

Biplot 이론과 타이어 제조공학에의 응용¹⁾

조 완 현²⁾

요 약

일반적으로 타이어 제조공학에 있어서 측정되는 자료는 여러개의 실제인자와 성능 특성치가 사용된 다변량 자료행렬로 주어지는데, 이러한 자료행렬의 중요한 특성중의 하나는 각 반응값들이 서로 다른 것들과 높은 상관관계를 가질 수 있다는 것이다. 따라서 본 연구는 이러한 복잡한 성격을 갖는 자료행렬의 분석에 적합한 biplot 작성의 수학적 이론을 알아보고, 또한 각 변수들의 구조적 특성 및 내재한 상호 관련성을 다변량 분산분석 biplot과 다반응치 회귀모형 biplot을 이용하여 포괄적으로 고찰하였다.

1. 서론

제조기업에 있어서 제품의 품질 향상은 기업이 생존하기 위한 절대 절명의 목표이다. 과거에는 기업이 제품의 제조, 개발방법의 연구에 중점을 두었지만, 오늘날에는 보다 더 품질이 좋은 제품, 생산비용이 저렴한 제품, 소비자 요구에 완전한 만족을 줄 수 있는 제품을 생산할 수 있는 신 기술이 요구되고 있다.

또한 산업의 전문화에 의해 각 영역에서 점차 다양한 종류의 타이어를 요구하게 되고, 소비자들이 고품질의 타이어를 희망하게 되면서, 회사는 이러한 요구수준에 맞는 제품을 신속하게 개발해 내지 않으면 안되게 되었다. 따라서 이러한 시대의 목적에 부합하기 위해서 타이어의 재원, 구조 설계인자와 성능 특성값들의 측정자료에 대하여 최적의 분석을 실시하고, 예측모형을 위한 통계 시스템을 개발하여 이를 현장에 신속히 적용함으로써 고품질 제품의 타이어를 생산 제조할 수 있는 기술 향상에 기여하고자 본 연구를 시작하였다.

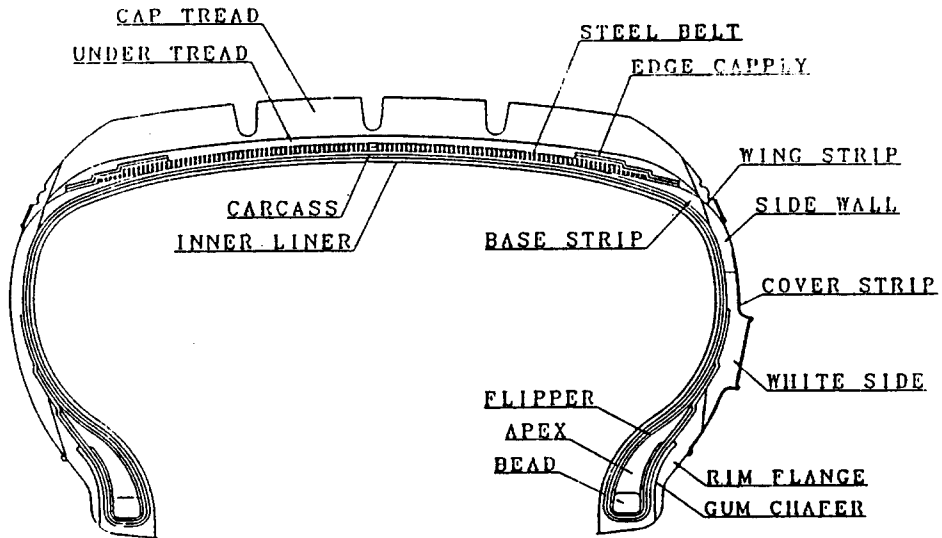
현재 여러 회사에서 생산하고 있는 타이어의 구조와 특성을 간단히 알아보면(백세종 (1993), 남일(1993)) 일반적으로 타이어 종류는 크게 두가지로 구분할 수 있는데, 이들은 각각 바이어스(BIAS)와 래디얼 플라이(RADIAL PLY) 타이어이다. 유럽에서 개발된 래디얼 타이어의 장점이 인식되기 시작한 1960년대까지 바이어스 타이어는 미국 자동차 산업의 표준으로 사용되어 왔다. 그 이후로 수 십년동안 승용차산업에서 바이어스 타이어가 래디얼 타이어로 대체되면서 오늘날에는 래디얼 타이어가 표준이 되었다. 그러나 아직도 승객이나 화물을 많이 실어나르는 버스나 트럭에서는 두 종류의 타이어가 동등하게 사용되고 있다. 또한 두 타이어의 중간단계로 바이어스 벨트(BIAS-BELT) 타이어가 단기적으로 사용되었지만 오늘날에는 거의 찾아볼 수가

1) 이 연구는 1994년도 전남대학교 학술진흥재단 연구조성비에 의하여 연구되었음.
2) (500-757) 광주광역시 북구 용봉동 300 전남대학교 통계학과 교수.

없다. 여기서 래디얼 타이어의 단면도를 그려보면 다음 <그림 1.1>과 같으며, 이것의 구조와 특성을 간단히 고찰해 보자.

주어진 단면도에서 표시된 바와 같이 타이어를 구성하는 부품은 크게 트레드(TREAD), 벨트(BELT), 카카스(CARCASS), 캐플라이(CAPPLY), 쇼울더(SHOULDER), 사이더 월(SIDE-WALL), 비드(BEAD), 플리퍼(FLIPPER) 등으로 나눌 수 있다.

먼저 카카스는 양측단에 비드 와이어(BEAD-WIRE)를 가지고 림 플랜지(RIM PLANGE)와 림 베이스(RIM BASE)를 부착하여 내압을 유지함과 동시에 하중을 지탱하는 가장 중요한 부분이며, 트레드는 타이어가 전동할때에 노면과 접촉하는 부분으로, 카카스를 보호함과 동시에 타이어나 노면간의 마찰계수를 확보하여 유효한 제동력이나 조정성을 발휘시키게 한다. 또한 벨트는 카카스 코오드(CORD)가 반경방향으로 향하고 있기 때문에 주방향의 강도가 높은 벨트를 조여 붙이지 않으면 평병비를 적게할 수가 없다. 따라서 벨트는 단일 브레이크(BRAKER)는 아니지만 반경방향으로 압축하는 역할을 하고 있으며, 캐플라이는 벨트위에 부착되는 코드지로 타이어의 구조적 강도를 지탱해주는 역할을 하고 있다. 사이드 월은 타이어의 측면을 보호하는 것이 주목적이거나 주행중에 끊임 없는 굴곡이나 신축에 대하여 일그러짐에 강하고, 승차감을 향상시키는 역할을 한다. 비드는 공기의 누설을 방지하여 내부공기압을 유지하도록 하고, 펑크(PUNCTURE)가 났을때 타이어가 림에서 떨어져 나가는 것을 방지하며 림 플랜지와 마찰로 인한 타이어의 마모를 방지한다. 끝으로 플리퍼는 비드에 가해지는 충격을 완화시켜주며, 성형시 비드부에 공기 입력을방지하는 역할을 한다.



