

타이어 설계 인자들에 대한 회귀모형의 수립

박정수 · 황현식 · 조완현
전남대학교 자연과학대학 통계학과

Building Regression Models for Tire Design Factors

Jeong-soo Park · Hyun-sik Hwang · Wan Hyun Cho
Dept. of Statistics, Chonnam National University

Abstract

Two regression models for explaining the tire performances (especially conering coefficients) by tire design and experimental factors are built. One is the ordinary regression model, and the explaining variables in the model are selected by a stepwise method. The other model is built by a modern nonparametric regression technique, called projection pursuit regression. Then two models are compared and combined, so that the relationship between the tire performances and design factors are well figured out. The optimal experimental design issue and future research ideas are also discussed.

1. 서론

타이어의 설계시에 고려되는 제반 설계인자 및 실험실에서 고려되는 시험조건들과 타이어의 성능을 평가하는 여러가지 특성값들 간의 관계를 규명하는 것이 본 연구의 일차목표이다. 나아가 일단 규명된 관계로부터 실험되지 않은 조건 값에 대하여 타이어의 성능을 예측하고자 한다. 이를 통하여 최적의 설계요인 및 시험조건을 파악할 수 있을 뿐만 아니라 타이어 성능 실험의 횟수를 줄이는 등의 다목적의 효과를 얻을 수 있어서, 결국 경비를 줄이고 타이어의 품질향상에 도움을 줄 수 있다 (정 성룡, 1994).

이러한 목표를 위해서 본 연구에서는 과거 수년 동안의 타이어 특성에 관한 실험자료에 바탕하여 전통적인 회귀분석을 이용한 회귀모형을 수립하였고, 또한 비모수적 방

법이며 컴퓨터의 집중사용 방법 (computer intensive method) 인 projection pursuit 회귀분석 (ppreg) 모형을 수립하였다. 전자의 방법은 중요한 변수들이 무엇인가를 규명하는 역할을 한다면 후자의 방법은 새로운 입력에 대한 타이어 성능의 예측에 유용하다고 하겠다. 나아가서 이들 두가지 회귀분석 방법에 의해서 구해진 모형을 비교함으로써 매우 중요한 변수들과 불필요한 변수들을 골라냈다. 또한 차후의 다변량 회귀모형의 수립을 위해 축차적 최적 실험계획법에 관해 논의하였다.

본 연구는 주식회사 금호타이어 내의 금호연구소의 의뢰와 협조에 의해서 이루어졌음을 밝혀둔다.

2. 변수들의 특성과 자료의 성격

먼저 타이어의 설계시에 고려되는 제반 설계인자 및 실험실에서 고려되는 시험조건들과 타이어의 성능을 평가하는 여러가지 특성값에 대응되는 변수들의 특성과 금호연구소에서 얻은 실험자료의 성격에 대하여 알아 보겠다. 특히 타이어의 제품사양상 flipper 미사용이면서 capply 가 없는 경우의 제품이 현재 금호타이어에서 주로 취급하고 있으므로 이러한 제품사양 (제 6 사양이라고 불림)에 대해서만 연구를 국한시켰다.

제반 설계인자 및 실험조건에 관련되는 변수가 모두 27개 있는데 이를 독립변수로 하고, 타이어의 성능과 관련된 변수가 10개 있는데 이를 종속변수로 하는 회귀모형의 추정을 염두에 두고, 각 변수를 대략 분류하면 다음과 같다 (그림 2.1의 타이어의 구조와 각 부분별 명칭을 참고하기 바란다. 백세종, 1993)

(1) 타이어의 성능 특성값과 이에 대응하는 변수명 (10개의 종속변수)

cc1, cc4: conering coefficients

atc1, atc4: aligning torque coefficients 의 5 종류

rr: rolling resistance

(2) 타이어 설계인자 및 실험조건 항목과 이에 대응하는 변수명 (괄호 속에 27 개의 독립변수명이 쓰여 있음)

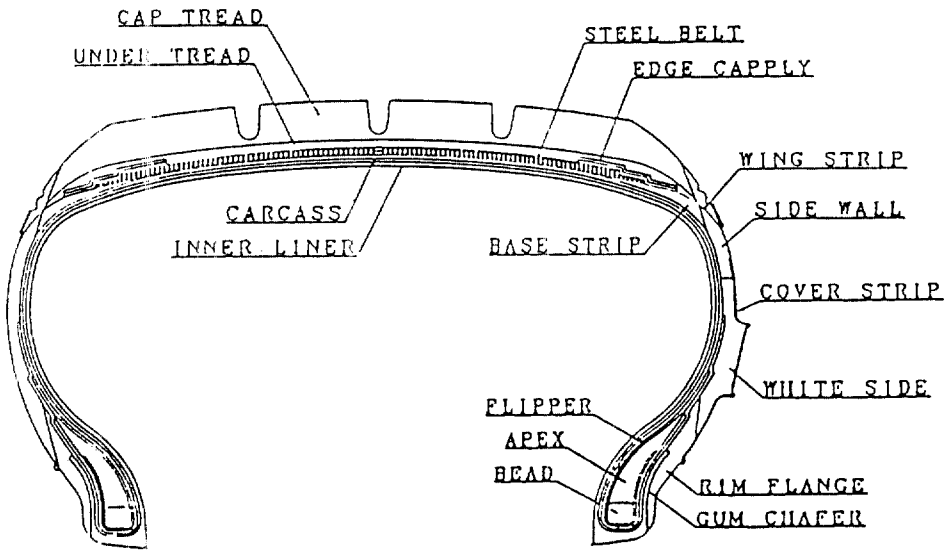
- * 실험조건 항목 (7개 변수) ----- 립폭(rimwd), 하중(load), 실험온도(temp), 공기압 (infl), 립반경 (rimdi), 트레드반경 (tdr), 트레드경도 (hard).

- * 압출류 조립 항목 (7개 변수) --- base strip 의 폭 (aw1), base strip 의 길이 (all),

black sidewall 의 폭 (aw2),
 림 flange 의 폭 (aw3),
 cap-tread 의 폭 (aw6),
 cap-tread 의 길이 (al6),
 under-tread 의 폭 (aw7).

- * liner 조립 항목 (4개 변수) --- gum chafer 의 폭 (lw1),
 gum chafer 의 길이 (ll1),
 안쪽 라이너의 폭 (lw2),
 shoulder strip 의 폭 (lw3).
- * 단독부품 항목 (9개 변수) ---- 벨트 gage (cg1), 벨트의 폭 (cw1),
 벨트의 길이 (cl1), ply 의 각도 (ca4),
 ply gage (cg4), ply 의 폭 (cw4),
 ply 의 길이 (cl4) 외 2 종류.

