

델파이를 이용한 회귀진단 툴 개발

현미진¹ · 박진표² · 박희창³

요약

본 논문에서는 시각화된 회귀진단 툴을 소개하고자 한다. 시각화된 회귀진단 툴은 회귀진단에 필요한 여러 가지 통계량들과 다양한 잔차그림을 메뉴방식에 의해서 쉽게 구할 수 있게 하였으며, 또한 프로그램을 한글화 하여 일반 사용자들이 보다 쉽게 사용할 수 있도록 개발된 프로그램이다. 특히 다양한 잔차그림을 쉽게 그릴 수 있게 하여 데이터의 분석결과를 시각화할 수 있도록 하므로써 그래프를 통한 회귀분석 교육에 활용하고자 한다. 이 툴을 개발하기 위해서 시각화 언어인 델파이 3.0을 사용하였다.

주제어 : 델파이, 회귀진단, 통계그래픽스, 객체지향언어.

1. 서론

회귀분석(Regression Analysis)은 어떤 현상을 규명하기 위하여 종속변수와 독립변수들 간의 관계를 연구하는데 널리 사용되고 있는 통계적 분석기법 중 하나이다. Box(1980)는 통계적 모형 구축 과정을 다음과 같이 설명하였다. (1) 문제 설정 및 데이터수집, (2) 모형 선정 및 가정 수립, (3) 모형 추정 및 모수에 대한 추론과 검정 그리고 (4) 선정된 모형과 가정에 대한 진단으로 설명하였다. 위에서 과정 (1)(2)(3)은 설정한 모형과 가정들이 만족한다는 가정하에서 이루어진다. 그러나 실제로 데이터 분석에서 설정한 모형과 가정은 만족되지 않는 경우가 많다. 따라서 통계적 모형구축에서 선정된 모형과 가정에 대한 진단 과정은 중요하다. 회귀분석에서도 분석에 사용된 회귀모형 및 가정에 문제점이 있는지를 검토하고, 각각의 관측값이 모형 및 가정에 어떠한 영향을 미치는지를 진단하는 과정인 회귀진단 (regression diagnostics)은 중요한 과정이다.

일반적인 다중회귀모형은 $y = \mathbf{X}\beta + \epsilon$, $\epsilon \sim \mathbf{N}(0, \sigma^2\mathbf{I})$ 이다. 즉 이 모형에 대한 기본가정은 (1)선형성, (2)등분산성, (3)정규성 그리고 (4)독립성이다. 회귀진단을 통해서 이 모

¹(631-701) 경남 마산시 합포구 월영동 449번지, 경남대학교 컴퓨터공학과 대학원

²(631-701) 경남 마산시 합포구 월영동 449번지, 경남대학교 컴퓨터공학과 부교수

³(641-773)경남 창원시 사림동 9, 창원대학교 통계학과 부교수

형과 기본가정이 타당한지를 진단하고, 또한 각 관측값과 각 변수가 모형의 결정에 미치는 영향을 검토한다.

최근 컴퓨터 환경의 발달로 그래프를 쉽게 그릴 수 있게 되자 많은 통계학자들이 통계 그래픽스를 이용한 데이터분석 방법을 연구하고 있다. 특히 Cook과 Weisberg(1994)는 회귀분석 시스템인 R-code를 Xlisp-Stat를 응용하여 개발하였다. 또한 국내 많은 학자들이 한국형 통계패키지를 개발·연구하고 있다. 이정진과 강근석(1994)은 한국형 통계패키지를 개발·연구하였고, 이정진과 강근석(1995) 등은 전문가용 한국형 통계패키지 개발·연구하였다. 조신섭(1996) 등은 엑셀을 이용한 시계열분석 툴을 개발하였고, 이외에도 많은 학자들이 한글화된 통계패키지를 만들려는 노력을 하고 있다. 그러나 이러한 소프트웨어들은 회귀분석에서 아주 중요한 과정인 회귀진단 부분은 깊이 있게 취급하지 않았다.

본 연구에서는 통계전문가를 위한 회귀진단 툴의 개발보다는 비전문가인 일반 사용자들이 보다 쉽게 접근할 수 있는 회귀진단 툴을 개발하여 회귀분석 교육에 활용하고자 한다. 이를 위해 회귀진단에 필요한 여러 가지 통계량을 메뉴에 의해서 쉽게 구할 수 있고, 또한 다양한 잔차그림을 메뉴를 이용해서 쉽게 그릴 수 있도록 하여 분석의 결과를 시각화 하므로써 결과 해석을 용이하게 하였고, 또한 한글로 구현된 새로운 형태의 회귀진단 교육용 툴을 개발하였다. 개발에 사용한 언어는 프로그램을 시각화 하는데 편리하고, 비주얼베이직이나 파워빌더 보다 실행파일의 실행속도도 빠르고, 컴파일 속도도 빠른 델파이 언어를 사용하였다.

2. 회귀진단 툴의 구현

델파이는 오브젝티브 파스칼 언어에 기반을 둔 객체 지향적인 언어로써 프로그램을 시각화 하기에 편리한 언어이다. 또한 델파이는 다음과 같은 장점을 가지고 있다. 델파이는 컴파일 방식을 채택하였기 때문에 비주얼 베이직이나 파워빌더보다 실행파일의 실행속도가 빠르고, 컴파일 속도가 빠르며, 한글 사용에 별로 제한이 없어서 프로그램을 한글화 하는데 편리하고, 별도의 DLL이나 라이브러리가 필요없이 실행파일만 있으면되므로 프로그램사용자의 입장에서 볼 때 따로 델파이를 설치할 필요없이 개발된 프로그램을 편리하게 사용할 수 있다. 이러한 장점 때문에 본 논문에서는 델파이 3.0을 이용하여 회귀진단 툴을 개발하였다.

개발된 회귀진단 툴의 구조는 다음과 같다.

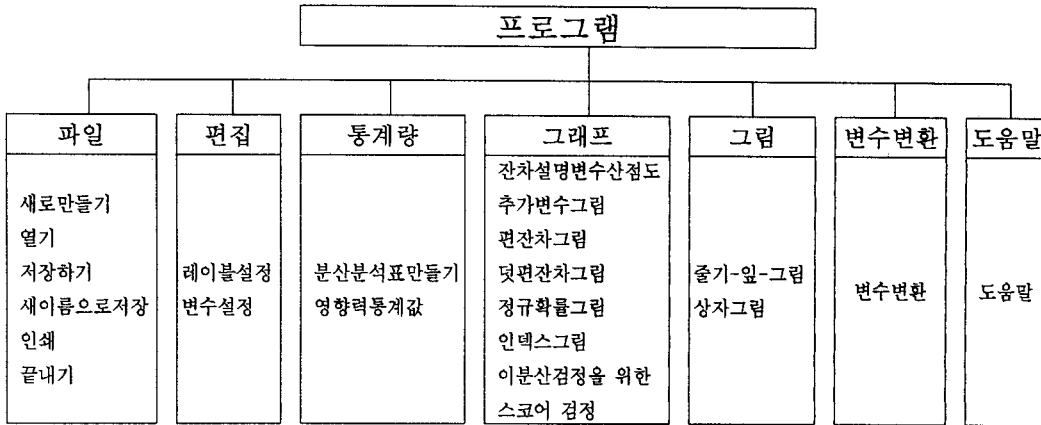


그림 2-1 프로그램의 구조

각 메뉴에 대한 기능은 다음과 같다.

