

應用論文

전자부품 검사에서 대용특성을 이용한 사례연구

이종설* · 윤원영**

* LG전자, ** 부산대학교 산업공학과 (기계기술연구소)

A Case Study on Electronic Part Inspection Based on Screening Variables

Jong Seol Lee * · Won Young Yun**

LG Electronics, ** Department of Industrial Engineering(Research Institute of Mechanical Technology), Pusan National University

Keywords: Screening variable, Performance variable, Inspection, Regression, Discriminant analysis.

Abstract

In general, it is very efficient and effective to use screening variables that are correlated with the performance variable in case that measuring the performance variable is impossible (destructive) or expensive. The general methodology for searching surrogate variables is regression analysis. This paper considers the inspection problem in CRT (Cathode Ray Tube) production line, in which the performance variable (dependent variable) is binary type and screening variables are continuous. The general regression with dummy variable, discriminant analysis and binary logistic regression are considered. The cost model is also formulated to determine economically inspection procedure with screening variables.

1. 서론

현대의 기업의 품질 관리에서 품질특성에 대한 규격의 결정과 사용은 절대적이며 항구적인 것이 아니라 고객의 요구수준에 따라 변화하게 된다. 규격의 요구에 만족하는 제품을 양품이라고 하며 그러하지 못한 제품을 불량품이라고 한다. 현장에서 양/불량을 구분하기 위해서는 검사를 실시하는데 품질 특성자체를 직접적으로 검사하기 곤란한 경우가 많다. 예를 들면 검사가 파괴 검사이거

나 비용이 매우 많이 드는 경우 우리는 품질 특성과 상관성을 가진 대용 특성을 찾고 이 대용 특성을 선정하고 이에 근거한 검사 (비파괴검사 등) 를 실시하게 된다. 냉장고의 소음이나, 텔레비전 화면의 밝기 같은 것은 데시벨 (dB) 이나 휘도 (brightness)로서 정량 측정 (대용특성)이 가능하지만, 고객은 그런 정량 측정 값에 아무런 관심 없다. 데시벨로 정량화 된 값 보다 소음의 종류나 상태가 오히려 더 크게 영향을 미칠 수 있다. 이러한 감성 품질의 규격이라는 것은 양품과

불량의 이분법적인 접근만이 가능한 것이 대부분이다. 그리고 이것은 컴퓨터나 로봇에 의한 자동 판정이 거의 불가능한 항목이라고 할 수 있다.

전적으로 인간에 의해서 검사되고 판정되어야 한다면 대량 생산 체계를 갖춘 기업에서는 피하고 싶은 상황이라고 할 수 있다. 거기에 들어가는 노력이나 자원이 너무 많기 때문일 것이다. 따라서 현재 관리되고 있는 정량화된 값으로 감성의 문제를 정확하게 판정하는 문제는 현장에서 매우 중요하며 이러한 문제는 대용 특성의 적용에 관한 문제로 귀결된다.

대용(代用) 특성이란 다른 하나 혹은 그 이상의 특성의 조합으로 원하는 품질특성을 표현하는 것이다. 본 연구에서 관심을 갖는 분야는 양품과 불량품을 정량화된 관리 값으로 표현하는 것이다. 즉, 종속변수가 이치형 변수이며, 독립변수가 연속 변수인 경우 두 변수간의 관계에 대한 수학적 함수 관계를 규명하는 것이다. 종속변수가 양품/불량품인 이치형 변수이며 독립변수가 연속형 변수인 경우, 이치형 변수를 가변수(dummy variable)로 변환하여 회귀분석 (regression)을 실행하거나, 연속 변수를 일정한 범위를 정하여 분류변수화 하여 분할표 검정 등을 통해 관계를 탐색하기도 한다.[1,2,3]

본 논문에서 이러한 경우 직접 적용 가능한 로지스틱 모형 (logistic model) 과 판별 분석 (discriminant analysis)을 다루고자 한다. 즉, 로지스틱 회귀분석(logistic regression)과 판별 분석을 통해 로지스틱 확률 함수와 판별식을 구하고, 얻어진 함수식들의 실제 적용을 위한 판정의 임계 값 (critical value)을 설정하는 과정을 적용 사례로서 고찰하고자 한다. 로지스틱 확률함수와 판별식

을 구하는 문제는 다양한 통계 패키지가 일반화 됨으로 하여 누구든지 쉽게 접근 가능하게 되었다. 본 논문에서는 통계 패키지로써 Minitab을 활용했다.