< 그림 1.1 > 래디얼 타이어의 단면도

또한 타이어의 역학적 성능을 평가하는 항목으로는, 일반적으로 6 항목이 사용되는데, 먼저 타이어의 안정성을 나타내는 코너링(CONERING) 계수와 코너링 파워(CONERING POWER)가 있으며, 조정성 평가는 얼라이팅 토그(ALIGNING TORQUE) 계수, 노이마틱 트레일(PNEUMATIC TRAIL)등이 사용된다. 또한 코너링시에 선회궤도의 안쪽과 바깥쪽 타이어들 사이에서 발생하는 하중전달에 대한 감도는 로드 트랜스퍼(LOAD TRANSFER)로 평가하고, 수직하중 자체의 증가나 감소에 대한 횡방향 안정성은 로드 감각도(LOAD SENSITIVITY)로 측정한다.

Biplot의 이론은 처음으로 Gabriel(1971)에 의해서 제안되었으며, 그는 이 논문에서 biplot의 작성요령과 주성분 분석에의 응용을 고찰하였으며, 또 다른 논문(Gabriel(1972))에서는 실제 측정자료를 이용한 다변량 분산분석에 관한 문제를 생각하였다. 그리고 Smith 와 Cornell(1993)등은 혼합실험에서 측정되는 다반응치 자료에 대한 회귀계수들의 biplot 표현방법을 고찰하였고, 최근에는 Gower(1992), Daigle 과 Rivest(1992)등에 의하여 biplot의 일반화 문제와 러버스트 biplot의 개념등이 제안되었다. 우리나라에서는 최용석(1991)에 의하여 저항 주성분 분석이론이 연구되고 biplot 및 대응 분석등에 대한 적용방법이 고찰되었으며, 최용석,심의경(1994)에 의해서 한국 프로야구 성적분석에 사용되었고, 또한 장대홍(1994)은 관측된 자료로부터 추정된 회귀계수에 대한 biplot에의 응용방법을 제시하고 있다.

끝으로 본 논문은 제2절에서 biplot의 수학적 이론을 고찰하였으며, 제3절에서는 생산현장에서 얻어진 실제 측정자료에 적용하여 자료분석에 어떻게 이용될 수 있는가를 생각하였다. 마지막으로 제4절에는 토의와 요약을 제시하였다.

2. Biplot의 수학적 이론과 통계적 성질

Biplot은 고차의 계수(rank)를 갖는 자료 행렬을 저차의 계수(보통 계수가 2이다)를 갖는 행렬로 근사적으로 표현하고, 2차원 평면상에 그림으로 도식하여 각 변수들과 관측치들 사이에 구조적 관련성을 파악할 수 있는 통계적 분석방법이다(Daigle과 Rivest(1992), Gabriel(1971), Smith와 Cornell(1993)).

먼저 \mathbf{X} 를 계수가 r ($r \leq \min(n,m)$) 인 $(n \times m)$ 자료행렬이라 하자. 이 때 비정칙분해(Singular Value Decomposition: SVD) 를 사용하면 이 행렬을 다음과 같이 나타낼 수 있다.

$$\mathbf{X}_{n \times m} = \mathbf{P}_{n \times r} \mathbf{A}_{r \times r} \mathbf{Q}_{r \times m}^t, \quad (2.1)$$

여기서 행렬 $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_r)$ 의 열벡터는 $\mathbf{X}\mathbf{X}^t$ 의 정칙 고유벡터(orthonormal eigenvector)들이고, 행렬 $\mathbf{Q} = (\mathbf{q}_1, \dots, \mathbf{q}_r)$ 의 열벡터는 $\mathbf{X}^t\mathbf{X}$ 의 정칙 고유벡터들이다. 즉

$\mathbf{P}^t\mathbf{P} = \mathbf{Q}^t\mathbf{Q} = \mathbf{I}_{r \times r}$ 이다. 또한 $\mathbf{A}_{r \times r}$ 는 주대각선 원소가 크기 순서로 된 비정칙값(singular values) $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ 들을 갖는 주대각 행렬이며, 이들 비정칙값들은 정방행렬 $\mathbf{X}^t\mathbf{X}$ 또는 $\mathbf{X}\mathbf{X}^t$ 의 고유치들의 양의 제곱근이다. 따라서, 행렬 $\mathbf{X}_{n \times m}$ 를 다음과 같은 합의 형태로 다시 표현할 수 있다.

$$\mathbf{X} = \sum_{k=1}^r \lambda_k \mathbf{p}_k \cdot \mathbf{q}_k^t \quad (2.2)$$

그러면 우리는 행렬 $\mathbf{X}_{n \times m}$ 을 다음과 같이 정의되는 순위가 2 인 $\mathbf{X}_{n \times m}^*$ 로 근사시킬 수 있다.

$$\begin{aligned} \mathbf{X}^* &= \lambda_1 \mathbf{p}_1 \cdot \mathbf{q}_1^t + \lambda_2 \mathbf{p}_2 \cdot \mathbf{q}_2^t \\ &= [\mathbf{p}_1, \mathbf{p}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^t \\ \mathbf{q}_2^t \end{bmatrix}. \end{aligned} \quad (2.3)$$

이 경우, 방정식(2.3)에서 정의된 행렬 $\mathbf{X}_{n \times m}^*$ 은 행렬의 유클리디안 거리(Euclidean norm)를

$$\| \mathbf{X}_{n \times m} - \mathbf{X}_{n \times m}^* \|$$

최소화하는 의미에서 행렬 $\mathbf{X}_{n \times m}$ 의 최적 근사행렬이다. 또한 여기서 최적성(the goodness-of-fit)의 정도는 $(\lambda_1^2 + \lambda_2^2) / \text{tr}(\mathbf{X}^t \mathbf{X})$ 로 측정될 수 있다.

