

의학적 의사결정 지표의 고찰 및 해석에 기초한 품질통계기법의 적용

최성운*

*가천대학교 산업공학과

Application of Quality Statistical Techniques Based on the Review and the Interpretation of
Medical Decision Metrics

Sungwoon Choi*

*Department of Industrial Engineering, Gachon University

Abstract

This research paper introduces the application and implementation of medical decision metrics that classifies medical decision-making into four different metrics using statistical diagnostic tools, such as confusion matrix, normal distribution, Bayesian prediction and Receiver Operating Curve(ROC).

In this study, the metrics are developed based on cross-section study, cohort study and case-control study done by systematic literature review and reformulated the structure of type I error, type II error, confidence level and power of detection.

The study proposed implementation strategies for 10 quality improvement activities via 14 medical decision metrics which consider specificity and sensitivity in terms of α and β .

Examples of ROC implication are depicted in this paper with a useful guidelines to implement a continuous quality improvement, not only in a variable acceptance sampling in Quality Control(QC) but also in a supplier grading score chart in Supplier Chain Management(SCM) quality.

This research paper is the first to apply and implement medical decision-making tools as quality improvement activities. These proposed models will help quality practitioners to enhance the process and product quality level.

Keywords: Medical Decision Statistics, Quality Improvement Tools, Confusion Matrix, normal Distribution, Bayesian Prediction, ROC, Specificity, Sensitivity, Variable Acceptance Sampling, Supplier Grading Scoring.

1. 서론

인간을 대상으로 하는 임상연구에서는 안전성(Do Not Harm)과 윤리성(Ethics)이 우선되어야 하나 의사의 체계적이지 못한 경험, 신념, 통념과 병태생리학 검사의 한계로 인해 최선의 의료행위가 실시되고 있지 않을 수 있다. 따라서 최근 체계적인 문헌검토와 메타

분석(Systematic Review and Meta Analysis), RCT(Randomized Clinical Trial)에 의한 최적 근거를 임상통계학을 이용해 비판적으로 평가(Critical Appraising)하여 임상의학의 불확실성에 대처하기 위한 문제해결형 의사결정 방식의 EBM(Evidence-Based Medicine)에 대한 필요성이 제기되고 있다.[2]

† 이 논문은 2013년 가천대학교 교내연구비 지원에 의한 것임(GCU-2012-R008)

† Corresponding Author : Sungwoon Choi, 1342 Seongnamdaero ,Sujeong-gu, Seongnam-si, Gyeonggi-do, Korea MP: 011-256-0697, E-mail: swchoi@gachon.ac.kr

Received January 9, 2013; Revision Received March 5, 2013; Accepted March 13, 2013.

이러한 EBM에서는 PICO(Population, Intervention, Comparison Outcome) 또는 PEO(Population, Exposure, Outcome)에 의한 연구의 타당성 검토, 병인인 위험요인의 평가, 관측값의 요약정리, 진단, 치료, 예측의 평가를 위한 의사결정 방법으로 임상기술통계학, 통계적 추론, 혼돈행렬(Confusion Matrix, Contingency Table), 베이지안 예측(Bayesian Prediction), ROC(Receiver Operating Characteristics)곡선, 오즈비(Odds Ratio) 등을 사용하고 있다.[1-2,5,13] 이러한 의학 의사결정(Medical Decision Making) 방법을 신약에 대한 생동등성(Bioequivalence) 분야[3,9], 범죄통계[4,16], 부도(Default) 회사에 대한 신용평가(Credit Scoring)[8,15]에서 응용 연구되어 오고 있다.

그러나 품질평가 및 개선에서 활용되는 로트의 합부 판정을 위한 샘플링 검사(Acceptance Samplings)[12,14], 공정의 이상모니터링을 위한 SPC(Statistical Process Control) 관리도(Control Chart)[6,17], 품질개선의 효과과약을 위한 품질가설검정(Hypothesis Testing)[7,11]에서 판정의 판별력을 향상시키기 위해 의학적 의사결정 기법을 적용한 연구는 전혀 이루어지고 있지 않다.

연구목적(Purpose)은 임상연구의 의학적 의사결정지표(Medical Decision Making Metrics)를 정규분포, 베이지안 예측(Bayesian Prediction), ROC(Receiver Operating Characteristics)곡선의 네 가지 관점에서 유형화하고 품질통계 분야에 대한 응용방안과 적용사례를 제시한다.

연구방법(Design, Methodology, Approach)은 진단조사를 위한 횡단연구(Cross-Sectional Study), 임상결과 관찰을 위한 코호트 연구(Cohort Study), 병력추적을 위한 환자-대조군 연구(Case-Control Study) 등 3가지 임상연구의 체계적인 문헌고찰(Systematic Literature Review)을 통하여 의학적 의사결정지표를 유형화하고 샘플링검사, 협력업체 품질등급평가에서의 ROC 적용사례(Case Study)를 제시한다.

연구결과(Findings)는 의학적 의사결정지표를 네 가지 관점에서 유형화하고, Specificity에 영향을 주는 α 와 Sensitivity에 영향을 주는 β 의 해석을 통한 통계기법에서의 적용방안을 제시한다. 또한 계량샘플링검사에서 상한규격, 하한규격, 양쪽규격 등의 세 가지 스페셜 ROC의 최적 절단점 결정방안과 공급자등급평가지 로트판정의 변별력(Discriminative Power)를 향상할 수 있는 적용방안을 제시한다.

연구한계(Limitations)는 적용가능한 품질통계기법 중 샘플링검사와 공급자 평가에 대한 적용사례만 제시한 데 있다.

실무적용(Practical Implications)은 검수시 효율적인

샘플링검사를 이용하는 구매품질부서와 SCM(Supply Chain Management) 품질의 구매품질부서에서 장기적인 협력관계를 유지하기 위해 공급자 품질등급평가를 하는 품질담당자의 이해를 돕기 위한 ROC 적용사례를 제시한다.

연구차별성(Originality, Value)은 의학적 의사결정지표를 EBM(Evidence-Based Medicine)의 체계적 고찰(Systematic Review)를 통해 새로운 4가지 관점(Viewpoint, Perspective)에서 유형화(Classification)하고, 통계기법 용도별 FP(False Positive)와 FN(False Negative)의 해석으로 지표에 대한 적용방안(Application Strategy)을 처음으로 제안하여 샘플링검사와 협력업체평가에 대한 사례(Case Study)를 제시한 데 있다.