2. 이치형 특성분석방법

2.1 가변수를 이용한 회귀분석

명목척도나 서열척도로 측정된 범주형 자료를 회귀식에 포함시키고자 할 때에는 가변수의 형태로 부호화(Coding)하는 방법이 널리 이용된다. 범주형 자료가 독립변수 혹은 종속변수라도 모두 부호화하여 회귀분석을 실행할 수 있다. 여기서는 종속변수가 범주형 자료인 종속 가변수의 경우만을 다루기로 한다. 범주형 변수 중 성별이나 양품/불량품처럼 이치형 변수를 회귀분석에 포함시키기 위해서는 이를 가변수로 부호화한다. 가변수의 부호화는 1과 0의 값을 부여한다.

일반적으로 특정 속성을 지니고 있으면 1, 없으면 0의 값을 부여한다. 일반선형회귀모형에서 종속변수가 가변수라면,

$$Y = \beta X + \varepsilon, Y = 0, 1 \tag{1}$$

이 되는데, Y는 1이나 0 이므로, 종속 변수 Y는 다음과 같은 확률 분포는 $P(Y=1)=p$ 그리고 $P(Y=0)=1-p$ 이고 $E(Y)=\beta X=p$ 이므로 독립변수가 어떤 값을 취할 때 Y의 기대 반응은 단순히 Y=1이 되는 확률이다. 그러나 종속변수가 0과 1만을 갖는 가변수인 경우에는 몇 가지 문제점이 발생한다. 특히 다음의 세가지 문제점은 통계적인 추정과 검정에 어려움을 야기한다.[1,2,3,4,5] 첫째, $\varepsilon = Y -$

βX 는 단지 두개의 값만 가진다. 즉 $Y=1$ 인 경우 $\varepsilon = 1 - \beta X$ 이고, $Y = 0$ 인 경우 $\varepsilon = -\beta X$ 이 되므로 오차 항이 정규분포를 하고 있지 않음은 명백하다. 둘째, β 가 되어, 일반적으로 등분산을 갖고 있지 않다. 셋째, 반응함수 $E(Y)$ 는 확률이므로 이 함수의 범위는 $0 \leq E(Y) = p \leq 1$ 이 되며, 회귀선을 아무런 제약 조건 없이 적합시키고 해석하는 것은 타당하지 않을 것이다.[3,4,6]

2.2. 로지스틱 회귀모형

종속변수가 범주형 변수인 경우 가변수를 적용하는 경우 표본의 수가 충분히 크면 회귀분석과 관련된 각종의 검사에서 오차항이 정규분포를 하고 있을 때처럼 하여도 크게 잘못을 범하지는 않는다.[1,3] 그러나 종속변수가 범주형 변수이고 독립변수가 연속형 변수인 경우 로지스틱 모형으로 보다 더 잘 적합시킬 수 있다. n 개의 독립변수가 있는 경우, 로지스틱 모형은 다음과 같이 표현된다.

$$\pi_i = \frac{\exp(\beta' X_i)}{1 + \exp(\beta' X_i)} \quad (2)$$

여기서 β 는 모형계수이며 X_i 는 독립변수들이다. 모형계수의 추정에는 최대 우도법에 의하는 경우 $n+1$ 개의 우도 방정식이 구해지고, 해는 수치적 방법(예 Newton-Raphson 방법)으로 구해진다.[3,4,6]

2.3. 판별분석

판별분석(discriminant analysis)은 여러 개의 독립 변수들을 사용하여 하나의 종속변수가 어떤 집단에 속하는가를 판별하는 통계

기법이다. 이것은 종속 변수의 집단이나 범주를 독립변수들의 선형 조합(linear combination)으로 판별 방정식을 구하고, 이것을 이용해서 새로운 사례들이 속하는 집단을 예측하는데 활용된다. 이러한 선형 조합을 판별 방정식이라 부른다. 판별 방정식은 집단 내 산포를 최소화하고, 집단 간 산포를 최대화하는 방향으로 구해지는데, 다음과 같은 형식으로 표시된다.

$$Z = A + W_1 X_1 + \dots + W_n X_n \quad (3)$$

여기서 Z 는 판별점수, A 는 상수, W_i 는 판별 계수이며, X_i 는 각 독립변수들이다. 이 논문에서는 종속변수가 이치형이므로 부분 집단이 두개인 경우이다. 이 경우 Fisher의 판별방법을 이용하여 분석 가능하다.[1,3,4]