행렬 $\mathbf{X}_{n \times m}^*$ 에 대한 biplot를 작성하기 위하여, $\mathbf{X}_{n \times m}^*$ 에 대하여 두가지 종류의 행렬분해(Matrix Factorization)를 실시할 수 있다. 첫번째 방법은 $\mathbf{G} \cdot \mathbf{H}^t$ 분해로서 이것을 \mathbf{h} -biplot 라고 하는데 다음과 같이 정의된다. 즉

$$\mathbf{X}_{n \times m}^* = \mathbf{G} \cdot \mathbf{H}^t$$

로 주어지고, 여기서

$$\mathbf{G}_{n \times 2} = [\mathbf{p}_1, \mathbf{p}_2] \quad , \quad \mathbf{H}_{m \times 2} = [\lambda_1 \mathbf{q}_1, \lambda_2 \mathbf{q}_2] \quad (2.4)$$

이다. 이때 $\mathbf{G}\mathbf{H}^t$ 분해에서 biplot을 위한 행표시자(또는 벡터) $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$ 는 행렬

$\mathbf{G}_{n \times 2}$ 의 n 개의 행벡터들이고, 열표시자(또는 벡터) $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_m$ 은 행렬 $\mathbf{H}_{m \times 2}$ 의 m 개의 행벡터이다. 그러면, 이 경우 작성된 \mathbf{h} -biplot으로 부터 얻을 수 있는 통계적 성질은 먼저 이 것이 변수들의 분산공분산 행렬에 대한 그래프적 표현이다. 왜냐하면 행렬 \mathbf{G} 의 각 열들이 정칙벡터이므로 다음이 성립하고,

$$\mathbf{X}^{*t} \cdot \mathbf{X}^* = \mathbf{H} \cdot \mathbf{H}^t .$$

따라서, 만약에 행렬 \mathbf{X} 의 각 열들이 중심화되어 있으면, 그때 각 열벡터 $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_m$ 들의 길이는 대응되는 변수들의 표준편차에 근사되며, 또한 이 들 임의의 두 벡터 \mathbf{h}_i 와 \mathbf{h}_j 의 각에 대한 코사인(cosine) 값은 대응되는 두 변수의 상관계수에 근사적으로 일치한다.

다음으로 주어진 그림에서 각 관측치들 사이의 거리는 마하라노비스(Mahalanobis) 거리의 의미를 갖는다. 왜냐하면

$$\mathbf{G} \cdot \mathbf{G}^t = \mathbf{X}^* \mathbf{Q} \mathbf{A}^{-2} \mathbf{Q}^t \mathbf{X}^{*t} \approx \mathbf{X} \boldsymbol{\Sigma}^{-1} \mathbf{X}^t$$

이 성립하고, 따라서 두 행표시자 g_i 와 g_j 간의 거리는 원래 주어진 측정 자료에서 두 관측치들의 마하라노비스 거리를 나타낸다.

두번째 방법은 $J \cdot K^t$ 분해를 이용하는 것으로, 이때 주어진 행렬 $X_{n \times m}^*$ 는 다음과 같이

$$X_{n \times m}^* = J \cdot K^t$$

로 분해되고, 여기서

$$J_{n \times 2} = [\lambda_1 p_1, \lambda_2 p_2] , \quad K_{m \times 2} = [q_1, q_2] \quad (2.5)$$

으로 주어진다. 이 그림을 j -biplot로 부른다. 그리고, 위의 행렬분해에서 행표시자 j_1, j_2, \dots, j_n 는 행렬 $J_{n \times 2}$ 의 행벡터이고, 열표시자 k_1, k_2, \dots, k_m 은 행렬 $K_{m \times 2}$ 의 행벡터이다. 이 그림은 주어진 관측치들에 대한 고전적 다차원 척도 (multidimensional scaling) 의미로 해석될 수 있다. 즉 행렬 J 의 임의의 행벡터 j_i 는 주어진 자료행렬 X 에 서 관측치 i 에 대응되는 biplot 그림에서 좌표값을 나타내고, 행렬 K 의 열들이 정칙벡터이므로 다음을 만족한다.

$$X^* \cdot X^{*t} = J \cdot J^t$$

따라서 JJ^t 의 각 원소들은 biplot 상의 점에 대응되는 관측치들의 유클리디안 거리에 근사적으로 일치한다.

끝으로 우리는 자료행렬 X 의 각 원소 x_{ij} 는 벡터의 내적의 정의에 의하여 h -biplot 와 j -biplot 에서 대응되는 행표시자와 열표시자의 내적으로 주어진다.

$$x_{ij}^* = \begin{cases} \| g_i \| \cdot \| h_j \| \cos \theta_{ij}, & G \cdot H^t \text{ 분해인 경우} \\ \| j_i \| \cdot \| k_j \| \cos \theta_{ij}, & J \cdot K^t \text{ 분해인 경우} \end{cases}$$

따라서 이러한 의미로 만약에 벡터 g_i (또는 j_i)가 벡터 h_j (또는 k_j)와 같은 방향이고 이들 벡터의 길이가 크면 원소 x_{ij}^* 의 값은 양의 큰 수로 주어진다. 그러나, 역으로 g_i (또는 j_i) 의 방향이 h_j (또는 k_j)의 방향과 반대방향이면 음수의 큰값을 갖으며, 반면에 두 벡터가 직교하면 이값은 영에 가까워 진다.

3. 타이어 실험자료에 대한 biplot의 적용

3.1 다변량 분산분석에서 biplot의 적용

만약 m 개의 반응변수를 사용하여 k 개의 모집단의 평균이나 처리효과(treatment effect)의 비교를 실시할때 우리는 다음과 같은 확률벡터들을 얻게된다.

$$\mathbf{X}_{11}, \mathbf{X}_{12}, \dots, \mathbf{X}_{1n_1}, \dots, \mathbf{X}_{k1}, \mathbf{X}_{k2}, \dots, \mathbf{X}_{kn_k}$$

이때 주어진 측정된 자료에 대하여 biplot을 그리기 위하여, 먼저 i 번째 표본에 대한 평균벡터와 전체 평균벡터를 아래식과 같이 계산하고,

$$\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}, \quad \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbf{x}_{ij}, \quad n = \sum_{i=1}^k n_i$$

이를 이용하여 전체평균과 각 평균의 차를 나타내는 $\mathbf{X}_{(k \times m)}$ 행렬과 표본내의 제곱합과 교차합을 나타내는 $\mathbf{S}_{(m \times m)}$ 행렬을 다음과 같이 정의하자.

$$\mathbf{X}_{(k \times m)} = \begin{bmatrix} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}})^t \\ \vdots \\ (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^t \end{bmatrix}, \quad \mathbf{S}_{(m \times m)} = \sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^t$$

그리고 $\mathbf{N}_{(k \times k)}$ 행렬은 주대각선원소가 표본의 크기인 대각행렬이라 하자.

그러면 다변량 분산분석에서 biplot은 행렬 \mathbf{X} 의 행은 표본크기 행렬 $\mathbf{N}_{(k \times k)}$ 에 의해 가중되고, \mathbf{X} 의 열은 변수들의 제곱합과 교차합행렬 $\mathbf{S}_{(m \times m)}$ 의 역행렬에 의해 가중되는 다음과 같은 형태의 행렬에 실행된다.

$$\mathbf{Y} = \mathbf{N}^{\frac{1}{2}} \mathbf{X} \mathbf{S}^{-\frac{1}{2}}, \quad (3.1)$$

여기서 $\mathbf{S}^{-\frac{1}{2}} \mathbf{S}^{-\frac{1}{2}} = \mathbf{S}^{-1}$ 이다.

