經營分析의 合理性提高를 위한 主成分分析과 判別分析의 連繫模型

Linkage Model of PCA and DA for Rational Business Analysis

張 有 喆*

《目 次》

Ⅰ. 緖 論

우리가 어떠한 意思決定을 하게 될 경우, 반드시 어떠한 要素로 評價할 것인가 그리고 그와 같은 要素가 결정되었다 하더라도 어느 요소를 더 重視해야 될 것인가를 고려해야만 한다. 經營分析의 경우에도 이와 마찬가지로 이들을 合理的으로 결정하느냐 그렇지 않으냐에 따라 分析의 信賴性은 크게 차이가 나게 된다. 따라서 經營分析을 할 경우 經營分析의 目的에 비추어, 評價項目(criteria)의 選定과 加重値(weight)의 算定이 가장 중요한 것이라 하겠다.

그런데 既存의 經營分析 方法을 보면, 評價項目의 設定과 이에 대한 加重値의 算定에 있어 統計的 檢定을 거치지 않고 거의 直觀 내지는 경험에 의해 이들이 결정되어 經營分析을 客觀的으로 할 수 없다는 결점이 있다.

本硏究에서는 이러한 問題點을 제거하기 위해, 즉 經營分析을 하는데 있어 合理的인 評價項目의 설정과 加重値의 산정을 위해, 統計的 技法인 主成分分析(principal component

* 聖心女子大學 經營學科 專任講師

analysis)과 判別分析(discriminant analysis)을 활용하고자 한다. 經營分析을 할 때 情報의 喪失을 방지하기 위해 단순히 많은 評價項目을 취하면 의사결정을 위한 變數가 많아진다. 그렇게 되면 問題의 次元이 높게 되어 복잡하게 될 뿐 아니라, 統計的으로 誤差가 많이 발생하는 등 價値性이 없는 結果가 나오므로 合理的인 經營分析「시스템」을 만드는 것이 곤란하다.

그런데 主成分分析을 활용하면 비록 評價項目의 數가 많아도 각각의 情報를 가급적 상실함이 없인 될 수 있는대로 少數의 相互獨立的인 主成分으로 變換된다. 즉, 복잡한 問題의 次元을 낮게 해 줌으로써 問題가 單純하게 되고 나아가서 이해하기 쉽게 된다. 더구나 무리없이 判別分析과 連繫지어 合理的인 經營分析模型을 설정할 수 있다. 그리고 判別分析은 統計集團이 여러가지 共通特性을 갖는 單位에 의하여 측정되어 있을 때, 이 統計集團을 두개 혹은 그 이상의 部分集團으로 分類하는 방법이다. 分析의 對象이 기업의 경우에도 優良企業과 不良企業의 標本을 客觀的으로 抽出하면 역시 이와 같은 判別模型을 만들 수 있다.

따라서 本硏究의 目的은 主成分分析과 判別分析을 連繫的으로 활용함으로써 經營分析의 合理性을 提高할 수 있는 評價模型의 設定方案을 모색하는데 있다 하겠다.

II. 經營分析의 理論的 考察

一般的으로 經營分析은 財務分析, 企業分析, 企業評價 等의 이름으로 실시되고 있으며, 또한 이것은, 會社分析, 經營比較, 企業比率分析이라고도 한다.(1)

經營分析은 豫算統制, 標準原價計算 등과 더불어 管理會計의 한 分野로 발달한 것으로 財務諸表나 原價計算 등 諸會計資料에 표시되는 會計數値를 비교분석하여 기업의 財務狀態와 經營成果의 良否를 판단하고 나아가서는 그 原因을 규명하는 하나의 技術 또는 方法이라 할 수 있다.(2) 즉, 經營分析은 그 目的이 分析의 主體가 누구이든간에 기본적으로는 企業의 實態를 해명하여 그 기업이 지니는 機會와 危險을 파악함으로써 장래에 대처하려는 것이라고 할 수 있다. 그러나 실제에 있어서는 기업을 分析·評價하는 경우에는 누가, 무엇때문에 행하는가에 따라 分析의 視角 내지 重點을 두는 곳에 差異가 있게 되며 適用하는 分析技法도 特色이 있게 된다. 그러므로 보다 客觀的인 分析을 기대하기 위해서는 무엇인가 共通되는 評價基準을 설정함과 아울러 分析方法과의 關聯을 명백히 하여 둘 필요가 있다.

初期의 美國에 있어서의 經營分析은 銀行家들에 의하여 貸付先의 債務支給能力을 판단하

(1) 高松和男, 改訂經營分析の實務, 日本生産性本部, 1979, p. 2.
(2) 金熙執, 新經營分析, 博英社, 1978, p. 15.

는 信用分析(credit analysis)의 수단으로 출발하게 되었다. 그러나 經營이 발달함에 따라 經營分析에 對한 방법도 점차로 발전하게 되었으며 그 적용 범위도 넓어졌다. 즉 一部 專門投資家들은 投資對象企業의 業績과 內容을 분석하는 소위 投資分析(investment analysis)를 실시하게 되었으며, 그 內容에 있어서도 收益性分析을 중심으로 다루게 되었다. 그리고 外部的 目的으로 출발한 經營分析은 다시 기업을 직접 담당하고 있는 經營者들에 의하여 企業內部의 經營活動의 良否를 스스로 판단하고 나아가서는 合理的인 改善을 위한 手段의 하나로 이용하게 되었다.

이와 같이 企業의 分析·評價는 기업을 둘러싼 經濟社會의 環境變化에 따라 그 方法과 目的이 점차로 변천되어 왔고 또 변화되어 가고 있다. 특히 現代企業의 經營規模擴大와 生產方法의 複雜化, 企業의 國際化로 기업의 分析·評價는 從來의 傳統的 經營分析方法에만 의존할 수 없게 되었으며 새로운 方法들을 硏究·開發하지 않을 수 없는 段階에 이르고 있다.

經營分析의 最近動向은 종전의 經營分析이 주로 過去와 現在의 狀態에 대한 分析에 중점을 두고 있는데 반해, 未來指向的인 方向으로 그 초점을 바꾸게 된 것이 특징적으로 나타나고 있다. 또한 美國 등 先進國에서는 統計的 方法을 活用, 會計外的인 資料까지 처리하는 것이 요청되고 있다. 즉, 財務諸表 이외에 경영에 영향을 주는 組織의 構造를 분석할 필요가 있고, 經營者의 思考方式과 姿勢도 중요한 요소가 된다. 이밖에 기업의 資源配分問題, 企業全體에 대한 有機性, 그리고 市場占有率이나 앞으로의 可能性 뿐만 아니라 그 企業의 經營技術의 方向 등 여러 가지 要因도 고려해야 한다.

이와 같이 볼 때 經營 分析은 기업의 過去와 現在 그리고 未來의 計劃分野에 이르는 광범위한 財務資料를 토대로 企業의 財務狀態와 經營成果를 분석하고 그 原因을 규명함으로써 株主를 비롯한 利害關係者들에게 보다 正確한 企業情報를 제공하는 모든 分析方法을 총칭하는 것이라 하겠다.

III. 主成分分析에 의한 經營分析

1. 序

일반적으로 經營分析을 할 경우, 어떠한 評價項目으로 또한 어떠한 加重値로 분석할 것인가 하는 것이 가장 중요한 問題로 대두된다.

