
         { C[i,3] <-  a[i,3] + C[i,2]
           Flag <- 1  }

    while (Flag == 0){
    if  (((a[i,3]>=  0)  &  (a[i,3]<=  0.2)  &  (C[i,2]  >=  (-
0.75*a[i,3])) & (C[i,2] <= (3*a[i,3])))  |
       ((a[i,3]>=  0.2)  &  (a[i,3]<=  0.8)  &  (C[i,2]  >=  (-
0.75*a[i,3])) & (C[i,2] <= (0.75 - 0.75*a[i,3]))) |
       ((a[i,3]>=  0.8)  &  (a[i,3]<=  1)  &  (C[i,2]  >=  (-3  +
3*a[i,3])) & (C[i,2] <= (0.75 - 0.75*a[i,3]))) )
      { C[i,3] <-  a[i,3] + C[i,2]
        Flag <- 1
      }
       else { ind_res_aleat <- trunc((runif(1,0,1)*(nrow(xyz)- 1))
+ 1)
              C[i,2] <- residuo[ind_res_aleat, 2 ]


        }
      }
    }
    for(i in 1:nrowA)  {
```

```
if ( (C[i,3] < 0)  )
     { C[i,3] <-  0 }

if ( (C[i,3] > 1)  )
     { C[i,3] <-  1 }
}
C
}


###########
############  Interpolação com os arqquivos B e A doadores


   Sintese_BA_AB_Interpolacao <- function (A, B, XYZ, Residuo
)
{
    nrowA <-  nrow(A)
    nrowXYZ <-  nrow(XYZ)
    nrowB <-  nrow(B)
    nrowR <-  nrow(Residuo)

                      ### 1. Gera número ALEAT e residuo ALEAT

   C  <-  Soma_residuo_e_percentil(A, XYZ,Residuo)

  A <- data.frame(cbind(A,C[,3]))
  colnames(A )<- c( "X","Y","Py","Py + res_aleat")


 ###########
 ### Intervalo P_anterior e P_posterior)

   Intervalo <- Intervalo(A, B)
                                      ### - Interpolação
   P_est_obs_A  <- Inverte_Interpolacao(A,B, Intervalo)
   Sintese <- data.frame(cbind(A[ ,1:2],P_est_obs_A [ ,2]))

 Sintese
}
```

```
#####################


     Sintese_BA_AB_Intervalo <- function (A, B, XYZ, Residuo
)
{

    nrowA <-  nrow(A)
    nrowA <-  nrow(XYZ)
    nrowB <-  nrow(B)
    nrowR <-  nrow(Residuo)

     ### 1. Gera número ALEAT e residuo ALEAT

     C  <-  Soma_residuo_e_percentil(A, XYZ,Residuo)

  A <- data.frame(cbind(A,C[,3]))
  colnames(A )<- c( "X","Y","Py","Py + res_aleat")


                      ### Intervalo P_anterior e P_posterior)

   Intervalo <- Intervalo(A, B)
                               ### - Intervalar
   P_est_obs_A  <- Inverte_Intervalar(A,B, Intervalo)
   Sintese <- data.frame(cbind(A[ ,1:2],P_est_obs_A [ ,2]))

 Sintese
}
```

9.4.Apêndice D

As distribuições marginais são preservadas conforme mostradas nas tabelas 13 e 14, onde B e A , são os respectivos receptores.

***B* doador - Transformação não-paramétrica relacional**							
		TNRIr			***TNRIo***		
Valores originais			***Viés***	**EQM**		**Viés**	**EQM**
cor(X,Y)	**0,580047**	**0,580736**	0,000690	0,000000	**0,580736**	0,000690	0,000000
cor(X,Z)	**0,489605**	0,405740	0,083865	0,007033	0,403536	0,086070	0,007408
cor(Y,Z)	**0,735542**	0,743708	0,008167	0,000067	0,743833	0,008291	0,000069

Tabela 13 - Estimativas das correlações do emparelhamento estatístico *B* doador.

***A* doador - Transformação não-paramétrica relacional**							
		TNRIr			***TNRIo***		
Valores originais			***Viés***	**EQM**		***Viés***	**EQM**
cor(X,Y)	**0,580047**	0,470259	0,109788	0,012053	0,469120	0,110927	0,012305
cor(X,Z)	**0,489605**	**0,484469**	0,005136	0,000026	**0,484469**	0,005136	0,000026
cor(Y,Z)	**0,735542**	0,775334	0,039792	0,001583	0,776041	0,040499	0,001640

Tabela 14 - Estimativas das correlações do emparelhamento estatístico *A* doador.