2.1 초기화면

초기화면은 아래 그림 (그림 2-1)과 같이 스프레드 시트 형태처럼 외부 데이터 화일을 받아들이거나 직접 입력한 후에 프로그램을 실행할 수 있다.

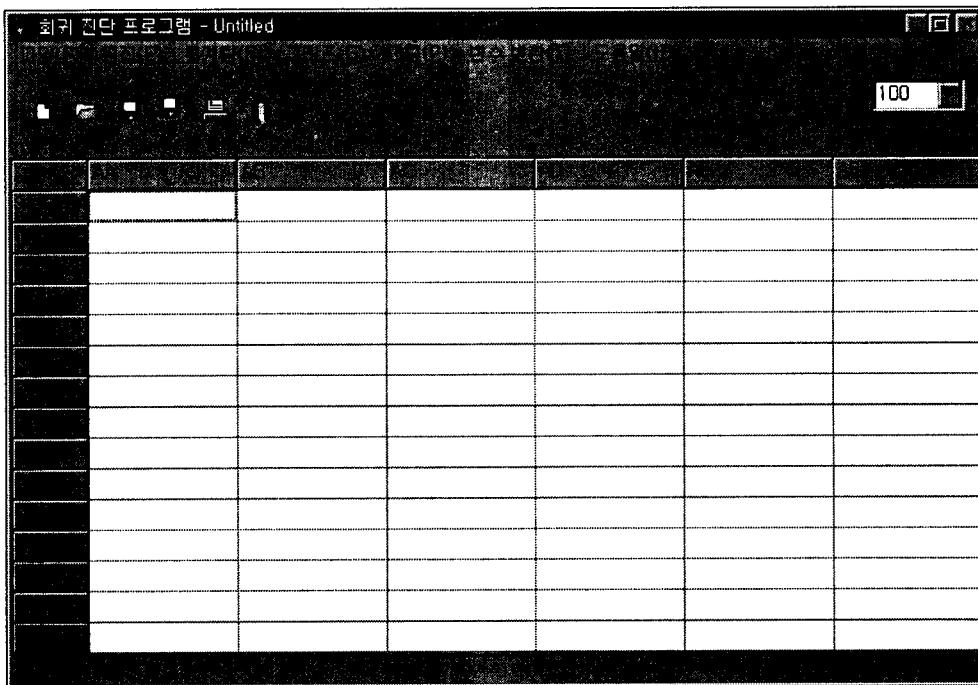


그림 2-2 초기화면

2.2 시스템의 메뉴에 대한 설명

2.2.1 파일메뉴

파일메뉴는 일반적인 윈도우즈 프로그램과 같은 형태이며, 파일메뉴에는 새로만들기, 열기, 저장, 새이름으로저장, 인쇄, 그리고 끝내기메뉴가 있다. 새로만들기는 새로운 데이터를 입력할 수 있는 새로운 데이터 입력화면을 만드는 기능이다. 열기는 외부에 있는 데이터 파일을 불러올 수 있는 기능이고, 저장은 입력된 데이터를 지정된 파일로 저장하는 기능이고, 새이름으로저장은 입력된 데이터를 수정 했을 경우 기존의 데이터를 그대로 보존하면서 수정된 데이터를 지정된 파일로 저장하는 기능이다. 인쇄는 입력된 데이터를 출력할 수 있는 기능이고, 그리고 끝내기는 사용중 인 회귀진단 시스템에서 완전히 벗어나도록 하는 기능이다.

2.2.2 편집메뉴

편집메뉴에는 레이블설정과 변수설정메뉴가 있다. 레이블설정은 입력된 데이터에 변수 이름을 지정하고, 그 변수에 대한 설명을 붙이는 기능이다. 변수설정은 분석에 필요한 종속 변수와 독립변수를 선택할 수 있는 기능이다.

2.2.3 통계량메뉴

통계량메뉴에는 분산분석표만들기와 영향력통계값메뉴가 있다. 분산분석표만들기메뉴에서는 추정된 회귀모형의 타당성을 검토할 수 있는 분산분석표와 평균제곱오차의 제곱근 값, 결정계수 및 수정된 결정계수 그리고 종속변수의 변동계수를 얻을 수 있다. 또한 모회귀계수의 최소사승추정치과 추정치의 표준오차, 모회귀계수 $\beta_i (i = 1, 2, \dots, k)$ 가 0인가를 검정할 수 있는 검정통계값 t , 유의확률(p -값)을 얻을 수 있다. 영향력통계값메뉴에서는 회귀진단에 사용할 수 있는 통계량들을 얻을 수 있다. 즉 적합값, 잔차, 햇행렬(hat matrix)의 대각선원소, 마하라노비스(mahalanobis)거리, 내면스튜던트화잔차(internally studentized residuals), 외면스튜던트화잔차(externally studentized residuals), Cook의 통계량, 수정된 Cook의 통계량, DFFITS, COVATIO 그리고 DFBETAS를 얻을 수 있다.

2.2.4 그래프메뉴

그래프메뉴에는 잔차설명변수산점도, 추가변수그림, 편잔차그림, 덧편잔차그림, 정규확률그림, 인덱스그림 그리고 이분산검정을 위한 스코어검정그림메뉴가 있다. 잔차설명변수산점도메뉴에서는 입력된 변수는 물론이고 통계량메뉴에서 계산한 모든 통계량의 값들을 x -축 혹은 y -축으로 선택할 수 있게 하였기 때문에, 회귀진단에 사용할 수 있는 다양한 그래프를 그릴 수 있다. 추가변수그림메뉴에서는 추가변수그림은 다중회귀모형 $y = X\beta + \epsilon$ 에 새로운 설명변수 Z 를 추가시키고자할 때 새로운 변수의 추가 여부와 모형의 비선형성을 시각적으로 확인할 수 있는 추가변수그림(added variable plots)을 그릴 수 있다. 또한 한 화면에 여러 개의 추가변수그림을 동시에 그릴 수 있게 되어 있어 상호 비교도 가능하다. 편

잔차그림메뉴에서는 종속변수와 독립변수와의 정확한 관계와 경우에 따라서 변수변환의 필요성을 진단할 수 있는 편잔차그림(residual plus component plots)을 그릴 수 있다. 덧편잔차그림메뉴에서는 비선형성의 효과를 편잔차그림보다 좀더 잘 검토할 수 있다고 알려진 덧편잔차그림(augmented partial residual plot)을 그릴 수 있다. 정규확률그림에서는 정규성 가정을 검토할 수 있는 정규확률그림을 그릴 수 있다. 인덱스그림에서는 x -축을 데이터의 관측번호로 하고 y -축을 통계량메뉴에서 계산된 통계량으로 하여 산점도를 그리는 인덱스그림을 그릴 수 있다. 인덱스그림은 이상치 검출이나, 독립성 가정을 검토할 때 많이 사용된다. 이분산 검정을 위한 스코어 검정법에서는 등분산 가정을 검정할 수 있는 그림을 그릴 수 있다.