RADIAL TIRE 단면도



< 그림 21 > 타이어의 구조와 각 부분별 명칭: RADIAL 타이어의 단면도

본 논문에서는 특히 안정성의 계수인 cc1에 대해서만 회귀모형을 세우고 독립변수들의 cc1에 영향을 미치는 정도를 알아 보았다. 다른 종속변수에 대해서도 본 논문에서

제시한 방법과 유사한 방법을 이용하여 같은 종류의 분석을 할 수 있겠다.

독립변수의 한 특징으로는 cg4 와 cal 을 포함한 4개의 변수가 2수준 값 만을 갖는다는 것이다. 이런 변수에 대해서는 2차식의 회귀항을 적합시키기가 불가능하다. 또한 오직 3수준만을 가지는 변수도 4개 있다. 이들은 본질적으로는 연속적인 값을 가질 수 있지만 여러가지 이유로 2-3 수준에서만 설계되거나 시험되어 지고 있다. 따라서 이들 변수들을 가변수로 처리하는 대신 연속형변수로 취급하여 회귀모형의 독립변수로 처리하였다 (조완현 등 (1994)은 이들을 가변수로 처리하는 회귀모형을 수립하였다).

제 6사양에는 총 385개의 실험자료가 있는데, 이중 상당수는 한 개 이상의 결측치를 포함하고 있어서 통계적 분석상의 어려움을 가지고 있다. 실제로 종속변수 cc1 과 25개의 독립변수에 결측치가 전혀 없는 자료는 114개에 불과하였고, 나머지 271개의 자료는 적어도 한 변수 이상에서 결측치를 가지고 있었다. 이들 114개 관측치에 바탕하여 볼 때, 독립변수 cal, aw8 은 오직 1 수준 값 만을 가지므로 이들은 회귀모형에 포함시키지 않았다.

3. 회귀모형의 수립

114개의 관측자료를 가지고 종속변수 cc1 과 25개의 독립변수들 간의 적절한 회귀모형을 수립하고, 이를 바탕으로 실험되지 않은 점에서의 종속변수 값의 예측이 본 절의 목적이다.

먼저 cc1 의 기술통계를 보면 다음과 같다.

평균	표준편차	최소	최대
0.20953	0.01439	0.164	0.241

종속변수의 정규분포 가정이 만족되는지를 보기 위하여 cc1 의 줄기와 잎 (stem-leaf) 그림을 그렸다 <그림 3.1>. 또한 콜모고로프-스미르노프의 정규분포에 대한 적합도 검정에서 검정통계량 W 는 0.977 이고 유의확률 (p-value)은 0.325 로서 정규분포를 따르고 있음을 알 수 있고, 따라서 종속변수에 대한 변수변환은 고려하지 않았다. 독립변수들 간에 상관계수를 구해 보았으나, 절대치 0.8 이상의 상관계수를 가지는 변수들은 없었으며, 따라서 다공산성은 없는 것으로 보았다.

줄기 잎	빈도
24 011	3
23 8	1
23 00023333	8
22 558889	6
22 0001344	7
21 55666667777788889	17
21 00000011112222444	17
20 556667778889	12
20 1222333334444	14
19 555555777899	12
19 1112223344	10
18 78999	5
18	

< 그림 3.1 > 종속변수 cc1 의 112개의 자료에 대한 줄기와 잎 그림
(줄기와 잎의 값에 100을 곱하여 구함)

적절한 회귀모형을 찾기 위하여 먼저 제 1차 회귀모형, 즉

$$cc1 = \beta_0 + \sum_{j=1}^p \beta_j x_j$$

(여기서 x_j 는 독립변수를 뜻함) 을 적합시키기 위하여 25개의 독립변수들에 대하여 단계적 선택 (stepwise selection)을 실시하였다. 그 결과 8 개의 변수가 포함되는 모형이 세워졌는데, 결정계수가 0.37 로 매우 낮은 편이었다.

결정계수 값이 매우 작은 값일 뿐만 아니라 일부 독립변수는 종속변수에 대하여 2차식의 관계와 독립변수들 간에 교호작용이 예상되므로, 이번에는 제 2차 회귀모형, 즉

$$cc1 = \beta_0 + \sum_j \beta_j x_j + \sum_j \beta_{jj} x_j^2 + \sum_i \sum_k \beta_{ik} x_i x_k$$

을 고려하였다. 그러나 이 경우 전체 회귀항의 수는 $25+25+\binom{25}{2} = 350$ 개로써

SAS/PC 에서는 한 번에 단계적선택이 불가능했으므로 (out of memory), 다음과 같은 단계를 거쳤다. 회귀분석 상의 제반 개념과 변수선택의 기법들은 Weisberg (1985) 또는 박성현 (1981) 및 SAS 매뉴얼 (1987)을 참고하기 바란다.

(1) 먼저 25개의 1차항과 25개의 2차항에 대해서 전진적 선택 (forward selection) 을 실시하였다. 이때 모형내 변수를 포함하는 기준인 F 값 (F-in)을 작은 값으로 좁 으로서 (SAS에서는 선택의 유의확률로서 slentry라는 option을 주는데 이 경우 0.25 를 이용했다) 될 수 있는대로 많은 변수가 포함되도록 하였다. 그 결과 20개의 회귀 항이 선택되었는데, 이는 14개의 서로 다른 변수로 이루어졌다. 이 때의 결정계수 (R-square)는 0.57 이었다.

(2) 이제 교호작용을 고려하기 위해서, (1)에서 선택된 14개의 변수들 간의 교호작용 항 91개와, (1)에서 선택 안 된 변수 11개와 선택된 변수들 (14개) 간의 교호작용 항 154개 (11*14)를 만들었다. 다음으로 (1)에서 포함된 20개의 회귀항은 변함없이 포함 시킨 상태에서 이들 교호작용 항들에 대해서만 전진적 선택을 실시하였다. 그 결과 5 개의 교호작용 항이 선택되었으며, 결국 25개의 회귀항으로 이루어진 모형 (18개의 서로 다른 변수로 구성됨) 을 만들게 되었다. 이때의 결정계수는 0.63 이었으며, 이를 “모형1” 이라고 부르자.