우선 合理的인 評價項目, 즉 가능한 한 變數의 數가 적으면서도 全體를 나타내 줄 수 있는 變數의 발견이 요망되는데, 이를 위해 統計的 技法인 主成分分析을 활용한다. 그리고 이

들 評價項目에 대한 相對的 重要度인 加重値 또한 중요한 意味를 가지는데, 이러한 問題의 해결도 역시 主成分分析에 의해 가능하다.

이와 같은 主成分分析을 經營分析에 활용하는 理由는 다음과 같다. 經營分析을 하는데 중요하다고 생각되는 變數는 많이 있다. 그것들은 서로 相關關係가 있지만 그렇다고 해서 단순히 한 쪽에 다른 쪽을 代替할 수는 없다. 서로가 補完關係가 있는 것이므로 經營分析을 할때 될 수 있으면 많은 變數를 사용한다. 비록 變數의 數가 많아도 主成分分析을 활용하면 보다 적은 數의 相互獨立된 變數로 變換이 될 뿐 아니라 이들 變數에 대한 合理的인 加重値 역시 산정될 수 있다.

이제 이러한 두 가지 問題點을 해결하고 分析의 精度를 높이는 模型設定의 방법으로 主成分分析에 대해 고찰해 보고자 한다.

2. 主成分分析法

1) 主成分分析의 意義

主成分分析은 多變量分析法(multivariate statistical analysis)이라 하는 統計分析方法의 일종으로, 多量의 變數를 몇 개의 主成分(principal components)으로 분해하는 방법이다.

즉, P개의 相關關係가 있는 變數(評價項目)들간의 關係를 알아보기 위해서, 이 變數들의 集合을 主成分이라 하는 相關關係가 전혀 없는 새로운 變數들의 集合으로 轉換하는 방법이다. 이러한 새로운 變數들은 原變數들의 一次結合(linear combination)으로 표시되며, 이를 사용함으로써 問題의 次元을 줄일 수 있다. 이러한 轉換은 실제 P次 空間에서 直角轉換(orthogonal rotation)이 된다. 이러한 전환을 발견하기 위한 기법이 主成分分析이다. 이것은 多重回歸分析(multiple regression analysis)에서 처럼 變數들을 從屬變數와 몇 개의 獨立變數간의 關係로 파악하는 것이 아니라 모든 變數를 동등한 것으로 취급하여 변수들간의 관계를 파악하려는 分析技法이다.

이 方法의 目的은 중요한 몇 개의 主成分들이 變數들의 分散을 어느 정도 설명하는가를 살펴보고 問題의 次元을 낮추는데 있다. 다시 말하면 만약에 變數들이 높은 相關關係를 가지고 있는 몇 개의 集團으로 구성되어 있다면, 각 集團들은 유사한 성질들을 가지므로 그 變數들에 대해서는 거의 線型에 가까운 制約式이 存在한다 하겠다. 이러한 경우에 몇 개의 主成分만이 意味가 있으며, 이것들이 變數보다 資料를 더 잘 이해할 수 있도록 해 준다. 그런데 이러한 主成分들의 意味를 발견하고 命名한다는 것이 결코 쉬운 것은 아니다.

우리는 이 方法의 目的은 분석을 간단하게 하기 위해서 資料의 次元을 낮추는데 있다고 할 수 있다. 主成分分析은 이와같이 相關關係가 있는 變數들의 集合을 相關關係가 없는 새

로운 變數들의 集合으로 轉換하는 것이다. 그러므로 만약 變數들이 거의 相關關係가 없다면 主成分分析을 할 필요가 없게 된다. 그런데 主成分分析은 變數들에 대한 確率分布에 관한 假定을 하지 않는 數學的 方法이다. 그러나 觀察值들이 多變量正規分布(multivariate normal distribution)(3)을 한다는 假定을 할 경우에는 主成分들에 더 의미를 부여할 수 있다.

2) 主成分分析의 資料와 主成分導出

主成分分析에 사용되는 資料는 n개의 기업에 대한 經營分析을 위한 P 種類의 變數(評價項目)들로써 구성된다. 이것은 X_{rj} $(r=1,2,\cdots,n;\ j=1,2,\cdots,p)$로 표현될 수 있으며, 다음 〈表 1〉에서 보는 바와 같다.

〈表 1〉 主成分分析의 資料

企業 \ 變數	X_1	X_2	$\cdots$	X_j	$\cdots$	X_p
1	X_{11}	X_{12}	$\cdots$	X_{1j}	$\cdots$	X_{1p}
2	X_{21}	X_{22}	$\cdots$	X_{2j}	$\cdots$	X_{2p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		
r	X_{r1}	X_{r2}	$\cdots$	X_{rj}	$\cdots$	X_{rp}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		
n	X_{n1}	X_{n2}	$\cdots$	X_{nj}	$\cdots$	X_{np}

이제 主成分分析과 判別分析 등 多變量分析을 위해, 線型代數(linear algebra)를 이용하면 간단히 표기할 수 있다.(4) 즉, $\boldsymbol{x}^T=[X_1, X_2, \cdots, X_p]$가 平均과 共分散行列이 각각

$$\boldsymbol{\mu}^T=[\mu_1, \mu_2, \cdots, \mu_p]$$

$$\Sigma=\begin{pmatrix} \sigma_{11} & \sigma_{12}\cdots\cdots\sigma_{1p} \\ \sigma_{21} & \sigma_{22}\cdots\cdots\sigma_{2p} \\ \\ \sigma_{p1} & \sigma_{p2}\cdots\cdots\sigma_{pp} \end{pmatrix}$$

인 p次 確率變數라 하자.

우리의 目的은 이 變數들로부터 相關關係가 전혀 없는 변수들의 새로운 集合, 즉 $Y_1, Y_2, \cdots, Y_p$를 발견하려는 것이다.(5) 이 變數들의 分散은 Y_1에서 Y_p로 갈수록 점점 작아진다. 각 Y_j는 $\boldsymbol{x}$의 一次結合으로 표현되므로

$$Y_j=a_{1j}X_1+a_{2j}X_2+\cdots\cdots+a_{pj}X_p$$
$$=\boldsymbol{a}_j{}^T\boldsymbol{x}$$

(3) P次元의 確率變數 $\boldsymbol{x}$는 만약에 $\boldsymbol{x}$의 모든 一次結合이 單一變數正規分布(univariate normal distribution)를 하는 경우에 多變量正規分布를 한다고 한다.

(4) 여기에서 「보울드」體(boldface)로서 大文字는 行列을, 그리고 小文字는 「벡터」(vector)를 각각 나타낸다.

(5) 主成分分析에 대한 최근 경향은 統計的 模型을 假定하지 않는 數學的 技法으로 간주한다. 그러므로 主成分分析에서는 標本 그 자체를 母集團으로 간주한다. 구체적인 것은 C. Chatfield and A.J. Collins, *Introduction to Multivariate Analysis*, Chapman and Hall, 1980, p. 62.

로 나타내어 진다. 이 때 $\boldsymbol{a}_j^T=[a_{1j}, a_{2j}, \cdots, a_{pj}]$로서 常數를 元素로 한 「벡터」이다. 위의 $\boldsymbol{a}_j$ 「벡터」는 그 길이(norm, length)가 확정된 것이 아니라 임의적인 것이다. 그러므로 우리는 $\boldsymbol{a}_j^T\boldsymbol{a}_j=\sum_{k=1}^{p} a_{kj}^2=1$인 條件을 첨가한다. 즉, $\boldsymbol{a}_j$ 「벡터」의 길이를 1이 되도록 標準化(normalization)해 주어야 한다. 이러한 標準化로 p次 空間에서의 轉換은 直角轉換이 됨을 볼 수 있을 것이다.