2.2.5 그림메뉴

그림메뉴에서는 줄기-잎그림과 상자그림메뉴가 있다. 데이터의 구조와 특징을 탐색하기 위해서 많이 사용하고 있는 줄기-잎그림과 상자그림을 이용하여 통계량메뉴에서 계산한 영향력통계량값들의 구조와 특징을 파악하여 회귀진단에 활용하고자 한다. 줄기-잎그림메뉴에서는 통계량메뉴에서 계산한 각 통계량의 줄기-잎그림을 그릴 수 있다. 각 통계량의 줄기-잎그림을 통해서 이상치 검출과 정규성을 검정할 수 있다. 상자그림메뉴에서는 통계량메뉴에서 계산된 각 통계량의 상자그림을 그릴 수 있다. 각 통계량의 상자그림에서 안쪽 울타리(inner fence)를 벗어나는 관측값을 이상치로 정의한다. 따라서 각 통계량의 상자그림을 이용하여 이상치를 용이하게 검출할 수 있다. 한편 표준정규분포를 따르는 데이터에 대한 상자그림의 안쪽 울타리는 $[-2.69, 2.69]$ 이므로, 표준정규분포를 따르는 데이터의 99%가 상자그림의 안쪽 울타리 안에 위치하기 때문에 안쪽 울타리를 벗어나는 관측치를 이상치로 간주해도 무방하다.

2.2.6 변수변환메뉴

변수변환메뉴에서는 Box-Cox 변환에서 멱승(power)의 값에 따라 변수를 변환할 수 있는 15가지 방법을 제공하고 있다. 본 논문에서는 멱승의 값을 미리 지정하였지만, 사용자가 멱승값을 입력할 수 있게할 수도 있다.

2.2.7 도움말메뉴

도움말메뉴에서는 아래 (그림2-3)와 같이 개발된 툴의 사용방법과 개발된 툴에 사용된 여러 가지 영향력통계량과 여러 가지 그림 각각에 대한 도움말을 얻을 수 있고 또한 변수 변환방법에 대한 도움말을 얻을 수 있다.

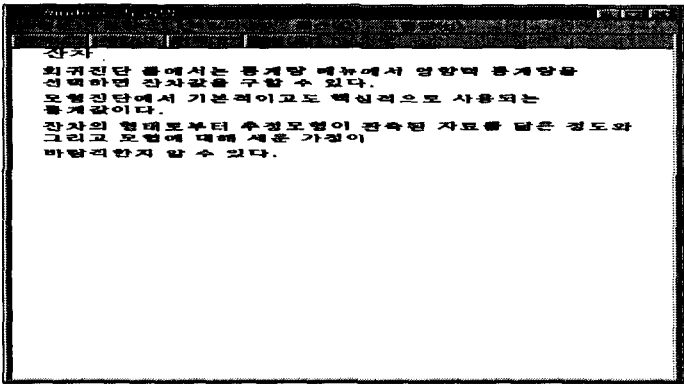
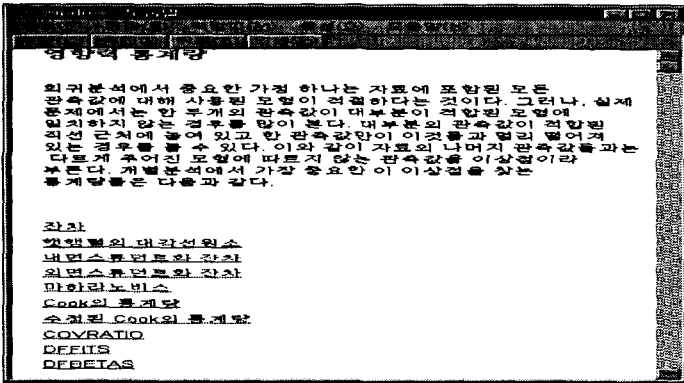
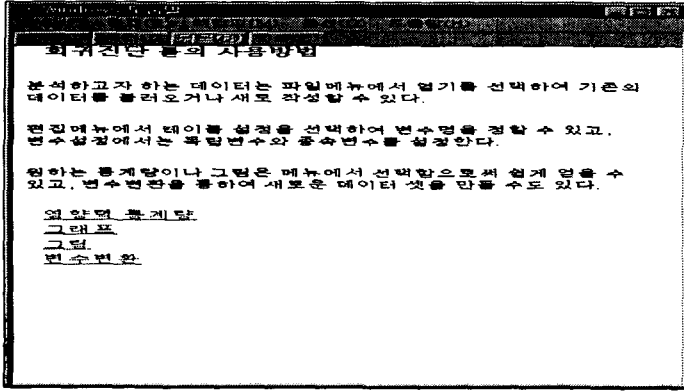


그림2-3 도움말화면

3. 실제 데이터를 이용한 프로그램의 실행 예

다음 데이터는 미니탭 통계패키지 중에서 발췌한 흑체리나무 데이터이다. 이 데이터는 쓰러진 31개의 흑체리나무에서 나무의 높이(단위:피트), 지표면 4.5피트에서의 나무의 지름(단위:인치) 및 부피(단위:입방피트)를 측정한 데이터이다(강명욱(1996)).

(그림 2-2)은 외부에서 데이터를 불러드리거나 혹은 데이터를 직접 입력할 수 있는 초기 화면이다. (그림 3-1)은 편집메뉴의 레이블설정 메뉴를 이용하여 입력된 데이터에 변수명을 지정하고, 그 변수에 대한 설명을 붙인 화면이다. (그림 3-2)는 편집메뉴의 변수설정메뉴를 이용하여 변수리스트에서 회귀분석에 이용할 독립변수(나무의 지름, 나무의 높이)와 종속변수(나무의 부피)를 선택한 화면이다. (그림 3-3)은 그래프메뉴의 잔차설명변수메뉴를 이용하여 종속변수인 나무의 부피와 독립변수인 나무의 지름 사이의 산점도를 그린 화면으로서 부피와 지름사이에 밀접한 관계가 있음을 보여주고 있다. (그림 3-4)는 그래프메뉴에서 잔차설명변수메뉴를 이용하여 종속변수인 나무의 부피와 독립변수인 나무의 높이 사이의 산점도를 그린 화면으로서 높이의 증가에 따라 분산이 증가하는 경향을 보여주고 있어 두 변수의 관계가 명백하게 나타나 있지 않다. (그림 3-5)는 통계량메뉴에서 분산분석표만들기메뉴를 이용하여 추정된 회귀모형의 타당성을 검토할 수 있는 분산분석표와 평균제곱오차의 제공근값, 결정계수 및 수정된 결정계수 그리고 종속변수의 변동계수가 출력된 화면이다. 또한 모회귀계수의 최소자승추정치과 추정치의 표준오차, 모회귀계수 $\beta_i (i = 1, 2, \dots, k)$ 가 0인가를 검정할 수 있는 검정통계값 t , 유의확률(p-값)이 출력된 화면이다. 이 결과에서 p-값이 0.0001이기 때문에 다중회귀모형이 주어진 데이터를 설명하는데 유의함을 알 수 있다. (그림 3-6)은 통계량메뉴에서 영향력통계값메뉴를 이용하여 회귀진단에 사용할 수 있는 통계량값들, 즉 적합값, 잔차, 헛행렬의 대각선원소, 마하라노비스거리, 내먼스튜던트화잔차, 외먼스튜던트화잔차, Cook의통계량, 수정된 Cook의통계량, DFFITS, COVATIO 그리고 DFBETAS 가 출력된 화면이다. 여기서 계산된 통계량값을 이용하여 관측값들이 모형에 미치는 영향력을 분석한다. 출력결과를 검토해보면 31번째 관측값에 대한 대부분의 통계량값이 크게 출력이 되었다. 이것은 31번째의 관측값을 다시 검토해 볼 필요성이 있다고 판단된다. 각 통계량에 대한 인덱스그림이나 줄기-잎-그림 그리고 상자그림을 이용하면 해석을 더욱더 용이하게 할 수 있다. (그림 3-7)은 그래프메뉴에서 추가변수그림메뉴를 이용하여 나무의 높이와 지름에 대한 추가변수그림을 그린 화면이다. 화면에서 왼쪽 그림은 먼저 나무의 지름이 모형에 있을 때 나무의 부피를 나무의 지름에 회귀시킨 후 만들어진 잔차 $e(\text{나무부피}|\text{나무지름})$ 을 y -축으로 하고, 나무의 높이를 나무의 지름에 회귀시킨 후 만들어진 잔차 $e(\text{나무높이}|\text{나무지름})$ 를 x -축으로 하는 산점도를 그린 나무의 높이에 대한 추가변수그림이다. 그림에서 두 잔차 사이에는 양의 상관관계가 아주 미약하게 나타나 있음을 알 수 있다. 이것은 나무의 높이를 모형에 추가했을 때 추가적인 설명력이 미비하다는 것을 나타낸다. 오른쪽 그림은 $e(\text{나무부피}|\text{나무높이})$ 를 y -축으로 하고 $e(\text{나무지름}|\text{나무높이})$ 를 x -축으로 하는 산점도를 그린 나무의 지름에 대한 추가변수그림

