

Quantile-based Nonparametric Test for Comparing Two Diagnostic Tests

Young-Min Kim¹⁾ and Hae-Hiang Song²⁾

Abstract

Diagnostic test results, which are approximately normal with a few number of outliers, but non-normal probability distribution, are frequently observed in practice. In the evaluation of two diagnostic tests, Greenhouse and Mantel (1950) proposed a parametric test under the assumption of normality but this test is inappropriate for the above non-normal case. In this paper, we propose a computationally simple nonparametric test that is based on quantile estimators of mean and standard deviation, instead of the moment-based mean and standard deviation as in some parametric tests. Parametric and nonparametric tests are compared with simulations under the assumption of, respectively, normality and non-normality, and under various combinations of the probability distributions for the normal and diseased groups.

Keywords: Diagnostic tests; nonparametric test; quantile estimator.

1. 서론

의학기술의 발전과 함께 각 분야에서 지속적으로 새로이 개발되는 진단검사의 평가는 질병의 유무가 확인된 질병군 (diseased group)과 정상군 (non-diseased group)에 속한 개체에게 기준의 진단검사와 새로이 개발된 진단검사를 모두 적용하여 수집된 자료로부터 알아낸다. 진단검사를 비교하는 목적으로 제시된 임상논문을 살펴보면 더 나은 진단검사를 결정함에 있어서 여러 가지 다른 평가척도 또는 검정통계량에 근거하고 있다. 그러나 진단검사의 여러 평가척도 중에서도 가장 빈번히 사용되는 척도는 민감도 (sensitivity)와 특이도 (specificity)라 할 수 있다. 민감도는 질병군에서 진단검사 결과가 병이 있는 것으로 진단 내려질 확률이며, 특이도는 정상군에서 진단검사 결과가 정상으로 진단 내려질 확률이다.

1) Graduate Student, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea.
E-mail : white-08@hanmail.net

2) Professor, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea.
Correspondence : hhsong@catholic.ac.kr

현실적으로 특이도와 민감도는 진단검사의 결과가 연속형이라 하더라도 이산형 자료로 변환하여 계산하며, 특이도와 민감도 각각의 비교는 McNemar 검정에 근거한다 (Bennett, 1972). 진단검사의 결과가 연속형일 때에도 McNemar 검정법보다도 연속형 자료에 근거한 검정법이 더욱 바람직하다. 일반적으로 임상의는 민감도에 더 큰 관심이 있으며, 이러한 측면에서 Greenhouse와 Mantel (1950)은 진단검사의 결과가 연속형일 때 높은 특이도를 전제로 한 민감도를 비교하는 검정통계량을 제시하였다. 따라서 두 진단검사를 모두 정해진 특이도, 일반적으로 90% 또는 95%에 해당하는 특이도 하에서 민감도를 비교하여 더 나은 진단검사를 알아낸다. Greenhouse와 Mantel (1950)에서와 같이 일정 특이도에 해당하는 질병군에서의 표준정규변량 (standard normal deviate)을 비교할 수도 있고, 민감도를 직접 추정하여 비교할 수도 있다. 민감도를 추정하는 과정은 정규변량의 추정에서 한 단계 더 나아간 과정이므로 검정통계량이 한층 복잡하다 (Wieand 등, 1989). 반면에 정규변량의 추정은 비교적 단순하기 때문에 정해진 높은 특이도를 전제로 한 민감도에 해당하는 정규변량의 비교가 더욱 장점으로 대두된다.

Greenhouse와 Mantel (1950)의 정규변량을 이용한 검정법은 진단검사 결과가 정규분포한다는 가정하에서의 제시된 모수적 검정법이다. 본 논문에서는 정규분포 가정이 성립하지 않는 경우에 적용할 수 있는 비모수적인 정규변량 통계량을 제안한다. 이는 구체적으로 분위수 (quantile)에 근거한 평균과 표준편차를 이용한 정규변량의 비모수 추정량이며, 이를 비교하는 검정통계량을 제안한다. 두 진단검사의 결과가 정상군과 질병군에서 서로 매우 다양한 분포 형태로 맞물려 있는 여러 경우에 대해 모의실험으로 모수적 검정법과 비모수적 검정법을 비교한다.

2. 방법

두 진단검사를 비교하는 검정법들을 설명하기 위해 우선 정상군을 X 군이라 하고, 질병군을 Y 군이라 하자. X 군의 개체는 m 명, Y 군의 개체는 n 명으로 총 $N (= m + n)$ 명에게 두 가지 다른 진단검사를 실시한다. 진단검사 1로부터 X_1 과 Y_1 을, 진단검사 2로부터 X_2 와 Y_2 를 수집한다.

정상군에 속한 각 개체의 진단검사 자료는 다음과 같다.

$$\mathbf{X}_i = (X_{1i}, X_{2i})', \quad i = 1, 2, \dots, m.$$

여기서 X_{ti} 는 i 번째 개체에게 실시한 진단검사 t ($t = 1, 2$)로부터 얻어진 측정값이다. 질병군에 속한 각 개체의 진단검사 자료는 다음과 같다.

$$\mathbf{Y}_j = (Y_{1j}, Y_{2j})', \quad j = 1, 2, \dots, n.$$

여기서 Y_{tj} 는 j 번째 개체에게 실시한 진단검사 t ($t = 1, 2$)로부터 얻어진 측정값이다. 각 군의 개체들은 서로 독립이라고 가정하고, 진단검사의 결과가 연속형으로 측정될 때, 질병군이 당연히 진단검사에서 질병이 있는 것으로 판단될 가능성이 높아야 하므로 정상군보다 높은 수치를 가진다고 가정한다.

