

## The System for Checking Multivariate Normality and Outliers

강명래<sup>1)</sup>, 최용석<sup>2)</sup>

### 요약

다변량분석 기법을 사용하기 위해서는 자료가 정규성(normality)가정을 만족해야 한다. 본 연구에서는 GUI(graphic user interface)환경 하에서 일변량(univariate)과 다변량자료(multivariate data)의 정규성검정, 이상치(outliers)제거 및 변수변환(variable transformation)을 지원하는 시스템을 구축하여 사용자들이 보다 편리하게 사용할 수 있음을 소개 하고자 한다.

주요용어 : 다변량 정규성검정, 이상치 제거, Visual Basic

### 1. 서론

실생활에서 우리가 접하는 대부분 자료(data)들은 하나 이상의 변수들로 이루어진 다변량자료이고 다변량 분석기법(인자분석, 주성분분석, 판별분석 등)을 요구하는 자료이다. 이 다변량 분석기법을 사용하기 위해서는 자료가 정규성 가정을 만족하여야 하며, 이 가정을 만족한다면 우리는 정규분포의 유용한 점들을 이용하여 편리하게 분석을 할 수 있다(Johnson and Wichern, 1992, p. 157).

기존의 통계 패키지(SAS, SPSS등)에서는 일변량 자료에 대한 정규성검정은 지원하지만 다변량 자료에 대한 정규성검정을 지원하는 프로시저(procedure)가 없어 통계전문가가 아니면 다변량자료의 정규성검정을 할 수 없었고, 통계전문가 또한 정규성가정을 경시하고 분석하여 통계적 오류를 야기시키는 경우가 많았다.

본 논문은 일변량 자료에 대한 정규성검정으로 Q-Q(quantile-quantile) plot, 왜도(skewness) 및 첨도(kurtosis)에 의한 방법과 샤피로-윌크 검정(Shapiro-Wilk test)을 소개하며, 다변량자료의 정규성검정으로 카이제곱(chi-square) plot과 왜도 및 첨도에 의한 방법을 소개한다. 더불어 정규성가정을 만족하지 않는 자료에 대해서는 마할라노비스거리(mahalanobis distance)를 이용한 방법과 Wilk's 통계량을 이용한 방법으로 이상치를 제거하며 여러가지 변수변환을 통하여 정규성가정을 만족하도록 하는 시스템을 구축하여 그 주된 구성과 예를 소개하고자한다.

### 2. 시스템의 구성

이 시스템은 Visual Basic 언어를 사용하여 사용자들이 GUI환경 하에서 편리하게 사용할 수 있도록 하였으며, 자료를 처리하는 순서 아래와 같다.

---

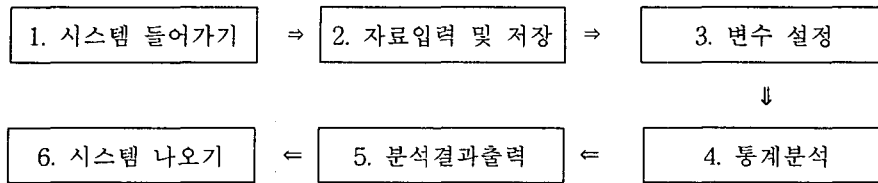
1) 부산대학교 통계학과 대학원 석사과정

2) 부산대학교 통계학과 부교수 <yschoi@hyowon.pusan.ac.kr>

609-735, 부산광역시 금정구 장전동 산 30번지

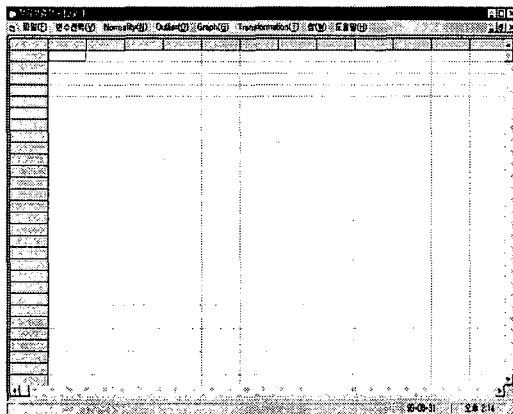
Home Page : <http://home.pusan.ac.kr/~yschoi>