기 관측된 자료를 통해 얻은 판별함수와 판별기준을 사용하여 소속 그룹이 알려지지 않은 관측 개체를 분류할 때, 그 분류가 정당했는지는 당장 평가가 가능한 것은 아니다. 따라서 판별함수를 도출할 때 사용한 자료를 분류 규칙에 적용하여 분류의 정확성을 통계적으로 평가해 볼 수 있다. 즉, 재대입 분류를 실시하므로 정확도, 총 오류율, 그룹별 오류율을 구할 수 있을 것이다.[3,6]

3. 적용 사례 연구

본 적용 사례는 L사의 CRT 공장에서 특정 항목의 양품과 불량품을 판정하는데 두 가지 연속 변수의 선형 조합에 의해서 실행한 사례이다. 즉, 이분변수의 대응 특성을 찾고자 하는 것이라고 할 수 있다. CRT

(Cathode Ray Tube)는 1897년 Braun에 의해 개발된 이래, T.V와 컴퓨터 Monitor용 표시 소자로서 널리 활용되고 있다. 브라운관의 제조 공정은 검사 공정을 포함하여 대부분이 자동화되어 있다. 본 사례에서 검출하고자 하는 불량은 사람의 육안에 의해서는 검출이 가능하나, 기존의 자동화된 검사 체계로는 완벽한 검출이 불가능했다. 이 불량과 관련된 인자는 여러 가지가 알려져 있었으나, 선행 연구조사 결과 전압차(Voltage Difference, X1)와 전류차(Current Difference, X2) 두 가지로 압축되었다. 한 브라운관의 전압차와 전류차는 자동으로 검사관리가 가능하므로, Y를 X1과 X2의 조합에 의해서 관리하고자 하는 것이 본 연구의 목적이다. 즉, 양품/불량품인 이치형 변수 Y와 연속형 변수 X₁과 X₂ 사이의 함수관계를 규명하고 도출된 함수식을 활용하여 Y의 검사관리를 X₁, X₂로서 간접적으로 양/불량을 구분하되 경제성을 고려한 최적의 기준점을 찾고자 한다. 분석 기초데이터는 <부록 1>과 같으며, 분석용 통계 패키지는 Minitab을 활용하고자 한다.

3.1 분석 모형 선정

종속변수가 이치형 변수이고, 독립변수가 두개의 연속형 변수이므로 분석 모형은 다음과 같이 교호작용항 까지 고려한 2차 함수 모형으로 선정한다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 + \varepsilon \quad (4)$$

통계적 분석방법은 ①일반회귀분석(general regression analysis)에서 종속변수

를 가변수로 적용한 경우 ② 선형 및 곡선 판별분석(linear and quadratic discriminant analysis) ③ 이치형 로지스틱 회귀분석(binary logistic regression analysis)을 차례로 적용해 보고, 그 결과를 상호 비교한다. 이 중에서 가장 우수한 모형에 대해서 적용을 위한 경제성을 검토하고자 한다.[1,2,3,4,6]

3.2 분석 및 결과

3.2.1. 종속 가변수를 적용한 분석

양품을 1, 불량을 0으로 부호화하여 일반 선형회귀 분석을 실시한 결과 Minitab 출력 값은 부록 2와 부록 3에 첨부되어 있다. 부록 2는 2차 및 교호작용 까지를 고려한 분석이며, 부록 3은 부록 2에서 분석 결과 유의한 것으로 판단되는 인자만을 가지고 재분석한 것이다. 여기서는 분석 시 분석 인자간의 독립성을 보완하기 위해 표준화 하였다. 즉, $CV_i = CV - CV_i$, $DV_i = DV - DV_i$ 로서 CV_i , DV_i 는 표준화된 인자 값이며, CV, DV는 각인자의 평균으로 각각 323.85, 4.8125이다. 부록 2, 3에는 분석결과 유의하다고 (유의수준 90%) 인정되는 두개의 인자에 대한 회귀 방정식이 주어지고, 분산분석표에서 해당 회귀방정식은 유의한 것으로 나타난다. 각각의 인자의 유의성도 알 수 있다. 그러나 결정계수가 각각 39.1%, 38.5%로서 설명력이 떨어진다. 일반적으로 결정계수의 65% 이상의 값은 되어야 충분한 설명력이 있는 것으로 판단하는 것에 비추어 볼 때 모형의 타당성은 부족한 것으로 판단된다.[1,2,3] 부록 8,9는 선형모형에서의 잔차의 히스토그램(Histogram)과 정규확률그림을 각각 나타낸다. 그림에서 보듯이 일반선형회귀의 기본 가정인 오차의 정규성 가정을 만족

시킨다고 판단하기는 곤란하다.