모형1에서 잔차분석을 실시한 결과 (잔차, 표준화 잔차, student 잔차, h_i , $Dffits$, $Cook's D$ 등을 고려함), 3개의 자료가 이상치로 간주되어 이들을 제외하고 111개의 자료에 대하여 위의 모형을 재적합시켰다 (이상치 추출에 관한 자세한 내용은 Neter, Wasserman & Kutner (1989)나 Weisberg (1985)를 참조하기 바람). 그 결과 결정계 수가 0.82로 급격히 증가하였다. 즉, 3개의 이상치가 결정계수를 누르고 있었다는 점 을 알 수 있다. 결정계수 이외에 모형선택에 사용되는 다른 기준들의 값은 다음과 같 았다.

AIC (Akaike Information Criterion) = -1102.7

PRESS (Predicted Residual Sum of Squares) = 0.0169

Adjusted R-square = 0.78.

결측치 때문에 분석에서 제외된 자료를 될 수 있는대로 이용하고 또한 관측치의 수 가 회귀항의 수에 비하여 상대적으로 작으므로 다음과 같이 자료를 보완하였다. 제 6 사양의 자료 385개 중, 모형1에 포함된 18개의 변수에는 결측치가 없지만 모형1에 포 함되지 않은 나머지의 변수에서 결측치가 있음으로 인하여 처음의 114개의 관측치에 포함되지 않았던 자료가 26개 확인되어, 이를 현재의 111개 자료에 추가시킴으로서, 이 제 137개의 관측치가 분석대상이 되었다.

실제로 모형1은 전진적 선택만을 했고, 따라서 너무 많은 회귀항을 포함하고 있음으 로 인하여 과도적합 (overfitting) 되었을 가능성이 크므로, 모형의 경제성 (parsimony) 원칙에 입각하여 회귀항의 수를 줄이는 것이 바람직하다. 따라서 모형1에 대해서 후 진제거법 (backward elimination)을 실시하였다. 이 때 변수제거 기준인 F-out (또는 SAS 에서는 slstay)을 조정하여 될 수 있는대로 많은 변수가 제거되도록 하였다 (위 와 같이 전진선택과 후진제거법을 순서적으로 적용한 일련의 변수선택 전략은 Wiesberg (1985, pp. 212-213) 에도 간단히 소개되어 있다).

이러한 모형 선택의 과정에서 7개의 이상치를 다시 제거시키고, 전체 130개의 자료에 대해서 최종적으로 14개의 회귀항 (11개의 서로 다른 변수로 구성) 으로 이루어진 모형을 구하게 되었다 (각 변수명의 의미는 제 2장에서 설명되어 있다). 이를 “모형2” 라고 부르자. 모형2에 대한 분산분석표 및 회귀추정 결과가 아래 <표 3.1>에 나와있다 (예를 들어, aw1_s 는 aw1 의 제곱항을 의미하며, lw1cg4 는 lw1 과 cg4 의 교호작용 항을 의미한다).

< 표 3.1 > 모형2에 대한 분산분석표 및 회귀모수의 추정 결과

분 산 분 석 표 (종속변수: CC1)

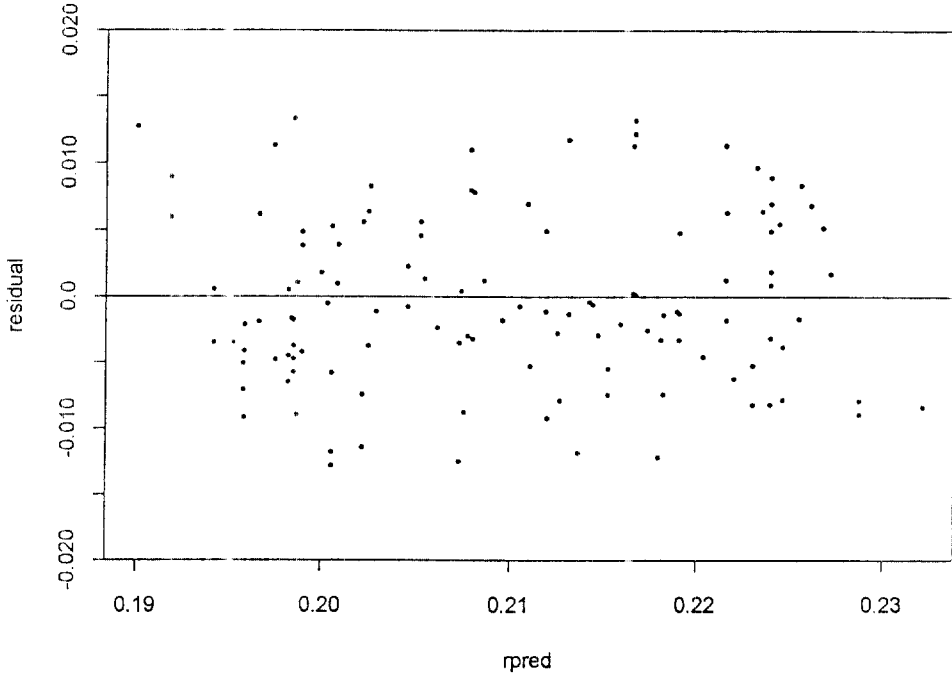
요인	자유도	평방향	불편분산	F 값	유의확률
모형	14	0.0150	0.0010	22.562	0.0001
잔차	115	0.0054	0.0001		
총계	129	0.0204			
표준오차	0.0068		결정계수	0.7331	

회귀모수의 추정치

변수명	모수 추정치	표준오차	Ho:모수=0 T값	유의확률
INTERCEP	-6.883	1.6921	-4.068	0.0001
LW1	-0.002	0.0024	-1.089	0.2784
CL1	0.010	0.0017	5.713	0.0001
AW6	0.017	0.0019	8.698	0.0001
LOAD	-0.0007	0.0001	-5.535	0.0001
CA4	-0.015	0.0088	-1.717	0.0887
AW1_S	-0.00002	0.00001	-3.932	0.0001
AW6_S	-0.00004	0.00001	-8.876	0.0001
AW7_S	-0.00001	0.00001	-6.771	0.0001
RIMWD_S	-0.0473	0.0097	-4.866	0.0001
TEMP_S	-0.00003	0.00001	-1.893	0.0609
CL1_S	-0.00001	0.00001	-5.824	0.0001
LW1CG4	0.0039	0.0019	1.999	0.0480
LOADLW2	0.00001	0.00001	5.420	0.0001
LOADRIMW	0.00012	0.00001	5.295	0.0001

모형2는 모형1에 비하여 거의 반만의 회귀항의 수를 가졌음에도 결정계수는 별로 감소하지 않았고, 그 외의 선택기준도 AIC=-1280.0, PRESS=0.0055, Adjusted R-sqaure=0.7006 으로서 모형1에 대한 모형2의 상대적 우위를 보여주고 있다. <그림 3.2>는 130개의 자료를 모형2에 적합시킨 뒤의 예측치 대 잔차의 그림으로서, 적합치

의 최소값과 최대값 근처에서 적합이 잘 안되고 있지만, 대체로 양호하다고 하겠다. 결국 모형2를 최종적인 회귀모형으로 수립하게 되었다.