The System for checking Multivariate Normality and Outliers

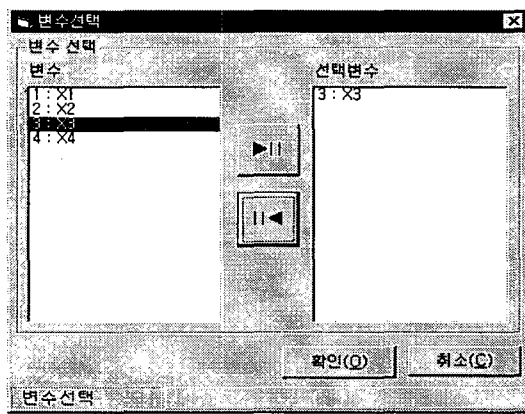


3. 시스템의 예

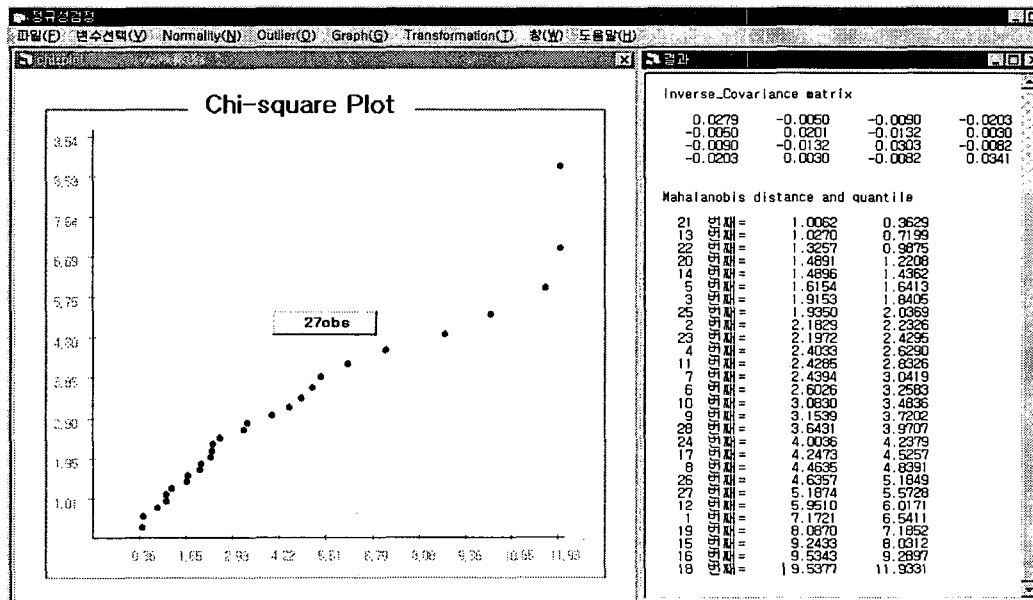
시스템의 주화면은 파일, 변수선택, 정규성검정, 이상치제거, 변수변환, 그래프, 창, 도움말로 되어있으며 풀다운(pull down) 형식의 메뉴로 구성되어있다.



<그림 1> 주화면



<그림 2> 변수선택 화면



<그림 3> 통계분석 결과창

<그림 1>의 주화면으로부터 자료를 입력받거나 저장을 할 수 있고, <그림 2>의 변수선택 화면에서와 같이 입력받은 변수중 분석에 사용할 변수를 선택한다. 그후 풀다운 메뉴에서 정규성 검정, 이상치 제거, 변수변환중 필요한 통계분석을 선택하면 아래의 <그림 3>의 통계분석 결과와 같은 결과 창을 얻을 수 있다.

#### 4. 결론

본 논문은 다변량자료의 정규성검정을 여러 가지 방법으로 검정할 수 있게 하였으며, 윈도우 환경에 많이 익숙해져 있는 일반인들을 위하여 GUI환경 하에서 클릭(click)만으로 다변량자료의 정규성 검정을 할 수 있도록 하였다.

#### 참고문헌

- [1] Johnson, R. A. and Wichern, D. W. (1992) Applied Multivariate Statistical Analysis. Prentice-Hall, New York.
- [2] Jobson, J. D (1992) Applied Multivariate Data Analysis. Springer-Verlag, New York.
- [3] Rencher, A. C. (1995) Methods of Multivariate Analysis. John Wiley & sons, New York.
- [4] Barnett, V. and Lewis. T. (1982) Outlier in Statistical Data. John-Wiley & Sons, New York.
- [5] Shammas, N. C. (1996) Mathematical Algorithms in Visual Basic for Scientists & Engineers. McGraw-Hill, Inc.

※ 본 시스템은 해상도 1024×768 화면을 기준으로 구축하였다.