이다. 그림에서 두 잔차 사이에 선형관계가 뚜렷이 나타나있다. 이것은 지름의 추가설명력은 아주 크다는 것을 나타낸다. <그림3-8>은 그래프메뉴에서 편잔차그림메뉴를 이용하여 나무의 높이와 지름에 대한 편잔차그림을 그린 화면이다. 화면에서 왼쪽 그림은 $e(\text{나무부피}|나무지름, \text{나무높이}) + \hat{\alpha}$ 나무지름을 y -축으로 하고 나무지름을 x -축으로 하는 산점도를 그린 지름을 위한 편잔차그림이다. 여기서 $\hat{\alpha}$ 은 변수가 모두 포함된 모형에서 나오는 추정량이다. 그림의 모양이 곡선의 형태를 나타나 있는 것으로 보아 분명하지는 않지만 이 변수에 대한 변환이 필요하다고 여겨진다. 오른쪽 그림은 $e(\text{나무부피}|나무지름, \text{나무높이}) + \hat{\beta}$ 나무높이를 y -축으로 하고, 나무 높이를 x -축으로 하여 산점도를 그린 높이를 위한 편잔차그림이다. <그림3-9>는 그래프메뉴에서 덧편잔차그림메뉴를 이용하여 나무의 높이와 지름에 대한 덧편잔차그림을 그린 화면이다. 화면의 왼쪽 그림은 $e(\text{나무부피}|나무지름, \text{나무높이}, \text{나무높이}^2) + \hat{\beta}$ 나무높이 + $\hat{\theta}$ 나무높이² 을 y -축으로 하고, 나무높이를 x -축으로 하여 산점도를 그린 높이를 위한 덧편잔차그림이다. 여기서 $\hat{\beta}$ 와 $\hat{\theta}$ 은 나무지름, 나무높이 그리고 나무높이²을 포함하는 모형에서 구한 추정량이다. 오른쪽 그림은 $e(\text{나무부피}|나무지름, \text{나무지름}, \text{나무지름}^2) + \hat{\alpha}$ 나무지름 + $\hat{\gamma}$ 나무지름²을 y -축으로 하고, 나무지름을 x -축으로 하여 산점도를 그린 지름을 위한 덧편잔차그림이다. 이 그림은 편잔차그림보다 비선형의 효과를 좀더 잘 탐색할 수 있는 것으로 알려져 있으나, 여기서는 편잔차그림보다 개선된 점이 별로 보이지 않는다. <그림3-10>은 그래프메뉴에서 잔차설명변수메뉴를 이용하여 내면스튜던트잔차와 지름과의 산점도이다. 이 그림에서 나무의 지름이 증가함에 따라 내면스튜던트잔차가 곡선 형태를 보이므로 모형에 이차항을 포함하거나 변수변환이 필요한 것으로 나타났다. 이러한 곡선형태는 내면스튜던트잔차와 높이와의 산점도를 나타낸 <그림3-11>에서도 미약하게 나타난다. <그림3-12> 그래프메뉴에서 잔차설명변수메뉴를 이용하여 내면스튜던트잔차와 적합값(fitted value)과의 산점도이다. 이 그림에서 역시 적합값이 증가함에 따라 내면스튜던트잔차가 곡선형태로 나타남을 알 수 있다. <그림3-13>은 내면스튜던트잔차와 관측값의 관측번호의 산점도를 그린 인덱스그림이다. 이 그림에서 31번째 관측값의 내면스튜던트잔차값이 다른 관측값에 비해서 큰값을 가진다는 것을 알 수 있다. 따라서 이 관측값을 주의 깊게 조사할 필요성이 있다고 여겨지고, 또한 그림의 형태가 일정한 주기를 가지고 있지 않기 때문에 독립성 가정은 타당성이 있는 것으로 판단된다. <그림3-14>는 그래프메뉴에서 이분산검정을 위한 스코어 검정메뉴를 이용하여 각 독립변수에 대한 스코어 검정그림을 그린 것이다. 그림에서 나무 높이와 지름이 커짐에 따라 분산이 증가 하는 형태의 현상이 뚜렷하게 드러난다. <그림3-15>은 그래프메뉴에서 정규확률그림메뉴를 이용해서 정규성을 검정할 수 있는 정규확률그림을 그린 것이다. 이 그림이 거의 선형으로 나타났다기 때문에 정규성 가정은 타당하다고 판단할 수 있다. <그림3-16>는 그림메뉴에서 줄기-잇-그림메뉴를 이용하여 내면스튜던트잔차에 대한 줄기-잇-그림을 그린 것이다. 이 그림에서 대부분의 관측값의 내면스튜던트잔차의 절대값이 0에서 1.7사이의 값을 가지고 어떤 한 개의 관측값의 내면스튜던트잔차값은 2.4라는 사실을 알 수 있다. 인덱스그림에서 이것이 31번째 관측값이라 것을 알 수 있다. <그림3-17>는 그림메뉴에서 상자그림메뉴를 이

용하여 각종 통계량에 대한 상자그림을 그린 것이다. 이 그림의 안쪽 울타리를 벗어나는 관측값을 이상치로 정의 하였다. 이 관측값은 그림에 표시가 된다. 따라서 이상치를 찾는 데 이 그림은 유용하게 사용된다. (그림3-18)은 BOX-COX 변환에서 멱승(power)의 값에 따라 변수를 변환할 수 있는 화면이다.