2. 의학적 의사결정 지표의 고찰 및 해석

2.1 Confusion Matrix에 의한 이해

의학적 의사결정 지표(Medical Decision Making Metrics)는 횡단면 연구(Cross-Sectional Study)에 의해 <Table 1>과 같은 2×2 Confusion Matrix에 의해 결정되어 +(Positive, 양성)는 질병(Disease), 비정상(Nonnormal)이고 -(Negative, 음성)는 비질병(Non-Disease), 정상(Normal)이다. 열(Column)은 Gold Standard에 의한 기준(Reference) 또는 질병의 존재(Presence of Disease), 실제질병(Actual Disease)을, 행(Row)은 의사결정(Decision) 또는 시험결과(Test Result), 시험예측(Test Prediction)을 나타낸다. <Table 1>에서 $1 - \beta$ 는 진양성(True Positive, TP), 적중(Hit), 검출력(Power of Detection)이고 β 는 가음성(False Negative, FN), 실수(Miss), 제2종 오차(Type II Error), 소비자 위험(Consumer's Risk)이다. α 는 가양성(False Positive, FP), 실수경보(False Alarm), 제1종 오차(Type I Error), 유의수준(Significance Level), 생산자위험(Producer's Risk)이고, $1 - \alpha$ 는 진음성(True Negative, TN), 옳은 기각(Correct Rejection), 신뢰수준(Certainty Level)이다. 여기서 α , β 는 목표 샘플링오차로 실제 샘플링오차는 확률 또는 확률분포에 의해 구해진다.

첫 번째 열의 관계에서 1행 1열은 TPR(True Positive Rate), Sensitivity, 이익, 적중률(Hit Rate), 리콜(Recall)이고 2행 1열은 FNR(False Negative Rate)이다. 두 번째 열의 관계에서 1행 2열은 FPR(False Positive Rate), 폴아웃(Fall-Out), $1 - \text{Specificity}$ 이고, 2행 2열은 TNR(True Negative Rate), Specificity이다. 이를 <Table 1>에 의해 표시하면 식(1)-(4)와 같다.

<Table 1> 2×2 Confusion Matrix

Presence of Disease \ Test Result	<i>D+</i>	<i>D-</i>	Total
	<i>TP</i> $P(T+ D+) = 1 - \beta$	<i>FP</i> $P(T+ D-) = \alpha$	<i>TP + FP</i>
<i>T-</i>	<i>FN</i> $P(T- D+) = \beta$	<i>TN</i> $P(T- D-) = 1 - \alpha$	<i>FN + TN</i>
Total	<i>TP + FN</i>	<i>FP + TN</i>	

$$Sensitivity = \frac{TP}{TP + FN} = \frac{1 - \beta}{(1 - \beta) + \beta} \quad (1)$$

$$FNR = \frac{FN}{TP + FN} = \frac{\beta}{(1 - \beta) + \beta} \quad (2)$$

$$FPR = \frac{FP}{FP + TN} = \frac{\alpha}{\alpha + (1 - \alpha)} \quad (3)$$

$$Specificity = \frac{TN}{FP + TN} = \frac{1 - \alpha}{\alpha + (1 - \alpha)} \quad (4)$$

첫 번째 행의 관계에서 1행 1열은 검사 후 확률인 PPV(Positive Prediction Value, 양성예측도), 정밀도(Precision)이고, 1행 2열은 NPV(Negative Prediction Value, 음성예측도)이다. 두 예측도의 비(%)와 차(-)로 비교하기 위한 위험도를 각각 RRR(Relative Risk Reduction, 상대위험도감소), ARR(Absolute Risk Reduction, 절대위험도감소)라고 한다. NNT(Number Needed to Treat)는 ARR의 역수로 몇 명을 치료해야 1명의 Risk를 예방할 수 있는지의 지수이다.

$$PPV = \frac{TP}{TP + FP} = \frac{1 - \beta}{(1 - \beta) + \alpha} \quad (5)$$

$$NPV = \frac{FN}{TP + FP} = \frac{\alpha}{(1 - \beta) + \alpha} \quad (6)$$

$$RRR = \frac{PPV}{NPV} \quad (7)$$

$$ARR = PPV - NPV \quad (8)$$

임상연구는 시간적 분류에 따라 현재의 위험(Risk, Exposure) 요인 및 치료(Intervention, Etiology)의 유무(행의 두 그룹)가 미래의 시간을 추적해 발생 가능한 질환 및 증상의 결과지표(Outcome, Event) 유무(열의 두 그룹)를 평가하는 전향적 방법인 코호트 연구(Cohort Study)가 있다. 이와 반대로 현재 질환 및 증

상의 결과지표(Outcome) 유무에 따라 실험 환자(Experimental Case)와 대조군(Control, Placebo)으로 나누어 과거의 시간을 추적해 과거위험요인 및 치료유무를 분류하는 후향적 연구와 환자-대조군 연구(Case-Control Study)가 있다. 또한 현재 시점에서 위험요인 및 치료의 유무와 질환 및 증상의 결과지표 유무의 관련성이나 관측값을 조사하는 횡단연구(Cross-Sectional Study)가 있다.

<Table 1>의 열(Column)을 질환 및 증상의 결과지표의 유무로 행(Row)을 위험요인의 유무(실험군과 대조군)로 고려할 경우 위의 3가지 임상연구에 따라 위험도 계산 방법이 달라진다. 진단적 조사를 위한 횡단연구에서의 진단 위험도는 식(7), (8)과 같고, 임상경과 관찰을 위한 코호트 연구에서의 예측 위험도는 EER(Experimental Event Rate), CER(Control Event Rate), RR(Relative Risk), AR(Attributable Risk)로 식(9)-(12)와 같다.

