

Povzetek

Vstavljanje je postopek, pri katerem manjkajoče oziroma napačne vrednosti v statističnih anketah nadomestimo s primerno vrednostjo. V zadnjem času je vstavljanje vse bolj priljubljeno, saj lahko z uporabo metod vstavljanja precej zmanjšajo stroške pri zbiranju podatkov. V tem delu so predstavljene tako metode enkratnega kot tudi večkratnega vstavljanja. V prvem delu tega besedila se posvetimo metodam enkratnega vstavljanja, kot so: deterministična metoda, družina hot-deck metod, vstavljanje s povprečjem, razmernostno vstavljanje in metode regresije. Drugi del začnemo z osnovami Bayesove statistike. Sledi predstavitev Bayesovega pristopa k vstavljanju in osnovnih načel večkratnega vstavljanja. Zaključimo pa s predstavitvijo augmentacije podatkov, ki spada med metode MCMC sedaj najbolj popularne statistične tehnike.

Math. Subj. Class. (MSC 2000): 62J05, 62F15

Ključne besede:

vstavljanje, deterministično vstavljanje, hot-deck metode, linearna regresija, Bayesovo sklepanje, model manjkajočih podatkov, večkratno vstavljanje

Keywords:

imputation, deterministic imputation, hot-deck methods, linear regression, Bayes' inference, missing data model, multiple imputation

Literatura

- [1] J. Bernier, D. Haziza, K. Nobrega, P. Whitridge, *Handling Missing Data*, <http://www.math.yorku.ca/Who/Faculty/Ng/ssc2002/MissingMain.htm>, 2002.
- [2] J. Carpenter, *Statistical modelling with missing data using multiple imputation*, London School of Hygiene & Tropical Medicine, November 2009.
- [3] J. Carpenter, M. Kenward, *Missing Data*, <http://www.missingdata.org.uk/>, 2005.
- [4] Spletna enciklopedija, *Wikipedia (ang)*, [http://en.wikipedia.org/wiki/\(Mehalanobis_distance, /Taxicab_geometry, /Jaccard_index, /Imputation_\(statistics\), /Linear_regression, /Least_squares, /Regression_analysis#Linear_regression, /Bayesian_method, /Bayesian_linear_regression, /Conjugate_prior, /Scaled-inverse-chi-square_distribution, /Multivariate_normal_distribution\)](http://en.wikipedia.org/wiki/(Mehalanobis_distance,_/Taxicab_geometry,_/Jaccard_index,_/Imputation_(statistics),_/Linear_regression,_/Least_squares,_/Regression_analysis#Linear_regression,_/Bayesian_method,_/Bayesian_linear_regression,_/Conjugate_prior,_/Scaled-inverse-chi-square_distribution,_/Multivariate_normal_distribution)).
- [5] Spletna enciklopedija, *Wikipedia (slo)*, [http://sl.wikipedia.org/wiki/\(Linearna_regresija, /Študentova_t_porazdelitev, /Porazdelitev_hi-kvadrat\)](http://sl.wikipedia.org/wiki/(Linearna_regresija,_/Študentova_t_porazdelitev,_/Porazdelitev_hi-kvadrat)).
- [6] I. P. Fellegi, D. Holt, *A Systematic Approach to Automatic Edit and Imputation*, Journal of the American Statistical Association **71** (1976), 17–35.
- [7] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, *Bayesian Data Analysis (second edition)*, Chapman & Hall/CRC, 2004.
- [8] P. Giles, *A model for generalized edit and imputation of survey data*, The Canadian Journal of Statistics **16** (1988), 57–73.

-
- [9] O. Harel, X. H. Zhou, *Multiple imputation - Review of theory, implementation and software*, <http://www.bepress.com/uwbiostat/paper297>, 2006.
 - [10] L. V. Homecillo, *Treatment of Missing Data in the Annual Survey of Establishments*, 9th National Convention on Statistics (NCS), 2004.
 - [11] U. Jensen, H. Gartner, S. Rässler, *Measuring overeducation with earnings frontiers and multiply imputed censored income data*, IABDiscussionPaper **11** (2006).
 - [12] J. Jesenko, *Multivariatne statistične metode*, Moderna organizacija, 2007.
 - [13] C. T. Kelley, *Iterative methods for optimization*, Siam, 1999.
 - [14] S. Laaksonen, S. Rässler, C. Skinner, *Imputation and Non-Response*, DACSEIS IST-2000-26057, Workpackage 11, Deliverable 11.2, 2004.
 - [15] M. Lavine, *What is Bayesian statistics and why everything else is wrong*, ISDS, Duke University, Durham, North Carolina, 2000.
 - [16] J. A. Rice, *Mathematical Statistic and Data Analysis (Second Edition)*, Duxbury Press, 1999.
 - [17] D. B. Rubin, *Multiple Imputation for Nonresponse in Surveys*, John Wiley, 1987.
 - [18] J. L. Schafer, *The multiple imputation FAQ page*, <http://www.stat.psu.edu/~jls/mifaq.html>, 1998.
 - [19] J. L. Schafer, *Multiple Imputation: a Primer*, Statistical Methods in Medical Research **9** (1999), 3–15.
 - [20] J. L. Schafer, M. K. Olsen, *Multiple imputation for multivariate missing-data problems: a data analyst's perspective*, The Pennsylvania State University, March 1998.
 - [21] D. Schunk, *A Markov Chain Monte Carlo Algorithm for Multiple Imputation in Large Surveys*, Advances in Statistical Analysis **92** (2008), 101–114.
 - [22] M. M. Seastrom, *NCES Statistical Standards*, <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003601>, 2002.

-
- [23] R. Seljak, *Statistično urejanje podatkov: Osnovni pojmi in metode*, Statistični Urad RS, 2005.
- [24] R. Seljak, *Banff-aplikacija za avtomatsko urejanje podatkov*, Statistični Urad RS, 2007.
- [25] M. A. Tanner, W. H. Wong, *The Calculation of Posterior Distributions by Data Augmentation*, Journal of the American Statistical Association **82** (1987), 528–540.
- [26] L. Wasserman, *All of nonparametric statistics*, Springer, 2007.