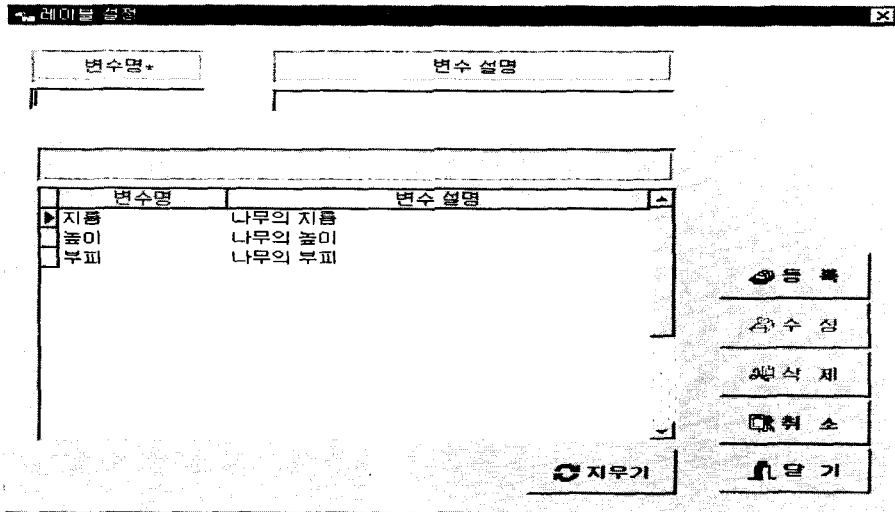


그림 3-1 레이블 설정

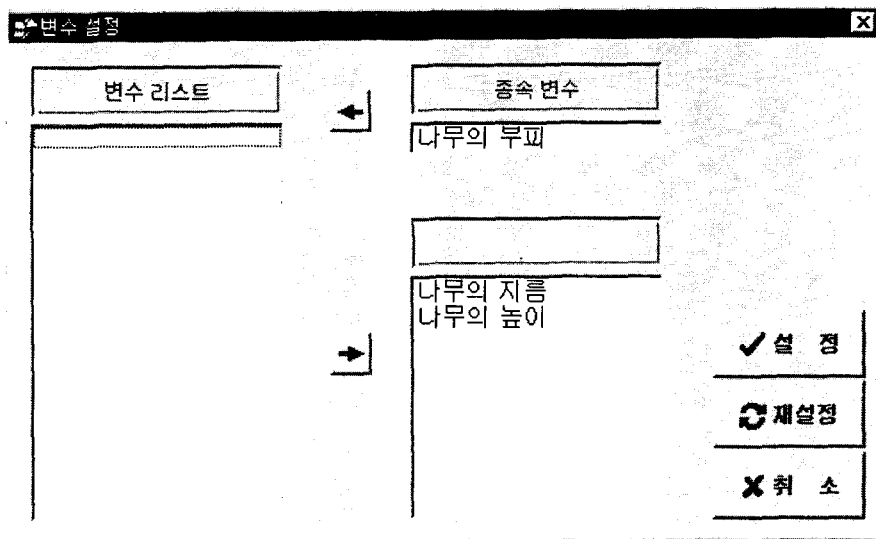


그림 3-2 변수 설정

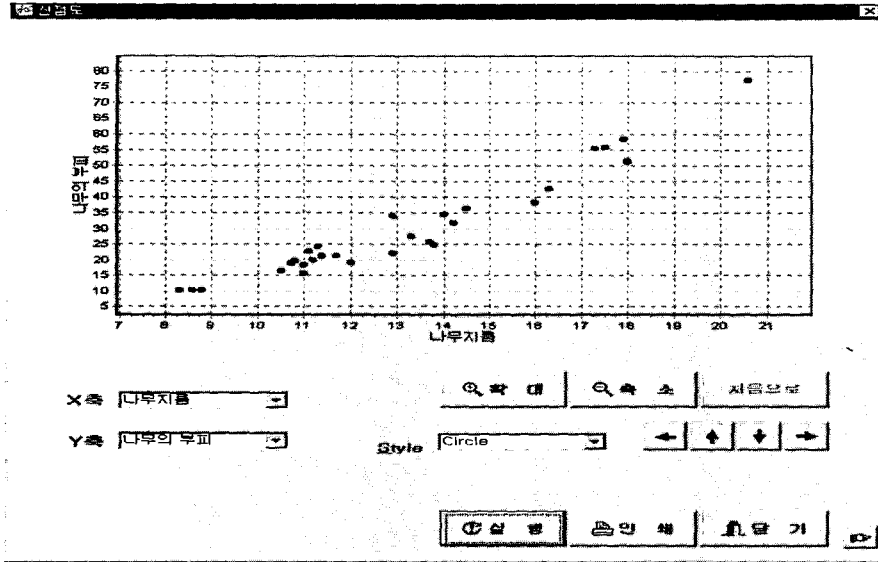


그림 3-3 나무의 지름과 부피와의 산점도

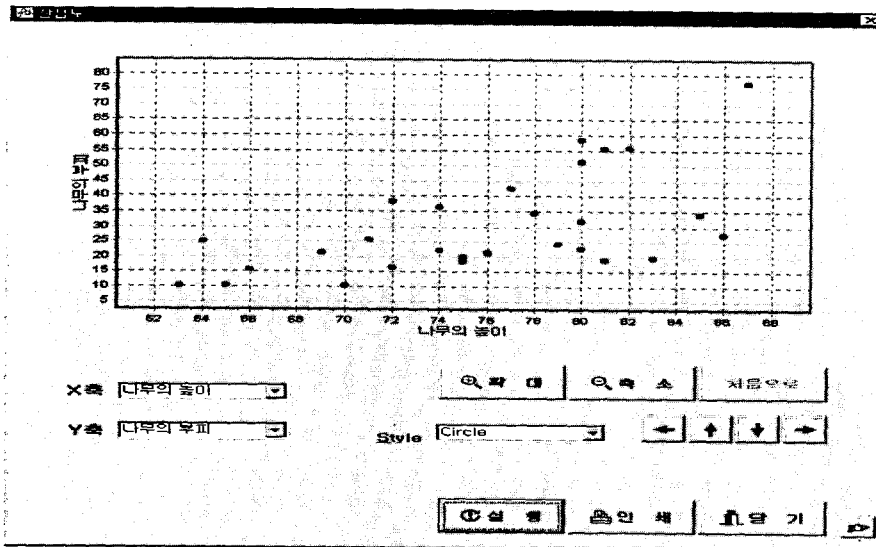


그림 3-4 나무의 높이와 부피와의 산점도

회귀 분석 결과
분산 분석표

요인	자승합	자유도	평균 자승합	F-Value	P-Value
회기	7684.16251	2	3842.08125	254.972	0.0001
간차	421.92136	28	15.06862		
계	8106.08387	30			

Root MSE	3.88183	결정 계수	0.94795
변동 계수	12.86612	수정 결정 계수	0.94423

모수 추정값

	추정값	표준 오차	t-value	p-value
절편	-57.98766	8.63823	-6.713	0.0001
지름	4.70816	0.26426	17.816	0.0001
높이	0.33925	0.13015	2.607	0.0144

확인 세 | 닫기

그림 3-5 분산분석표

영향력 통계값

순서	나무의 부피	추정치	잔차	내연표준잔차	외연표준잔차
1	10.3000	4.8377	5.4623	1.4965	1.5321
2	10.3000	4.5539	5.7461	1.6029	1.6517
3	10.2000	4.8170	5.3830	1.5285	1.5677
4	16.4000	15.8741	0.5259	0.1397	0.1372
5	18.8000	19.8690	-1.0690	-0.2937	-0.2888
6	19.7000	21.0183	-1.3183	-0.3696	-0.3638
7	15.6000	16.1927	-0.5927	-0.1623	-0.1594

순서	COVRATIO	DFRITS	INTERCEPT	나무의 지름_DFB	나무의 높이_DFB
1	0.9623	0.5545	0.2260	-0.3787	-0.0428
2	0.9798	0.6862	0.5353	-0.2234	-0.3657
3	1.0433	0.7267	0.6299	-0.1322	-0.4814
4	1.1830	0.0344	0.0148	-0.0166	-0.0052
5	1.2566	0.1070	0.0639	0.0802	-0.0794
6	1.3018	0.1563	0.0936	0.1143	-0.1272
7	1.2564	0.0574	-0.0506	-0.0036	0.0433