2.1. 모수적 검정법

Greenhouse와 Mantel (1950)의 모수적 검정법은 두 연속형 진단검사의 검정법으로 진단검사 결과가 정상군과 질병군의 두 군 모두에서 정규분포하거나, 적절한 변환으로 정규분포하는 것을 가정한다. 그러나 등분산을 가정하지는 않는다. 식 (2.1)과 식 (2.2)의 정상군과 질병군 각각 연속형 자료의 분포는 구체적으로 다음과 같다.

$$\mathbf{X}_i = \begin{pmatrix} X_{1i} \\ X_{2i} \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \mu_{X_1} \\ \mu_{X_2} \end{pmatrix}, \begin{pmatrix} \sigma_{X_1}^2 & \rho_X \sigma_{X_1} \sigma_{X_2} \\ \rho_X \sigma_{X_1} \sigma_{X_2} & \sigma_{X_2}^2 \end{pmatrix} \right], \quad i = 1, 2, \dots, m, \quad (2.1)$$

$$\mathbf{Y}_j = \begin{pmatrix} Y_{1j} \\ Y_{2j} \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{pmatrix}, \begin{pmatrix} \sigma_{Y_1}^2 & \rho_Y \sigma_{Y_1} \sigma_{Y_2} \\ \rho_Y \sigma_{Y_1} \sigma_{Y_2} & \sigma_{Y_2}^2 \end{pmatrix} \right], \quad j = 1, 2, \dots, n. \quad (2.2)$$

Greenhouse와 Mantel (1950)의 모수적 검정법은 다음과 같은 절차로 시행한다. 우선 두 진단검사의 비교에서 특이도를 최소 p_0 이상으로, 다시 말하면 최대로 허용할 수 있는 위양성을 (false positive rate) 수준인 $1 - p_0$ 을 결정한다. 다음으로 정해진 $1 - p_0$ 에 해당하는 정상군의 임계점 (critical point) ξ_{p_0} 을 각 진단검사의 정상군의 결과로부터 추정한다. 정상군의 진단검사 결과가 정규분포 $N(\mu_X, \sigma_X^2)$ 에 따른다는 가정 하에서 참 (true) 임계점은 $\xi_{p_0} = \mu_X + z_{p_0} \sigma_X$ 로 표현되며, 이때 z_{p_0} 은 표준정규분포에서의 p_0 분위수이다. 이제 질병군의 정규분포 $N(\mu_Y, \sigma_Y^2)$ 에 따른 진단검사 결과에서 임계점 이상인 확률이 바로 특이도 p_0 에 해당하는 민감도로 다음과 같이 표현된다. 편의상 첨자를 생략하고 한 군의 경우로 설명하며, 민감도를 간략히 Sens로 표시한다.

$$\begin{aligned} \text{Sens} &= \Pr [Y > \xi_{p_0}] \\ &= \Pr \left[\frac{Y - \mu_Y}{\sigma_Y} > \frac{\xi_{p_0} - \mu_Y}{\sigma_Y} \right] \\ &= \Pr \left[Z < \frac{\mu_Y - \mu_X - z_{p_0} \sigma_X}{\sigma_Y} \right]. \end{aligned} \quad (2.3)$$

Greenhouse와 Mantel (1950)은 두 진단검사의 평가에 있어 위와 같은 민감도를 비교하지 않고 민감도에 해당하는 질병군의 정규분포 하에서의 분위수, 다시 말하면 식 (2.3)의 마지막 항인 표준정규변량인 $G \equiv (\mu_Y - \mu_X - z_{p_0} \sigma_X) / \sigma_Y$ 를 추정하여 비교한다. 민감도가 적어도 50% 이상인 경우에 G 는 양수 값을 가지며, 진단검사 1과 2의 비교에서 G_1 과 G_2 중 더욱 큰 값을 가지는 진단검사가 좋은 진단검사라 할 수 있다.

이제 $G_1 - G_2$ 를 살펴보면,

$$G_1 - G_2 = \frac{\mu_{Y_1} - \mu_{X_1}}{\sigma_{Y_1}} - \frac{\mu_{Y_2} - \mu_{X_2}}{\sigma_{Y_2}} - z_{p_0} \left(\frac{\sigma_{X_1}}{\sigma_{Y_1}} - \frac{\sigma_{X_2}}{\sigma_{Y_2}} \right) \quad (2.4)$$

가 되며, 일반적으로 마지막 항의 z_{p_0} 으로 인해 어떤 특이도에 해당하는 민감도를 비교하는가에 따라서 검정의 결과가 달라진다. 그러나 $\sigma_{X_1}/\sigma_{Y_1} = \sigma_{X_2}/\sigma_{Y_2}$ 의 조건하에서는 z_{p_0} 에 대한 항이 사라져 두 진단검사를 비교하는 검정의 결과는 위양성 수준과 무관하게 된다. 즉, 두 진단검사를 90%, 95% 또는 다른 특이도 하에서 G_1 과 G_2 를 비교하여도 두 진단검사의 우위에 있어 변함이 없다.

이제 하나의 진단검사에 대한 G 의 분산을 설명한다. G 의 표본추정량은 X 군과 Y 군에 해당하는 모평균 μ_X, μ_Y 와 모표준편차 σ_X, σ_Y 를 각각 표본추정량인 \bar{X} 와 \bar{Y} , s_X 와 s_Y 를 대입하여

$$\hat{G} = \frac{1}{s_Y}(\bar{Y} - \bar{X} - z_{p_0}s_X) \quad (2.5)$$

가 된다. Greenhouse와 Mantel (1950)은 \hat{G} 통계량의 분산을 다음과 같이 제시하였다.

$$\hat{V}(\hat{G}) = \frac{\hat{G}^2}{2n} + \frac{1}{n} + \frac{s_X^2}{ms_Y^2} + \frac{(z_{p_0}s_X)^2}{2ms_Y^2}. \quad (2.6)$$

각 개체들에게 두 진단검사 모두를 실시하여 분산 $V(\hat{G}_1 - \hat{G}_2) = V(\hat{G}_1) + V(\hat{G}_2) - 2\text{Cov}(\hat{G}_1, \hat{G}_2)$ 는 식 (2.6)과 Greenhouse와 Mantel (1950)이 제시한 공분산

$$\text{Cov}(\hat{G}_1, \hat{G}_2) \cong \frac{\rho_X^2 G_1 G_2}{2n} + \frac{\rho_Y}{n} + \frac{\rho_X \sigma_{X_1} \sigma_{X_2}}{m \sigma_{Y_1} \sigma_{Y_2}} + \frac{(z_{p_0} \rho_X)^2 \sigma_{X_1} \sigma_{X_2}}{2m \sigma_{Y_1} \sigma_{Y_2}} \quad (2.7)$$

를 이용하여 구한다. 여기서 ρ_X 와 ρ_Y 는 각각 X 군과 Y 군에 해당하는 두 진단검사에 의한 측정값 사이의 모상관계수이며, 표본추정량으로서 Pearson의 적률 상관계수 (product-moment correlation coefficient)인 r_X 와 r_Y 를 사용한다.

결과적으로 두 진단검사의 비교는 표본의 크기가 충분히 크다면 $Z = (\hat{G}_1 - \hat{G}_2)/\sqrt{\hat{V}(\hat{G}_1 - \hat{G}_2)}$ 가 점근적으로 (asymptotically) $N(0, 1)$ 임을 이용하여 검정한다.