$$EER = \frac{TP}{TP + FP} = \frac{(1 - \beta)}{(1 - \beta) + \alpha} \quad (9)$$

$$CER = \frac{FN}{FN + TN} = \frac{\beta}{\beta + (1 - \alpha)} \quad (10)$$

$$RR = \frac{EER}{CER} \quad (11)$$

$$AR = EER - CER \quad (12)$$

코호트 연구에서는 위험도를 추적할 시 행의 위험요인에 따라 열의 질환의 결과지표의 사건발생률을 모두 계산했지만, 병력 추적을 위한 환자-대조군 연구에서는 열의 두 질환결과별 행의 위험요인의 유무의 상대적 OR(Odds Ratio)로 식(13)과 같으며 이는 1행과 2행의 위험요인에 1열과 2열의 상대적 비라는 의미이다.

$$OR = \frac{\frac{TP}{FP}}{\frac{FN}{TN}} = \frac{\frac{1-\beta}{\alpha}}{\frac{\beta}{1-\alpha}} \quad (13)$$

$$= \frac{TP \times TN}{FP \times FN} = \frac{(1-\beta) \times (1-\alpha)}{\alpha \beta}$$

첫 번째 행의 관계에서 1행 1열과 1행 2열은 검사 후 확률(Post-Test Probability)인 LR+(Positive Likelihood Ratio, 양성가능비)이고 대각선의 관계에서 1행 2열과 2행 1열은 검사 후 확률인 LR-(Negative Likelihood Ratio, 음성가능비)이다. 전체와 1행 1열과 1행 2열 대각선의 관계는 Accuracy(정확성, 효율성)이고, 전체와 두 번째 열의 관계는 검사 전 확률(Pre-Test Probability) Prevalence(유병률)을 나타낸다. 이를 <Table 1>에 의해 표시하면 식(14)-(17)과 같다.

$$LR+ = \frac{TP}{FP} = \frac{Sensitivity}{1 - Specificity} = \frac{(1-\beta)}{\alpha} \quad (14)$$

$$LR- = \frac{FN}{FP} = \frac{1 - Sensitivity}{1 - Specificity} = \frac{\beta}{\alpha} \quad (15)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} = \frac{(1-\beta) + (1-\alpha)}{(1-\beta) + \beta + \alpha + (1-\alpha)} \quad (16)$$

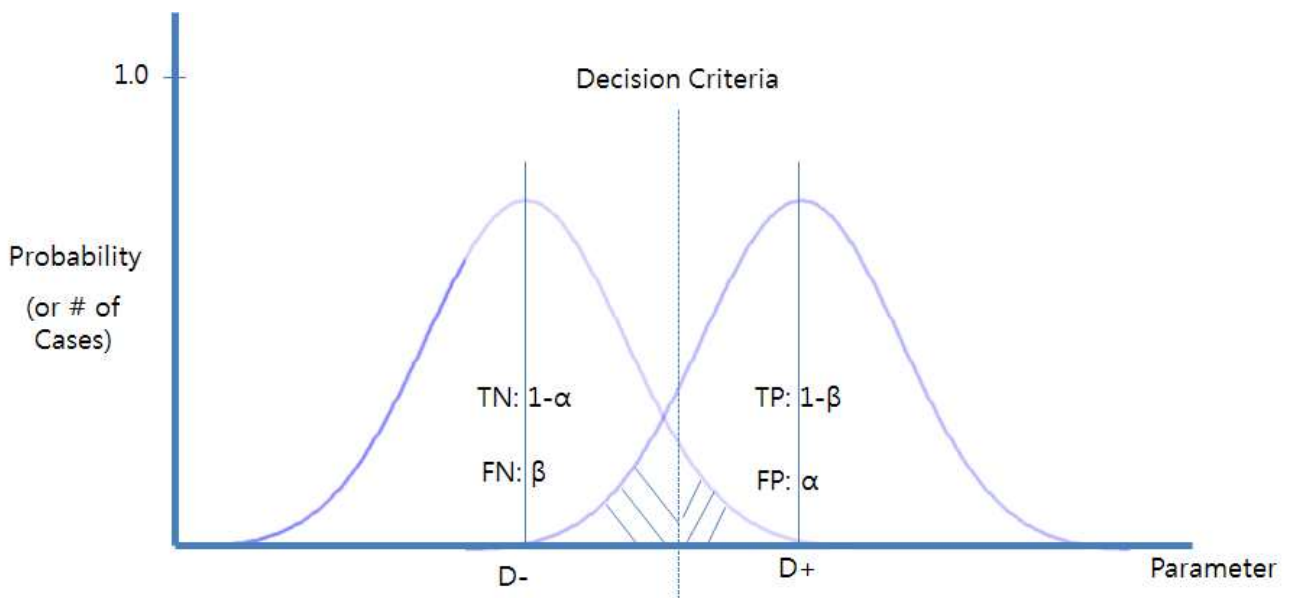
$$Prevalence = \frac{TP + FN}{TP + FN + FP + TN} = \frac{(1-\beta) + \beta}{(1-\beta) + \beta + \alpha + (1-\alpha)} \quad (17)$$

(9)식에서 Sensitivity가 매우 큰 경우 즉 (15)식에서 LR-가 작은 경우 1-Specificity가 크므로 Negative(음성)이면 진단을 Rule-Out(제외)할 수 있다. (10)식에서 Specificity가 매우 큰 경우 즉 (14)식에서 LR+가 큰 경우 Sensitivity가 커지므로 시험결과가 Positive(양성)이면 진단을 Rule-In(확정)할 수 있다.

