

Improving the Power of Outlier Detection through Reasonably Choosing Ratio Composition

Cheng Bangwen^{1*}, Yang Hongjin¹, Shi Linfen¹, Wang Yali¹, He Rong² and Xu Ji¹

¹School of Management, Huazhong University of Science and Technology, Wuhan, China

²School of Management, Wuhan Textile University, Wuhan, China

In socio-economic statistics, the outlier detection based on ratio edits is widely used and play an important role when detecting the abnormal points in survey data from basic statistical units. Socio-economic survey data are multi-dimensional, therefore, a statistical item can be tested in many ratio edits by selecting different denominators. Different ratio edits have different power of detecting outliers. Obviously, studying how to reasonably choose an appropriate denominator to improve the detection power is quite significant. Besides, it is impossible and unnecessary to test survey data in all possible ratio edits and we need to choose those important and efficient ones from many ratio edits. The paper discusses the power of ratio edits in detecting outliers and the way to improve the power by reasonably choosing ratio composition.

At present, the study on ratio edits is mainly focused on the design of ratio statistics, the determination of the upper and lower bounds of tolerances and the comparison of different ratio edit approaches, while there is little research on how to reasonably choose the ratio composition. Generally, socio-economic data are highly right skewed, but after a power transformation, they usually can be turned into a normal distribution. It is the aim of this paper to discuss the power of outlier detection by ratio edits under the situation that the two variables of ratios have marginal correlated normal distribution.

The study shows the power of ratio edits is mainly determined by the variation coefficients of ratio variables and the power can be improved if an appropriate denominator is chosen which will make the variation coefficient of the ratio variable be small, and when a certain condition is not satisfied the ratio test result will become worse. Reasonably selecting ratio's denominator is the key to effectively detect outliers. Firstly, the correlation coefficient between the numerator and the denominator should be as large as possible. Secondly, the variation coefficient of the denominator should be less than the one of the numerator, but the difference between them should not be too large. Besides, the critical conditions with which we can judge if the detection power is improved and the improved degree are given, which is useful for judging if the composition of ratio variables is reasonable.

Keywords: quality of statistical data; ratio edit; power of outlier detection;

Taylor approximation of the variance of the ratio variable