확인 세 | 닫기

결과치를 변수로 지정

그림 3-6 영향력 통계값

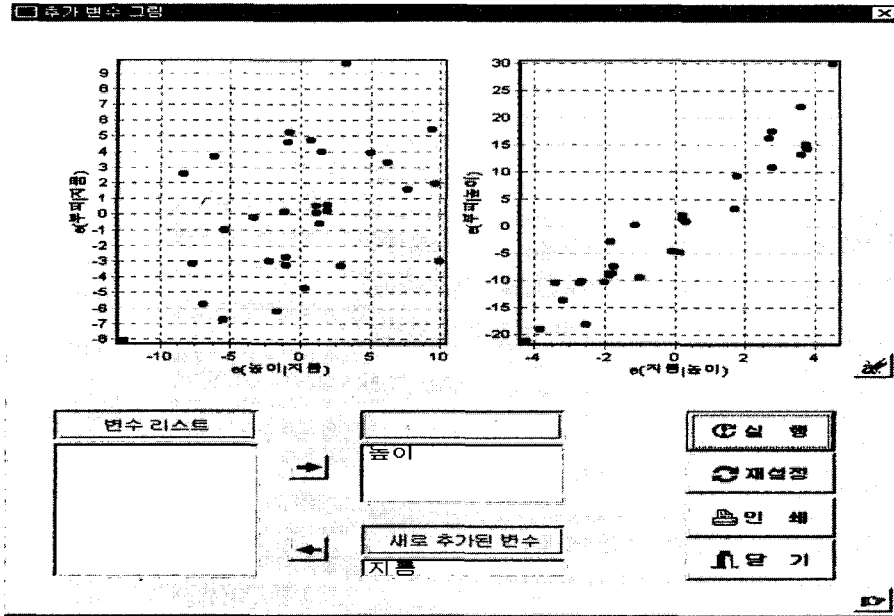


그림 3-7 추가변수그림

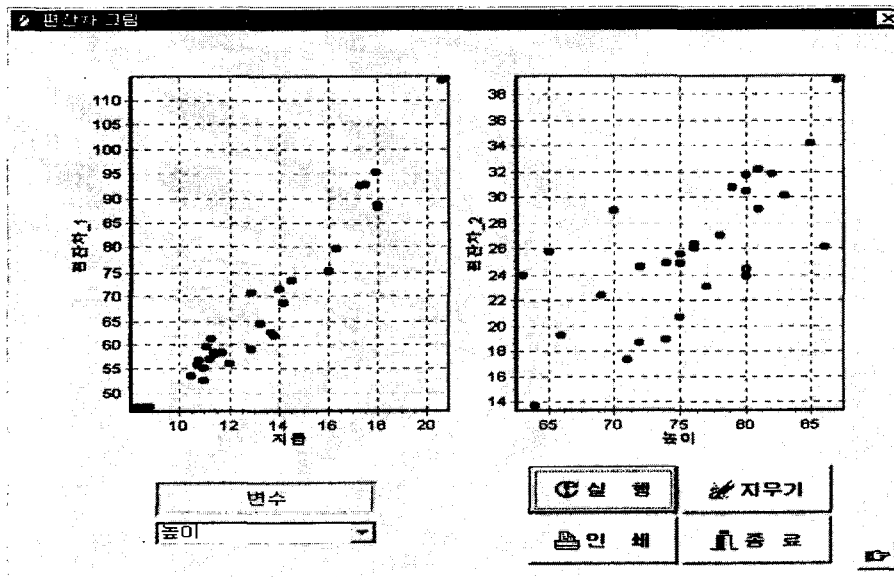


그림 3-8 편잔차그림

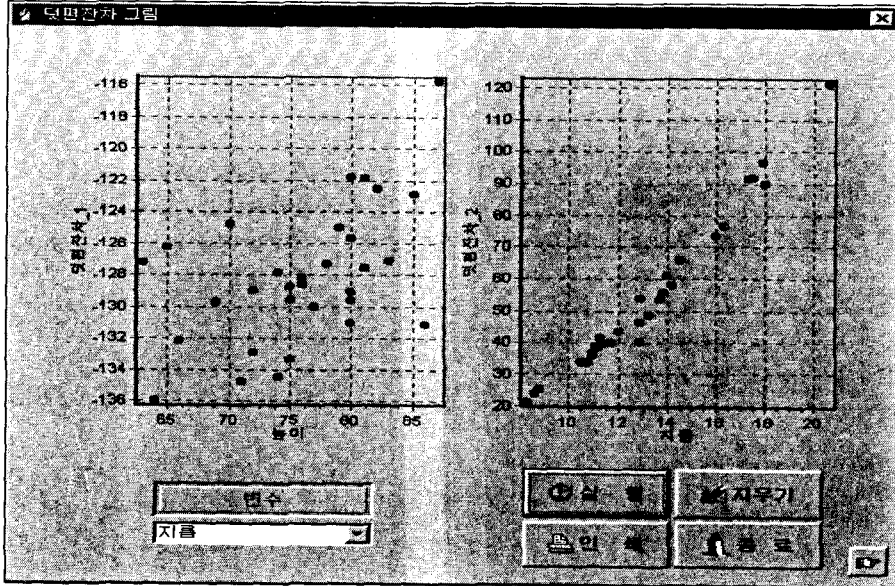


그림 3-9 덧편잔차그림

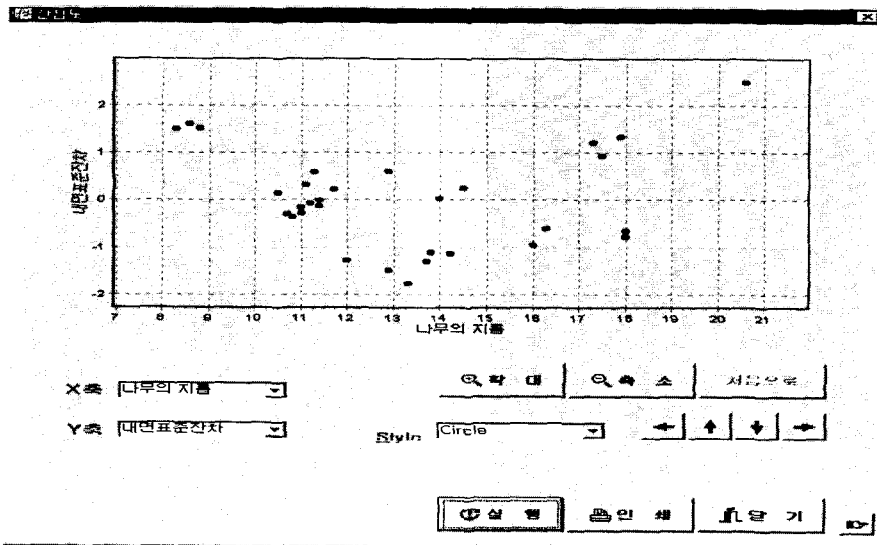


그림 3-10 내면표준잔차와 지름과의 산점도

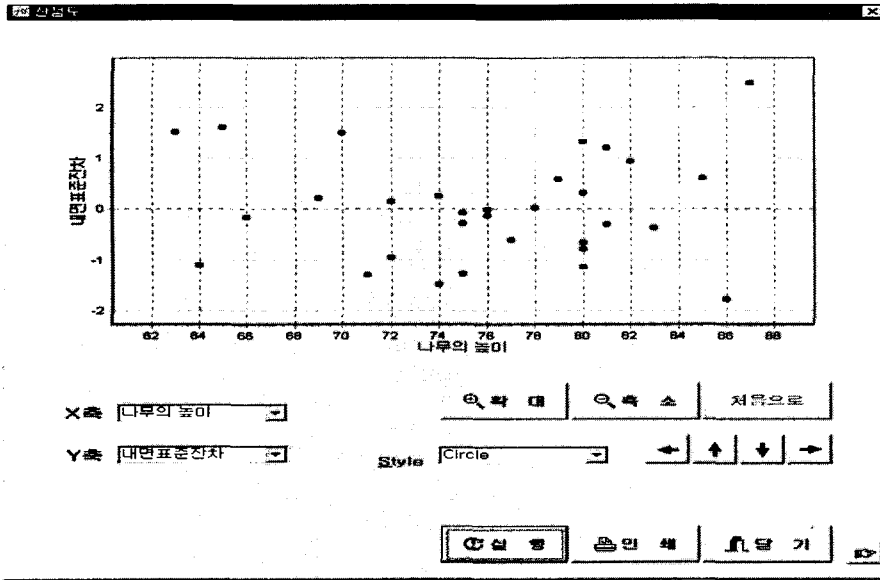


그림 3-11 내면표준잔차와 높이와의 산점도

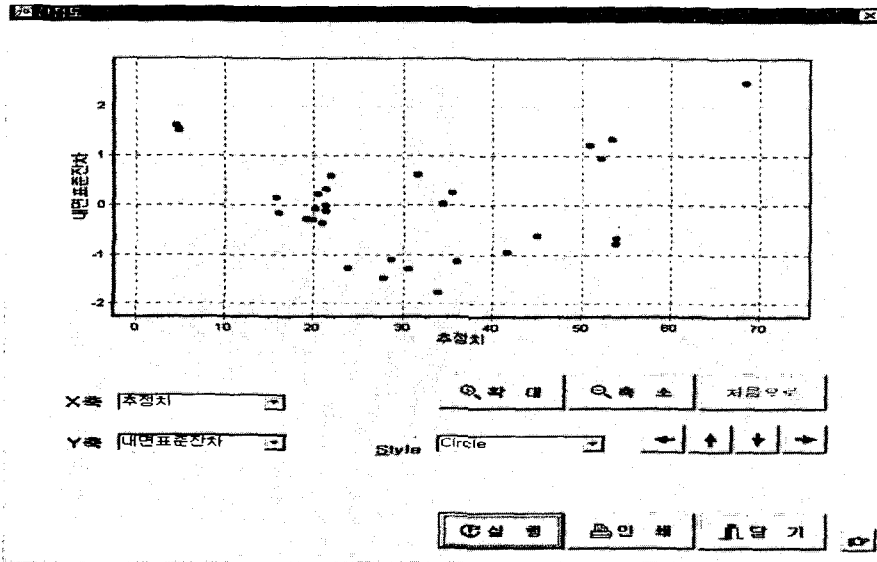


그림 3-12 내면표준잔차와 적합값과의 산점도

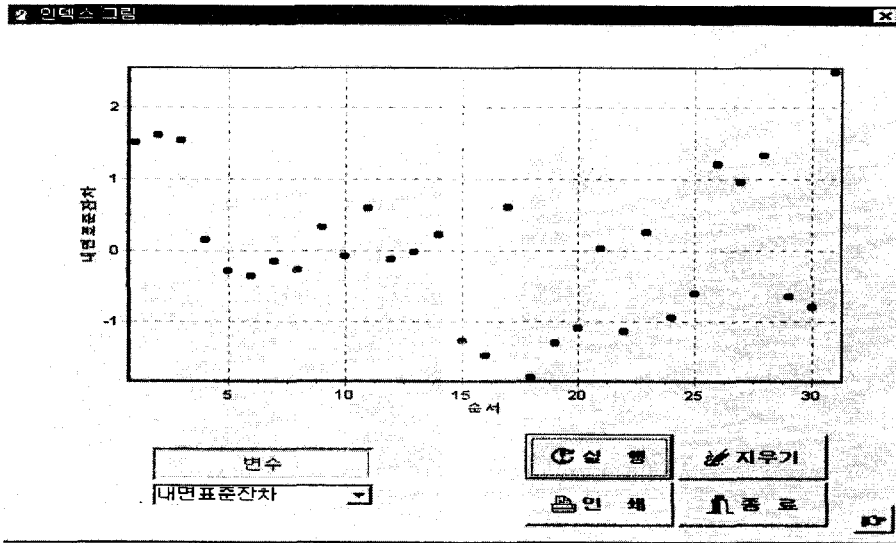


그림 3-13 인덱스그림

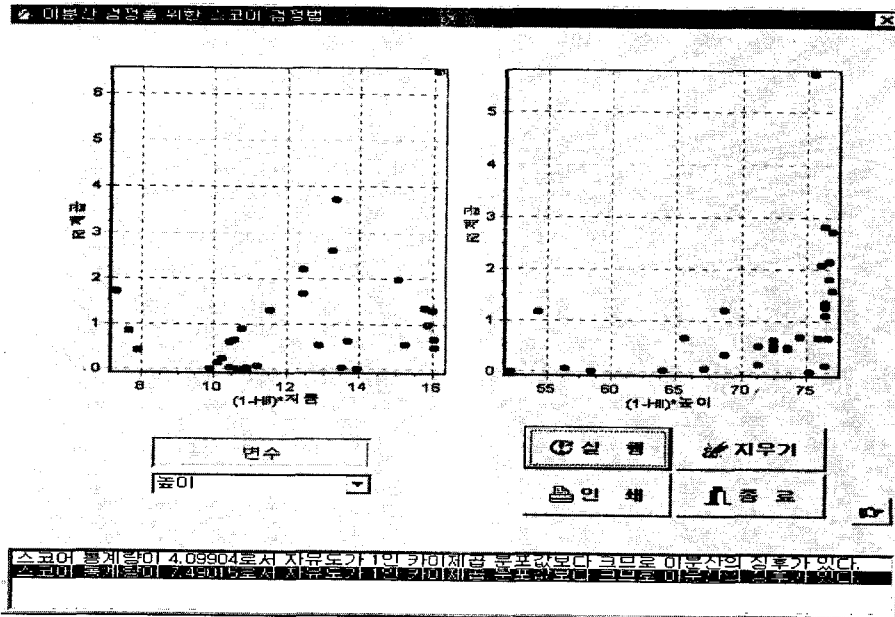


그림 3-14 이분산 검정을 위한 스코어 검정그림

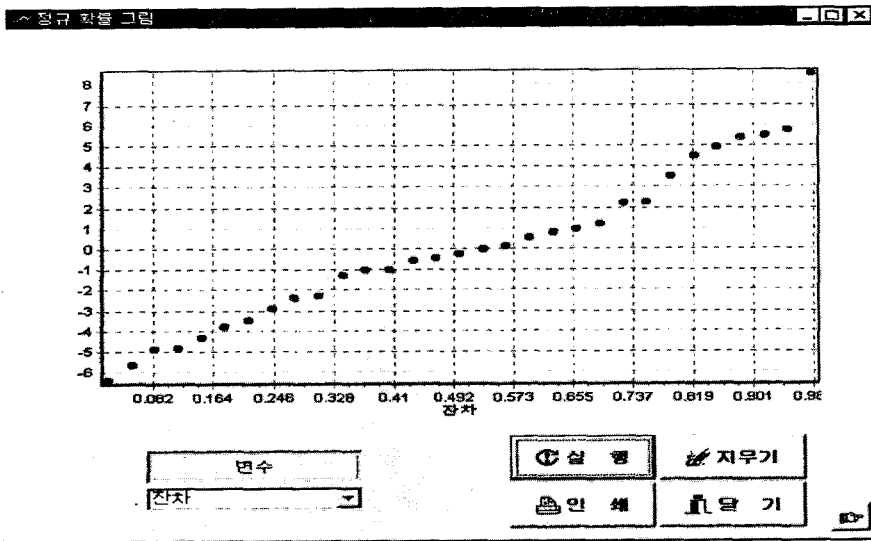


그림 3-15 정규확률그림

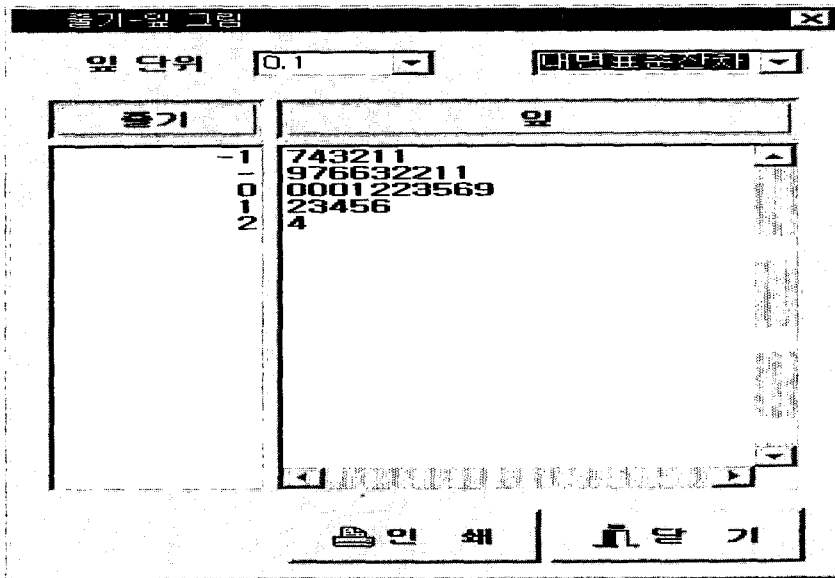


그림 3-16 잔차의 즐기-앞-그림

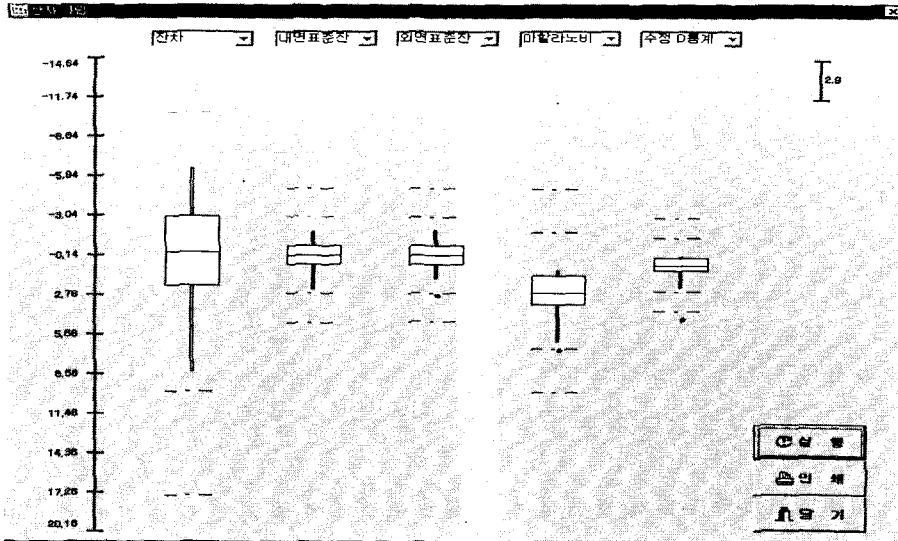


그림 3-17 상자그림

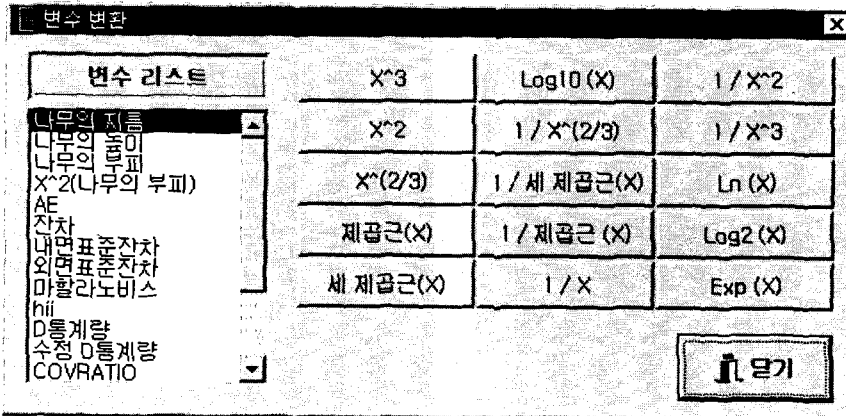


그림 3-18 변수변환표

4. 결론