第1主成分인 Y_1은 $\boldsymbol{a}_1$을 도출함으로써 발견될 수 있으며 이 때 Y_1이 가장 큰 分散을 갖게 된다. 즉, 우리는 $\boldsymbol{a}_1^T\boldsymbol{a}_1=1$ 인 制約條件下에서 $\boldsymbol{a}_1^T\boldsymbol{x}$의 分散을 極大化하는 $\boldsymbol{a}_1$을 도출해야 된다. 그리고 第2主成分은 $\boldsymbol{a}_2$를 도출함으로써 발견되는데, Y_2는 Y_1과 相關關係가 없으면서 두번째로 큰 分散을 갖는다. 마찬가지로 서로 相關關係가 없고 分散이 점점 더 작은 主成分 $Y_3, Y_4, \cdots, Y_p$도 도출해 낼 수 있다.

우선 第1主成分 Y_1을 도출해 보자, 우리는 標準化 條件인 $\boldsymbol{a}_1^T\boldsymbol{a}_1=1$인 制約下에서 Y_1의 分散을 極大化하는 $\boldsymbol{a}_1$ 「벡터」를 도출하기를 원한다. 그런데 Y_1의 分散은

$$\begin{aligned}\mathrm{Var}(Y_1)&=Var(\boldsymbol{a}_1^T\boldsymbol{x})\\&=E[\{\boldsymbol{a}_1^T(\boldsymbol{x}-\boldsymbol{\mu})\}^2]\\&=E[\boldsymbol{a}_1^T(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\ \boldsymbol{a}_1]\\&=\boldsymbol{a}_1^T\Sigma\boldsymbol{a}_1\end{aligned}$$

으로 된다. 그러므로 $\boldsymbol{a}_1^T\Sigma\boldsymbol{a}_1$을 우리의 目的函數(objective function)로 취한다.

한 개 혹은 몇 개의 制約條件이 주어지고, 目的函數를 極大化하는 대표적 방법은 Lagrange 乘數法이 있다. 즉, $f(X_1, X_2, \cdots, X_p)$를 目的函數, $g(X_1, X_2, \cdots, X_p)=C$를 制約條件이라 할 때, 이를 만족하는 均衡點은

$$\frac{\partial f}{\partial X_j}-\lambda\frac{\partial g}{\partial X_j}=0,\ \ j=1, 2, \cdots, p$$

로부터 구해진다. 이 때 λ는 Lagrange 乘數(multipliers)이다. 이러한 p개의 函數들은 目的函數를 極大化하는 必要條件(necessary conditions)들이다.[6]

우리의 目的을 위해 새로운 函數

$$L(\boldsymbol{x})=f(\boldsymbol{x})-\lambda[g(\boldsymbol{x})-C]$$

---

(6) $Y=f(X_1, X_2)$의 函數가 極大値를 갖기 위한 充分條件은

1) $\frac{\partial Y}{\partial X_1}=\frac{\partial Y}{\partial X_1}=0,$ 2) $\frac{\partial^2 Y}{\partial X_1^2}<0$ 그리고 $\frac{\partial^2 Y}{\partial X_2^2}<0,$

3) $\frac{\partial^2 Y}{\partial X_1^2}\cdot\frac{\partial^2 Y}{\partial X_1^2}>\frac{\partial^2 Y}{\partial X_1\partial X_2}$

등 세 條件이 만족해야 된다. 그러나 極大化 函數는 당연히 極大値를 가지므로 1) 條件만으로도 사실 충분하다.

를 만든다. 이 때 괄호 안은 당연히 0의 값이 된다. 그러면 均衡點을 찾기 위한 函數式들은

$$\frac{\partial L}{\partial \boldsymbol{x}}=0$$

로 됨을 알 수 있다.

우리의 문제에 이 방법을 적용하기 위해, Lagrange 函數

$$L(\boldsymbol{a}_1)=\boldsymbol{a}_1\Sigma\boldsymbol{a}_1-\lambda(\boldsymbol{a}_1{}^T\boldsymbol{a}_1-1)$$

을 설정한다. 이로부터 우리는

$$\frac{\partial L}{\partial \boldsymbol{a}_1}=2\Sigma\boldsymbol{a}_1-2\lambda\boldsymbol{a}_1$$ [7]

을 얻고, 이를 0으로 놓으면 다음과 같은 式을 얻는다.

$$(\Sigma-\lambda\boldsymbol{I})\boldsymbol{a}_1=0$$

이 때 괄호 안의 項을 $(P\times P)$인 Σ와 같은 次元을 만들기 위해 單位行列(identity matrix) $\boldsymbol{I}$를 삽입하는데 주의해야 된다.

만약 위의 式에서 $\boldsymbol{a}_1$이 0이 아닌 解를 갖기 위해서는 $(\Sigma-\lambda\boldsymbol{I})$가 正則行列(nonsingular matrix)이 되어서는 안된다. 즉, 行列$(\Sigma-\lambda\boldsymbol{I})$의 行列式이 0이 되어야 한다. 그러므로 λ는

$$|\Sigma-\lambda\boldsymbol{I}|=0$$

로 부터 구할 수 있다. 따라서 $\boldsymbol{a}_1$이 0이 아닌 解를 갖기 위해서는 λ가 Σ의 固有値(eigenvalue)일[8] 경우 뿐이다.

이 때 Σ는 半正定形行列(positive semidefinite matrix)[9]이므로 陰數가 아닌 p개의 固有値를 갖게 된다. 이 때 p개의 固有値를 $\lambda_1, \lambda_2, \cdots, \lambda_p$라 하면 $\lambda_1>\lambda_2>\cdots>\lambda_p\geq 0$가 된다.

그러면 이들 중 어느 것이 第1主成分을 결정하는 것일까? 第1主成分의 分散은

$$\begin{aligned} Var(\boldsymbol{a}_1{}^T\boldsymbol{x})&=\boldsymbol{a}_1{}^T\Sigma\boldsymbol{a}_1\\ &=\boldsymbol{a}_1{}^T\lambda\boldsymbol{I}\boldsymbol{a}_1 \end{aligned}$$

---

(7) Σ가 $(P\times P)$인 對稱行列(symmetric matrix)일 때 二次形態(quadratic form) $Y=\boldsymbol{x}^T\Sigma\boldsymbol{x}$를 $\boldsymbol{x}$에 대해 微分하면

$$\frac{\partial Y}{\partial \boldsymbol{x}}=2\Sigma\boldsymbol{x}$$가 된다.

(8) 이것은 characteristic roots, latent roots라고도 하며, Σ가 $(P\times P)$인 行列일 때 λ의 P次 多項式인 $|\Sigma-\lambda\boldsymbol{I}|=0$ 函數式의 根이다.

具體的인 것은 F. Max Stein, *Introduction to Matrices and Determinants*, Wedsworth Publishing Company, Inc., California, 1967, pp. 138-139 參照.