2.2. 비모수적 검정법

임상의 현장에서는 진단검사의 결과가 정규분포하지 않거나, 분포를 알 수 없거나, 또는 매우 크거나 매우 작은 이상치 (outlier)가 포함된 경우가 종종 발견된다. 따라서 정규분포를 가정하지 않는 진단검사의 비교에 사용될 수 있는 비모수적 검정법이 요구되며 이를 본 논문에서 제안하고자 한다.

Greenhouse와 Mantel (1950)은 정규분포 가정하에서 표준정규변량의 차를 비교하는 통계량 $\hat{G}_1 - \hat{G}_2$ 를 제안하면서 또한 정규분포 가정이 요구되지 않는 경우의 비모수 검정법을 제시하였다. 그러나 Greenhouse와 Mantel (1950)의 비모수적 검정법은 정규변량에 근거한 검정법이 아니고 정해진 특이도에 해당하는 민감도를 추정하여 비교한다. 반면에 본 논문에서 제안하는 비모수적 검정법은 Greenhouse와 Mantel (1950)의 모수적 검정법에서 언급된 표준정규변량 G 에 포함된 평균과 표준편차의 추정량으로서 분위수를 이용한 비모수적 위치와 분산의 추정량으로 대치하여 비모수적 추정량인 $\tilde{G}_1 - \tilde{G}_2$ 이다.

우선 비모수적 추정량 $\tilde{G}_1 - \tilde{G}_2$ 에서 사용하게 되는, 모집단 분포가 정규분포가 아닌 다른 분포라 하더라도 크게 어긋나지 않는 위치와 분산의 로버스트한 (robust) 추정량인 분위수 추정량 (quantile estimator)은 여러 장점을 가졌다 (Benson, 1949; Brown, 1981). Cramer (1946, p. 181)는 분위수 추정량을 다음과 같이 정의한다. X 군의 누적 확률 분포함수 (cumulative distribution function)를 $F(x)$ 라 하고 p 가 0과 1 사이의 임의

의 값일 때, p 분위수는 $F(\zeta_p) = p$ 에서 ζ_p 이다. 앞 절에서 정규분포에서의 p 분위수를 기호 ξ_p 로 표기하였으나, 일반적인 분포에서의 p 분위수는 ζ_p 로 구분하여 표기한다. 예를 들어서 X 군의 자료에서 만약 mp 가 정수가 아니라면, p 분위수의 추정량은 표본 값들을 순서대로 나열한 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(m)}$ 에서 $x_{([mp])}$ 를 선택한다. $[mp]$ 는 mp 값을 초과하지 않는 최대의 정수이다. 만약 mp 가 정수라면, p 분위수의 추정량은 $x_{(mp)}$ 와 $x_{(mp+1)}$ 의 사이의 값을 가진다. 본 논문에서는 $x_{(mp)}$ 와 $x_{(mp+1)}$ 의 사이의 값들 중에서 $(x_{(mp)} + x_{(mp+1)})/2$ 을 사용하였다 (Benson, 1949). Y 군의 누적확률 분포함수를 $H(y)$ 라고 할 때, 분위수 추정량도 np 가 정수인가 아닌가에 따라 X 군과 같은 방법으로 추정한다.

대칭인 임의의 분포의 평균은 μ 이고 표준편차는 σ 일 때, 평균에서 분위수들의 거리를 $k\sigma$ 라고 하면, $0 < p < 1 - p < 1$ 의 조건하에서 $\zeta_p = \mu - k\sigma$, $\zeta_{1-p} = \mu + k\sigma$ 이 되므로 다음이 성립한다.

$$\zeta_{1-p} + \zeta_p = 2\mu, \quad \zeta_{1-p} - \zeta_p = 2k\sigma.$$

위의 관계식에서 표본수가 커지면 임의의 분포는 정규분포에 가까워지므로 k 는 표준정규 분포표에서 p 에 대응하는 값을 찾을 수 있다 (Benson, 1949). 이제 Y 군 자료에서 p 분위수인 ζ_p 의 추정량을 $\tilde{\zeta}_p$ 라 할 때, 평균 분위수 추정량은

$$\tilde{\mu}_Y = \frac{(\tilde{\zeta}_{1-p_m} + \tilde{\zeta}_{p_m})}{2}$$

이 되며, 표준편차 분위수 추정량은

$$\tilde{\sigma}_Y = \frac{(\tilde{\zeta}_{1-p_s} - \tilde{\zeta}_{p_s})}{2k}$$

가 된다. 여기서 평균 분위수 추정량에서의 p 는 p_m 으로, 표준편차 분위수 추정량에서 는 p_s 로 구분하여 표시한다. 평균과 표준편차의 추정량에 사용되는 p_m 과 p_s 는 Pearson (1920)에 의하면 여러 표본 추정량들과 비교했을 때 분위수 추정량의 분산이 최소가 되는 p 값이 가장 바람직하며, 평균과 표준편차 각각의 경우에 그 값은 p_m 은 0.2703이고, p_s 는 0.0692로 알려져 있다 (Benson, 1949). 한편 k 는 표준정규 분포표에서 $1 - p_s$ 에 대응되는 값으로 k 는 1.4818이다.

분위수를 이용한 위와 같은 평균과 표준편차 추정량은 비교적 계산이 간편하고 이 상치에 큰 영향을 받지 않는 통계량이다. 이를 이용하여 본 논문에서 다음과 같이 비모수적 G 의 추정량을 제안한다.

$$\tilde{G} = \frac{1}{\tilde{\sigma}_Y} (\tilde{\mu}_Y - \tilde{\zeta}_{p_0}). \quad (2.8)$$

우선 식 (2.8)에 포함되어 있는 p_0 은 미리 정해진 특이도이며, 모수적 검정법과 비모수적 검정법에서 동일 값으로 정한다. 진단검사의 결과가 정규분포하지 않더라도 대칭에 가까운 분포라면, $\tilde{G}_1 - \tilde{G}_2$ 를 이용한 검정이 앞 절에 제시된 모수적 검정통계량 $\hat{G}_1 - \hat{G}_2$ 에 근거한 검정보다 더 좋은 결과가 나올 것으로 기대된다.