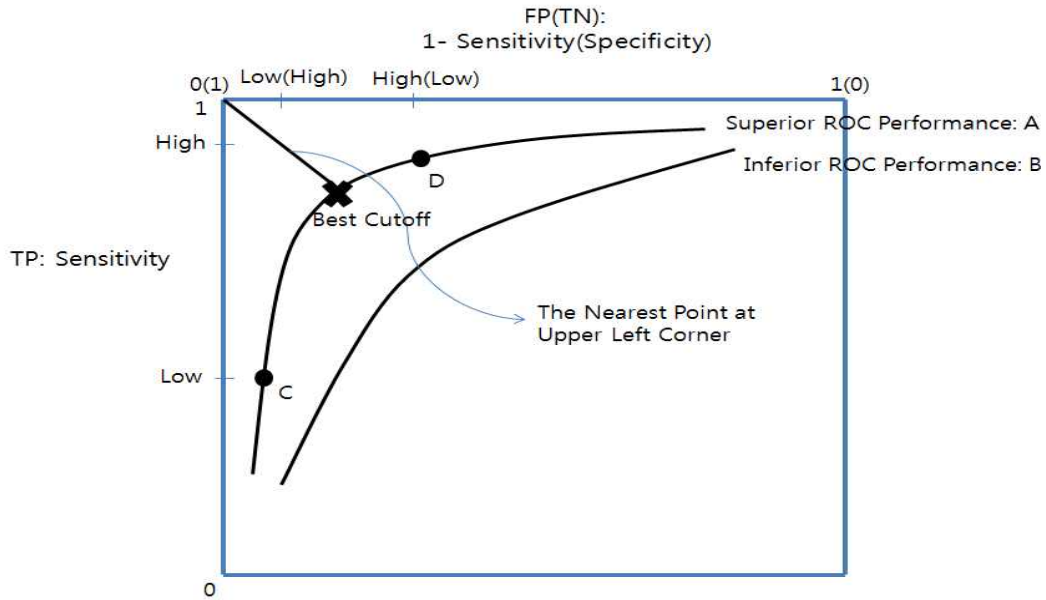
2.2 정규확률분포에 의한 이해

<Figure 1>은 정규확률분포(Normal Probability Distribution)로, 분포(Distribution)는 이론적으로 가정된 수학 모형으로 실제 데이터와 다르지만 다루기가 편해서 <Table 1>과 같은 숫자를 그림으로 표시해주고 효율적인 수학 함수로 사용하기 때문에 반드시 사용 전 정규성 검정을 실시해야 한다. 확률(Probability)은 y축을 나타내는 1.0점 만점의 상대비교 척도로 원래는 Case의 개수이나 기준 개수가 다를 경우 이를 비교하기 위해 전체 개수를 나누어 주는 %와 같은 개념이다.

숫자에 의한 대수적인 표현인 <Table 1>과 식 (1)-(12)을 그림에 의한 기하학적인 표현인 <Figure 1>의 정규확률함수로 이해하면 의사결정지표인 Proportion(전체의 비), Ratio(같은 종류의 서로 다른 비), Rate(다른 종류의 서로 다른 비)의 차이를 쉽게 알 수 있다. 따라서 LR-는 Ratio가 아닌 Rate로 표시해야 한다.



<Figure 1> Normal Probability Distribution



<Figure 2> Properties of ROC

2.3 ROC에 의한 이해

통계학에서 샘플링오차(Sampling Error)인 α , β 를 동시에 줄이려면 샘플의 크기를 크게 해야 하고 이를 제로(Zero)로 만들고 싶으면 전수검사를 시행하면 된다. 그러나 의학통계학에서는 Test의 수가 늘어날 경우 환자, 의사 모두에게 검사비용과 조기치료기회의 상실로 인해 부담이 되므로 TP(True Positive)와 FP(False Positive) 간에 판단의 정확성을 내릴 수 있는 ROC(Receiver Operating Curve) 방법으로 의학진단의 의사결정을 내린다. 의학통계학에서는 진단방법별, 진단병명별 판독자의 양성(Positive)에 대한 주관적인 판정을 내리는 경계값(Subjective Decision Threshold)에 의해 진단결과가 영향을 주므로 ROC에 의한 최적판단점(Best Cutoff)에 의한 객관적인 증거기준임상(Evidence-Based Clinics)이 필요하다.

ROC는 질병의 원인을 파악하는 치료용이 아닌 진단 방법과 진단병명에 대한 객관적인 판독기준을 위한 용도로, Receiver 즉 판독자가, Operating 즉 양성(Positive)을 판독하는 데, Curve 즉 Sensitivity와 Specificity를 시각적으로 절충(Tradeoff)하면서 자신의 경험인 암묵지(Implicit Knowledge)와 연계해서 통계에 의한 형식지(Explicit Knowledge)를 통섭(Consilience)하자는 의미이다. ROC는 <Figure 2>와 같이 y축에 오진에 민감한 중병을 잘 진단하는 Sensitivity와 x축에 진단실수가 없어 불필요한 검사를 하지 않게 하는 옳은 기각(Correct Rejection)을 갖는

전문적인 특이성을 갖는 Specificity(실제는 1-Specificity)에 의한 AUC(Area Under Curve, 판단의 정확성)가 1.0에 가까운 상단좌측곡선(Upper Left Curve)의 최적절단점(Best Cutoff)을 주라는 곡선이다.

<Figure 2>에서 x축은 FP인 1-Specificity 혹은 TN인 Specificity 둘 중 하나로 표현 가능하므로 후자는 괄호로 표시하였다. ROC 중 우측 상단 구석(Upper Left Corner)에 있는 A가 B보다 성능(Performance)이 더 우수한(Superior) 진단방법이며, A 곡선에서 C점은 Low Sensitivity LowFP(or High Specificity), D점은 High Sensitivity High FP(or Low Specificity)으로 두 지점 사이 AUC가 최대가 되는 점이 최적절단점(Best Cutoff)이 된다.

2.4 Bayes 이론에 의한 이해

<Table 1>에서 $P(D+)$, $P(D-)$ 는 사전확률(Prior Probability), $P(T-|D+)$, $P(T+|D+)$, $P(T-|D-)$, $P(T+|D-)$ 는 우도(Likelihood), $P(T- \cap D+)$, $P(T+ \cap D+)$, $P(T- \cap D-)$, $P(T+ \cap D-)$ 는 동시확률(Joint Probability), $P(D+|T-)$, $P(D-|T-)$, $P(D+|T+)$, $P(D-|T+)$ 는 사후확률(Posterior Probability)로 베이시안 예측(Baysian Prediction)에 해당한다.

Diffusion Matrix에 대한 Bayes 공식을 알기 쉽게 그림으로 표시하면 <Figure 3>과 같다.

<Figure 3>에서 주어지는(Given) 사전확률(Prior

Probability)과 우도(Likelihood)로 각 열(Column)의 면적 동시확률(Joint Probability) I, II와 III, IV를 구한 후 예측하려는(Find) 각 행(Row)의 사후확률(Posterior Probability)을 전체면적에 대한 해당 면적의 비로 구한다.