(9) $\boldsymbol{A}=(a_{ij})$, $i, j=1, 2, \cdots, n$일 때 모든 $\boldsymbol{x}\neq\boldsymbol{0}$ 대해 $\boldsymbol{A}$行列을 $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}<0$이면 負定形行列, $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}>0$이면 正定形行列이라 하며, $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\leq 0$이면 半負定形行列, $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\geq 0$이면 半正定形行列이라 한다.

$$=\lambda \boldsymbol{a}_1^T \boldsymbol{a}_1$$
$$=\lambda$$

가 된다. 우리는 이 分散을 極大化하는 것을 원하기 때문에 가장 큰 固有値인 λ_1을 선택한다. 그리고 $(\Sigma-\lambda \boldsymbol{I})\boldsymbol{a}_1=0$을 사용함으로써, 우리가 찾고 있는 第1主成分 $\boldsymbol{a}_1$는 가장 큰 固有値 λ_1에 대응되는 Σ의 固有「벡터」(eigenvector)[10]임을 알 수 있다.

第2主成分($Y_2=\boldsymbol{a}_2^T \boldsymbol{x}$)도 위와 같은 방법으로 구해진다. 이 때 $\boldsymbol{a}_2^T\boldsymbol{a}_2=1$ 이란 制約條件 이외에 Y_2는 Y_1과 相關關係가 없어야 된다는 또 하나의 制約條件을 추가해야 한다.[11]

$$\begin{aligned} \mathrm{Cov}(Y_2, Y_1) &= \mathrm{Cov}\ (\boldsymbol{a}_2^T \boldsymbol{x},\ \boldsymbol{a}_1^T \boldsymbol{x}) \\ &= E[\boldsymbol{a}_2^T(\boldsymbol{x}-\mu)(\boldsymbol{x}-\mu)^T \boldsymbol{a}_1] \\ &= \boldsymbol{a}_2^T \Sigma \boldsymbol{a}_1 \end{aligned}$$

즉, 이것이 0이 되어야 됨을 요구한다. 그런데 $\Sigma \boldsymbol{a}_1=\lambda_1 \boldsymbol{a}_1$이므로 이를 위 式에 代入하면, 위 條件과 동등한 것으로 $\boldsymbol{a}_2^T\boldsymbol{a}_1=0$를 구할 수 있는데 이를 대신 사용한다. 이로부터 $\boldsymbol{a}_1$과 $\boldsymbol{a}_2$는 서로 直角「벡터」(orthogonal vector)임을 알 수 있다.

위의 두 制約條件下에서 Y_2의 分散을 極大化하기 위하여 우리는 두 개의 Lagrange 乘數를 필요로 한다. 이것을 λ와 δ라 하면 Lagrange 函數는

$$L(\boldsymbol{a}_2)=\boldsymbol{a}_2^T \Sigma \boldsymbol{a}_2-\lambda(\boldsymbol{a}_2^T \boldsymbol{a}_2-1)-\delta \boldsymbol{a}_2^T \boldsymbol{a}_1$$

과 같이 된다. 이로부터 均衡點은

$$\frac{\partial L}{\partial \boldsymbol{a}_2}=2(\Sigma-\lambda \boldsymbol{I})\boldsymbol{a}_2-\delta \boldsymbol{a}_1=0$$

로부터 구할 수 있다. $\delta \boldsymbol{a}_1$를 이항하고 양변에 $\boldsymbol{a}_1^T$를 곱해 주면 $\boldsymbol{a}_1^T\boldsymbol{a}_2=0$ 이드로 $2\boldsymbol{a}_1^T\Sigma\boldsymbol{a}_2-\delta=0$를 얻는다. 한편 「스칼라」(scalar)인 $\boldsymbol{a}_2^T\Sigma\boldsymbol{a}_1$가 0이므로 이것의 轉置(traspose)인 $\boldsymbol{a}_1^T\Sigma\boldsymbol{a}_2$가 0이 된다. 따라서 δ가 0이 된다. 그러므로

$$\frac{\partial L}{\partial \boldsymbol{a}_2}=(\Sigma-\lambda \boldsymbol{I})\boldsymbol{a}_2=0$$

가 쉽게 유도된다. 따라서 第2主成分은 Σ의 두 번째로 큰 固有値를 선택하면 된다. 그리고 $\boldsymbol{a}_2$는 이 固有値에 대응된 固有「벡터」이다.

일반적으로 j째 主成分은 j째로 큰 固有値에 대응된 固有「벡터」임을 알 수 있다.

이제 $(P\times P)$인 固有「벡터」들의 行列을

---

(10) $\Sigma C_i=\lambda_i C_i$의 各 固有値에 對應되는 C_i「벡터」들을 固有「벡터」라고 한다. 이들 「벡터」들의 길이는 임의적인 것으로 일반적으로 길이를 1로 標準化한다.

(11) $\rho_{ij}=\sigma_{ij}/\sigma_i\sigma_j$인 關係로부터 相關係數는 共分散을 標準化해 준 것이다. 그러므로 Correlation(Y_2, Y_1) 대신 Cov(Y_2, Y_1)을 사용한다.

$$\boldsymbol{A}=[\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_p]$$

라 하고, $(P\times 1)$인 主成分「벡터」를 $\boldsymbol{y}$라 하자. 그러면 모든 主成分은

$$\boldsymbol{y}=\boldsymbol{A}^T\boldsymbol{x}$$

로 표현될 수 있다. $\boldsymbol{y}$의 $(P\times P)$인 共分散行列을 Λ라 하면

$$\Lambda=\begin{pmatrix} \lambda_1 & 0 \cdots\cdots 0 \\ 0 & \lambda_2 \cdots\cdots 0 \\ \vdots & \\ 0 & \cdots\cdots\cdots \lambda_p \end{pmatrix}$$

가 된다. 즉. 主成分들 간에는 相關關係가 없는 것이므로 Λ는 對角行列(diagonal matrix)이 된다. 우리는 $\mathrm{Var}(\boldsymbol{y})$를

$$\Lambda=\boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A}$$

로 나타낼 수 있다. 이로부터 $\boldsymbol{x}$의 共分散行列과 대응된 主成分들간의 중요한 關係를 알 수 있다. 위 式은 $\boldsymbol{A}$行列이 直角行列[12] (orthogonal matrix)이므로 다음과 같이 변형될 수 있다.

$$\Sigma=\boldsymbol{A}\Lambda\boldsymbol{A}^T$$

우리는 固有値들이 각기 다른 主成分들의 分散들 임을 알았다. 그러면 이러한 分散들의 合은

$$\sum_{i=1}^{p}\mathrm{Var}(Y_j)=\sum_{i=1}^{p}\lambda_j=\mathrm{trace}(\Lambda)$$

이다. 또한

$$\begin{aligned}\mathrm{trace}(\Lambda)&=\mathrm{trace}(\boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A})\\&=\mathrm{trace}(\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{A}^T)^{(13)}\\&=\mathrm{trace}(\Sigma)\\&=\sum_{i=1}^{p}\mathrm{Var}(X_j)\end{aligned}$$

이다. 그러므로 이들로부터 X_j들의 分散의 合과 主成分 Y_j들의 分散의 合이 같다는 중요한 결과를 얻을 수 있다. 이로부터 i째 主成分이 모든 變數들의 總分散에 대한 설명하는 比率은 $\lambda_i/\sum_{j=1}^{p}\lambda_j$로 나타낼 수 있다. 그리고 처음 m개 主成分의 總分散에 대한 說明力은 $\sum_{j=1}^{m}\lambda_j/\sum_{j=1}^{p}\lambda_j$임을 알 수 있다.