이때 행렬 \mathbf{Y} 에 대한 biplot을 작성하기 위하여 $\mathbf{G} \cdot \mathbf{H}^t$ 분해를 실시해야 되는데, 이것은 다음과 같은 행렬의 고유치와 고유벡터가 필요하다.

$$\mathbf{Y}' \mathbf{Y} = \mathbf{S}^{-\frac{1}{2}} \mathbf{X}' \mathbf{N} \mathbf{X} \mathbf{S}^{-\frac{1}{2}},$$

그런데 우리는 계산을 간단히 하기 위해 $\mathbf{Y}' \mathbf{Y}$ 대신에 $\mathbf{X}' \mathbf{N} \mathbf{X} \mathbf{S}^{-1}$ 을 사용할 수 있고, 이 행렬의 가장 큰 두개의 고유치 λ_1^2, λ_2^2 을 구하여 이들에 대응되는 고유벡터 $\mathbf{q}_1, \mathbf{q}_2$ 를 계산하고,

이를 이용하여 벡터 $\mathbf{w}_1, \mathbf{w}_2$ 를 다음과 같이 정의하자.

$$\mathbf{w}_i = \mathbf{S}^{\frac{1}{2}} \cdot \mathbf{q}_i, \quad i=1, 2.$$

그때 $\mathbf{G} \cdot \mathbf{H}^t$ 분해에서 행렬 \mathbf{G} 와 \mathbf{H} 를 다음의 행렬로 취하고,

$$\mathbf{G} = \mathbf{X} \mathbf{S}^{-1} (\mathbf{w}_1, \mathbf{w}_2), \quad \mathbf{H} = (\mathbf{w}_1, \mathbf{w}_2). \quad (3.2)$$

최종적으로 얻어진 행렬 \mathbf{G} 의 행벡터 $\mathbf{g}_1, \dots, \mathbf{g}_k$ 와 행렬 \mathbf{H} 의 행벡터 $\mathbf{h}_1, \dots, \mathbf{h}_m$ 를 이차원

평면상에 도식하면 다변량 분산분석 biplot을 얻게된다. 이 biplot 으로부터 우리는 다음과 같은 통계적 성질을 얻을 수 있다.

첫째로 각 열표시자 벡터 \mathbf{h}_j 는 각 변수들의 그룹내 자승합의 성격을 나타낸다. 즉 벡터 \mathbf{h}_j 의 길이는 각 변수의 표준편차를 표현하고, 또한 임의의 두 벡터 \mathbf{h}_i 와 \mathbf{h}_j 의 두 사이의 각의 코사인 값은 두 변수의 상관계수에 근사된다. 왜냐하면

$$\mathbf{H}\mathbf{H}^t = (\mathbf{w}_1, \mathbf{w}_2)(\mathbf{w}_1, \mathbf{w}_2)^t = \mathbf{S}^{\frac{1}{2}}(\mathbf{q}_1\mathbf{q}_1^t + \mathbf{q}_2\mathbf{q}_2^t)\mathbf{S}^{\frac{1}{2}} \approx \mathbf{S}.$$

둘째로 각 행표시자 벡터 \mathbf{g}_i 들은 6가지 주어진 사양들을 표현하는데, Biplot 상에서 임의의 두 벡터 \mathbf{g}_i 들간의 거리는 대응하는 사양들의 마하라노비스 거리에 근사된다. 왜냐하면

$$\mathbf{G}\mathbf{G}^t = \mathbf{X}\mathbf{S}^{-1}(\mathbf{w}_1, \mathbf{w}_2)(\mathbf{X}\mathbf{S}^{-1}(\mathbf{w}_1, \mathbf{w}_2))^t = \mathbf{X}\mathbf{S}^{-1}(\mathbf{w}_1\mathbf{w}_1^t + \mathbf{w}_2\mathbf{w}_2^t)\mathbf{S}^{-1}\mathbf{X}^t \approx \mathbf{X}\mathbf{S}^{-1}\mathbf{X}^t.$$

셋째로 각 행표시자 벡터 \mathbf{g}_i 와 열표시자 벡터 \mathbf{h}_j 의 내적은 각 요인들에 대응하는 처리효과와의 추정값이다. 왜냐하면 \mathbf{g}_i 는 다음 행렬 \mathbf{G} 의 i 번째 행이고, \mathbf{h}_j 는 \mathbf{H} 의 j 번째 행이므로

$$\mathbf{G}\mathbf{H}^t = \mathbf{X}\mathbf{S}^{-1}(\mathbf{w}_1\mathbf{w}_1^t + \mathbf{w}_2\mathbf{w}_2^t) \approx \mathbf{X}.$$

지금까지 고찰한 이론을 실제로 측정된 타이어 자료에 적용해 보자. 일반적으로 타이어의 제조사양은 크게 캐플라이의 종류(Full cappy, Edge cappy, No cappy)와 플리퍼(Flipper)의 사용 유·무에 따라 6 가지로 분류되며, 또한 타이어의 역학적 특성은 서론에서 고찰한 것처럼 코너링 계수, 얼라이링 토그 계수등 6 가지 항목으로 평가된다.

따라서 이들 각 제조사양들과 특성항목들에 대한 측정자료에 대하여 각 수준별 평균을 계산하여 다음 <표 3.1> 과 같이 요약하였다. 이 <표 3.2> 는 또한 각 특성항목들간의 상관계수를 보여 주고 있다. 이 결과로부터 우리는 대략적으로 다음과 같은 결론을 얻게 된다.

- (가) 플리퍼를 사용하는 경우에 코너링 계수와 얼라이링 토그 계수, 코너링 파워워등의 성능 특성값을 높게 만드는 반면, 사용하지 않는 경우에는 로드 트랜스퍼와 뉴우메틱 트레일을 높게 함을 알 수 있다.
- (나) 캐플라이는 Edge캐플라이를 사용하는 경우가 뉴우메틱 트레일을 제외하고 대부분의 성능을 높게한다.
- (다) 종합적으로 Edge 캐플라이, 플리퍼 사용의 제조사양이 추천됨을 알 수 있다.

또한 주어진 자료는 두가지 요인에 대한 6가지 항목에 대해서 측정된 것이므로 이원분류 다변량 분산분석법(2-way MANOVA)을 적용하여 분석을 실시하면, 다음 <표 3.3> 과 <표 3.4> 같은 제곱합 행렬과 검정결과를 얻는다.