<Figure 3>을 이용한 <Table 1>의 베이지안 예측(Bayesian Prediction)의 사후확률(Posterior Probability)은 식(18)-(21)과 같다.

$$P(D- | T-) = \frac{P(T- \cap D-)}{P(T- \cap D+) + P(T- \cap D-)} = \frac{P(D-)P(T- | D-)}{P(D+)P(T- | D+) + P(D-)P(T- | D-)} \quad (18)$$

$$P(D+ | T-) = \frac{P(T- \cap D+)}{P(T- \cap D+) + P(T- \cap D-)} = \frac{P(D+)P(T- | D+)}{P(D+)P(T- | D+) + P(D-)P(T- | D-)} \quad (19)$$

$$P(D- | T+) = \frac{P(T+ \cap D-)}{P(T+ \cap D+) + P(T+ \cap D-)} = \frac{P(D-)P(T+ | D-)}{P(D+)P(T+ | D+) + P(D-)P(T+ | D-)} \quad (20)$$

$$P(D+ | T+) = \frac{P(T+ \cap D+)}{P(T+ \cap D+) + P(T+ \cap D-)} = \frac{P(D+)P(T+ | D+)}{P(D+)P(T+ | D+) + P(D-)P(T+ | D-)} \quad (21)$$

		D+	D-		
	$T- D+$	I $P(T- \cap D+)$	III $P(T- \cap D-)$		$T- D-$
	$T+ D+$	II $P(T+ \cap D+)$	IV $P(T+ \cap D-)$		$T+ D-$

I + II: Prior Probability = Area = $P(D+)$

I: Joint Probability = Prior Probability \times Likelihood = Column Area \times Proportion of Column Area = $P(T- \cap D+) = P(D+)P(T- | D+)$

II: Joint Probability = $P(T+ \cap D+) = P(D+)P(T+ | D+)$

III+IV: Prior Probability = $P(D-)$

III: Joint Probability = $P(T- \cap D-) = P(D-)P(T- | D-)$

IV: Joint Probability = $P(T+ \cap D-) = P(D-)P(T+ | D-)$

Posterior Probability = $P(D+ | T-) = \frac{I}{I + III} = \frac{\text{Each Area of Row}}{\text{Total Area of Row}}$

$$P(D- | T-) = \frac{III}{I + III}$$

$$P(D+ | T+) = \frac{II}{II + IV}$$

$$P(D- | T+) = \frac{IV}{II + IV}$$

<Figure 3> Understanding of Bayes Formula

2.5 의학적 지표를 이용한 통계기법의 해석

의학적 이론 또는 임상경험에 의한 사고 실험에 대한 연구가설은 영향(Effect), 관계(Relation), 유(Presence), 무(Absent)에 따라 RHPT(Research Hypothesis Present Type)와 RHAT(Research Hypothesis Absent Type)의 둘 중 하나를 설정하고, Popper의 반증원리(Falsification)에 따라 이론을 지지(Support)하는 실제적 유의성(Practical Significance)을 열린 관점으로 계속 업데이트해 나간다.

이러한 연구가설은 통계가설로 번역되어 귀류법(Proof By Contradiction)에 의해 대립가설 H_1 을 증명하기 위해 이에 대한 Negation인 귀무가설 H_0 의 유의성을 검정하게 된다. 통계가설에 대한 검정에서는 2개의 샘플링 오차 α, β 중 H_0 에 대한 False Positive인 α 유의수준(Significance Level)만을 고려한다. 따라서 <Table 2>에서 샘플링검사 판정, 문제해결, 관리도 모니터링 같은 품질통계학과 질병진단과 같은 의료통계학에서는 소비자, 환자에 대한 문제가 더욱 심각하므로 이에 대한 False Negative 오류인 β 가 α 유의성 검정(NHST: Null Hypothesis Significance Test)과 다른 관점에서 고려되어야 한다.

그러나 의료통계학에서는 α, β 를 동시에 고려할 수

있는 ROC나 2결과 같은 다양한 의학 의사결정지표를 활용하여 의사, 환자가 모두 만족할 수 있는 진단시스템을 구축하고 있다.

이와 다르게 품질통계를 적용하는 제조 및 서비스기업에서는 생산자 위험인 α 관점에서 개선이 이루어지고 있어 소비자 위험인 β 관점을 고려하지 않을 경우 비록 6시그마 혁신운동을 추진하더라도 소비자에게는 잠재적인 클레임이 계속 발생한다는 사실을 명심해야 한다.

<Table 2>에서 이론 샘플링검사는 생산자위험 $\alpha = 5\%$ 가 소비자위험 $\beta = 10\%$ 보다 작게 설계되어 있는 이유는 2차 세계대전 때 무기가 갑자기 필요해진 구입자인 미국국방성이 군납을 하려고 하는 민간 생산업자에게 인센티브를 주기 위한 방안이었기 때문이다. 관리도 역시 장기적인 공정모니터링의 용도로 많은 샘플의 크기를 요구하기 때문에 생산자의 부담을 덜어주기 위해 정상을 이상으로 잘못 판단하여 생산성을 저해시키는 생산자위험 α 를 0.27%(3 σ Shewhart 관리도)로 아주 작게 설계하는 방법이기 때문이다. 가설검정의 경우도 소비자의 클레임에 대한 생산기술조건외 확인과 같이 문제해결을 하는 경우 귀류법의 NHST에 의해 생산자위험 α 관점에서만 검정이 실시되므로 $\alpha = 1\%$ 의 결과가 $\alpha = 5\%$ 보다 더 유의적이어서 의미가 있다는 해석은 생산자관점의 일방적인 해석이다.