(12) 直角行列 $\boldsymbol{A}$는 $\boldsymbol{A}\boldsymbol{A}^T=\boldsymbol{I}$인 특성을 갖는다. 그러므로 $\boldsymbol{A}^T=\boldsymbol{A}^{-1}$이 된다.
(13) $\boldsymbol{A}$와 $\boldsymbol{B}$가 正方行列(square matrix)이 아니더라도 $\boldsymbol{AB}$가 正方行列이면, $\mathrm{trace}(\boldsymbol{AB})=\mathrm{trace}(\boldsymbol{BA})$가 성립한다.

3) 相關係數로부터 主成分導出

실제로 主成分은 變數 X_j들을 標準化해[14] 주어 分散이 1이 되도록 한 후에 도출하는 것이 일반적인 方法이다. 즉 이것은 共分散行列 Σ가 아니라 相關係數行列

$$\boldsymbol{P}=\frac{\sigma_{ij}}{\sigma_i\sigma_j} \qquad \boldsymbol{i}, \boldsymbol{j}=1, 2, \cdots \boldsymbol{p}$$

$$=\begin{pmatrix} 1 & \rho_{12}\cdots\cdots\rho_{1p} \\ \rho_{21} & 1 \quad \cdots\cdots\rho_{2p} \\ \vdots & \vdots \qquad\quad \vdots \\ \rho_{p1} & \rho_{p2}\cdots\cdots 1 \end{pmatrix}.$$

로 부터 主成分들을 구한다는 것을 의미한다.

數學的 過程은 Σ로부터 主成分을 도출하는 것과 같다. 그러므로 이 때 主成分들은 $\boldsymbol{P}$의 固有「벡터」들이 된다. 그런데 $\boldsymbol{P}$의 固有値들과 固有「벡터」들은 일반적으로 Σ의 그것들과 같지 않다는 것을 주의해야 된다. Σ보다는 $\boldsymbol{P}$로써 분석하는 것은 모든 變數들의 중요성이 같음을 뜻한다.

相關係數行列은 對角元素들이 모두 1이므로 對角元素의 合은 p가 될 것이다. 그러므로 $\boldsymbol{P}$의 固有値의 合은 역시 p가 되고 j번째 主成分에 의해 설명되는 比率은 λ_j/p가 된다.

3. 主成分評價模型

위에서 살펴 본 $\boldsymbol{y}=\boldsymbol{A}^T\boldsymbol{x}$는 變數 $\boldsymbol{x}$와 主成分 $\boldsymbol{y}$간의 關係를 나타낸다. 그런데 이 때 $\boldsymbol{y}$는 平均이 0이 아닌 값을 가진다. 여기에 적절한 常數를 추가함으로써 主成分의 平均이 0이 되도록 하는 것이 일반적이다. 만약 $\boldsymbol{x}$가 平均 $\boldsymbol{\mu}$, 共分散行列 Σ일 때, 이에 대한 固有「벡터」行列을 $\boldsymbol{A}$라면,

$$\boldsymbol{y}=\boldsymbol{A}^T(\boldsymbol{x}-\boldsymbol{\mu})$$

로 轉換될 수 있다. 이것은 直角回轉에 의한 轉換이다.[15]

임의의 r企業의 經營分析을 위한 主成分評點(principal component score)을 구하기 위해서 觀察値 $\boldsymbol{x}_r$을

$$\boldsymbol{y}_r=\boldsymbol{A}^T(\boldsymbol{x}_r-\boldsymbol{\mu})$$

에 代入하면 된다. 이 때 $\boldsymbol{y}_r$를 r企業의 主成分評點이라 한다.

여기서 注意해야 할 것은 만약 $\boldsymbol{A}$가 相關係數行列 $\boldsymbol{P}$에 대한 固有「벡터」의 行列이라면, 위式은 $(\boldsymbol{x}_r-\boldsymbol{\mu})$를 分散이 1이 되도록 標準化한 후에 사용해야 된다.

---

(14) 標準化는$(X_j-\bar{X})/S_x$로 하며, 이렇게 하게 되면 變數의 單位가 없어지고, 平均 0, 分散 1 그리고 共分散은 相關係數와 일치하게 된다.

(15) $\boldsymbol{A}\boldsymbol{A}^T=\boldsymbol{I}$이므로 $\boldsymbol{A}$는 直角行列이며, $|\boldsymbol{A}|=1$이므로 $\boldsymbol{y}$는 $\boldsymbol{A}$를 매개변수로 한 直角回轉에 의한 $\boldsymbol{x}$의 「매핑」(mapping)이 된다.

그런데 主成分의 數는 p개이나 p개 모두 중요한 意味를 갖는 것은 아니다. 이제 p개의 主成分 중 중요한 意味를 가진 $m(<p)$개의 主成分을 추출하기 위해 固有値의 形態(pattern)을 살펴보아야 한다.

만약 變數들 중 一次從屬인 관계가 몇 개 있다면, Σ의 固有値 중 몇 개는 0이 된다. 觀察値들이 포함된 空間의 次元은 Σ의 「랭크」(rank)와 같으며, 이것은 ($p-0$인 固有値의 數)와 같다. 만약 k개의 0인 固有値가 있다면 次元은 $(p-k)$로 줄어들며, 우리는 變數에 대한 k개의 獨立制約式을 발견할 수 있다. 이런 制約式을 우리는 構造的 關係(structural relationship)라 한다. [16]

그런데 현실적으로 완전한 一次從屬인 관계는 드물다. 실제 더 중요한 것은 거의 一次從屬인 관계를 갖는 경우이다. 만약 가장 작은 固有値인 λ_p가 0에 가깝다면, p개 主成分 $\boldsymbol{a}_p^T\boldsymbol{x}$는 거의 일정하게 되고(constant) $\boldsymbol{x}$의 次元은 p보다 적다고 간주할 수 있다. 만약 마지막 $(p-m)$개의 固有値들이 작다고 판단되면, 次元은 m으로 줄어들게 할 수 있고 이들에 대한 效率은 $\sum_{i=1}^{m}\lambda_j/\sum_{i=1}^{p}\lambda_j$으로 나타낼 수 있다.

만약 λ_{m+1}, ……, λ_p가 작은 값을 보인다면 主成分의 값을 그들의 平均値인 0으로 대체한다고 해도 많은 情報를 잃어 버리지 않는다. 그러므로 우리는 固有値의 形態를 살려봄으로써 중요한 主成分을 발견할 수 있으며, [17] 이 때 r企業에 대한 主成分評點은

$$\boldsymbol{y}_r^T=[Y_{r1},\ Y_{r2}, \cdots,\ Y_{rm},\ 0, \cdots,\ 0]$$

임을 알 수 있다.

이제 m개의 중요한 主成分이 발견되면, 이들의 意味를 발견하는 것이 중요하다. 이를 위해 固有「벡터」인 $\{\boldsymbol{a}_j\}$ 대신에

$$\boldsymbol{a}_j^*=\lambda_j^{1/2}\boldsymbol{a}_j,\ \ j=1, 2,\ \cdots,\ p$$

를 살펴봄으로써 우리는 主成分이 함축하고 있는 意味를 발견할 수 있고, 필요한 경우에는 命名도 가능하다. [18]

이상에서 본 바와 같이 먼저 p개의 主成分 $\boldsymbol{a}_j$를 導出하고 이들 중 m개의 중요한 主成分

---

(16) P. Sprent, *Models in Regression*, London, Hethuen, 1969, p. 32.