< 그림 3.2 > 모형2로부터의 예측치 대 잔차의 그림

4. Projection pursuit 회귀모형

컴퓨터를 집중적으로 사용하는 비모수적 회귀분석 방법인 Projection pursuit 회귀분석 (ppreg)은

$$y \approx \bar{y} + \sum_{m=1}^{M_0} \beta_m \phi_m (a'_m x) \quad (4.1)$$

을 만족하는 회귀계수 β_m , $m=1, 2, \dots, M_0$, 비선형 변환 함수 ϕ_m , $m=1, 2, \dots, M_0$ 과 방향벡터인 a_1, a_2, \dots, a_{M_0} 를 찾아준다. 이 때 함수 항의 수를 의미하는 M_0 는 과도한 적합 (overfitting) 과 부족한 적합 (underfitting) 사이에서 적절한 값을 결정할 수 있다 (뒤에 결정하는 방법이 기술되어 있다). 실제로 (4.1)에서의 모수들 β_m , ϕ_m , a_m , $m=1, 2, \dots, M_0$ 는 평균제곱오차 (mean squared error)

$$E [y - \mu_y - \sum_{m=1}^{M_0} \beta_m \phi_m (a_m^t x)]^2$$

를 최소로 하는 값으로서 구해진다. 이들을 구하는 구체적 과정은 Friedman과 Stuetzle (1981) 또는 Thisted (1987)을 보기 바란다.

위 모형 (4.1)의 장점은 독립변수들 사이의 교호작용을 충분히 고려한 회귀모형을 세울 수 있다는 점이다. 또한 특별한 경우에 가법모형 (additive model)과 보통의 선형모형 (ordinary linear model)을 포함하고 있어서 매우 일반화된 회귀모형으로서 유용하다고 하겠다. 그러나 (4.1)의 형태로는 어떤 독립변수가 종속변수에 어떤 영향을 미치는지를 알 수 없다. 즉 기존의 전통적 회귀모형에 비하여 독립변수들과 종속변수와의 관계를 규명하기 어렵다. 반면 식 (4.1)로부터 새로운 독립변수의 값에 대해서 종속변수의 값을 잘 예측할 수 있는 장점을 가진다.

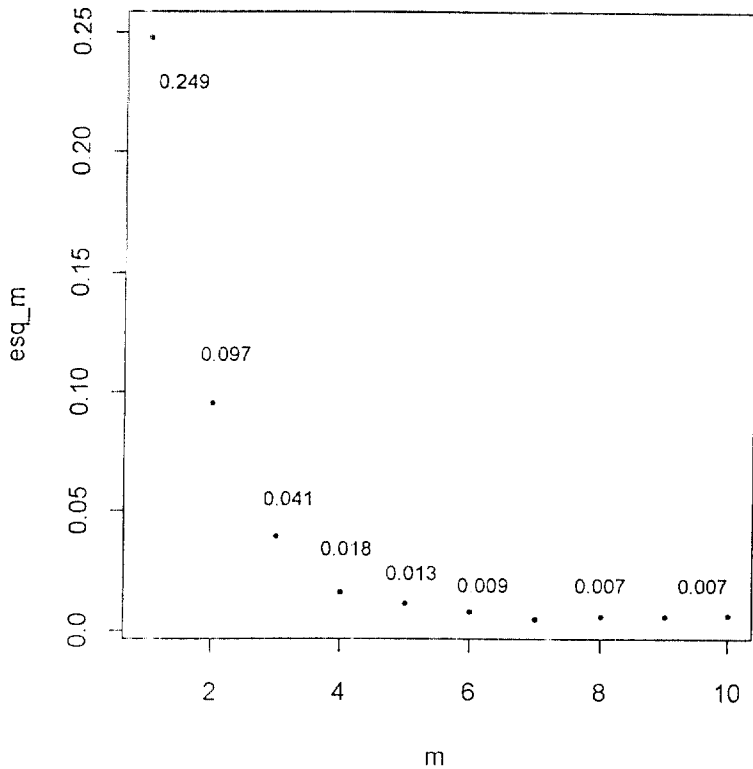
(4.1)에서 모형 선택의 주요 기준인 M_0 는 아래와 같은 방법으로 선택된다 (S-plus, 1993). 각 함수항의 수 m 에 대하여 pprog는 회귀모형 (4.1)에 의해서 설명되지 않은 분산의 비율 ($e^2(m)$ 으로 표시함)을 계산해 준다. 먼저 $m=1$ 에서 부터 적당한 크기까지의 m (변수 수가 많을 때는 변수보다 적게, 변수 수가 적을 때는 변수 수보다 많게 잡아줌)에 대하여 $e^2(m)$ 을 계산한다. $e^2(m)$ 은 m 에 대하여 감소하는데 그것을 그림으로 그려 보아서 <그림 4.1>, 급격히 감소하다가 완만해지기 전의 m 을 M_0 로 선택한다.

본 연구에서는 제 3절에서와 같은 변수들 (25개의 독립변수와 1개의 종속변수)에 대해서, 2개의 극단값을 제외한 112개의 관측치를 가지고 S-plus에서 pprog를 실행하였다.

<그림 4.1>은 M_0 를 선택하기 위한 $e^2(m)$ 의 그림이다. 이 그림에 바탕하여 M_0 를 3으로 결정하고, 다시 pprog를 실행했다 (M_0 를 4로 선택할 수도 있겠으나 결정계수의 증가량이 0.023으로 미미한 정도여서 M_0 를 3으로 선택했다). 그 결과로서 방향벡터 a_1, a_2, a_3 , 회귀계수 $\beta_1, \beta_2, \beta_3$ 그리고 ϕ_1, ϕ_2, ϕ_3 가 구해졌다. 이 때의 결정계수는 0.9594 이었다.

pprog에 의해서 추정된 β 값은 $\widehat{\beta}_1 = 0.0126$, $\widehat{\beta}_2 = 0.0086$, $\widehat{\beta}_3 = 0.0061$ 이었다. 여기서 ϕ_1 에 대한 계수 (또는 가중치)가 ϕ_3 에 대한 계수보다 약 2배의 값이 주어지고 있음을 알 수 있다. 여기서 방향벡터 a_1, a_2, a_3 의 값의 제시는 단지 숫자의 나열에 불과하므로 생략한다.

