Lee, Shen-Ming; Lukusa, T. Martin; Li, Chin-Shang

Estimation of a zero-inflated Poisson regression model with missing covariates via nonparametric multiple imputation methods. (English) Zbl 07225402

Summary: Zero-inflated Poisson (ZIP) regression is widely applied to model effects of covariates on an outcome count with excess zeros. In some applications, covariates in a ZIP regression model are partially observed. Based on the imputed data generated by applying the multiple imputation (MI) schemes developed by D. Wang and S. X. Chen [Ann. Stat. 37, No. 1, 490–517 (2009; Zbl 1155.62021)], two methods are proposed to estimate the parameters of a ZIP regression model with covariates missing at random. One, proposed by Rubin (in: Proceedings of the survey research methods section of the American Statistical Association, 1978), consists of obtaining a unified estimate as the average of estimates from all imputed datasets. The other, proposed by R. E. Fay [J. Am. Stat. Assoc. 91, No. 434, 490–498 (1996; Zbl 0869.62015)], consists of averaging the estimating scores from all imputed data sets to solve the imputed estimating equation. Moreover, it is shown that the two proposed estimation methods are asymptotically equivalent to the semiparametric inverse probability weighting method. A modified formula is proposed to estimate the variances of the MI estimators. An extensive simulation study is conducted to investigate the performance of the estimation methods. The practicality of the methodology is illustrated with a dataset of motorcycle survey of traffic regulations.

MSC:
65C60 Computational problems in statistics (MSC2010)

Keywords:
zero-inflated Poisson regression; nonparametric multiple imputation; inverse probability weighting (IPW); missing at random; count data

Full Text: DOI

References:

[14] Horvitz, DG; Thompson, DJ, A generalization of sampling without replacement from a finite universe, J Am Stat Assoc, 47, 663-685 (1952) · Zbl 0047.38001
[15] Hsieh, SH; Lee, SM; Shen, PS, Semiparametric analysis of randomized response data with missing covariates in logistic
Johnson, NL; Kemp, AW; Kotz, S., Univariate discrete distributions (2005), New York: Wiley, New York
Lambert, D., Zero-inflated Poisson regression, with an application to defects in manufacturing, Technometrics, 34, 1-14 (1992)  · Zbl 0850.62756
Lee, JH; Han, G.; Fulp, WJ; Giuliano, AR, Analysis of overdispersed count data: application to the human papillomavirus infection in men (HPM) study, Epidemiol Infect, 140, 1087-1094 (2012)
Lee, SM; Li, CS; Hsieh, SH; Huang, LH, Semiparametric estimation of logistic regression model with missing covariates and outcome, Metrika, 75, 621-653 (2012)  · Zbl 1362.60108
Lee, SM; Hwang, WH; Tapsoba, JD, Estimation in closed capture-recapture models when covariates are missing at random, Biometrics, 72, 1294-1304 (2016)  · Zbl 1390.62280
Li, CS, Score test for semiparametric zero-inflated Poisson model, Int J Stat Probab, 1, 1-7 (2012)
Little, RJA, Regression with missing X's: a review, J Am Stat Assoc, 87, 1227-1237 (1992)
Lu, SE; Lin, Y.; Shih, WCI, Analyzing excessive no changes in clinical trials with clustered data, Biometrics, 60, 257-267 (2004)  · Zbl 1130.62376
Lukusa, TM; Lee, SM; Li, CS, Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates, Metrika, 79, 457-483 (2016)  · Zbl 1349.62137
Paelinck, BT; Preisser, JS; Stearns, SC; Rozier, RG, Multiple imputation of dental caries data using a zero-inflated Poisson regression model, J Public Health Dent, 71, 71-78 (2011)
Reilly, M.; Pepe, MS, A mean score method for missing and auxiliary covariates data in regression methods, Biometrika, 82, 299-314 (1995)  · Zbl 0828.62097
Robins, JM; Rotnitzky, A.; Zhao, LP, Estimation of regression coefficients when some regressors are not always observed, J Am Stat Assoc, 89, 846-866 (1994)  · Zbl 0815.62011
Rubin, DB, Inference and missing data, Biometrika, 63, 581-592 (1976)  · Zbl 0349.62034
Rubin, DB, Multiple imputation for nonresponse in surveys (1987), New York: Wiley, New York
Rubin, DB; Schenker, N., Multiple imputation for interval estimation from simple random samples with ignorable nonresponse, J Am Stat Assoc, 81, 366-374 (1986)  · Zbl 1015.62011
Schafer, JL; Graham, JW, Missing data: our view of the state of the art, Psychol Methods, 7, 147-177 (2002)
Singh, S., A note on inflated Poisson distribution, J Indian Stat Assoc, 1, 140-144 (1963)
Wang, CY; Wang, S.; Zhao, LP; Ou, ST, Weighted semiparametric estimation in regression with missing covariate data, J...
[51] Wang, CY; Chen, JC; Lee, SM; Ou, ST, Joint conditional likelihood estimator in logistic regression with missing covariate
data, Stat Sin, 12, 555-574 (2002) · Zbl 1137.62330

[52] Xiang, L.; Lee, AH; Yau, KKW; McLachlan, GJ, A score test for overdispersion in zero-inflated Poisson mixed regression

[53] Yau, KKW; Lee, AH, Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically
matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original
paper as accurately as possible without claiming the completeness or perfect precision of the matching.