\tilde{G} 의 근사분산의 추정식은

$$\hat{V}(\tilde{G}) = \frac{p_s(1-2p_s)\hat{\sigma}_Y^2\tilde{G}^2}{2n(kg_s\tilde{\sigma}_Y)^2} + \frac{1}{\tilde{\sigma}_Y^2} \left[\frac{p_m\hat{\sigma}_Y^2}{2ng_m^2} + \frac{p_0(1-p_0)\hat{\sigma}_X^2}{mg_{p_0}^2} \right] \quad (2.9)$$

이다. 이 분산 공식의 유도과정이 부록에 제시되었다. 여기서 g_m 과 g_s 는 각각 p_m 과 p_s 에 대응되는 표준정규분포의 확률밀도함수의 값이다. 또한 모분산 σ_X^2 와 σ_Y^2 의 추정량으로 표본분산 s_X^2 와 s_Y^2 를 사용한다. 분위수 추정량에서 p_m 과 p_s 가 각각 0.2703과 0.0692일 때 최적이 되므로, 이에 해당하는 확률밀도함수의 값 g_m 과 g_s 는 각각 0.3308과 0.1331이 된다.

두 진단검사를 각 개체에게 모두 적용하여 두 진단검사를 비교할 때, $\tilde{G}_1 - \tilde{G}_2$ 의 분산은 다음과 같이 구한다. 이제 $V(\tilde{G}_1 - \tilde{G}_2) = V(\tilde{G}_1) + V(\tilde{G}_2) - 2\text{Cov}(\tilde{G}_1, \tilde{G}_2)$ 에서 $\text{Cov}(\tilde{G}_1, \tilde{G}_2)$ 의 근사 추정식은

$$\hat{\text{Cov}}(\tilde{G}_1, \tilde{G}_2) = \frac{r_Y^2 \hat{\sigma}_{Y_1} \hat{\sigma}_{Y_2} \tilde{G}_1 \tilde{G}_2}{2(n-1)\tilde{\sigma}_{Y_1} \tilde{\sigma}_{Y_2}} + \frac{r_Y \hat{\sigma}_{Y_1} \hat{\sigma}_{Y_2}}{n \tilde{\sigma}_{Y_1} \tilde{\sigma}_{Y_2}} + \frac{r_S(x) p_0 (1-p_0) \hat{\sigma}_{X_1} \hat{\sigma}_{X_2}}{mg_{p_0}^2 \tilde{\sigma}_{Y_1} \tilde{\sigma}_{Y_2}} \quad (2.10)$$

이며 이 근사 추정식의 유도과정이 부록에 제시되었다. 여기서 $r_s(x)$ 은 X 의 Spearman의 순위상관계수이다.

결과적으로 동일 개체에게 적용한 두 진단검사의 비교는 검정통계량 $Z_{\text{Non}} = (\tilde{G}_1 - \tilde{G}_2)/\sqrt{\hat{V}(\tilde{G}_1 - \tilde{G}_2)}$ 의 정확한 분포는 알 수 없으나 표본의 크기가 충분히 크다면 중심극한 정리에 의해서 Z_{Non} 이 점근적으로(asymptotically) $N(0, 1)$ 을 따름을 이용하여 검정한다.

3. 사례와 모의실험

3.1. 적용 사례

모수적인 검정법과 함께 본 논문에서 제안한 비모수 검정법을 이제 Obuchowski 등 (2004)의 방사선분야 자료 중에서 유방암 환자들을 대상으로 실시한 두 진단검사, 즉 필름 유방촬영술 (film-screen mammography)과 디지털 유방촬영술 (digitized mammography)의 비교에 적용한다. 이 자료의 자세한 설명은 Powell 등 (1999)에서 찾아볼 수 있다. 이 진단검사 자료의 특징은 질병군에 속하는 유방암 환자 30명과 정상군에 속하는 유방암이 없는 것으로 확인된 30명에게 실시한 유방암 진단검사 결과는 정규분포를 따르지 않는다는 점이다. 진단검사 결과는 0과 100 사이의 값을 가진다. 여섯 명 방사선과 의사 (radiologists)가 두 진단검사에 의해 판독한 결과 중 본 논문에서는 한 방사선과 의사의 결과만을 분석한다.

필름 유방촬영술을 진단검사 1로 디지털 유방촬영술은 진단검사 2라고 둔다. 두 검정법을 적용했을 때의 결과는 표 3.1에 제시되었다. 귀무가설은 특이도 95% 하에서 두 진단검사의 민감도에 해당하는 정규변량이 서로 동일하다는 것이다. 모수적 검정법을 적용했을 때의 Z 값은 -0.776이고 비모수적 검정법을 적용했을 때의 Z_{Non} 값은 -1.859이

며, 통계량의 값이 음수인 것은 G_1 과 G_2 중 G_2 의 값이 더욱 커서 진단검사 2가 더욱 바람직하다는 것을 뜻한다. 두 검정법을 비교해 보면 비모수적 Z_{Non} 값의 절대값이 더욱 큰 값을 가져 정규분포하지 않는 이 예제 자료에서 비모수적 검정법이 더욱 유의함을 말해 준다. 따라서 이 예제 자료에서는 비모수적 검정법의 적용이 더욱 바람직하다.

표 3.1: 예제 적용 결과

모수적 추정량	추정값	비모수적 추정량	추정값
\bar{X}_1	25.500	$\tilde{\zeta}_{1;p_0}$	70.000
\bar{X}_2	26.500	$\tilde{\zeta}_{2;p_0}$	60.000
\bar{Y}_1	54.667	$\tilde{\mu}_{Y_1}$	50.000
\bar{Y}_2	61.000	$\tilde{\mu}_{Y_2}$	60.000
s_{X_1}	18.632		
s_{X_2}	18.343		
s_{Y_1}	36.054	$\tilde{\sigma}_{Y_1}$	32.056
s_{Y_2}	33.020	$\tilde{\sigma}_{Y_2}$	28.682
\hat{G}_1	-0.041	\tilde{G}_1	-0.624
\hat{G}_2	0.131	\tilde{G}_2	0.000
$\hat{V}(\hat{G}_1 - \hat{G}_2)$	0.049	$\hat{V}(\tilde{G}_1 - \tilde{G}_2)$	0.113
Z	-0.776	Z_{Non}	-1.859

3.2. 모의실험의 계획

본 논문에서 제안하는 Greenhouse와 Mantel (1950)의 모수적 검정법과 비모수 검정법을 비교하기 위해, 정규분포 자료와 또한 정규분포 양쪽 꼬리에 이상치가 포함된 오염된 (contaminated) 정규분포 자료를 생성한다. 질병군과 정상군의 분포의 다양한 형태는 Beam과 Wieand (1991)의 모의실험을 참고하여 실행하였다.

모의실험에 대한 가설은 다음과 같다.

$$\begin{aligned} H_{01} : z \geq 0 & \quad vs \quad H_{A1} : z < 0 \quad \text{와} \\ H_{02} : z \leq 0 & \quad vs \quad H_{A2} : z > 0. \end{aligned}$$

귀무가설 하에서 각각 $z > 1.645$ 이고, $z < -1.645$ 이면 기각한다.