< 표 3.1 > 각 사양별 반응치들에 대한 평균

반응변수 flipper capply		Cornering Coefficient	Aligning Coefficient	Cornering Power	Load Transfer	Load Sensitivity	Pneumatic Trail	Sample Size
		CC4	ATC1	CP42	GF4	HF1	PT4	
사 용	Full	0.66672	8.1062	64.52896	0.28604	0.14464	21.88708	25
	Edge	0.77379	10.180177	68.7228	0.29515	0.15798	19.24008	34
	No	0.63355	8.028107	59.90091	0.272314	0.12388	21.9627	121
미 사 용	Full	0.65561	8.41415	53.10615	0.29731	0.12853	22.0919	13
	Edge	0.637186	7.19834	51.19535	0.30024	0.11813	20.92046	209
	No	0.615588	6.780855	48.31895	0.29958	0.10177	20.473225	222
총평균		0.637806	7.434887	53.3891	0.2936106	0.116877	20.9350	624

< 표 3.2 > 반응변수들에 대한 표준편차 및 상관계수 행렬

반응변수	CC4	ATC1	CP42	GF4	HF1	PT4
표준편차	0.0662	1.4783	12.2320	0.0556	0.0542	5.2133
반응변수	CC4	ATC1	CP42	GF4	HF1	PT4
CC4	1.000	0.3118	-0.3141	0.0201	0.5683	-0.5228
ATC1		1.000	0.3290	0.7534	-0.5066	0.5475
CP42			1.000	0.1893	-0.3937	0.6002
GF4				1.000	-0.7180	0.6393
HF1					1.000	-0.8448
PT4						1.000

< 표 3.3 > 다변량 분산분석에서 제곱합 행렬

㉑ Flipper 사용에 대한 제곱합 행렬

CC4	ATC1	CP42	GF4	HF1	PT4
0.17	3.96	40.9	-0.04	0.08	-0.4
3.96	93.8	9670	-1.04	1.87	-9.5
40.9	970	10024	-10.7	19.3	-97.9
-0.04	-1.04	-10.7	0.01	-0.02	0.11
0.08	1.87	19.3	-0.02	0.04	-0.19
-0.4	-9.5	-97.9	0.11	-0.19	0.96

㉒ Cappy 종류에 대한 제곱합 행렬

CC4	ATC1	CP42	GF4	HF1	PT4
0.56	9.09	41.76	0.08	0.18	-7.34
9.09	148.7	690	1.31	3.03	-107
41.76	690	3222	6.05	14.24	-460
0.08	1.31	6.05	0.01	0.03	-0.04
0.18	3.03	14.42	0.03	0.06	-1.87
-7.34	-107	-460	-0.04	-1.87	147.3

㉓ 두 요인의 교호작용에 대한 제곱합 행렬

CC4	ATC1	CP42	GF4	HF1	PT4
0.31	5.01	15.6	0.05	0.04	-7.8
5.01	96.6	248.1	0.67	0.83	103.7
15.6	248	778	2.69	2.4	-390
0.05	0.67	2.69	0.01	0.008	-1.6
0.05	0.83	2.4	0.008	0.008	-1.13
-7.8	-103.7	-390.6	-1.6	-1.13	219.4

㉔ 잔차에 대한 제곱합 행렬

CC4	ATC1	CP42	GF4	HF1	PT4
2.71	18.87	-157	0.05	1.26	-112
18.87	1351	3677	38.27	-25	2607
-157	3677	97266	79.58	-161	23652
0.05	38.27	79.58	1.91	-1.34	114.5
1.26	-25	-161	-1.34	1.8	-148
-112	2607	23650	114.5	-148	219.4

< 표 3.4 > 각 요인에 대한 유의성검정

요 인	검 정 통 계 량	임계치	자유도	p-value
플리퍼 주효과	Wilks' Lamda	0.7458	613	0.0001
	Pillai's Trace	0.2541	613	0.0001
	Hotelling-Lawley Trace	0.3406	613	0.0001
캐플라이 주효과	Wilks' Lamda	0.7063	1226	0.0001
	Pillai's Trace	0.3079	1228	0.0001
	Hotelling-Lawley Trace	0.3956	1224	0.0001
교호작용	Wilks' Lamda	0.7521	1226	0.0001
	Pillai's Trace	0.2633	1228	0.0001
	Hotelling-Lawley Trace	0.3092	1224	0.0001

따라서 위의 결과로부터 두가지 요인(Flipper 와 Cappy)의 주효과 및 교호작용이 유의하므로, 이들 효과들에 대한 각 반응변수들의 추정치 행렬을 계산하면 다음 < 표 3.5 > 와 같다.

< 표 3.5 > 각 처리효과에 대한 추정치 행렬

반응변수 처리효과	CC4	ATC1	CP42	GF4	HF1	PT4
a_1	0.018	1.247	11.582	-0.027	0.022	1.47
a_2	-0.018	-1.247	-11.582	0.027	-0.022	-1.47
b_1	0.040	-1.633	4.787	-0.0024	0.027	1.616
b_2	0.022	0.142	2.877	0.0004	0.016	0.447
b_3	-0.062	-1.775	-7.664	0.002	-0.043	-2.066
$a_1 b_1$	-0.007	-1.555	-0.159	0.016	-0.006	-1.694
$a_1 b_2$	0.119	1.734	5.945	0.022	0.018	-3.17
$a_1 b_3$	-0.112	-0.179	-5.786	-0.038	-0.012	4.864
$a_2 b_1$	0.007	1.555	0.159	-0.016	0.006	1.694
$a_2 b_2$	-0.119	-1.734	-5.945	-0.022	-0.018	3.17
$a_2 b_3$	0.112	0.179	5.786	0.038	0.012	-4.864

* a_1 : 플리퍼 사용 a_2 : 플리퍼 미사용
 b_1 : Full capply b_2 : Edge capply b_3 : No capply

여기서 우리는 이 추정치행렬에 대한 biplot을 작성하여 보자. 먼저 행렬 $\mathbf{E}^t \mathbf{E}$ 의 가장 큰 두개의 고유치 λ_1^2, λ_2^2 을 구하고, 이들에 대응되는 고유벡터 $\mathbf{q}_1, \mathbf{q}_2$ 를 계산하여 벡터 $\mathbf{w}_1 = \mathbf{S}^{\frac{1}{2}} \mathbf{q}_1, \mathbf{w}_2 = \mathbf{S}^{\frac{1}{2}} \mathbf{q}_2$ 얻는다. 여기서 \mathbf{S} 는 < 표 3.2 >에서 주어진 상관계수 행렬이다.

이때 행표시자 행렬 \mathbf{G} 와 열표시자 행렬 \mathbf{H} 를 각각 다음과 같이 취하고,

$$\mathbf{G} = \mathbf{E} \cdot \mathbf{S}^{-1} \cdot (\mathbf{w}_1, \mathbf{w}_2), \quad \mathbf{H} = (\mathbf{w}_1, \mathbf{w}_2)$$

이들의 행벡터들로 이차원 그림을 작성하면 다음 <그림 3.1> 과 같은 다변량 분산분석 biplot을 얻게된다.