(17) 固有値들의 合이 p인 相關係數行列로부터 分析할 때, 일반적으로 固有値가 1보다 작은 主成分은 고려하지 않는 方法을 사용한다. 이러한 방법은 실제 유용한 방법이지만 주먹구구식(rule of thumb)으로 理論的인 근거는 없다. 그러므로 固有値의 形態를 보고 意思決定하는 것이 더 合理的이다.
C. Chatfield and A.J. Collins, *op. cit.*, p. 72 參照.

(18) $\{\boldsymbol{a}_j^*\}$는 두 가지 解釋이 가능하다. 첫째 共分散行列 Σ로부터 主成分이 導出되었을 때 $\boldsymbol{a}_j^*$를 因子負荷量(component loadings)이라 하며, 둘째 相關係數行列 $\boldsymbol{P}$로부터 主成分이 導出되었을 때 因子相關(component correlations)이라 하는데, 모두 變數 X_j와 主成分 Y_j와의 關係를 나타낸다.

을 선택한다. 그리고 因子負荷量 혹은 因子相關에 의해 이들의 의미가 발견되면, 우리는 主成分評價模型을 설정할 수 있다.

즉, m개의 主成分評價模型은

$$
\begin{aligned}
Y_1 &= a_{11}(X_1-\bar{X}_1)+a_{21}(X_2-\bar{X}_2)+\cdots+a_{p1}(X_p-\bar{X}_p)=\boldsymbol{a}_1^T(\boldsymbol{x}-\bar{\boldsymbol{x}})\\
Y_2 &= a_{12}(X_1-\bar{X}_1)+a_{22}(X_2-\bar{X}_2)+\cdots+a_{p2}(X_p-\bar{X}_p)=\boldsymbol{a}_2^T(\boldsymbol{x}-\bar{\boldsymbol{x}})\\
&\vdots\\
Y_m &= a_{1m}(X_1-\bar{X}_1)+a_{2m}(X_2-\bar{X}_2)+\cdots+a_{pm}(X_p-\bar{X}_p)=\boldsymbol{a}_m^T(\boldsymbol{x}-\bar{\boldsymbol{x}})
\end{aligned}
$$

으로 된다. 이들로부터 各 企業의 主成分評點을 계산하여 經營分析을 할 수 있다.[19]

IV. 主成分分析과 判別分析의 連繫

1. 序

判別分析(discriminant analysis)은 하나의 現象을 合理的으로 분류하는 방법이다.[20] 지금 하나의 統計集團이 여러 가지 共通特性을 갖는 單位에 의하여 측정되어 있을 때, 이 統計集團을 두 개의 部分集團으로 분류하는 방법으로써 우리는 이 特性들의 良否, 有無, 濃淡, 大小 등을 비교해 볼 수 있다. 그러나 이것은 명확한 數量的 基準을 세우는 方法이라고 할 수는 없다. 이와 같은 問題點을 해결해 주는 것이 判別分析이다.

分析의 對象이 企業의 경우에도 역시 이와 같은 判別模型을 만들 수 있다. 判別分析을 이용하여 綜合評價模型을 設定할 경우, 變數의 選定이 필요한데 이는 Ⅲ章에서 구한 m개의 主成分으로 한다. 왜냐하면 原變數의 情報를 거의 함축하면서도 이들은 次元을 낮출 수 있을 뿐더러 相互獨立인 關係를 갖고 있기 때문이다. 그리고 또한 判別分析에서 대두되는 중요한 問題는 加重値의 算定인데 이를 위해 명확히 優良企業, 그리고 명확히 不良企業인 標本企業을 選定해야 할 필요성이 있다.[21]

이와 같은 變數와 標本企業이 선정되면, 判別分析을 하기전 먼저 「프로파일」分析(profile analysis)에 의해 標本企業의 選定이 잘되었는지 檢定해야 한다. 檢定結果 有意하다는 결론을 얻으면 判別分析을 이용하여 綜合評價模型을 설정한다. 이렇게 함으로써 主成分分析과

---

(19) 主成分評點으로부터 우리가 일반적으로 인식하기 쉬운 100點 滿點, 그리고 50點을 平均點으로 바꾸어 줄 경우에는 $Y_j'=\dfrac{(Y_j-\bar{Y}_j)}{\sigma_{Yj}}\times\dfrac{100}{6}+50$으로 變換시켜 준다.

(20) R.A. Fisher, The Use of Multiple Measurement in Taxonomic Problems, *Annals of Eugenic*, Vol. 7, 1936이 이 方法의 起點을 形成한다.

(21) 線型判別函數와 線型回歸函數와의 주요한 差異는 從屬變數의 성격에 기인한다. 線型回歸函數에서는 1次函數의 係數를 정하기 위해 獨立變數와 從屬變數의 값을 사용하며, 반면에 線型判別函數에서는 이러한 數値나 變數를 갖지 않으며 資料의 二元分類를 사용함으로써 係數를 추정한다.

判別分析의 連繫的活用이 가능하며 經營分析의 合理性이 提高된다 하겠다.

2. 「프로파일」分析法

判別分析을 이용하여 綜合評價模型을 設定하기 위해서는 標本企業이 合理的으로 추출되었는지에 대한 統計的 有意性을 검정할 필요성이 있다. 즉, 標本으로 선정된 優良企業群과 不良企業群間에 各 變數(主成分)에 대한 平均値의 差異가 우연으로 일어난 것인지 혹은 差異가 명확히 있다고 할 수 있는지를 검정해야 한다.

다음 〈表 2〉와 같은 資料로부터 各 變數의 平均値의 差異가 統計的으로 有意한가를 보기 위한 방법으로는 「프로파일」 分析을 활용할 수 있다.

〈表 2〉 標本企業의 資料

集團 \ 企業 \ 變數		Y_1	Y_2	...... Y_m
1 (優 良)	C_{11}	Y_{111}	Y_{211}	...... Y_{m11}
	C_{12}	Y_{121}	Y_{221}	...... Y_{m21}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	C_{n1}	Y_{1n1}	Y_{2n1}	...... Y_{mn1}
2 (不 良)	C_{12}	Y_{112}	Y_{212}	...... Y_{m12}
	C_{22}	Y_{122}	Y_{222}	...... Y_{m22}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	C_{n2}	Y_{1n2}	Y_{2n2}	...... Y_{mn2}

즉, 「프로파일」分析은 共分散行列이 같은 Σ를 갖는 두 母集團으로부터 추출한 獨立標本들에 대한 다음 假說을 검정하고자 할 때 사용된다. 즉, 歸無假說(null hypothesis)

$$H_0: \boldsymbol{C}^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=\boldsymbol{\phi}$$

$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$: 각각 $\boldsymbol{y}_1, \boldsymbol{y}_2$의 母平均

을 검정하는 경우에 필요한 統計的 分析方法이다. 이를 檢定하는 統計量(statistic)은

$$\mathcal{T}^2=\frac{n_1\ n_2}{n_1+n_2}[\boldsymbol{C}^T(\boldsymbol{y}_1-\boldsymbol{y}_2)-\boldsymbol{\phi}]^T(\boldsymbol{C}^T\boldsymbol{S}\boldsymbol{C})^{-1}[\boldsymbol{C}^T(\boldsymbol{y}_1-\boldsymbol{y}_2)-\boldsymbol{\phi}]$$

이다. 이 때

$$\boldsymbol{S}=\frac{(n_1-1)\boldsymbol{S}_1+(n_2-1)\boldsymbol{S}_2}{n_1+n_2-2}$$

로 두 集團의 結合母分散 Σ의 推定値(pooled within groups estimate of Σ)이다. 그리고 $\boldsymbol{C}$는 「랭크」가 $q(<m)$인 $(m\times q)$인 특정한 行列이다. 例컨대, 變數(主成分)의 數가 5개인 경우

$$\boldsymbol{C}^T=\begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

로 놓으면 된다.