이상의 결과를 종합해 볼 때 가변수를 적용한 분석은 양품과 불량품을 판정하는 민감한 문제에 직접 적용하기는 어려운 것으로 판단된다.

3.2.2. 판별분석 모형

Minitab에서는 선형판별분석과 곡선판별분석을 지원해 주며, 우도비 원리에 의한 판별식을 제공해준다. 부록 4은 선형판별분석을 통해 출력된 결과이다. 양품인 경우 $Y=1$, 불량인 경우 $Y=0$ 으로 부호화했다. 분석을 통해 얻어진 판별식은 다음과 같다.

$$L_0(DV, CV) = -14.307 + 1.489DV + 0.052CV \quad (5)$$

$$L_1(DV, CV) = -7.790 + 0.712DV + 0.044CV \quad (6)$$

새로운 관측값이 얻어지면 위 식에 대입하여 $L_0 > L_1$ 이면 불량품으로 판별하고, 아니면 양품으로 판별한다. 이때 판별의 오류율을 고려해야 한다. 재대입 분류를 실시하면 불량을 양품으로 분류하는 오류율은 20%이며, 양품을 불량으로 분류하는 오류율은 20.5%이고, 총 오류율은 20.3% (= 전체 분류 정확도 79.7%)이다. 일반적으로 재대입 분류의 정확도는 과다하게, 오류율은 과소하게 잡히기 마련이라는 것도 감안되어야 한다.[3,6] 80% 정도의 분류 정확도를 가지고 있다면 어느 정도의 효과는 있다고 판단된다. 그러나 이것을 생산 Line의 검사 공정에 적용하기 위해서는 오류율에 대한 경제적 검토가 요구된다.

부록 5는 곡선판별분석을 적용했을 때의 Minitab 출력 결과이다. 여기서도 분류정확

도는 79.7%이다. 단지 불량을 양품으로 판정하는 비율이 16%로서 다소 줄어든다. 대신 양품을 불량으로 판정하는 비율은 23.1%로 다소 늘어난다. 종합적으로 보아 직선판별분석과 곡선판별분석은 대동소이한 결과를 보여준다. 이차항과 교호작용항을 포함하여 분석한 결과도 거의 동일했다.

3.2.3. 이치형 로지스틱 모형 분석

Minitab 패키지를 활용한 이치형 로지스틱 회귀분석 결과는 부록 6, 7과 같다. Minitab에서의 회귀계수는 최대우도법으로 추정된 것이다. 모형계수에 대한 검정은 우도비 검정에서 곡선 로지스틱 회귀분석과 직선 로지스틱 회귀분석 결과 모두가 95% 유의수준에서 귀무가설은 기각된다. 즉, 회귀계수 모두가 0은 아니다. 또한 Pearson 카이제곱 통계량으로 적합성을 검정 할 수 있다. Pearson 카이제곱 통계량은 곡선의 경우 48.831로서 자유도 56에서 P-값이 0.731로서 적합결여를 나타내지 않는다. 또한 직선회귀분석의 경우는 Pearson 카이제곱 통계량이 46.884로서 자유도 59에서 P-값이 0.873이므로, 적합결여는 나타내지 않는다. 즉, 두 모델 모두 적용은 가능하다. 그러나, 곡선의 경우 제곱항 및 교호작용항은 회귀계수가 유의하지 않다. 이것은 로지스틱 분석표에서 z값과 그에 따르는 Wald 통계량(z^2)의 P-값으로 확인 가능하다. 즉, 굳이 복잡한 모델로 분석할 필요는 없이, 부록 7과 같이 선형 로지스틱 모형으로도 충분히 설명 가능하다. 따라서 모형의 구조는

$$P\{Y=1|X_1, X_2\} = \pi(X_1, X_2) = \frac{\text{Exp}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{1 + \text{Exp}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)} \quad (7)$$

에서, $\beta_0=8.347$, $\beta_1=-0.7784$, $\beta_2=-0.011554$ 이므로, 분석 모형의 선형식은 다음과 같다.

$$P\{Y = 1|DV, CV\} = \pi(DV, CV) = \frac{\text{Exp}(8.347 - 0.7784DV - 0.011554CV)}{1 + \text{Exp}(8.347 - 0.7784DV - 0.011554CV)}$$

그리고 이것은 DV(전압차)와 CV(전류차)로 해석되는 양품이 될 확률함수이다.