정규분포에 대한 모의실험 자료는 다음과 같이 정상군의 X 와 질병군의 Y 를 각각 이변량 정규분포 (bivariate normal distribution)로 생성한다.

$$\begin{aligned} \mathbf{X}_i &= \begin{pmatrix} X_{1i} \\ X_{2i} \end{pmatrix} \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right], \quad i = 1, 2, \dots, m, \\ \mathbf{Y}_j &= \begin{pmatrix} Y_{1j} \\ Y_{2j} \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right], \quad j = 1, 2, \dots, n. \end{aligned}$$

편의상 첨자를 생략하고 한 군의 경우로 설명하면, 정상군의 X 는 $N(0, 1)$ 을 따르며, 질병군의 Y 는 $N(\mu_Y, 1)$ 을 따르도록 생성한 두 진단검사의 관측값 사이의 상관계수는 X 와 Y 의 경우 모두 ρ 로 같은 값을 가정한다. 여기서 μ_Y 의 결정은 모의실험에서 미리 정하는 특이도와 민감도에 따라 결정된다. 다시 말하면 정해진 특이도 $p' = \Pr[X \leq c]$ 에 의해 절사점 (cut-off value)인 c 는 $c = \Phi^{-1}(p')$ 에 의해 결정된다. 여기서 Φ^{-1} 는 표준정규 누적분포함수의 역함수이다. 또한 평균 μ_Y 는 미리 정해진 민감도인 $q' = \Pr[Y > c]$ 에 의해 결정되며 구체적으로 $\mu_Y = c - \Phi^{-1}(1 - q')$ 이 된다.

오염된 정규분포는 정상군의 X 에 대해서 $(1 - \delta)\Phi(x) + \delta\Phi(x/\theta)$ 로서 오염의 정도인 $\delta = 0.025$ 와 $\theta = 5$ 로 두어, 전체 표준 정규분포 자료의 2.5%는 평균이 0이고 표준편차가 5인 정규분포로 생성한다. 질병군의 Y 에 대해서도 X 와 마찬가지로 $(1 - \delta)N(\mu_Y, 1) + \delta N(\mu_Y, \theta^2)$ 으로서 $\delta = 0.025$ 와 $\theta = 5$ 로 정하여 전체 정규분포 자료의 2.5%는 평균이 μ_Y 이고 표준편차가 5인 정규분포로 생성한다.

유의수준과 검정력을 알아보는 모의실험에서 각 군의 관측값의 개수는 X 와 Y 에서 각각 m 과 n 으로 정하고 $m = n = 50, 100, 200$ 을 시도하고, 10,000번 반복으로 두 가지 검정법을 비교한다. 또한 민감도는 각각 0.6, 0.8, 0.9인 경우로, 특이도는 각각 0.6, 0.8, 0.9인 경우, 두 진단검사의 관측값 사이의 상관계수는 각각 0.1, 0.5로 모의실험을 실행한다.

3.3. 모의실험 결과

표 3.2에 특이도가 0.9일 때의 모수적 검정법과 비모수적 검정법의 귀무가설 하에서의 검정력, 즉 추정된 유의수준을 제시하였다. 정규분포의 경우에 모수적 검정법은 0.05에 가까운 값들이지만, 비모수적 검정법은 두 진단검사의 관측값 사이의 상관계수가 0.1일 때에는 0.05에 가까운 값이고, 상관계수가 0.5일 때는 0.05 수준보다 약간 큰 0.07 안팎의 값들이다. 그러나 오염된 정규분포의 경우에는 모수적 검정법의 추정된 유의수준은 매우 큰 값으로서 0.16~0.26 수준이며 이와 달리 비모수적 검정법의 유의수준은 상관계수가 0.1일 때는 0.05에 가까운 값이고 상관계수가 0.5일 때는 0.05보다 약간 큰 0.05~0.07의 값으로 나타났다. 오염된 정규분포에서는 모수적 검정법은 표본수가 증가함에 따라 유의수준이 약간 증가하는 경향을 보이지만, 비모수적 검정법은 표본수가 증가함에 따라 유의수준이 약간 감소하는 경향을 보인다.

표 3.3에 특이도가 0.9일 때의 모수적 검정법과 비모수적 검정법의 검정력 결과를 제시하였다. 두 가지 방법론의 검정력을 추정된 유의수준이 동일한 상태에서 비교하기 위해서 모수적 검정법의 유의수준이 0.05 정도가 되도록 보정해 준 후에 검정력을 비교하였다. 구체적으로 모의실험에서 이에 대응하는 유의한계 (significance limit)는 0.00076이었다. 모수적 검정법과 비모수적 검정법 모두에서 표본수가 증가할수록, 두 진단검사의 관측값 사이의 상관계수가 높을수록 검정력이 증가한다. 또한 민감도의 차이가 더 클수록, 즉 0.1의 차이보다는 0.3의 차이일 때 검정력이 증가한다. 검정력은 특이도와 큰 관계가 없는 것으로 드러났다. 예상하는 바와 같이 정규분포에서는 모수적 검정법의 검정력이 더 높고, 오염된 정규분포에서는 비모수적 검정법의 검정력이 더 높다.

표 3.2: 특이도가 0.9일 때의 두 검정법의 추정된 유의수준

ρ	민감도		표본 수	정규분포				오염된 정규분포				
	진단 검사1	진단 검사2		모수적 검정법		비모수적 검정법		모수적 검정법		비모수적 검정법		
				< -1.645	> 1.645	< -1.645	> 1.645	< -1.645	> 1.645	< -1.645	> 1.645	
0.1	0.6	0.6	50	0.050	0.049	0.050	0.049	0.167	0.154	0.046	0.042	
			100	0.049	0.048	0.048	0.049	0.191	0.192	0.039	0.038	
			200	0.048	0.051	0.053	0.054	0.212	0.199	0.039	0.036	
0.5	0.6	0.6	50	0.049	0.049	0.074	0.076	0.174	0.178	0.058	0.061	
			100	0.051	0.049	0.074	0.070	0.201	0.215	0.055	0.057	
			200	0.048	0.048	0.076	0.074	0.218	0.229	0.058	0.057	
0.1	0.8	0.8	50	0.053	0.050	0.050	0.051	0.176	0.167	0.042	0.043	
			100	0.047	0.053	0.048	0.050	0.195	0.181	0.043	0.041	
			200	0.053	0.052	0.056	0.051	0.216	0.199	0.039	0.037	
0.5	0.8	0.8	50	0.055	0.050	0.074	0.072	0.190	0.180	0.063	0.065	
			100	0.053	0.050	0.071	0.074	0.221	0.210	0.059	0.056	
			200	0.051	0.046	0.078	0.073	0.239	0.212	0.058	0.057	
0.1	0.9	0.9	50	0.054	0.047	0.058	0.050	0.175	0.176	0.049	0.046	
			100	0.051	0.048	0.051	0.047	0.206	0.194	0.041	0.040	
			200	0.051	0.049	0.050	0.049	0.224	0.196	0.039	0.039	
0.5	0.9	0.9	50	0.049	0.050	0.066	0.072	0.205	0.178	0.068	0.064	
			100	0.050	0.050	0.069	0.067	0.233	0.201	0.058	0.054	
			200	0.051	0.049	0.076	0.073	0.261	0.207	0.055	0.051	