<Table 2> Meaning of α and β for Statistical Techniques

Use	Object	FP: $\alpha(H_0 \rightarrow H_1)$		FN: $\beta(H_1 \rightarrow H_0)$	
		Meaning	Interest	Meaning	Interest
샘플링검사 판정	Lot	합격 \rightarrow 불합격 Accept \rightarrow Reject $\alpha = 5\%$	공급자 합격로트 폐기	불합격 \rightarrow 합격 $\beta = 10\%$	구입자 클레임*
문제해결 및 개선효과 파악	문제: 부정 개선: 긍정	No Problem \rightarrow Problem 문제 \rightarrow 문제有 No Improvement \rightarrow Improvement 개선 \rightarrow 개선有	생산자 생산성지연 소비자 잠재클레임	문제有 \rightarrow 문제 \rightarrow 개선 \rightarrow 개선無	소비자 잠재클레임 생산자 생산성지연
관리도 문제해결	공정	정상 \rightarrow 이상 Random Cause \rightarrow Assignable Cause	생산자 생산성지연	이상 \rightarrow 정상	소비자 잠재클레임*
질병진단	병	병 \rightarrow 병有 Negative \rightarrow Positive	의사과잉진료	병有 \rightarrow 병無	환자치료시기상실
생동등성 판정	동등성	동등성 \rightarrow 동등성有 Nonequivalence \rightarrow Equivalence ·단측검정보다 양측검정을 실시권고	규제기관 소비자보호실패	동등성 \rightarrow 동등성有 ·규제기관이 허용가능한 큰 α 에서 제약회사가 유리하게 $(1 - \beta)$ 를 크게 유지	제약회사 허가 못 받음
형사재판	죄	죄 \rightarrow 죄有 Innocent \rightarrow Guilty	피고인 무고*	죄有 \rightarrow 죄無	판사 범인을 놓치는 것
성적평가	답	정답 \rightarrow 오답 Right Answer \rightarrow Wrong Answer	학생 성적감점오류*	오답 \rightarrow 정답	선생님 채점가점 실수
신용평가	부도	정상 \rightarrow 부도 Nondefault \rightarrow Default	정상차주기업 부도평가오류*	부도 \rightarrow 정상	대출기관 회수불능

따라서 로트의 합부관정을 위한 샘플링검사, 공정의 장기적인 개선을 위한 관리도, 고객클레임의 원인파악을 위한 가설검정 등 품질기법에서는 생산자를 보호하려는 α 관점에서 설계되었다는 점을 명심하고 소비자 관점에서의 β 를 고려하는 통계적인 판단이 병용되어야 한다.

따라서 품질기법에서도 α , β 를 동시에 고려하는 ROC 등의 의학적 의사결정지표를 적극 도입해서 활용해야 관정의 신뢰도를 높일 수 있다.

3. 품질통계에서 ROC 적용방안

계량규준형 샘플링검사, 계량형 평균검정, 계량형 $\bar{x}-R$, $\bar{x}-s$ 관리도에서는 제품스펙(Specification)에 따라 판정방법이 달라진다. 계량형 스펙에서는 순도와 같이 클수록 좋은 경우(망대특성)의 상한규격(USL: Upper Specification Limit), 불순물의 함유량과 같이 작을수록 좋은 경우(망소특성)의 하한규격(LSL: Lower Specification Limit, 조립치수처럼 기준치±허용차의 양쪽규격(LSL~USL) 등 3종류가 있다.

계량규준형 샘플링검사에서는 합격판정치(Acceptable Limit), 계량형 평균검정에서는 기각치(Critical Value), 계량형 \bar{x} 관리도에서는 관리한계(Control Limit)가 ROC의 절단점(Cutoff)이 된다.

<Figure 4>의 (a)에서 상한규격인 경우 좋은 로트의 평균값을 m_N , 나쁜 로트의 평균값을 m_P 라고 할 경우 ROC의 TP와 FP를 구하는 단계는 다음과 같다.

단계 1: 주어진 샘플링 계약조건 공급자 (m_0, α) , 구입자 (m_1, β) , \bar{x}_u 에 대해 $n = ((Z_\alpha + Z_\beta) / (m_P - m_N))^2 \sigma^2$, $\bar{x}_u = ((m_N Z_\beta + m_P Z_\alpha) / (Z_\alpha + Z_\beta))$ 를 구한다.

단계 2: $m_P = \bar{x}_u + Z_\beta \frac{\sigma}{\sqrt{n}}$ 에서 $Z_\beta = \sqrt{n}(\bar{x}_u - m_P) / \sigma$ 를 유도하고 표준정규분포표를 이용하여 β 를 구한 후 $TP = 1 - \beta$ 로 놓는다.

단계 3: $m_N = \bar{x}_u - Z_\alpha \frac{\sigma}{\sqrt{n}}$ 에서 $Z_\alpha = \sqrt{n}(m_N - \bar{x}_u) / \sigma$ 를 유도하고 표준정규분포표를 이용하여 $FP = \alpha$ 를 구한다.

단계 4: TP와 FP를 이용하여 ROC Curve를 작성한다.

<Figure 4>의 (b)에서 하한규격인 경우 ROC의 TP와 FP를 구하는 단계는 다음과 같다.

단계 1: 주어진 샘플링 계약조건 공급자 (m_0, α) , 구입자 (m_1, β) , \bar{x}_L 에 대해 $n = ((Z_\alpha + Z_\beta) / (m_P - m_N))^2 \sigma^2$, $\bar{x}_L = ((m_N Z_\beta + m_P Z_\alpha) / (Z_\alpha + Z_\beta))$ 를 구한다.

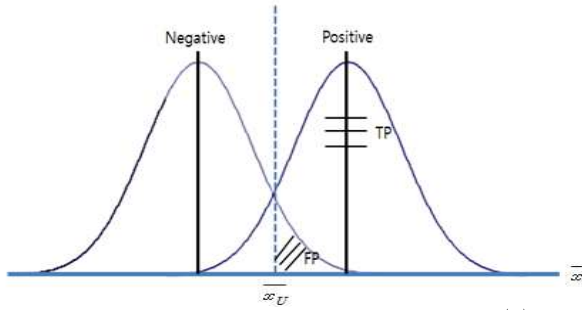
단계 2: $m_P = \bar{x}_L - Z_\beta \frac{\sigma}{\sqrt{n}}$ 에서 $Z_\beta = \sqrt{n}(\bar{x}_L - m_P) / \sigma$ 를 유도하고 표준정규분포표를 이용하여 β 를 구한 후 $TP = 1 - \beta$ 로 놓는다.

단계 3: $m_N = \bar{x}_L + Z_\alpha \frac{\sigma}{\sqrt{n}}$ 에서 $Z_\alpha = \sqrt{n}(m_N - \bar{x}_L) / \sigma$ 를 유도하고 표준정규분포표를 이용하여 $FP = \alpha$ 를 구한다.