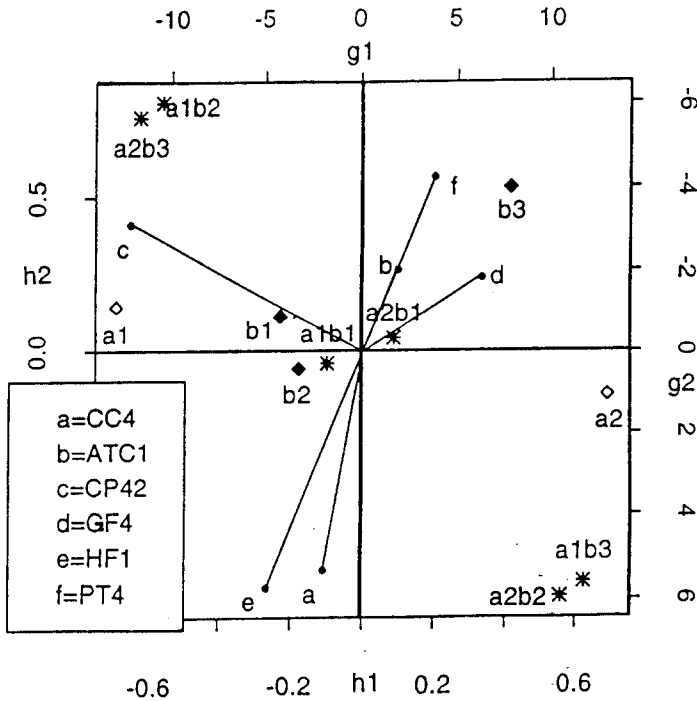
또한 이 biplot 으로부터 우리는 다음과 같은 결과를 얻을 수 있다.

(가) 코너링 계수(CC4)와 로드 센서티비티(HF1) 그리고 얼라이닝 토그 계수(ATC1), 코너링 파우워(GF4) 및 뉴우메틱 트레일(PT4)등 각 그룹 안에서는 정상관이 있으며 반면에 각 그룹 간에는 역상관이 있다. 그리고 코너링 파우워(CP42)은 이들 각 특성과 상관이 희박하다.

(나) 먼저 6가지 제조 사양은 크게 4가지 집락, 즉 $a_1 b_2$ (FTEC), $a_2 b_3$ (FLNC) 및 $a_1 b_3$ (FTFC), $a_2 b_3$ (FLNC) 그리고 $a_1 b_1$ (FTFC)와 $a_2 b_1$ (FLFC)로 주어지며, 이들간에는 통계적으

로 유의적인 성질을 갖는다.

(다) 코너링 파워워(CP42)는 플리퍼의 사용과 높은 관련이 있으며, 또한 Full 캐플라이와도 상관성이 높다. Edge 캐플라이는 코너링계수(CC4)와 로드 센서티비티(HF1)에 관련이 깊고, No 캐플라이는 엘라이닝 토크계수(ATC1), 뉴우메틱 트레일(PT4), 그리고 코너링파워워(GF4) 들과 양의 상관성이 있다.



< 그림 3.1 > 타이어 자료에 대한 다변량 분산분석 biplot

3.2 다반응 회귀 모형분석에서 biplot 의 이용

먼저 m 개의 반응변수에 대한 k 개의 설명변수들의 다중 회귀모형을 생각해 보자. X_{ij} , $i=1, \dots, n, j=1, \dots, k$ 는 j 번째 설명변수의 i 번째 수준값을 나타내고, Y_{ij} , $i=1, \dots, n, j=1, \dots, m$ 는 j 번째 반응변수의 i 번째 관측치를 나타낸다. 이때 설명변수와 반응변수들의 측정단위나 척도(scale)의 차이에 대한 영향을 제거하기 위해 다음과 같은 표준화가 필요하다.

$$Z_{ij} = \frac{X_{ij} - \bar{X}_{.j}}{\sqrt{\sum_1^n (X_{ij} - \bar{X}_{.j})^2}}, \quad Y_{ij}^* = \frac{Y_{ij} - \bar{Y}_{.y}}{\sqrt{\sum_1^n (Y_{ij} - \bar{Y}_{.y})^2}}$$

여기서 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_{ij}$, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_{ij}$ 이다.

그러면 이경우 다반응 회귀모형은 다음과 같은 형태로 쓰여진다.

$$\mathbf{y}_j^* = \mathbf{Z} \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, \quad j=1, 2, \dots, m$$

위의 방정식에서 \mathbf{y}_j^* 는 표준화된 j 번째 반응변수의 추정벡터이고, \mathbf{X} 는 $(n \times k)$ 행렬인데, 이것은 k 개의 설명변수들에 관한 n 개의 수준값을 포함하고 있으며, $\boldsymbol{\beta}_j$ 는 $(k \times 1)$ 모수벡터이고, $\boldsymbol{\varepsilon}_j$ 는 $(n \times 1)$ 오차벡터이다.

이때 관찰된 반응치 \mathbf{y}_j^* , $j=1, \dots, m$ 과 예측변수 행렬 \mathbf{Z} 의 값을 사용하여 최소제곱 추정량 $\mathbf{b}_j = (\mathbf{Z}^t \mathbf{Z})^{-1} \mathbf{Z}^t \mathbf{y}_j^*$, $j=1, \dots, m$ 을 얻을 수 있고, 이들로부터 다음과 같은 최소제곱 추정치 행렬 $\mathbf{B}_{k \times m}$ 을 만들 수 있다.

$$\mathbf{B}_{k \times m} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m]$$

그러면 여기서 행렬 \mathbf{B} 에 대한 biplot을 작성하여 보자. 이를 위해서는 제2절에서 소개한 두가지 분해를 고려할 수 있다. 먼저 λ_1, λ_2 를 행렬 \mathbf{B} 의 두개의 최대 비정칙값이라 하고, 이들에 대응하는 정칙 고유벡터를 $\mathbf{p}_1, \mathbf{p}_2$ 또는 $\mathbf{q}_1, \mathbf{q}_2$ 라 하자.

그때 첫째로 \mathbf{GH}^t 분해를 이용할 수 있으며, 이는 각 행렬을 다음과 같이 정의하고, 이들을

$$\mathbf{B}_{k \times m} \cong \mathbf{GH}^t, \quad \mathbf{G}_{k \times 2} = [\mathbf{p}_1, \mathbf{p}_2], \quad \mathbf{H}_{m \times 2} = [\lambda_1 \mathbf{q}_1, \lambda_2 \mathbf{q}_2]$$

이용하여 \mathbf{h} -biplot을 만들 수 있다. 이 그림의 통계적 성질은 행렬 \mathbf{G} 의 칼럼 벡터들이 서로 직교 함으로, $\mathbf{B}_{m \times k}^t \mathbf{B}_{k \times m} \cong \mathbf{H}_{m \times 2} \mathbf{H}_{2 \times m}^t$ 으로 주어진다. 고로 \mathbf{HH}^t 는 행렬

$\mathbf{B}_{k \times m}$ 의 분산-공분산 행렬의 $(n-1)$ 곱에 근접하게 된다. 따라서 반응변수들이 표준화에 의해 적절히 취해졌기 때문에 각 \mathbf{h}_j 벡터들 사이의 코사인인 반응변수들 사이의 상관관계의 추정치에 근사함을 알 수 있다. 또한 \mathbf{h} -biplot의 행표시자 벡터와 열표시자 벡터를 함께 그려나면, 우리는 각 반응변수들에 대한 특정한 독립변수들의 회귀관계의 크기와 방향이 서로 어떤 관계가 있는지를 대략적으로 파악할 수 있을 것이다.