4. 결론 및 토의

임상의 현장에서 진단검사의 결과가 정규분포한다면 모수적 검정법이 적절하지만, 여러 가지 요인들로 인해 진단검사의 결과에 이상치가 있게 되면 모수적 검정법을 적용하는 것이 옳지 않다. 이와 같이 이상치가 있는 진단검사의 결과를 분석할 때에는 본 논문에서 제안한 이상치의 영향을 많이 덜 받는 평균과 표준편차의 분위수를 이용한 비모수적 검정법이 더욱 적절하다. 검정력은 표본수가 클수록, 민감도의 차이가 클수록, 두 진단검사의 관측값 사이의 상관계수가 높을수록 더 좋다는 결론이 나왔다. 특히 오염된 정규분포에서 모수적 검정법의 유의수준이 매우 큰 값인 것에 반하여 비모수적 검정법은 상대적으로 안정한 값이다. 한편 비모수적 검정법은 표본수가 작은 경우에는 검정력이 비교적 낮으므로 어느 정도의 표본수가 확보되어야 한다. 장래의 연구로서 진단검사의 결과가 비대칭적일 때의 비모수적 검정 통계량의 개발이 남아 있다.

부록

하나의 진단검사로부터 구한 \tilde{G} 의 근사분산의 추정식 (2.9)의 유도과정을 알아보자. Kendall (1969, p. 232)에 제시된 $\text{Cov}(a, b) = 0$ 일 때 성립하는 공식

$$V\left(\frac{a}{b}\right) = \left\{ \frac{E(a)}{E(b)} \right\}^2 \left\{ \frac{V(a)}{E^2(a)} + \frac{V(b)}{E^2(b)} \right\}$$

표 3.3: 특이도가 0.9일 때의 두 검정법의 검정력

ρ	민감도		정규분포		오염된 정규분포		
	진단 검사1	진단 검사2	표본 수	모수적 검정법	비모수적 검정법	모수적 검정법	비모수적 검정법
0.1	0.9	0.6	50	0.881	0.715	0.312	0.549
			100	0.991	0.932	0.516	0.816
			200	1.000	0.997	0.745	0.979
0.5	0.9	0.6	50	0.967	0.853	0.409	0.683
			100	1.000	0.982	0.609	0.913
			200	1.000	1.000	0.803	0.994
0.1	0.9	0.8	50	0.321	0.227	0.093	0.167
			100	0.520	0.376	0.161	0.259
			200	0.778	0.590	0.256	0.429
0.5	0.9	0.8	50	0.430	0.337	0.118	0.241
			100	0.662	0.509	0.205	0.364
			200	0.908	0.745	0.305	0.565
0.1	0.8	0.6	50	0.509	0.354	0.123	0.248
			100	0.748	0.559	0.239	0.409
			200	0.949	0.815	0.376	0.664
0.5	0.8	0.6	50	0.651	0.489	0.171	0.351
			100	0.894	0.727	0.313	0.555
			200	0.993	0.928	0.471	0.801
0.1	0.6	0.9	50	0.884	0.709	0.314	0.553
			100	0.993	0.929	0.536	0.816
			200	1.000	0.997	0.771	0.977
0.5	0.6	0.9	50	0.968	0.849	0.423	0.689
			100	1.000	0.982	0.631	0.913
			200	1.000	1.000	0.826	0.994
0.1	0.6	0.8	50	0.499	0.351	0.128	0.256
			100	0.747	0.558	0.248	0.415
			200	0.955	0.824	0.403	0.669
0.5	0.6	0.8	50	0.661	0.492	0.172	0.352
			100	0.897	0.733	0.312	0.558
			200	0.993	0.931	0.472	0.806
0.1	0.8	0.9	50	0.339	0.236	0.098	0.174
			100	0.510	0.367	0.176	0.257
			200	0.772	0.590	0.284	0.432
0.5	0.8	0.9	50	0.438	0.342	0.130	0.240
			100	0.665	0.505	0.237	0.364
			200	0.904	0.740	0.351	0.574

을 이용하여 $V(\tilde{G})$ 는 다음과 같이 표현된다.

$$V(\tilde{G}) = V\left(\frac{\tilde{\mu}_Y - \tilde{\zeta}_{p_0}}{\tilde{\sigma}_Y}\right) = \left[\frac{E(\tilde{\mu}_Y - \tilde{\zeta}_{p_0})}{E(\tilde{\sigma}_Y)}\right]^2 \left[\frac{V(\tilde{\mu}_Y - \tilde{\zeta}_{p_0})}{E^2(\tilde{\mu}_Y - \tilde{\zeta}_{p_0})} + \frac{V(\tilde{\sigma}_Y)}{E^2(\tilde{\sigma}_Y)} \right]. \quad (\text{A.1})$$

식 (A.1)에 나오는 다음의 4개의 부분식을 대입한다.

$$\begin{aligned} E(\tilde{\mu}_Y - \tilde{\zeta}_{p_0}) &= E\left(\frac{\tilde{\zeta}_{1-p_m} + \tilde{\zeta}_{p_m}}{2} - \tilde{\zeta}_{p_0}\right) = \frac{\zeta_{1-p_m} + \zeta_{p_m}}{2} - \zeta_{p_0}, \\ E(\tilde{\sigma}_Y) &= E\left(\frac{\tilde{\zeta}_{1-p_s} - \tilde{\zeta}_{p_s}}{2k}\right) = \frac{\zeta_{1-p_s} - \zeta_{p_s}}{2k}, \\ V(\tilde{\mu}_Y - \tilde{\zeta}_{p_0}) &= V(\tilde{\mu}_Y) + V(\tilde{\zeta}_{p_0}) - 2\text{Cov}(\tilde{\mu}_Y, \tilde{\zeta}_{p_0}) = \frac{p_m \sigma_Y^2}{2ng_m^2} + \frac{p_0(1-p_0)\sigma_X^2}{mg_{p_0}^2}, \\ V(\tilde{\sigma}_Y) &= \frac{p_s(1-2p_s)\sigma_Y^2}{2nk^2g_s^2}. \end{aligned}$$