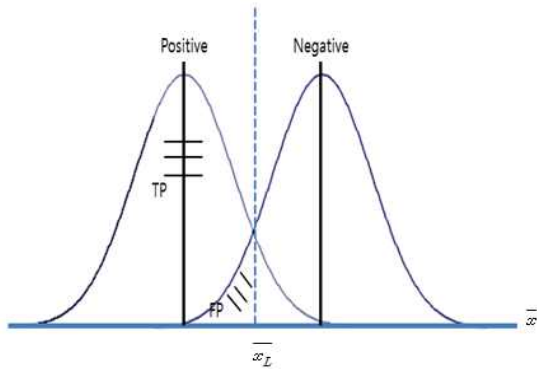
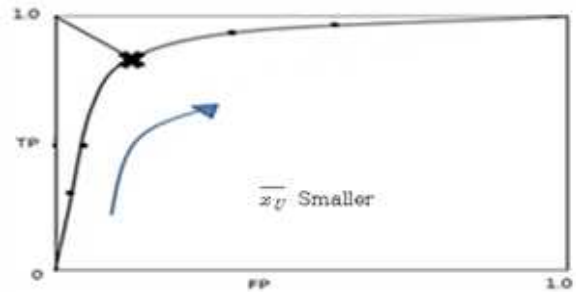
단계 4: TP와 FP를 이용하여 ROC Curve를 작성한다.

<Figure 4>에서 상한규격 USL인 상한합격판정치, 상한기각치, 관리상한인 \bar{x}_U 의 절단점이 작아짐에 따라 Sensitivity TP는 커지면 1-Specificity FP도 커진다. 하한규격인 LSL인 경우 하한합격판정치, 하한기각치, 관리하한인 \bar{x}_L 의 절단점이 커짐에 따라 Sensitivity TP는 커지고 Specificity TN은 작아진다. 양쪽규격인 경우 상하한 합격판정치, 상하한 기각치, 관리상하한인 \bar{x}_U 와 \bar{x}_L 은 좌우대칭(Symmetric)이므로 $|\bar{x}|$ 의 절단점이 작아짐에 따라 Sensitivity TP는 커지고 1-Specificity FP도 커진다. ROC의 최적절단점(Best Cutoff)은 <Figure 4>에서와 같이 상단 좌측 모서리(Upper Left Corner)에서 가장 가까운 X표시의 점이 된다. 이 절단점에서 불합격로트를 불합격로트로(샘플링검사), 대립가설을 대립가설로(가설검정), 이상원인의 공정을 이상원인의 공정(관리도)으로 옳게 판정하는 Sensitivity $(1 - \beta)$ 를 크게 하고 합격로트를 합격로트로(샘플링검사), 귀무가설을 귀무가설로(가설검정), 정상원인의 공정을 정상원인의 공정(관리도)으로 옳게 판정하는 Specificity $(1 - \alpha)$ 도 가장 크게 되는 최적 판단기준이 된다.

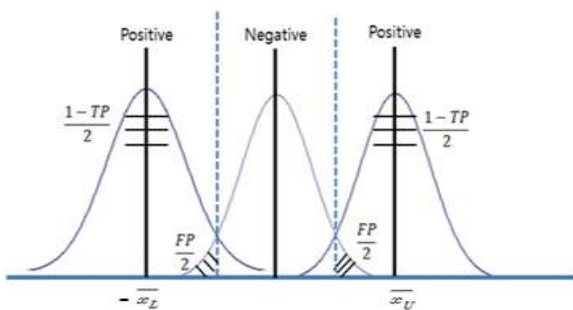
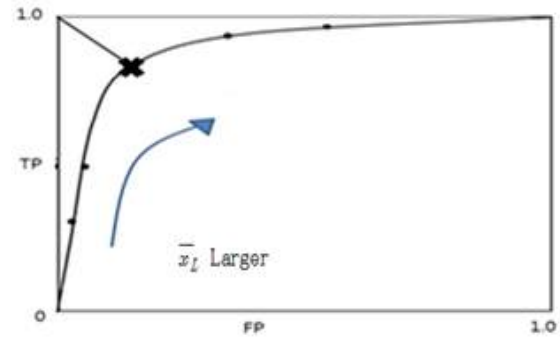
예를 들어 공급자·구입자의 품질협력관계를 유지하기 위해 샘플링검사의 계약으로 납품을 받는 공급업자에 대한 평가등급이 실제 로트를 납품받았을 경우 로트의 합격과 불합격으로 판정한 결과가 <Table 3>과 같다.



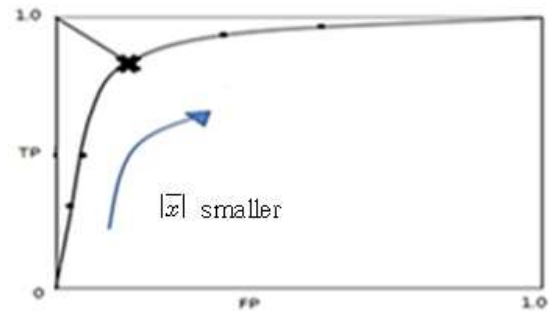
(a) ROC Under USL



(b) ROC Under LSL



(c) ROC Bilateral Specification



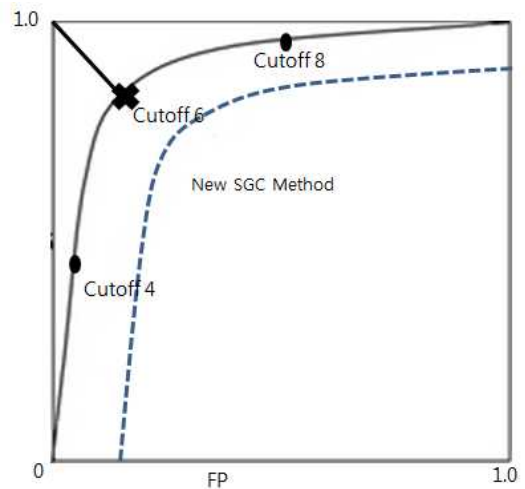
<Figure 4> ROC Under Bilateral Specification

<Table 3> Discrimination Number by Supplier Grade Scoring

Scoring	Lot	
	Accept	Reject
Less than 3rd Grade	19	2
4th~5th Grade	9	18
6th~7th Grade	6	38
More than 8th Grade	4	41
Total	38	99

<Table 4> Cutoff by TP and FP

Cutoff Values	TP	FP
4	0.50	0.02
6	0.74	0.20
8	0.89	0.59



<Figure 5> ROC Curve

<Table 3>에서 공급자 평가등급(SGC: Supplier Grade Scoring)의 변별력(Discrimination Power)을 측정하기 위해 <Table 4>와 같이 TP(True Positive, Hit Rate)와 FP(False Positive)를 계산하여 ROC(Receiver Operating Characteristics)를 작성하면 <Figure 5>와 같다. <Table 4>에서 Cutoff Value 4인 경우 TP=0.50은 <Table 3>에서 3등급 이하의 합격된 로트 수 19를 전체 합격로트 38로 나눈 값이고 FP=0.02는 불합격된 로트 수 2를 전체 불합격로트 99로 나눈 값이다. 여기서 Cutoff 4의 의미는 4등급 이상인 공급업자는 납품을 받지 않고 3등급에만 납품을 받을 경우의 평가등급이다.