결국 pprog를 통하여, 독립변수들과 종속변수와의 관계식 (4.1) 을 세우고, 실험자료에서 사용되지 않은, 독립변수들의 새로운 값에 대하여 cc1의 값을 잘 예측할 수 있게 되었다.



<그림 4.1> 각 m에 대해 (4.1)에 의해서 설명되지 않는 비율, $e^2(m)$ 의 그림

5. 두 회귀모형의 비교

본 절에서는 ppreg에서 변수를 선택하여 3절에서 수립한 모형2와 비교하고자 한다. 그러나 ppreg에서는 적절한 변수선택 방법이 알려져 있지 않기 때문에, 표준화된 자료를 바탕으로 방향벡터의 크기를 비교하는 방법을 택했다.

변수선택의 구체적인 과정은 다음과 같다.

(1) 독립변수 25개와 종속변수를 평균이 0 분산이 1이 되게끔 변수변환 (표준화)를 하였다.

(2) 이들 변수들에 대해서 앞 절에서 설명한 바와 같은 방법으로 구해진 $M_0=3$ 에 대해서 ppreg를 적용했다.

(3) 각 독립변수에 대해서, 가중 방향벡터의 합 ($\beta_1 a_1 + \beta_2 a_2 + \beta_3 a_3$)을 구해서, 이 값이 작을수록 종속변수에 영향력을 적게 미치는 변수로 생각했다.

모형2와 비교하기 위해서, 모형2에는 11개의 서로 다른 변수가 이용되고 있으므로, 위의 변수선택의 단계 (3)에서 가중 방향벡터의 합의 값이 작은 14개의 변수를 제거하였다. 이때 가중치로 사용된 회귀계수는 $\hat{\beta}_1=0.956$, $\hat{\beta}_2=0.559$, $\hat{\beta}_3=0.485$ 이었다. ppreg의 모형에서 제거되지 않은 변수 11개를 검토해 본 결과 (아래에 제시 됨), 8개의 변수가 모형2에서도 공통되게 이용되고 있었다.

ppreg에서 선택된 11개 변수들 (* 표시된 8개 변수는 모형2에서도 공통되게 사용됨).

* aw6	* lw1	* cl1	cl4	* ca4		
al6	* cg4	* aw7	* load	* rimwd	rimdi	

양 모형에서 공통되게 사용된 8개 변수는 일단 매우 중요한 변수로 간주된다. 또한 공통되게 사용되지는 않았더라도 양 모형에서 별도로 포함된 6개 변수들 (a_width), temp, l_width2, a_lenth6, c_lenth4, rimdi) 는 의미있는 변수들로 생각할 수 있다. 한편 양 모형에 전혀 포함되지 않은 나머지 11개의 변수들:

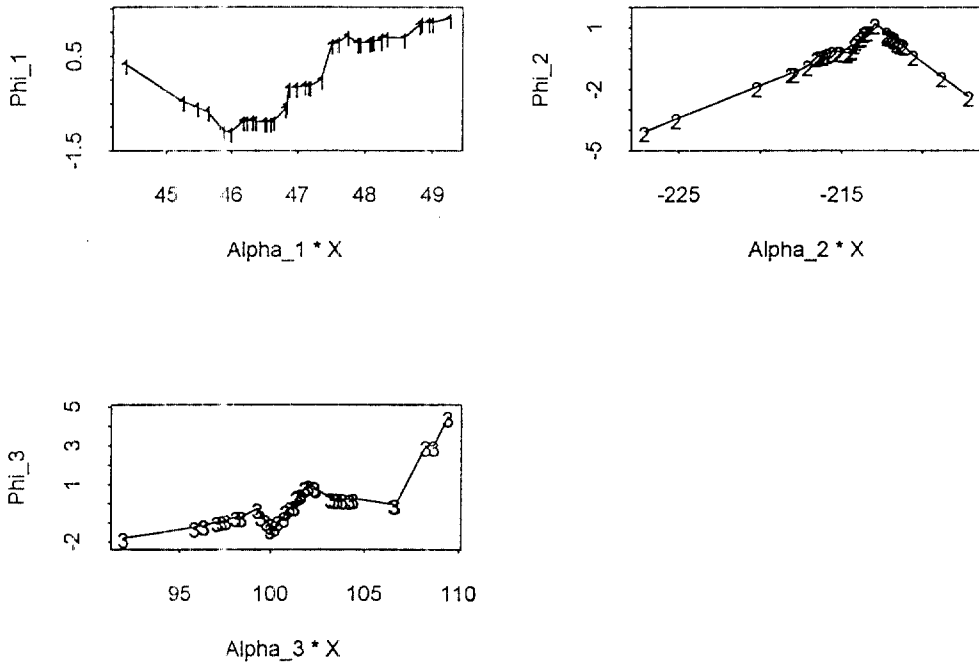
al1,	aw2,	aw3,	ll1,	lw3,	cg1,
cw1,	cw4,	infl,	tdr,	hard	

은 cc1 을 설명하거나 예측하는데 거의 불필요한 변수로 볼 수 있다.