세 번째와 네 번째 부분식에서 $\text{Cov}(\tilde{\mu}_Y, \tilde{\zeta}_{p_0}) = 0$ 을 이용하고, 또한 $0 < p < 1 - p < 1$ 의 조건하에서 Y 군의 분포에서 평균과 표준편차 분위수 추정량 모두에서 나타나는 $V(\hat{\zeta}_p) = V(\hat{\zeta}_{1-p}) = \{p(1-p)\sigma_Y^2\}/(ng^2)$ 와 $\text{Cov}(\hat{\zeta}_p, \hat{\zeta}_{1-p}) = p^2\sigma_Y^2/(ng^2)$ 을 이용하여 Benson(1949)이 제시한 평균과 표준편차 분위수 추정량의 분산인

$$V(\tilde{\mu}_Y) = \frac{p_m \sigma_Y^2}{2ng_m^2}, \quad V(\tilde{\sigma}_Y) = \frac{p_s(1-2p_s)\sigma_Y^2}{2nk^2g_s^2}$$

을 또한 이용한다.

이제 $V(\tilde{G})$ 의 추정량 $\hat{V}(\tilde{G})$ 는 다음과 같은 각 항의 추정량들을 대입하여 구한다.

$$\begin{aligned} \hat{E}(\tilde{\zeta}_{p_0}) &\cong \tilde{\zeta}_{p_0}, \\ \hat{E}(\tilde{\mu}_Y) &= \hat{E}\left(\frac{\tilde{\zeta}_{1-p_m} + \tilde{\zeta}_{p_m}}{2}\right) \cong \frac{\zeta_{1-p_m} + \zeta_{p_m}}{2}, \\ \hat{E}(\tilde{\sigma}_Y) &\cong \frac{\tilde{\zeta}_{1-p_s} - \tilde{\zeta}_{p_s}}{2k} = \tilde{\sigma}_Y. \end{aligned}$$

다음으로 두 진단검사의 비교에서 요구되는 식 (2.10)의 $\hat{\text{Cov}}(\tilde{G}_1, \tilde{G}_2)$ 의 유도과정을 알아보자. 우선 $\text{Cov}\{g(x), h(y)\} \cong \sum_{i=0}^k \sum_{j=0}^l g'_i(\theta) h'_j(\phi) \text{Cov}(x_i, y_i)$ 을 이용하면 $\text{Cov}(\tilde{G}_1, \tilde{G}_2)$ 는 다음과 같다.

$$\begin{aligned} \text{Cov}(\tilde{G}_1, \tilde{G}_2) &= \text{Cov}\left(\frac{\tilde{\mu}_{Y_1} - \tilde{\zeta}_{1;p_0}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\mu}_{Y_2} - \tilde{\zeta}_{2;p_0}}{\tilde{\sigma}_{Y_2}}\right) \\ &\cong \text{Cov}\left(\frac{\tilde{\mu}_{Y_1}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\mu}_{Y_2}}{\tilde{\sigma}_{Y_2}}\right) - \text{Cov}\left(\frac{\tilde{\mu}_{Y_1}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\zeta}_{2;p_0}}{\tilde{\sigma}_{Y_2}}\right) - \text{Cov}\left(\frac{\tilde{\zeta}_{1;p_0}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\mu}_{Y_2}}{\tilde{\sigma}_{Y_2}}\right) + \text{Cov}\left(\frac{\tilde{\zeta}_{1;p_0}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\zeta}_{2;p_0}}{\tilde{\sigma}_{Y_2}}\right) \end{aligned}$$

여기서 $\text{Cov}(\tilde{\mu}_{Y_1}, 1/\tilde{\sigma}_{Y_2}) = \text{Cov}(1/\tilde{\sigma}_{Y_1}, \tilde{\mu}_{Y_2}) = 0$ 을 이용하여 첫째항은

$$\begin{aligned} \text{Cov}\left(\frac{\tilde{\mu}_{Y_1}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\mu}_{Y_2}}{\tilde{\sigma}_{Y_2}}\right) &\cong E\left(\frac{1}{\tilde{\sigma}_{Y_1}}\right) E\left(\frac{1}{\tilde{\sigma}_{Y_2}}\right) \text{Cov}(\tilde{\mu}_{Y_1}, \tilde{\mu}_{Y_2}) \\ &\quad + E(\tilde{\mu}_{Y_1}) E(\tilde{\mu}_{Y_2}) \text{Cov}\left(\frac{1}{\tilde{\sigma}_{Y_1}}, \frac{1}{\tilde{\sigma}_{Y_2}}\right) \end{aligned}$$

이 된다. $\tilde{\mu}_{Y_t}$ ($t = 1, 2$)와 $\tilde{\sigma}_{Y_t}$ ($t = 1, 2$)은 $\tilde{\zeta}_{t;p_0}$ 와 서로 독립이고 이들의 합수도 서로 독립이라는 사실을 이용하여 나머지 세 항은 다음과 같다.

$$\begin{aligned}\text{Cov}\left(\frac{\tilde{\mu}_{Y_1}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\zeta}_{2;p_0}}{\tilde{\sigma}_{Y_2}}\right) &\cong E(\tilde{\mu}_{Y_1})E(\tilde{\zeta}_{2;p_0})\text{Cov}\left(\frac{1}{\tilde{\sigma}_{Y_1}}, \frac{1}{\tilde{\sigma}_{Y_2}}\right), \\ \text{Cov}\left(\frac{\tilde{\zeta}_{1;p_0}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\mu}_{Y_2}}{\tilde{\sigma}_{Y_2}}\right) &\cong E(\tilde{\zeta}_{1;p_0})E(\tilde{\mu}_{Y_2})\text{Cov}\left(\frac{1}{\tilde{\sigma}_{Y_1}}, \frac{1}{\tilde{\sigma}_{Y_2}}\right), \\ \text{Cov}\left(\frac{\tilde{\zeta}_{1;p_0}}{\tilde{\sigma}_{Y_1}}, \frac{\tilde{\zeta}_{2;p_0}}{\tilde{\sigma}_{Y_2}}\right) &\cong E\left(\frac{1}{\tilde{\sigma}_{Y_1}}\right)E\left(\frac{1}{\tilde{\sigma}_{Y_2}}\right)\text{Cov}(\tilde{\zeta}_{1;p_0}, \tilde{\zeta}_{2;p_0}) \\ &\quad + E(\tilde{\zeta}_{1;p_0})E(\tilde{\zeta}_{2;p_0})\text{Cov}\left(\frac{1}{\tilde{\sigma}_{Y_1}}, \frac{1}{\tilde{\sigma}_{Y_2}}\right).\end{aligned}$$