둘째, \mathbf{JK}^t 분해에 의하여 행렬 \mathbf{B} 를 다음과 같이 취하고, 이를 이용하여 \mathbf{j} -biplot을 작

$$\mathbf{B}_{k \times m} \cong \mathbf{JK}^t, \quad \mathbf{J}_{k \times 2} = [\lambda_1 \mathbf{p}_1, \lambda_2 \mathbf{p}_2], \quad \mathbf{K}_{m \times 2} = [\mathbf{q}_1, \mathbf{q}_2]$$

성하면 이로부터 우리는 다중 반응변수에 대한 독립변수들의 효과를 추정할 수 있다. 왜냐하면, 이 행렬분해에서 행렬 \mathbf{K} 의 열들은 서로 직교함으로, $\mathbf{B}_{k \times m} \mathbf{B}_{m \times k}^t \cong \mathbf{J}_{k \times 2} \mathbf{J}_{2 \times k}^t$ 을 얻는다. 고로 \mathbf{BB}^t 의 i 번째 주대각선 원소의 값은 i 번째 설명변수에 대한 j

< 표 3.9 > 반응변수들간의 상관계수

상관계수	CC4	ATC1	CP42	GF4	HF1	PT4
CC4	1.000	0.427	0.102	0.197	0.513	-0.314
ATC1		1.000	0.385	0.832	-0.469	0.630
CP42			1.000	0.385	-0.279	0.460
GF4				1.000	-0.641	0.703
HF1					1.000	-0.831
PT4						1.000

< 표 3.10 > 설명변수와 반응변수들간의 상관계수

상관계수	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_10	X_11
CC4	0.246	-0.148	0.262	0.155	0.110	0.493	0.081	-0.264	-0.338	0.332	-0.032
ATC1	0.092	0.102	0.308	0.191	0.337	0.266	0.151	-0.709	0.137	0.334	0.014
CP42	0.072	-0.002	0.240	0.782	0.764	0.251	0.791	0.073	0.813	0.200	-0.008
GF4	0.081	0.111	-0.04	0.184	0.200	-0.071	0.164	-0.582	0.094	-0.028	0.023
HF1	0.123	-0.222	0.180	-0.025	-0.099	0.449	-0.065	0.279	-0.325	0.153	-0.213
PT4	-0.026	0.153	0.169	0.245	0.401	-0.030	0.256	-0.466	0.481	0.161	-0.048

또한 주어진 자료로부터 추정된 회귀계수 행렬을 계산하면 다음의 <표 3.11> 과 같은 결과를 얻게 되고,

< 표 3.11 > 추정된 최소제곱 추정량의 행렬

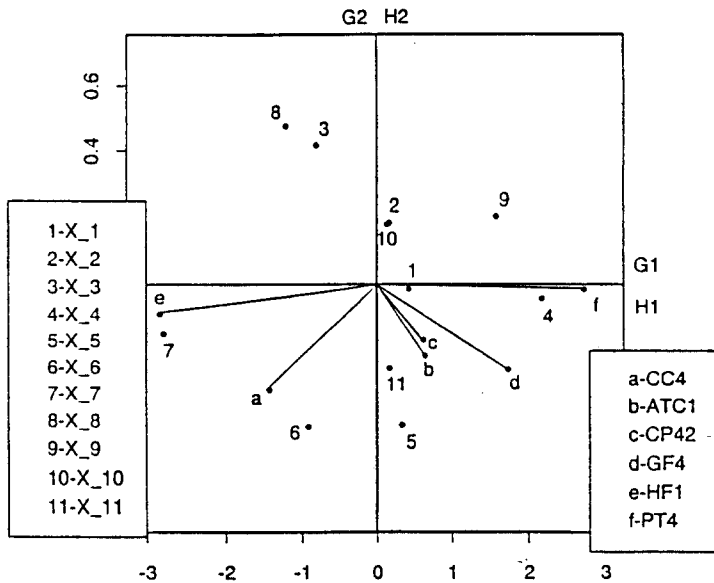
반응변수 설명변수	CC4	ATC1	CP42	GF4	HF1	PT4
X_1	-0.1159	0.0102	0.1027	0.2018	-0.3578	0.1965
X_2	-0.3545	-0.1104	-0.1481	-0.1144	-0.2153	0.0287
X_3	-0.2693	-0.4079	-0.4881	-0.8905	0.3532	-0.4790
X_4	-0.6201	0.2661	0.3152	0.9158	-1.4456	1.4464
X_5	0.5661	0.0649	0.7760	0.5729	-0.0858	0.2893
X_6	0.8754	0.3208	0.2469	0.0535	0.8500	-0.4624
X_7	1.0287	-0.1206	-0.2459	-0.9908	1.9061	-1.8441
X_8	-0.1791	-1.0420	-0.1111	-1.0432	0.6389	-0.7175
X_9	-1.0802	0.2687	0.4049	0.2721	-1.0300	0.9299
X_10	-0.2214	-0.1181	-0.2110	-0.2311	-0.1759	0.1714
X_11	0.3163	0.4153	0.2097	0.2373	-0.1502	0.0386

이때 계산된 행렬 **B** 에 대하여 \mathbf{GH}^t 분해를 실시하여 **h**-biplot 을 작성하면 다음의 <그림 3.2> 을 얻을 수 있고, 우리는 이 biplot으로부터 다음의 사실을 알 수 있다.