3.2.4 기준값의 결정

종합적으로 고려하여 적용 모델은 로지스틱모형을 적용하는 것으로 결정할 수있으며 따라서, 로지스틱확률함수를 적용하기 위해서는 양품이 될 확률이 어떤 값 이상일 때 양품으로 판정할 것인가 하는 문제를 결정해야 한다.[6,8] 로지스틱 확률 함수나, 판별 함수를 구하여 종속변수를 독립변수의 조합으로 관리한다면, 그것이 특히 양품과 불량품을 판정하는데 적용한다면 어떤 한계값을 양품과 불량품의 기준으로 삼느냐 하는 것은 매우 중요한 문제이다. 기준이 느슨하면 불량품의 유출이 심할 것이고, 기준이 너무 까다로우면 양품이 불량품으로 처리되는 손실이 커지게 된다. 결국 양자 사이의 최적점을 찾는 것이 필요해 진다. 로지스틱 회귀모형에서는 다음과 같은 방법으로 기준값을 결정할 수 있다.

$$P\{Y = 1|X_1, X_2\} = \pi(X_1, X_2) = \frac{\text{Exp}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{1 + \text{Exp}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}$$

여기서 $G(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ 라고 두면 $P\{Y = 1|X_1, X_2\}$ 은 $G(X_1, X_2)$ 의 증가함수

이다. 그러므로 $G(X_1, X_2) \geq \lambda$ 일때 Y 를 1로 분류하는 것은 타당하며 경제적인 λ 를 결정하는 것이 문제다. 이를 위해 $\alpha = P\{G(X_1, X_2) \geq \lambda\}$ 라고 두면 α 는 1로 분류하는 비율이다.

X_1, X_2 가 이변량 정규분포를 따른다고 가정하고 정규분포의 밀도함수를 $f(x_1, x_2)$ 라면

$$\alpha = \int_{G(X_1, X_2) \geq \lambda} f(x_1, x_2) dx_1 dx_2$$

$$\beta = P\{Y = 1, G(X_1, X_2) \geq \lambda\} = \int_{G(X_1, X_2) \geq \lambda} P\{Y = 1|X_1 = x_1, X_2 = x_2\} f(x_1, x_2) dx_1 dx_2 \quad (8)$$

여기서 β 는 양품을 바로 판정하는 비율이다. 대용특성으로 분류하므로 양품이 불량으로 판정되거나 (제1종 오류), 불량품이 양품으로 판정되는 오류(제2종 오류)가 존재하게 된다. 제1종 오류와 제2종 오류에 의해 발생되는 비용을 각각 C_r, C_a ($C_r < C_a$)라하면, 제1종 오류의 기대비용은

$$EC_1 = C_r P\{Y = 1, G(X_1, X_2) < \lambda\} = C_r (\gamma - \beta) \quad (9)$$

여기서, γ 는 대용특성 사용 전의 전체 제품의 평균 양품율이다. 그리고 제2종 오류의 기대비용은

$$EC_2 = C_a P\{Y = 0, G(X_1, X_2) \geq \lambda\} = C_a (\alpha - \beta) \quad (10)$$

그러므로 제품의 총 기대비용은

$$ETC = EC_1 + EC_2 = C_r(\gamma - \beta) + C_a(\alpha - \beta) \quad (11)$$

최적 기준값 λ 는 (11)식을 최소화하는 λ 값이 된다.

$$ETC = C_r \int_{G(X_1, X_2) < \lambda} \frac{1}{1 + \text{Exp}[-G(X_1, X_2)]} f(x_1, x_2) dx_1 dx_2 + C_a \int_{G(X_1, X_2) \geq \lambda} \left(1 - \frac{1}{1 + \text{Exp}[-G(X_1, X_2)]}\right) f(x_1, x_2) dx_1 dx_2 \quad (12)$$

만일 G 의 확률밀도함수를 $f(g)$ 라고 하면

$$ETC = C_r \int_{-\infty}^{\lambda} \frac{1}{1 + \text{Exp}[-g]} f(g) dg + C_a \int_{\lambda}^{\infty} \left(1 - \frac{1}{1 + \text{Exp}[-g]}\right) f(g) dg \quad (13)$$