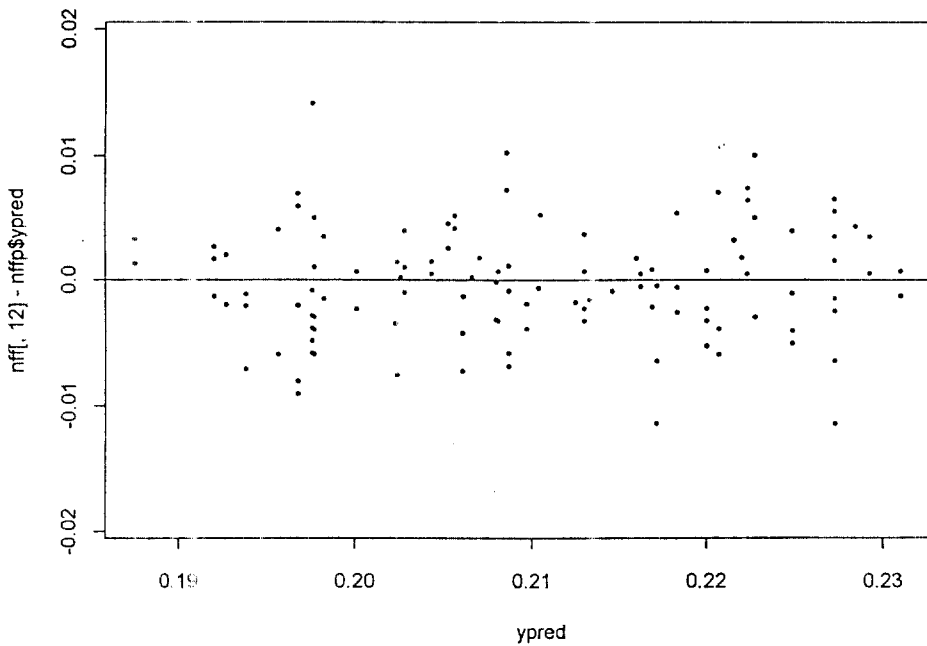
모형2와 ppreg를 비교하기 위해서, 모형2에서의 11개의 독립변수와 ppreg에서의 11개의 변수에 대해 130개의 주어진 자료 (표준화된 자료가 아닌 원자료)로 부터 각각 cc1을 적합해 보았고, 그 적합치의 차이를 비교해 보았다. ppreg에서의 결정계수는 0.8685 였고, 아래의 <그림 5.1>은 독립변수의 선형결합 ($a'_m x$) 대 ϕ_m 의 $m=1, 2, 3$ 에 대한 산포도이다. 그림에서 ϕ_m 의 일부분에서 부드럽지 못한 이유는 일부 변수의 실험점이 골고루 퍼져 있지 못하기 때문이다.

<그림 5.2>는 ppreg의 잔차그림이다. 이를 모형2의 잔차 <그림 3.2>와 비교해 볼 때, ppreg의 잔차가 모형2의 잔차보다 X 축 (Y=0 인 수평선)에 더 밀집되게 분포되어 있음을 알 수 있다.

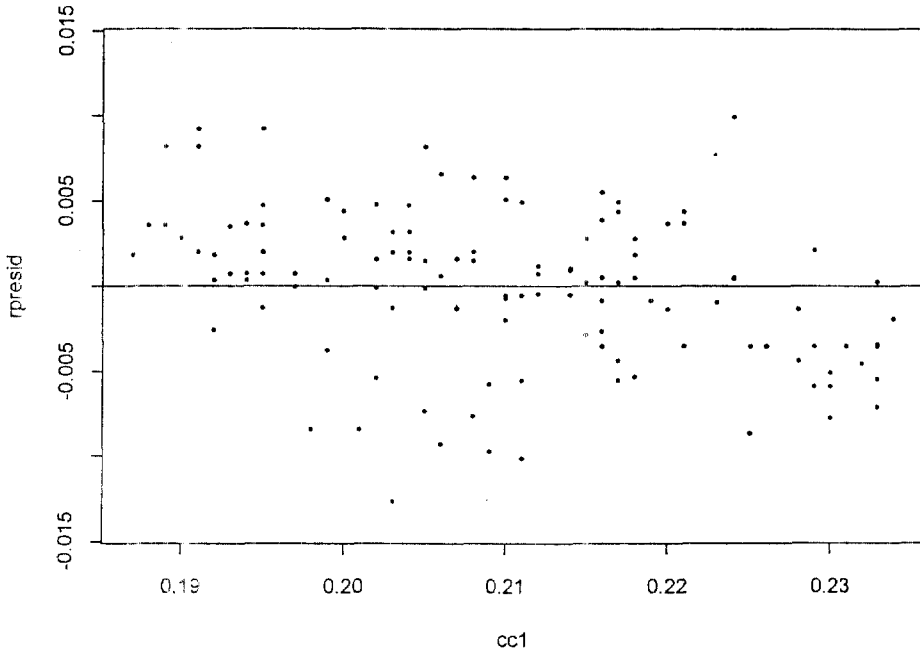
한편 모형2의 적합치와 ppreg의 적합치의 상관계수는 0.9202로 상당히 높게 나타나 밀접한 연관성을 보이고 있다. <그림 5.3>은 모형2의 적합치에서 ppreg의 적합치와 잔 값 (rpresid)을 cc1값에 대해서 산포도를 그린 것이다. 이 그림으로부터 알 수 있는 내용은, 첫째 rpresid의 수직 범위가 <그림 3.2>나 <그림 5.2>에서의 잔차의 수직 범위보다 작다는 점이다. 이는 모형2의 적합치와 ppreg의 적합치의 차이가 각 모형 (특히 모형2)에서의 잔차보다 작다는 뜻으로, 모형2와 ppreg의 유사성을 보여준다.



< 그림 5.1 > 독립변수의 선형결합 ($a_m^t x$) 대 ϕ_m 의 $m=1, 2, 3$ 에 대한 산포도



< 그림 5.2 > 11개의 변수에 대해 실행된 ppred의 잔차 그림



< 그림 5.3 > cc1의 값에 대한 rpresid (=모형2의 적합치-ppreg 의 적합치)의 산포도

둘째로 알 수 있는 점은, cc1의 작은 값에 대하여 rpresid가 영보다 대체로 크게 나타났는데 이는 cc1이 작은 값일 때 모형2가 ppreg 보다 큰 값으로 적합하고 있음을 알 수 있다. 또한 cc1의 큰 값에 대하여 rpresid가 영보다 대체로 작게 나타났는데 이는 cc1이 큰 값일 때 모형2가 ppreg 보다 작은 값으로 적합하고 있음을 알 수 있다. 이러한 사실들을 <그림 3.2>와 <그림 5.2>와 결부하여 해석해 볼 때, ppreg는 전반적으로 잘 적합하고 있는 반면 모형2는 cc1의 작은 값과 큰 값에서 상대적으로 잘 적합을 못하고 있음을 알 수 있다 (<그림 3.2>를 보기 바람).

ppreg에서 M_0 를 고정시켰을 때, 독립변수의 선택에 관한 문제는 아직 해결책이 제시되지 않았다. 저자는 일반 회귀분석에서 처럼 25개의 변수에 대한 ppreg에 뒤로 부터의 변수제거법 (backward elimination)을 적용하여 보았다. 즉 하나의 변수를 뺀 24개의 변수에 대해 ppreg를 실행하여 결정계수의 감소량을 파악하여 가장 적게 감소하는 경우에 제외된 변수를 제거하는 것이다. 이 경우 정지 원칙은 제외되는 변수에 대해 F-검정을 실시한다. 그러나 여기서의 문제점은 일반 회귀분석에서와는 달리 ppreg에서는 변수 수가 줄어들며 따라 항상 결정계수가 감소하지 않는다는 것이다. 이러한 문제점 때문에 저자는 표준화된 자료의 방향벡터와 회귀계수를 이용하였으나, 이 방법에 대한 어떠한 이론적 뒷받침도 없다. 다만 위의 비교에서 모형2 보다 좋은 잔차 그림이 양호하므로 이러한 접근이 나쁘지는 않다는 것을 알 수 있다. 이 방법의 합리성을 밝히기 위한 연구 (toy 모델을 이용한 시뮬레이션 연구 등)가 필요하다고 생각한다.