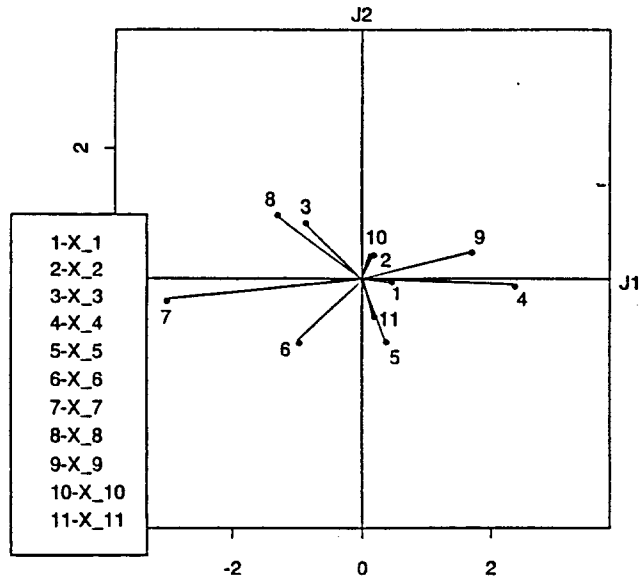
- (가) 주어진 6 가지 특성항목중 얼라이링 토크 계수(ATC1), 로트트랜스퍼(GF4), 코너링 파워어(CP42)등은 서로 밀접한 양의 상관관계를 갖고 있으며, 반면에 로드 센서티비티(HF1)와는 음의 상관관계가 있다. 또한 코너링 계수(CC4)는 로드 센서티비티(HF1)와는 양의 상관관이 있으나 뉴우메틱 트레일(PT4)과는 음의 상관관계를 갖고 있으며, 나머지 특성항목들과는 특별한 관계가 없는 것으로 주어진다.
- (나) 설명변수와 반응변수의 관계는 그림에서 보여주는 것과 같이 뉴우메틱 트레일(PT4)과 변수 X_1, X_7 등이 양의 관련이 있고, 얼라이링 토크 계수(ATC1), 코너링 파워어(CP42) 등은 X_5, X_11 등에 관련되며, 코너링 계수(CC4)는 X_6 에 깊은 양의 회귀관계를 갖는다.

다음으로 설명변수들의 복합효과를 알아 보기 위하여 \mathbf{JK}^t 분해를 실시하여 **j**-biplot을 작성한 결과가 다음 <그림 3.3> 과 같다.

이 그림으로부터 설명변수들 중에서 특성항목에 복합효과가 큰 것으로는 X_7, X_4, X_9, X_8, ... ,X10 순으로 주어진다. 또한 설명변수 X_4, X_9등은 X_7, X_6, X_8 등과 반응변수들에 대하여 서로 상반된 복합효과를 갖는다.



< 그림 3.2 > 다반응 회귀모형에 대한 **h**-biplot



< 그림 3.3 > 다반응 회귀모형에 대한 j-biplot

4. 토의와 결론

본 논문에서 고려한 biplot은 주어진 측정 자료행렬을 순위가 2인 행렬로 근사시켜 2차원 평면상에 그림으로 도식하는 통계적 그래픽의 한 방법으로 변수들과 관측치들 사이의 구조적 관계나 경향을 이해하는데 매우 유용한 방법으로 생각된다.

특히 다변량 분산분석에서 biplot은 다양한 반응변수들과 처리효과의 각 수준간에 존재하는 복잡한 다변량 관계를 파악할 수 있는 그림을 제공함과 동시에, 또한 각 수준별 처리효과의 차에 대한 통계적 검정을 대략적으로 실시할 수 있는 그림을 제시한다.

또한 여러개의 반응변수와 설명변수들을 갖는 회귀분석 모형에서 biplot은 반응변수들의 상호관계와 설명변수들의 복합효과를 파악할 수 있을 뿐만 아니라, 더불어 각 설명변수와 반응변수의 회귀관계도 알 수 있는 유용한 분석기법이다

마지막으로 이방법은 여러 다른 산업분야에서 얻어지는 측정자료의 분석에도 비슷하게 적용될 수 있을 것이다.

참고문헌

- [1] 백세종 (1993). 타이어 공학, 금호타이어 연구소.
- [2] 남일 (1993). 타이어 제조공정 일반, 금호타이어 주식회사.
- [3] 장대홍 (1994). Multiresponse Problem of Response Surface Methodology, 회귀분석을 통한 품질향상에 관한 워크샵, 한국통계 학회 공업통계 연구회편, 143-150.
- [4] 최용석 (1991). Resistant Principal Component Analysis, Biplot And Correspondence Analysis, 고려대학교 박사학위논문.
- [5] 최용석, 심의정 (1994). 82-92 한국 프로야구의 부분별 기록에 대한 각 팀의 연도별 추세분석, 부산대학교 자연대학 논문집, 제57권, 35-43.
- [6] Corsten, L.C.A. and Gabriel, K.L. (1976). Graphical Exploration in Comparing Variance Matrices. *Biometrics*, Vol. 32, 851-863.
- [7] Daigle, G. and Rivest, L.P. (1992). A Robust Biplot. *The Canadian Journal of Statistics*, Vol. 20, 241-255.
- [8] Gabriel, K.R. (1971). The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika*, Vol. 58, 453-467.
- [9] Gabriel, K.R. (1972). Analysis of Meteorological Data by Means of Canonical Decomposition and Biplots. *Jour. of Appl. meteorol.* Vol. 11, 1071 - 1077.
- [10] Gabriel, K.R. (1982). Biplot, *Encyclopedia of Statistical Science*, Vol. 1, 263- 270, Wiley, New York.
- [11] Gower, J.K. (1992). Generalized Biplots. *Biometrika*, Vol. 79, 475 - 493.
- [12] Gillespie, T.D. (1992). *Fundamentals of Vehicle Dynamics*. Society of Automotive Engineers Inc., U.S.A.
- [13] Heck, D.L. (1960). Charts of Some Upper Percentage Points of the Distribution of the Largest characteristic Root. *Ann. Math. Statist.*, Vol. 37, 625 - 642.
- [14] Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*. Prentice - Hall Inc., New Jersey.
- [15] Morrison, D.F. (1967). *Multivariate Statistical Methods*. Mcgraw-Hill Inc., USA.
- [16] Smith, W.F. and Cornell, J.A. (1993). Biplot Displays for Looking at Multiple Response Data in Mixture Experiments. *Technometrics*, Vol. 35, 337 - 350.

Biplot Method Algorithm and Application in Tire Engineering³⁾

Wan Hyun Cho⁴⁾

Abstract

It is essential in modern industry that quality and productivity are improved continuously. To accomplish this purpose, quality control must be maintained in all parts of a company. Recently, some tire manufacture companies are beginning to show interest in quality control. They have tried to achieve some results through the statistical analysis for the experimental data which has accumulated up to now and then they strive to determine the structural relationship between the design factors in tire construction and tire performance characteristics.

The measurement data obtained from tire construction engineering is given in multivariate form owing to the various properties found in tire design components as well as in performance. Also it may be existed the relationship among the multiple response variables. Thus we propose the use of the biplot graphical display as an analytic tool of data matrices with complex respects. The proposed biplots are also available to understand both the underlying structure of the data and the roles played by the different components. In particular, we consider the matter of how best to use the biplots in the multivariate analysis of variance and multiple response data. Finally we apply this method to analyze the actual data.

3) This research was supported by Chonnam National University Research Foundation.

4) Professor, Department of Statistics, Chonnam National University, Kwangju, 500-757, Korea.