(13) 식을 미분하여 0로 두면,

$$C_r \frac{1}{1 + \text{exp}(-\lambda)} f(\lambda) + C_a \left(1 - \frac{1}{1 + \text{exp}(-\lambda)}\right) f(\lambda) = 0$$

위의 식을 풀면 $\lambda = \ln(C_a / C_r)$

따라서 최소기대비용을 보증하는 분류방법은

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 \leq \ln(C_a / C_r) \quad (14)$$

그러므로 대용특성에 의한 검사 분류는 대용특성과 품질특성과의 관계식과 오류비용에 의해 결정되며 오류비용은 검사비용, 수리비용, 폐기 비용, 불량 유출에 따른 서비스 비용을 포함한 기타 유,무형의 비용 등을 종합적으로 고려해야 한다. L전자의 브라운관

생산 Line은 자동검사 공정이며, 자동검사에서 발견되는 불량에 대해서는 수동 재검사 기능을 가지고 있다. 즉, 양품을 불량으로 판정하는 경우의 비용은 재검사 비용이며, 불량을 양품으로 판정하는 경우 발생하는 비용은 고객 서비스 비용을 포함하여 여러 가지 유,무형의 손실을 생각할 수 있다. 최소 비용만 고려하여도 고객 서비스 비용이 재검사 비용에 비해서 1.5배 정도로 집계되었다. 따라서 구하는 기준값(λ)은 $\ln(1.5)=0.41$ 이 된다. 0.41을 기준으로 적용하는 경우 재분류 정확도는 다음과 같다. 양품을 불량품으로 분류하는 오류율은 12.8%이며, 불량품을 양품으로 분류하는 오류율은 36%이며, 총 오류율은 21.9% (전체 분류 정확도 78.1%)이다. 이것은 선형판별분석의 결과와 매우 유사하다.

<표 1> 이치형 로지스틱 모형에서의 재분류 정확도

		분류그룹	
		0	1
실제그룹	0	16	9
	1	5	34

제1종 오류와 제2종 오류에 따른 손실 비용의 비율에 따라서 판정 기준값은 달라진다. 정량적 집계가 어렵지만, 제 2종 오류에 의한 손실 비용은 제1종 오류에 의한 손실 비용의 1.5배를 훨씬 초과할 것으로 예상된다.

C_a 와 C_r 의 비가 2가 되는 경우 판정 기준값은 0.68이 되고, 이에 따른 전체 분류 정확도는 68.7%가 된다. 제1종 오류와 제2종 오류의 비가 2.7 이상이 되면 판정 기준값은

0.99가 되는데, 이 경우 모든 제품은 불량품으로 판정하게 된다. 즉, 로지스틱 판별함수의 적용이라는 것은 의미가 없고, 전수검사를 하는 것이 유리하다는 결론에 도달하게 된다. 제1종 오류와 제2종 오류의 비율에 따른 재대입 분류표에 의한 분류 정확도는 < 표 2>와 같다.

<표 2> 선형판별분석과 선형 로지스틱분석 결과 비교

구분	1종 오류	2종 오류	총 오류율
판별분석	20.5%	20.0%	20.3%
로지스틱 분석	12.8%	36.0%	12.9%

실제의 적용은 생산 공정의 특성을 고려하여 제1종 오류율이 낮은 로지스틱 분석 결과를 적용했다. 이것은 추가 자원의 투입 없이 재검사가 가능한 정도를 고려한 것이다.

4. 결론

측정이 불가능한 특성을 대응특성으로 판정한다는 것은 매우 경제적이며 효과적인 일이다. 그러나 이것이 양품과 불량품이라는 치명적인 특성을 풀어내는 것이라면 대응특성으로 판정할 때의 오류율에 대한 철저한 고려가 대단히 중요하다.

본 논문에서는 이러한 관점에서 독립변수가 이치형 변수이고 종속변수가 연속형변수인 경우 독립변수와 종속변수 간의 함수 관계식을 구하기 위해 가변수를 적용한 일반 선형모형, 판별분석에 의한 판별식, 그리고 이치형 로지스틱 회귀모형을 고찰하고, 판정의 오류율 관리를 위한 기준값 설정 문제를

경제적 관점에서 접근했다. 양산 검사 공정에 적용 결과 고객 불량은 50% 이상을 줄일 수 있었다. 반대 급부로서 재작업의 부하가 늘어난 것은 사실이지만 고객 불량의 50% 이상 감소는 충분히 가치있는 선택이다.