이제 첫 항에서 나타나는 $\text{Cov}(\tilde{\mu}_{Y_1}, \tilde{\mu}_{Y_2}) \approx \text{Cov}(\bar{Y}_1, \bar{Y}_2) = \rho_Y \sigma_{Y_1} \sigma_{Y_2} / n$ 과 모든 항에서 나타나는 $\text{Cov}(1/\tilde{\sigma}_{Y_1}, 1/\tilde{\sigma}_{Y_2})$ 은 Greenhouse와 Mantel(1950)에 제시된

$$\text{Cov}\left(\frac{1}{\tilde{\sigma}_{Y_1}}, \frac{1}{\tilde{\sigma}_{Y_2}}\right) \cong \left\{E\left(\frac{1}{\tilde{\sigma}_{Y_1}}\right)E\left(\frac{1}{\tilde{\sigma}_{Y_2}}\right)\right\}^2 \text{Cov}(\tilde{\sigma}_{Y_1} \tilde{\sigma}_{Y_2})$$

을 이용하고, 다시 $\text{Cov}(\tilde{\sigma}_{Y_1}, \tilde{\sigma}_{Y_2}) \approx \text{Cov}(s_{Y_1} s_{Y_2}) = \rho_Y^2 \sigma_{Y_1} \sigma_{Y_2} / \{2(n-1)\}$ 을 이용하여 정리한다. 마지막 항에 나타난 $\text{Cov}(\tilde{\zeta}_{1;p_0}, \tilde{\zeta}_{2;p_0})$ 는 상관계수는 단조 변환(monotone transformation) 함수 f 에 대해 근사적으로 불변하다는 사실, 즉

$$\begin{aligned}\text{Corr}\{f(X_i), f(X_j)\} &\approx \frac{[f'(\mu)]^2 \text{Cov}(X_i, X_j)}{\sqrt{[f'(\mu)]^4 V(X_i) V(X_j)}} \\ &= \text{Corr}\{X_i, X_j\}\end{aligned}$$

을 이용하여 (Zaykin 등, 2002), $\text{Cov}(\tilde{\zeta}_{1;p_0}, \tilde{\zeta}_{2;p_0}) \approx \text{Corr}(X_1, X_2) \sqrt{V(\tilde{\zeta}_{1;p_0}) V(\tilde{\zeta}_{2;p_0})}$ 을 유도한다. 이때 $\text{Corr}(X_1, X_2)$ 의 추정량으로는 비모수 Spearman의 순위상관계수인 $r_S(x)$ 을 사용하고, 각 요소의 추정량을 대입하여 $\text{Cov}(\tilde{G}_1, \tilde{G}_2)$ 의 추정량인 $\hat{\text{Cov}}(\tilde{G}_1, \tilde{G}_2)$ 을 구하면 다음과 같다.

$$\begin{aligned}\hat{\text{Cov}}(\tilde{G}_1, \tilde{G}_2) &\cong \hat{\text{Cov}}\left(\frac{1}{\tilde{\sigma}_{Y_1}}, \frac{1}{\tilde{\sigma}_{Y_2}}\right)(\tilde{\mu}_{Y_1} \tilde{\mu}_{Y_2} - \tilde{\mu}_{Y_1} \tilde{\zeta}_{2;p_0} - \tilde{\zeta}_{1;p_0} \tilde{\mu}_{Y_2} + \tilde{\zeta}_{1;p_0} \tilde{\zeta}_{2;p_0}) \\ &\quad + \frac{\hat{\text{Cov}}(\tilde{\mu}_{Y_1}, \tilde{\mu}_{Y_2})}{\tilde{\sigma}_{Y_1} \tilde{\sigma}_{Y_2}} + \frac{\hat{\text{Cov}}(\tilde{\zeta}_{1;p_0}, \tilde{\zeta}_{2;p_0})}{\tilde{\sigma}_{Y_1} \tilde{\sigma}_{Y_2}}.\end{aligned}$$

추가적으로 $\hat{E}(1/\tilde{\sigma}_Y) = \hat{E}(2k/(\tilde{\zeta}_{1-p_s} - \tilde{\zeta}_{p_s})) \cong 2k/(\tilde{\zeta}_{1-p_s} - \tilde{\zeta}_{p_s}) = 1/\tilde{\sigma}_Y$ 이 과정에서 필요하다.

참고문헌

- Beam, C. A. and Wieand, H. S. (1991). A statistical method for the comparison of a discrete diagnostic test with several continuous diagnostic tests. *Biometrics*, **47**, 907–919.
- Bennett, B. M. (1972). On comparisons of sensitivity, specificity and predictive value of a number of diagnostic procedures. *Biometrics*, **28**, 793–800.
- Benson, F. (1949). A note on the estimation of mean and standard deviation from quantiles. *Journal of the Royal Statistical Society, Ser. B*, **11**, 91–100.
- Brown, B. M. (1981). Symmetric quantile averages and related estimators. *Biometrika*, **68**, 235–242.
- Cramer, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.
- Greenhouse, S. W. and Mantel, N. (1950). The evaluation of diagnostic tests. *Biometrics*, **6**, 399–412.
- Kendall, M. G. (1969). *The Advanced Theory of Statistics*. 3ed, Charles Griffin & Co, London.
- Obuchowski, N. A., Beiden S. V., Berbaum K. S., Hillis S. L., Ishwaran H., Song H. H. and Wagner R. F. (2004). Multireader, multicase receiver operating characteristic analysis: an empirical comparison of five methods1. *Academic Radiology*, **11**, 980–995.
- Pearson, K. (1920). On the probable errors of frequency constants. *Biometrika*, **13**, 113–132.
- Powell, K. A., Obuchowski, N. A., Chilcote, W. A., Barry, M. M., Ganobcik, S. N. and Cardenosa, G. (1999). Film-screen versus digitized mammography: assessment of clinical equivalence. *American Journal of Roentgenology*, **173**, 889–894.
- Wieand, S., Gail, M. H., James, B. R. and James, K. L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, **76**, 585–592.
- Zaykin, D. V., Zhivotovsky, L. A., Westfall, P. H. and Weir, B. S. (2002). Truncated product method for combining *p*-values. *Genetic Epidemiology*, **22**, 170–185.

[Received September 2007, Accepted October 2007]