Correspondence Analysis and Perturbation of Chi-Squared Value

Masaaki Ida*

*National Institution for Academic Degrees and University Evaluation, 1-29-1 Gakuen-nishimachi, Kodaira, Tokyo, Japan

ida@niad.ac.jp

Abstract— **Correspondence analysis is a popular research technique in this Big Data era in order to analyze correspondence relation between data in multiple categories expressed by cross tabulation. Correspondence analysis is frequently utilized for text mining and for analyzing questionnaire survey with mutually exclusive choices, which is also applicable as a visualization technique for data mining. However, data addition or data fluctuation is one of the inevitable issues for correspondence analysis in practical situation. We have so far considered perturbation of ordinal and multiple correspondence analysis. These considerations has focused on perturbation of each element of cross tabulation. However, consideration on the whole influence of perturbation in cross tabulation due to data variation is required. This consideration is closely related to chi-squared value of test statistic. This article presents mathematical consideration and application example for perturbation of chi-squared value of cross tabulation.**

***Keyword*—Correspondence analysis, Perturbation, Chi-squared value, Data mining, Questionnaire analysis.**

Author Masaaki Ida was received the B.E., M.E, and Ph.D. degrees, all in engineering, from Kyoto University, Kyoto, Japan. He is currently a professor of National institution for Academic Degrees and University Evaluation of Japan. His current research interests are in computer and system science, especially in advanced database and communication technology in higher education.