본 논문에서 구현한 회귀진단 틀은 최근 윈도우즈 프로그램에 많이 이용되고 있는 비주얼 언어인 델파이 3.0을 이용하였다. 그러나 setup파일을 만들어 놓았기 때문에 델파이 3.0이 개인 컴퓨터에 인스톨이 되어있지 않더라도 사용할 수 있다. 또한 통계 비전문가들이

나 컴퓨터를 잘 모르더라도 사용할 수 있도록 메뉴를 이용하여 수행할 수 있도록 하였으며, 또한 프로그램을 한글화 하므로써, 회귀분석 교육 뿐만 아니라 실제 데이터분석이 이루어 지도록 구성하였다. 그리고 많은 통계 그래프를 쉽게 그릴 수 있게 하여 분석 결과를 시각화 하므로써, 결과 해석을 용이하게 하였다.

참 고 문 헌

1. Box, G.E.P.(1980). Sampling and Bayes' inference in scientific modelling and robustness(with discussion). *Journal of the Royal Statistical Society Series A*, 143, 383-430.
2. Cook, R.D., and Weisberg, S.(1994). *An Introduction to Regression Graphics*, John Wiley & Sons, Inc.
3. 이정진, 강근석(1994). 한국형 통계패키지 개발 연구, 응용통계연구 제 7권 2호, 279-288.
4. 이정진, 강근석, 이윤오, 김지현, 이창수, 김성철(1995). 전문가용 한국형 통계패키지 개발 연구, 한국통계학회논문집 제2권 2호, 434-444.
5. 강명옥 외 3명 (1996). 회귀분석, 을곡 출판사.
6. 조신섭, 성병찬, 강병국(1996). Excel을 이용한 시계열 분석틀의 개발, 한국통계학회 추계발표회 논문집, 187-188.

Development of a Regression Diagnosis Tool Using Delphi

Mi-Jin Hyun ⁴ · Jin-Pyo Park ⁵ · Hee-Chang Park ⁶

Abstract

In this paper we suggest the visualized regression diagnosis tool. The tool is developed by Hangul Delphi on the basis of windows, so users can easily make use of this tool though they do not have the expert knowledge about statistics and computer. Especially, to apply this tool to teaching regression analysis or data analysis, we offer various residual plots in the tool and show the results of analysis graphically.

Key Words and Phrases: Delphi, Regression diagnosis, Statistical graphics, Objective oriented language.

⁴Graduate School (631-701) Department of Computer Engineering, Kyungnam University

⁵Associate Professor, (631-701) Department of Computer Engineering, Kyungnam University

⁶Associate Professor, (641-773) Department of Statistics, Changwon National University