6. 최적 실험계획법에 관한 논의

실험자료를 살펴보거나 또는 지금까지의 논의를 통하여 짐작할 수 있듯이, 원자료는 적절한 실험계획법의 개념이나 방법을 이용하지 않고 구해짐으로 인하여 적절한 회귀모형의 추정이나 통계적 분석에 어려움이 많다. 이 어려움은 특히 많은 수의 독립변수와 50%를 넘는 결측치의 수 때문에 더욱 가중되었다. 이런 점에서 앞으로 얼마간의 실험이 더 요구되며, 이를 위해서 지금까지 주어진 자료의 특성을 최대한 살리면서 앞으로의 실험은 회귀모형 추정을 위해 최적이 되도록하는 축차적 최적 실험계획의 구축이 절실하다 (실제로 많이 사용되는 반응표면 실험계획법이 이 경우 사용되지 어려운 이유는 먼저 독립변수의 갯수가 27개로 너무 많아서 고도의 부분실험을 적용한다고 할지라도 많은 수의 실험점이 요구되며, 둘째로는 축차적 실험계획의 수립이 어렵다는 점이다.)

따라서 차후의 효과적인 통계분석 및 회귀모형의 수립을 위하여는 (여러가지 최적실험계획법을 생각할 수 있으나) D-최적기준에 기초한 축차적 최적 실험계획점을 수치적으로 구하여 차후 실험에 적용할 수 있도록 하여야 할 것이다. 본 문제에서 축차적 D-최적 실험계획법의 과정을 간단히 살펴보면 다음과 같다 (Atkinson & Donev 1992).

회귀모형이

$$y = X\beta + \epsilon$$

의 형태로 주어질 때, X 를 계획행렬이라고 하는데 이는 각 독립변수에 대한 실험점의 집합으로서, D-최적계획은 행렬식 $|X'X|$ 을 최대로 하는 X 를 말한다. 이를 위해서 보통 컴퓨터를 이용하여 수치적 비선형 최적 알고리즘을 실행 함으로서 최적해 X 를 구한다.

본 논문에서는 cc1에 대해서만 회귀모형을 추정하였으나, 종래에는 모든 종속변수에 대해서 회귀분석을 할 것이므로, 실험계획의 구축에는 10개의 종속변수와 27개의 독립변수 모두를 고려하였다. 또한 기존의 실험점을 반영하기 위해서 모든 변수에 대해 결측치가 전혀 없는 자료를 뽑아 본 결과 79개의 자료가 완전한 형태를 가졌다. 회귀모형은 일단 2차항이나 교호작용을 고려하지 않은 1차항 만의 모형으로 설정하였다.

독립변수들에는 몇 가지 제약조건이 있는데, 첫째는 aw6와 aw7은 거의 같은 값을 가진다는 것으로, 본 연구자는 aw7을 제외시켰다 (이제 26개의 변수가 된다.) 두번째의 제약조건으로는 금호연구소에서 많은 변수들에 대하여 2,3개 또는 4개의 수준 값 만을 표준으로 사용하기 때문에, 그 수준 값들이 이론적으로는 변경이 가능 할지라도 현실적으로는 몇 개의 수준 값 이외에서는 실험 불가능하다는 것이다. 따라서 이러한 몇 개의 (일부 변수는 정수 값의) 수준 값에서만 실험점을 잡아야 한다. 이러한 제약조건 하에서의 최적화는 이론적으로 완전하지 못하고 또한 현실적으로도 적합한 알고리즘을 적용하기 어려우므로, 이들 수준 값에 대한 제한을 일단 없앤 상태에서 비선형

최적화 알고리즘을 실행시켜 최적해를 구한 다음, 이 최적해로부터 가장 가까운 이산 수준 값으로 실험계획을 할당하는 방법을 채택할 수 있겠다.

기존에 존재하는 26개의 변수와 79개의 자료로 이루어진 계획행렬을 X_0 라고 하고, 앞으로 구해질 n_1 개의 자료에 의해 만들어지는 계획행렬을 X_1 이라고 하자. 여기서 차기 실험횟수 n_1 은 전체 실험횟수가 100이 되도록 21로 잡았다.

이제 X_0 와 X_1 을 합한 계획행렬을

$$X = \begin{bmatrix} X_0 \\ X_1 \end{bmatrix}$$

이라고 놓으면,

$$X'X = [X_0'X_0 + X_1'X_1]$$

이 되며, 이는 26×26 인 양반정치 행렬 (positive semi-definite matrix)이 된다 (박동권 1995). 따라서 이를 다음과 같이 콜레스키합성 (Cholesky decomposition)을 할 수 있는데, 즉 $X'X = LL'$. 여기서 L은 l_{ii} 대각선 원소로 갖는 하삼각 행렬이다. 따라서

$$|X'X| = |L| \times |L'| = \prod_{i=1}^{26} l_{ii}^2$$

가 되므로 실제로는 $-\log |X'X| = -2 \sum_{i=1}^{26} \log(l_{ii})$ 를 최소로 하는 X_1 을 찾는 문제가 된다.

그런데 이는 $21 \times 26 = 567$ 차원의 최적화 문제로서 매우 복잡하고 시간이 오래 걸리는 수치적 최적화 문제가 된다 (수퍼컴퓨터를 이용한 최적화가 바람직 하겠다). 이를 위하여 IMSL (1987)의 DUMINF 라는 부프로그램 또는 NAGLIB의 E04jbf 라는 부프로그램을 이용할 수 있겠다. 이들 부프로그램은 $|X'X|$ 의 일차 미분벡터를 유한차이 (finite difference) 로 근사시켜서 구하는 Quasi-Newton 최적 알고리즘이다.

