

Rungrawee Amnarttrakul 2012: Detection of Outliers for Multivariate Data Using the Minor Principal Components. Doctor of Philosophy (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Mrs. Ampai Thongteeraparp, Ph.D. 87 pages.

The objective of this study is to propose test statistics to detect outliers for multivariate data using the minor principal components when the outliers do not conform with the correlation structure of the remainder of the data. After studying principals and theories on principal component analysis and properties of normal distribution, the test statistics R_{2z}^2 and R_{3z}^2 , which use only 2 and 3 minor principal components respectively, were found. The study shows that the test statistics are Chi-square distribution with degree of freedom equal to 1. The comparison of the efficiency of the test statistics for detection of outliers in multivariate data with $d_{2i,(2)}^2$ and $d_{2i,(3)}^2$, which are proposed by Hawkins, and d^2 , which comes from Mahalanobis distance is tested by generating data through simulation technique using multivariate normal distribution under the condition of number of variables $p = 5, 10(2)20$ with a sample size of $n = 30, 40, 50, 75, 100$ and 4 levels of percentage of outliers 10%, 20%, 30% and 40%. There are 140 situations in total and each situation can be repeatedly simulated for 1,000 times.

The study shows that R_{2z}^2 is suitable for detection of outliers with $p \geq 16$ and $n \geq 50$. The efficient of R_{3z}^2 is lower than other test statistics in many situations, so it is not suitable to detect the outliers. The test statistic $d_{2i,(2)}^2$ is suitable for detection of the outliers with $p \leq 10$, whereas $d_{2i,(3)}^2$ is more suitable for detection of outliers with $12 \leq p \leq 16$. As for test statistic d^2 is suitable for detection of outliers with $p \geq 14$. However, it is discovered that all statistics are not much different, and it is found that R_{2z}^2 and R_{3z}^2 are not complicated and easily computed.

Student's signature

Thesis Advisor's signature