적용 사례 연구에서는 이러한 방법이 적용 가능성이 많음을 시사한다. 대량 생산 공정에서의 내외부 실패 비용의 격차 문제, 또한 인적(人的) 검사의 한계 - 비용, 시간등의 한계-를 극복하는 방안으로 채택 가능하기 때문이다. 또한 검사에의 적용 뿐만 아니라 감성 품질의 계량화를 통한 개선기회를 명확히 포착 할 수 있으며, 원류 관리에도 이러한 방법의 적용이 가능하다고 본다. 금번 연구에서는 명목변수(nominal variable) 중 이치형변수에 대해서만 접근했지만 다범주명목변수(multiple nominal variable)나 순서변수(ordinal variable)에도 개념의 확장이 가능하다고 본다.

참고문헌

[1] 김기영, 진명식(1994), 「다변량 통계 자료분석」, 자유아카데미.
 [2] 박성현(1999), 「회귀분석」, 3판, 민영사.
 [3] 정광모, 최용석(1999), 「범주형 자료 분석」, 자유아카데미.
 [4] Agresti, A.(1995), *An Introduction to Categorical Data Analysis*, Wiley-Interscience.
 [5] Cox, D.R.(1970), *The Analysis of Binary Data*, Chapman and Hall.
 [6] Kim, S.B. and Bai, D.S. (1990), "Economic Screening Procedures in Logistic and Normal Models", *Naval*

- Research Logistics*, Vol. 37 pp. 919-928
- [7] Neter, J., Wasserman, W. and Kutner, M.H.(1989), *Applied Linear Regression Models*, Irwin, 1989
- [8] Tang, K. and Tang, J.(1994), "Design of Screening Procedures: A Review", *Journal of Quality Technology*, Vol.26, pp. 209-226.
- [9] *Minitab User's Guide* (1998), Vol. 1 and Vol. 2.

<부록>

부록1. 사례의 측정 데이터

No	DV	CV	Y
1	5.2	370	0
2	7.4	360	0
3	7	330	1
4	4.6	380	1
5	7.8	340	0
6	8.6	360	0
7	2.2	240	1
8	7	320	1
9	6	260	1
10	4.2	390	1
11	5.4	300	1
12	6.6	320	1
13	4.2	380	1
14	2.8	120	1
15	6	220	0
16	8	350	0
17	5.2	380	0
18	6.6	320	0
19	7.4	300	1
20	5.8	430	0
21	4.4	420	1
22	8.8	420	0
23	4.2	350	0
24	3	320	0
25	5.6	280	1
26	7.4	290	0
27	8.2	360	0
28	8	360	0
29	8.8	630	0
30	4.2	410	1
31	2	210	1
32	3.2	200	1

No	DV	CV	Y
33	5.8	420	0
34	5.4	450	0
35	7.6	320	0
36	0.4	370	1
37	2	150	1
38	3.8	310	1
39	2.8	320	1
40	4.6	400	1
41	2.6	240	1
42	4.4	380	1
43	6.6	430	0
44	3.8	280	1
45	4.6	370	1
46	5.2	260	0
47	2	180	1
48	5.4	230	1
49	2.2	250	1
50	2.2	230	1
51	2.6	250	1
52	1.8	180	1
53	4.6	400	0
54	3.4	300	1
55	3.6	310	1
56	3.6	210	1
57	1.2	300	1
58	4	340	0
59	3.4	290	1
60	2.8	250	1
61	5.4	430	0
62	4.6	390	0
63	4.6	330	1
64	5.2	430	1

부록 2 가변수를 적용한 경우 모형결과(이차 모형)

Regression Analysis: Y versus DV, CV

The regression equation is

$$Y = 1.55 - 0.117 DV - 0.00117 CV$$

Predictor	Coef	SE Coef	T	P
Constant	1.5527	0.1938	8.01	0.000
DV	-0.11743	0.02781	-4.22	0.000
CV	-0.0011682	0.0006647	-1.76	0.084

S = 0.3921 **R-Sq = 38.5%** **R-Sq(adj) = 36.4%**

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	5.8585	2.9292	19.06	0.000

부록 3 가변수를 적용한 경우 모형결과(일차모형)

Regression Analysis: Y versus DV-, CV_, DV2_, CV2_, DCV_

The regression equation is

$$Y = 0.626 - 0.114 DV- - 0.00128 CV_ - 0.0051 DV2_ + 0.000004 CV2_ - 0.000245 DCV_$$

Predictor	Coef	SE Coef	T	P
Constant	0.62599	0.07448	8.41	0.000
DV-	-0.11393	0.02891	-3.94	0.000
CV_	-0.0012814	0.0007075	-1.81	0.075
DV2_	-0.00513	0.01305	-0.39	0.696
CV2_	0.00000361	0.00000837	0.43	0.668
DCV_	-0.0002451	0.0006003	-0.41	0.685