<Figure 5>에서 상단좌측 모서리점에서 가장 가까운 Cutoff 6인 경우 즉 5등급미만만 납품을 받았을 경우 검사로트의 판정에 대한 변별력을 가장 크게 하는 최적 절단점(Best Cutoff Value)이 된다. 따라서 협력업체 평가자는 6등급이상의 협력업체 평가기준에 대한 변별력을 위한 재검토가 요구된다. <Figure 5>에서 새로운 협력업체 평가(SGC)방법에 대한 ROC곡선이 점선과 같이 이루어질 경우 기존의 실선의 ROC곡선이 상단좌측 모서리에 위치에 있어 실제 불합격로트를 불합격으로 제대로 판정할 수 있는 TP와 실제 합격로트를 합격로트로 제대로 판정할 수 있는 TN이 최대가 된다. ROC곡선 변별력에 대한 우수성 판정기준은 AUC(Area Under Curve)의 값이 각각 0.91~1, 0.81~0.9, 0.71~0.8, 0.61~0.7, 0.5~0.6일 경우 Excellent, Good, Fair, Poor, Fail로 판정된다.

4. 결론

본 연구에서는 의학적 의사결정지표를 유형화하고 α 와 β 관점에서 해석하여 품질개선기법의 용도별 적용방안을 다음과 같이 제시하였다.

1) 샘플링검사, 관리도, 품질가설검정등의 품질기법에서 민감도, 특이도, 양성, 음성예측도, 상대, 절대 리스크 감소, 양성, 음성가능도비, 유병률, 정확도 등의 의학적 의사결정지표를 α 와 β 의 해석을 통한 적용방안을 제시하였다.

2) 계량형 샘플링검사에서 ROC 적용시 상한규격인 경우 절단점인 판정기준에 작아질수록 하한규격은 커질수록 민감도와 특이도가 커지고 양쪽규격인 경우 절단값의 판정기준이 작아질수록 민감도와 특이도가 커지며 최적 절단점은 상단좌측 구석점에서 가장 가까운 점에서 결정된다.

3) 협력업체품질등급평가에서 ROC 적용시 절단점인 평가등급이 낮아짐에 따라 민감도와 특이도가 커지며,

평가방법별 우수성은 AUC(Area Under ROC Curve)가 가장 큰 방법으로 선택된다.

5. 참고 문헌

- [1] Campbell M.J., Machin D., Walters S.J.(2007), Medical Statistics : A Textbook for the Health Sciences, 4 Edition, Wiley.
- [2] Felder S., Mayrhofer T.(2011), Medical Decision Making : A Health Economic Primer, Springer.
- [3] Hauschke D., Steinijans V., Pigeot I.(2007), Bioequivalence Studies in Drug Development : Methods and Applications, Wiley.
- [4] Kanane J.B.(2008), Statistics in the Law : A Practitioner's Guide, Cases, and Materials, Oxford University Press.
- [5] Lee B.(2006), "Medical Decision Making Tools: Bayesian Analysis and ROC Analysis", Imaging Science in Dentistry 36, 1-5.
- [6] Levinson W.A.(2010), Statistical Process Control for Real-World Applications, CRC Press.
- [7] Montgomery D.C.(2012), Statistical Quality Control, 7 Edition, Wiley.
- [8] Mukherjee A.(2010), Credit Scoring Model Using Data Mining Techniques : A Pragmatic Approach, VDM Verlag.
- [9] Patterson S., Jones B.(2005), Bioequivalence and Statistical in Clinical Pharmacology, Chapman and Hall/CRC.
- [10] Peat J., Barton B.(2005), Medical Statistics : A Guide to Data Analysis and Critical Appraisal, BMJ Books.
- [11] Ryan T.P.(2011), Statistical Methods for Quality Improvement, 3 Edition, 2011.
- [12] Shmueli G.(2011), Practical Acceptance Sampling : A Hands-On Guide, 2 Edition, Createspace Independent Publishing Platform.
- [13] Sox H., et al.(2006), Medical Decision Making, The American College of Physicians.
- [14] Stephens K.S.(2001), The Handbook of Applied Acceptance Sampling : Plans, Procedures & Principles, American Society for Quality.
- [15] Thomas L.C., Edelman D.B, Crook J.N.(2002), Credit Scoring & Its Applications, Society for Industrial and Applied Mathematics.

- [16] Walker J.T., Maddan S.(2008), Statistics in Criminology and Criminal Justice : Analysis and Interpretation, Jones & Bartlett Learning.
- [17] Wheele D.J.(2010), Understanding Statistical Process Control, 3 Edition, SPC Press.

저자 소개

최성운



현 가천대학교 산업공학과 교수.
한양대학교 산업공학과에서 공학사, 공학석사, 공학박사 학위를 취득하고, 1994년 한국과학재단 지원으로 University of Minnesota 에서 1년간 Post-Doc 을 수행했으며, 2002년부터 1년반 동안 University of Washington

에서 Visiting Professor를 역임하였음. 주요 관심분야는 자동화 생산 및 장치 산업에서의 품질관리이며, 통신, 정보시스템의 보안, 신뢰성 설계 및 분석, 서비스 사이언스, RFID시스템, Wavelet에도 관심을 가지고 있음.

주소: 경기도 성남시 수정구 성남대로 1342 가천대학교 산업공학과 ☎(031)750-5366,
swchoi@gachon.ac.kr