H_0下에서 $\mathcal{T}^2$은 自由度(degree of freedem)가 (n_1+n_2-2)인「호텔링」T^2分布(Hotelling T^2-distribution)(22)를 한다. 즉,

$$\mathcal{T}^2 \sim \mathcal{T}_m^{\,2}(n_1+n_2-2)$$

가 된다. 이의 有意性檢定을 위해서는 F分布로 轉換해 주어야 하는데, 이는 다음과 같이 해 주면 된다.

$$\frac{f-m+1}{mf}\mathcal{T}_m^{\,2}(n_1+n_2-2) \sim F(m, f-m+1)$$

이 統計量으로부터 두 集團의 平均値 간에 構造的 關係가 존재하는지에 대한 검정을 할 수 있다.

그 결과 有意한 것으로 판단되면, 標本抽出은 잘 이루어진 것으로 볼 수 있으며 다음의 判別分析을 함으로써 判別係數를 도출할 수 있다.

3. 判別分析法

判別의 問題는 主成分評點을 기본으로 해서 어떤 기업을 두 集團(優良企業과 不良企業)중 하나 또는 다른 集團에 分類할 때 발생한다. 判別分析을 활용한 綜合評價模型을 위한 變數(評價項目)는 Ⅲ章에서 발견한 m개의 主成分을 사용한다.

각 變數는 두 母集團에 대해 서로 다른 分布를 한다고 가정한다. i째 母集團의 密度函數(density function)를

$$f_i(\boldsymbol{y}),\quad i=1,2$$

1: 優良企業, 2: 不良企業

라 하자. 이 때 判別法則(decisicn rule)은, 만약

$$\frac{f_1(\boldsymbol{y})}{f_2(\boldsymbol{y})} \geq \frac{\pi_2 C(1|2)}{\pi_1 C(2|1)}$$

π_i: 母集團 i에 포함될 先驗的 確率(prior probability)

$C(i|j)$: j인 企業을 i로 誤分類할 費用.

일 경우에는 優良企業으로 분류하고 그렇지 않으면 不良企業인 것으로 판별한다. $\pi_2 C(1|2)=\pi_1 C(2|1)$일 때, 判別法則은 단순히 尤度(likelihood)가 더 큰 母集團에 분류하면 된다.

만약 正規分布를 假定하는 경우, 즉,

---

(22) Student t-分布의 多變量의 一般形態로, 만약
$\boldsymbol{x} \sim N_m(\mu,\ (1/k)\Sigma)$, $f\boldsymbol{S} \sim W_m(f, \Sigma)$라면
$k(\boldsymbol{x}-\mu)^T\ \boldsymbol{S}^{-1}\ (\boldsymbol{x}-\mu) \sim T^2_m(f)$이다.

$$\boldsymbol{y} \sim N_m\ (\boldsymbol{\mu}_i,\ \Sigma),\ i=1,2$$

일 때 結合密度函數(joint density function)는

$$f_i(\boldsymbol{y})=(2\pi)^{-m/2}|\Sigma|^{-1/2}\ \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_i)^T\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_i)\right\}$$ [23]

가 된다. 그리고 이로부터

$$\ln\left[\frac{f_1(\boldsymbol{y})}{f_2(\boldsymbol{y})}\right]=\boldsymbol{y}^T\Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)-\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\Sigma^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$$

가 유도 된다. 이 때

$$\boldsymbol{d}=\Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2),$$

$$\boldsymbol{k}=\ln[\pi_1 C(1|2)/\pi_2 C(2|1)]$$

라 놓으면 判別法則은 다음과 같이 된다. 만약 어떤 企業이

$$\boldsymbol{d}^T\boldsymbol{y}-\frac{1}{2}\boldsymbol{d}^T(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\geq \boldsymbol{k}$$

이면 優良企業에, 그리고 이렇지 않을 때는 不良企業에 속하는 것으로 分類한다.

이와 같은 判別法則은 「피셔」(R.A. Fisher)의 線型判別函數(Fisher's linear discriminant function)에 기초하고 있다.[24] 그리고 $\pi_2 C(1|2)=\pi_1 C(2|1)$일 때는 $\ln 1=0$이므로 $\boldsymbol{k}=0$가 된다.[25]

이제

$$U=\boldsymbol{d}^T\boldsymbol{y}-\frac{1}{2}\boldsymbol{d}^T(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$$

로 定義하고,

$$\boldsymbol{d}^T\boldsymbol{\mu}_1>\boldsymbol{d}^T\boldsymbol{\mu}_2$$

라 假定하자. 그러면 만약 $\boldsymbol{y}$가 優良企業으로부터 抽出되었다면,

$$U=U_1\sim N_1\left(\frac{1}{2}\alpha,\ \alpha\right)$$

이며, 반면에 $\boldsymbol{y}$가 不良企業으로부터 추출되었다면,

$$U=U_2\sim N_1\left(-\frac{1}{2}\alpha,\ \alpha\right)$$

이다.[26] 이 때,

---

(23) Donald F. Morrison, *Multivariate Statistical Methods*, McGraw-Hill, 1976, p. 85.

(24) C. Chatfield and A.J. Collins, *op. cit.*, p. 134.

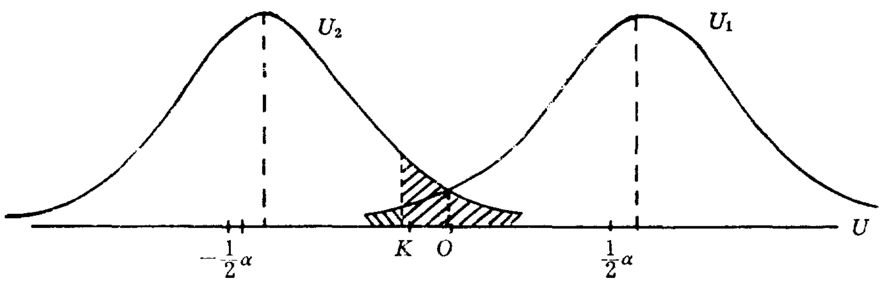
(25) 先驗的 確率이 알려져 있지 않을 때, 「미니맥스」(minimax)法則에 의해 $\boldsymbol{k}=0$로 간주한다. T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, New York, Wiley and Sons, 1958, pp. 132-133.

(26) $\boldsymbol{y}$가 多變量正規分布를 하면 이것의 一次結合인 U는 單一變數正規分布를 한다.

$$\alpha=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$$
$$=\boldsymbol{d}^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$$

는 두 母集團間의 거리로서 「마하라노비스」거리(Maharanobis distance)라고 한다.[27]

이를 그림으로 보면 다음 〔圖 1〕과 같다.