S = 0.3999 **R-Sq = 39.1%** **R-Sq(adj) = 33.9%**

Analysis of Variance

부록 4. 선형판별분석의 Minitab 출력 결과

Discriminant Analysis: Y versus DV, CV

Linear Method for Response: Y
 Predictors: DV CV

Group	0	1
Count	25	39

Summary of Classification
 Put intoTrue Group....

Group	0	1
0	20	8
1	5	31
Total N	25	39
N Correct	20	31
Proportion	0.800	0.795

N = 64 N Correct = 51
 Proportion Correct = 0.797

Linear Discriminant Function for Group

	0	1
Constant	-14.307	-7.790
DV	1.489	0.712
CV	0.052	0.044

부록 5. 곡선판별분석의 Minitab 출력 결과

Discriminant Analysis: Y versus DV, CV

Quadratic Method for Response: Y
 Predictors: DV CV

Group	0	1
Count	25	39

Summary of Classification
 Put intoTrue Group....

Group	0	1
0	21	9
1	4	30
Total N	25	39
N Correct	21	30
Proportion	0.840	0.769

N = 64 N Correct = 51
 Proportion Correct = 0.797

부록 6 곱항 및 교호작용항을 포함한 2차 로지스틱 회귀모형의
Minitab 출력 결과(각 인자는 표준화함)

Binary Logistic Regression: Y versus DV-, CV_, DV2_, CV2_, DCV_

Link Function: Logit
Response Information

Variable	Value	Count
Y	1	39 (Event)
	0	25
Total		64

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	0.8886	0.6676	1.33	0.183			
DV-	-0.8191	0.3297	-2.48	0.013	0.44	0.23	0.84
CV_	0.010969	0.005958	-1.84	0.066	0.99	0.98	1.00
DV2_	0.0202	0.1780	0.11	0.910	1.02	0.72	1.45
CV2_	-8.015E-06	0.00008701	-0.09	0.927	1.00	1.00	1.00
DCV_	-0.002748	0.006269	-0.44	0.661	1.00	0.99	1.01

Log-Likelihood = -26.918
Test that all slopes are zero: G = 31.799, DF = 5, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	48.831	56	0.741
Deviance	48.291	56	0.758
Hosmer-Lemeshow	10.726	8	0.218

부록 7 주인자만 고려한 선형 로지스틱 회귀모형의
Minitab 출력 결과

Binary Logistic Regression: Y versus DV, CV

Link Function: Logit
Response Information

Variable	Value	Count
Y	1	39 (Event)
	0	25
Total		64

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	8.349	2.401	3.48	0.001			
DV	-0.7784	0.2375	-3.28	0.001	0.46	0.29	0.73
CV	-0.011554	0.005631	-2.05	0.040	0.99	0.98	1.00

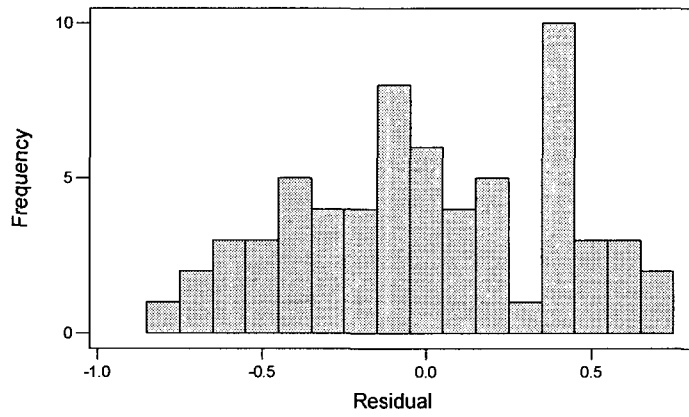
Log-Likelihood = -27.115
Test that all slopes are zero: G = 31.404, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	46.884	59	0.873
Deviance	48.686	59	0.829
Hosmer-Lemeshow	10.917	8	0.206

부록 8. 잔차의 히스토그램

Histogram of the Residuals
(response is Y)



부록 9. 잔차의 정규확률지결과

Normal Probability Plot of the Residuals
(response is Y)