7. 요약, 논의 및 연구과제

본 연구를 통하여 독립변수의 갯수가 많고 자료 중 결측치가 많으며, 또한 상당수의 독립변수가 2 내지 3개 수준에서만 값을 가질 때, 종속변수와 독립변수들 간에 적절한 회귀모형을 구축하는 단계적 방법을 알아 보았다. 즉 제 6사양에서의 자료 130개를 이용하여 (고전적 접근방법에 의해서) 종속변수 cc1과 타이어의 설계 및 실험 인자들 간의 회귀모형으로 모형2를 구축하였다.

또한 컴퓨터 집중사용법 중의 하나인 projection pursuit 회귀모형 (ppreg)을 수립하여, 모형2와 비교하였으며 중요한 변수들과 불필요한 변수들을 가려냈다. 모형2의 방법이 중요한 변수들이 무엇이고 종속변수와 독립변수들 간에 어떤 관계가 있는가를 규명하는 역할을 한다면, ppreg의 방법은 실험해 보지 않은 새로운 입력에 대한 타이어 성능의 예측에 유용하다고 하겠다. 더불어 차후의 실험과 다변량 회귀모형의 수립을 위하여 축차적 D-최적 실험계획법에 관한 논의를 덧붙였다.

본 논문에서는 특히 안정성의 계수인 cc1에 대해서만 회귀모형을 세우고 독립변수들의 cc1에 영향을 미치는 정도를 알아 보았다. 다른 종속변수에 대해서도 본 논문에서 제시한 방법과 유사한 방법을 이용하여 같은 종류의 분석을 할 수 있겠다.

모형2 및 ppreg 모형에 포함된 11개의 서로 다른 독립변수들이 cc1에 구체적으로 어떠한 영향을 미치는 지를 아는 것은, 타이어의 성능향상을 위한 설계인자 및 시험조건에 적절값 선택의 측면에서 매우 중요하다. 모형2에 교호작용항이 없다면, 각 변수들의 영향력을 대강 알 수 있다. 실제로 모형2에는 1, 2차항과 교호작용항이 혼재하고 있으므로, 모형2만으로 직접 독립변수들의 영향력을 알 수는 없다. 이러한 어려움에 대한 하나의 대안으로서, 본 연구자는 원자료와 모형2에 기초한 자료를 혼합한 자료에 대하여 비모수적 가법모형 (nonparametric additive model) 중의 하나인 Avs (Additivity and VAriance Stabilizing transformation)를 적용하여 보았으나, 이 방법 또한 문제가 있어서 적절한 영향분석 (sensitivity analysis)을 실시하지 못했다. 이들 독립변수들의 영향력을 그림으로 나타낸다면, 대략의 최적운전 조건을 설정할 수 있을 것이다.

혹자는 독립변수들의 수가 많으므로 이에 대해 요인분석 (factor analysis)를 먼저 실시하여 몇 개의 요인으로 줄인 다음 그 요인점수를 이용하여 회귀모형을 세울 것을 제안 할 수 있으나, 여기서의 독립변수들은 실험자에 의해서 계획적으로 잡어진 실험점들이기 때문에, 변수들 간의 상관관계에 기초한 요인분석을 적용할 수 없다.

실제 자료에는 종속변수들에서 뿐만 아니라 독립변수들에서도 결측치가 많이 있으므로 이들 결측치를 통계적으로 추정하는 방법에 대해서는 본 연구에서 다루지 못했으나, 모형2에 기초한 역회귀 (inverse regression) 방법으로 종속변수 값이 주어졌을 때 하나의 독립변수에서의 결측치를 구할 수도 있겠다. 또한 다른 여러가지 방법으로 결측치 추정이 가능할 것이나 (Little and Rubin, 1987), 이는 이들 결측치가 어떤 체계적인 이유에서 발생되었는지에 대한 조사가 선행된 다음의 일이다.

실제 자료에는 여러 개의 종속변수가 있고, 또한 종속변수들 간에 이론적으로 그리고 실제로 상호 상관관계가 존재하므로, 각 종속변수에 대해 각기 다른 회귀식을 세우기 보다는 다변량 회귀분석법 (multivariate regression analysis) 및 다변량 ppreg가 더 바람직 할 것이다. 본 논문에서는 이에 대한 고려는 없었으며 차후의 연구과제 중 하나이다.

한편 조완현, 이정환 (1996)에 의해서 진행되고 있는 Biplot에 의한 접근 또한 변수들 간의 복합적인 관계를 그림으로 보여줄 수 있는 방법으로서, 본 논문의 결과와 상호보완적으로 타이어 세 특성 간의 관계 규명에 큰 도움을 줄 것이다.

참고문헌

- [1] 박동권 (1995), 「실험계획법」, 자유아카데미.
- [2] 박성현(1981), 「회귀분석」, 대영사.
- [3] 백세종 역 (1993), 「타이어 공학」, (주)금호타이어 연구소.
- [4] 정성용 (1994), 「타이어 코너링 특성을 이용한 차량 동특성 연구」, 박사학위 논문, 전남대학교 기계공학과.
- [5] 조완현, 황현식, 정성용, 이정환 (1994), "설계인자와 타이어 제특성 간의 회귀방정식 추정," (주)금호타이어 연구소 연구보고서.
- [6] 조완현, 이정환 (1996), "Biplot 이론과 타이어 제조공학에의 응용," 「응용 통계연구」에 투고.
- [7] Atkinson, A. and Donev, A.(1992), *Optimum Experimental Designs*, Clarendon Press, Oxford.
- [8] Friedman, J. H. and Stuetzle, W.(1981), "Projection pursuit regression," *Journal of American Statistical Association*, 76, 817-823.
- [9] Little, R. and Rubin, D. (1987), *Statistical Analysis with Missing Data*. John Wiley, New York.
- [10] Neter, J., Wasserman, W. and Kutner, M. (1989), *Applied Linear Regression Models*, Irwin, Homewood, IL.
- [11] IMSL Inc. (1987), *User's Manual Math/Library*, Vol. 3., IMSL Inc., Houston
- [12] SAS Inc. (1987), *SAS/STAT Guide for Personal Computers, Version 6.03 Edition*, SAS Institute Inc., Cary, NC.
- [13] S-Plus (1993), *S-PLUS for Windows User's Manual*, Vol. 2, Statistical Sciences, Inc., Seattle.
- [14] Thisted R.(1987), *Element of Statistical Computing*, Chapman-Hall, New York.
- [15] Weisberg S.(1985), *Applied Linear Regression*, 2/e, John Wiley, New York.