〔圖 1〕 判別函數의 分布

〔圖 1〕에서 빗금친 부분은 誤分類의 條件附確率을 나타낸다. 즉,

$$P(1|2)=P(U_2>k)$$
$$P(2|1)=P(U_1<k)$$

가 된다. 위에서 두 母集團의 分散이 같다고 가정했으므로 $k=0$일 때 두 條件附確率은 같게 된다. 誤分類의 總確率은 이 두 條件附確率을 합친

$$\pi_1P(2|1)+\pi_2P(1|2)$$

로 된다. 그리고 $k=0$ 일 때 誤分類의 確率은

$$\pi_1P(U_1<0)+\pi_2P(U_2>0)$$
$$=P(U_1<0)$$
$$=P\left[\frac{U_1-\frac{1}{2}\alpha}{\sqrt{\alpha}}<\frac{-\frac{1}{2}\alpha}{\sqrt{\alpha}}\right]$$
$$=\Phi\left(-\frac{1}{2}\sqrt{\alpha}\right)$$

$\Phi(Z)$: 標準正規分布의 分布函數

로 부터 쉽게 구할 수 있다.

그런데 判別函數는 두 集團이 같은 共分散行列을 가질 때는 $\boldsymbol{y}$의 一次函數로 되지만, 共分散行列이 같지 않을 경우에는 $\boldsymbol{y}$의 二次函數가 된다. 이 경우 判別法則은 더 복잡해진다.[28]

(27) P.C. Maharanobis, On the Generalized Distance in Statistics, *Proceedings of the National Institute of Science of India*, Vol. 12, pp. 49-55 參照.

(28) C. Chatfield and A.J. Collins, *op. cit.*, p. 136.

지금까지 우리는 두 集團에 대한 $\boldsymbol{y}$의 母集團의 分布가 알려져 있다고 가정해 왔다. 실제로는 資料로부터 分布들의 形態가 가정되고, 母數들이 추정되어진다. 判別法則은 두 正規母集團인 경우 위의 方法을 사용하지만, 실제로 母數들은 標本으로부터 구한 推定値로 대체된다. 즉,

$$\boldsymbol{d}^T\boldsymbol{y}-\frac{1}{2}\boldsymbol{d}^T(\boldsymbol{y}_1+\boldsymbol{y}_2)\geq k$$

이면 優良企業에, 그렇지 않으면 不良企業에 속하는 것으로 분류한다. 이 때 $\boldsymbol{d}=\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}_1-\bar{\boldsymbol{y}}_2)$[29]이며, $\boldsymbol{S}$는 두 集團의 結合分散인 Σ이 推定値이다.

4. 綜合評價模型

綜合評價模型을 설정하기 위해서도 역시 主成分評價模型에서와 같이, 評價項目과 이들 評價項目에 대한 加重値를 결정하여야 한다. 綜合評價項目은 主成分으로써 하였으며, 評價項目別 加重値, 즉 判別函數에 대한 係數($\boldsymbol{d}$)는 判別分析을 이용하여 추정한다.

그 결과 綜合評價模型은

$$U=d_1Y_1+d_2Y_2+\cdots+d_mY_m-\frac{1}{2}[d_1(\overline{Y}_{11}+\overline{Y}_{21})+d_2(\overline{Y}_{12}+\overline{Y}_{22})+\cdots+(\overline{Y}_{1m}+\overline{Y}_{2m})]$$

$$=\boldsymbol{d}^T\boldsymbol{y}-\frac{1}{2}\boldsymbol{d}^T(\boldsymbol{y}_1+\boldsymbol{y}_2)$$

로 구해질 수 있다. 그래서 어떤 기업의 判別評點을 구하여 k와 비교함으로써 優良企業 혹은 不良企業 중 어디에 포함되는지 알 수 있다. 나아가서 기업의 綜合的인 經營分析을 할 수 있다.

그런데 여기서 한 가지 주의할 것은 判別函數의 加重値 $\boldsymbol{d}$는 定數倍해도 그 相對的 重要性이 변하지 않는다는 사실이다.[30] 그러므로 加重値에 定數倍하여 인식하기 쉬운 尺度로 바꾸어 평가하는 것이 좋다.

V. 結 論

本硏究의 目的은 主成分評價模型과 더불어 綜合評價模型에 대해, 合理的인 評價項目과 加重値를 설정함으로써 經營分析의 合理性을 提高하는데 있다. 이를 위해 主成分分析과 判別分析을 활용함으로써 經營分析模型을 설정하는 方法을 제시하였다.

이와 같은 本硏究에서 드러난 硏究結果를 요약하면 첫째, 既存의 여러 가지 經營分析方法

(29) $\boldsymbol{d}$은 두 平均을 比較하기 위한 t^2 統計量을 極大化하는 「벡터」로서, $(\bar{\boldsymbol{y}}_1-\boldsymbol{y}_2)$의 한 單位의 변화로 인한 t^2 統計量의 變化의 感度를 나타낸다.

(30) 金俊輔, 張寅植, 張秉志, 數理統計學, 高麗大學校出版部, 1972, p. 186.

은 統計的 處理에 있어 客觀性이 결여되었으나, 本硏究에서는 評價項目設定과 加重値算定에 있어 主成分分析과 判別分析을 활용함으로써 이들의 合理性이 提高되어 그 模型이 우수하다 하겠다. 둘째, 企業經營을 主成分別 및 綜合的으로 분석함으로써 그 「시스템」이 우수하다. 그러므로 이 「시스템」을 貸出企業評價, 企業內部統制, 投資基準設定등에 이용함으로써 業務의 改善에 크게 공헌할 수 있을 것이다. 세째, 比率分析의 合理性을 提高시킬 수 있다. 傳統的 比率分析은 그 自體가 비교적 정교하지 못한 방법이기 때문에 근래에는 중요하지 않은 分析技法으로 지적되어 낮게 평가되고 있는데, 이에 最近 統計的 方法을 適用함으로써 比率分析 技法을 배제하기보다는 오히려 有用하게 이용될수 있다는 可能性을 발견하였다.

結論的으로 本硏究는 主成分評價模型과 綜合評價模型에 대한 評價項目의 설정과 加重値의 산정에 있어, 主成分分析과 判別分析의 連繫的 活用으로 經營分析의 合理性을 提高시킬 수 있는 方案을 제시하였다는데 그 意義가 있다 하겠다.

參 考 文 獻

1. 權宅淵, 李禹翰, 線型代數學, 文運堂, 1972.
2. 金俊輔, 經濟統計論, 一潮閣, 1969.
3. 金俊輔, 張寅植, 張秉志, 數理統計學, 高麗大學校出版部, 1972.
4. 金熙執, 新經營分析, 博英社, 1978.
5. 張有喆, 企業評價모델 編制方案에 관한 硏究, 高大經營大學院, 1981
6. 高松和男, 改訂經營分析の實務, 日本生產性本部, 1979.
7. 奧野忠一, 多變量解析法, 日科技連出版社, 1978.
8. 奧野忠一, 山田文道, 情報化時代の經營分析, 東京大學出版會.
9. Anderson T.W., *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, Inc., New York, London, 1958.
10. Chatfield C. and A.J. Collins, *Introduction to Multivariate Analysis*, London, Chapman and Hall, 1980.
11. Morrison Donald F. *Multivariate Statistical Methods*, 2nd ed., McGraw-Hill, 1976.
12. Nie Norman H., C. Hadlai Hull, Jean G. Jenkins and Dale H. Bent, *Statistical Package for the Social Science:* SPSS, McGraw-Hill Book Campany, New York, 1970.

13. Sprent P., *Models in Regression*, London, Methuen, 1969.

14. Stein F. Max, *Introduction to Matrices and Determinants*, Wadsworth Publishing Company, Inc., Belmont California, 1967.