



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

내포 관계인 두 모형의 AUC
차이를 검정하는 방법 비교

연세대학교 대학원

의학전산통계학협동과정

의학통계학전공

이 호 재

내포 관계인 두 모형의 AUC 차이를 검정하는 방법 비교

지도 정 인 경 교수

이 논문을 석사 학위논문으로 제출함

2018년 6월 일

연세대학교 대학원
의학전산통계학협동과정
의학통계학전공
이 호 재

이호재의 석사 학위논문을 인준함

심사위원 정 인 경 인

심사위원 남 정 모 인

심사위원 한 경 화 인

연세대학교 대학원

2018년 6월 일

차 례

표 차례	ii
국문 요약	iii
제 1장 서론	1
1.1. 연구 배경 및 목적	1
1.2. 연구 내용 및 방법	2
1.3. 논문의 구성	3
제 2장 이론적 배경	4
2.1. AUC의 개념	4
2.2. AUC의 비교	5
2.2.1. Hanley & McNeil's method	5
2.2.2. DeLong's method	7
2.2.3. Bandos' method	8
2.2.4. Li's method	9
2.2.5. Heller's method	12
제 3장 모의실험	16
3.1. 모의실험 설계	16
3.2. 두 지표의 상관관계에 따른 AUC	17
3.3. 자료의 생성	21
3.4. 모의실험 결과	22
제 4장 실제 자료 적용	34
4.1. 자료 설명	34
4.2. 자료 분석 결과	34
제 5장 결론 및 고찰	35
참고 문헌	38
영문 요약	40

표 차례

표 1. 자료 구조	5
표 2. AUC의 차이가 없고 상관관계가 없는 경우	19
표 3. AUC의 차이가 없고 상관관계가 있는 경우	19
표 4. AUC의 차이가 있고 상관관계가 없는 경우	20
표 5. AUC의 차이가 있고 상관관계가 있는 경우	20
표 6. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 0, 0, 0)$	23
표 7. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0, 0)$	24
표 8. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.1, 0.1)$	25
표 9. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.1, 0.3)$	26
표 10. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.1, 0.5)$	27
표 11. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.3, 0.1)$	28
표 12. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.3, 0.3)$	29
표 13. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.3, 0.5)$	30
표 14. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.5, 0.1)$	31
표 15. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.5, 0.3)$	32
표 16. $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.5, 0.5)$	33
표 17. 실제 자료 분석 결과	35

국 문 요 약

내포 관계인 두 모형의 AUC 차이를 검정하는 방법 비교

의학적 기술의 발달로 새롭게 발견되는 생체지표나 새롭게 이루어지는 수술 등 기존에 알려진 지표 이외에도 활용 가능한 변수는 다양하다. 기존에 알려진 지표들로 이루어진 모형과 새로운 지표가 추가된 모형 사이에는 내포(nested) 관계가 형성되는데, 모형의 성능이 얼마나 증가하였는지 검정하는 것은 통계학적 관점과 임상적 관점에서 모두 중요하다. 모형의 성능을 평가하는 지표 중 ROC(receiver operating characteristic) 곡선 아래의 면적인 AUC(area under the curve)는 가장 널리 사용되는 척도이다.

실제 새롭게 발견된 지표가 통계학적으로 유의하더라도 AUC는 기존 모형에 비해 크게 증가하지 않는 연구 결과가 빈번하다. 최근 다양한 연구를 통해 DeLong의 방법이 너무 보수적이라고 보고되고 있어, 본 연구에서는 내포 관계에 있는 두 모형의 AUC 차이를 검정하기 위하여 기존의 Hanley와 McNeil의 방법(1983), DeLong의 방법(1988), 그리고 최근 제안된 정확한 붓스트랩 분산을 계산한 Bandos의 방법(2007), 일반화 추측량 개념을 이용한 Li의 방법(2008)과 최대순위상관 추정량을 활용한 Heller의 방법(2017)을 비교한다.

연구 결과 두 모형의 고정된 상관계수를 사용하며 AUC 평균이 적어도 0.70 이상이 되어야 하는 Hanley와 McNeil의 방법은 현실적으로 적용하기에 어려움이 존재한다. DeLong의 방법은 실제 모형의 성능을 증가시킬 수 있는 변수가 추가되었음에도 이를 잘 반영하지 못한다. Bandos의 방법은 DeLong 방법의 결과와 유사하며, 내포 관계에 있는 두 모형의 AUC 차이를 검정하기에 부적절하다. 최종적으로 Li의 방법이 AUC의 차이를 가장 잘 반영하며 전반적으로 신뢰구간 간격도 좁아, 비교한 5가지 방법 중에서 가장 좋은 검정 방법이다.

핵심되는 말 : 내포 관계, AUC 차이, 일반화 추측량, 최대순위상관 추정량, 붓스트랩

제 1장 서론

1.1 연구 배경 및 목적

임상과 역학 분야에서 질병 발생 여부에 대한 진단 및 예측모형 개발은 중요한 이슈이다. 의학적 기술의 발달로 새롭게 발견되는 생체지표는 다양하다. 새로운 생체지표가 발견되어 기존에 알려진 지표들로 이루어진 모형과 새로운 지표가 추가된 모형 사이에는 내포(nested) 관계가 형성되는데, 그렇다면 새로운 지표가 추가되었을 때 모형의 성능은 얼마나 증가하였을까? 이 논점에는 임상적 관점과 통계적 관점이 동시에 존재한다. 새로운 모형의 성능이 기존 모형과 비교하여 통계적으로 유의하게 증가하였다 하더라도, 새로운 생체지표를 탐색하는 과정이 비싸거나 환자에게 침습적인 절차가 동반되면서 그 차이가 미미할 때 임상적으로 수용되기 어렵다.

민감도와 특이도, 음성예측도, 양성예측도 뿐만 아니라 최근 널리 사용되어 온 NRI(net reclassification improvement)와 IDI(integrated discrimination improvement) 등 모형의 성능을 평가하는 다양한 지표가 있지만, 그 중 ROC(receiver operating characteristic) 곡선 아래의 면적인 AUC(area under the curve)는 가장 널리 사용되는 척도이다. 일반적으로 새로운 생체지표가 발견되면 기존 지표와 새로운 지표가 포함된 모형을 구축하고, Wald 검정이나 가능도비(likelihood ratio) 검정을 통해 이분형 결과와 새로운 지표 사이의 연관성을 파악한다. 추가된 지표가 통계적으로 유의하면, 기존 모형과 새로운 모형 각각의 AUC를 계산하고 두 AUC의 차이와 그 신뢰구간을 통해 새로운 지표의 수용 여부를 판단하게 된다.

두 모형의 AUC를 비교하기 위하여 기존에 개발된 통계적 방법은 질병이 있는 군과 없는 군 간의 상관관계를 이용한 Hanley와 McNeil의 방법(1983)이 있다. 이에 DeLong과 그 연구진(1988)은 Hanley와 McNeil의 방법이 두 AUC

의 분산을 추정함에 있어 Gaussian 모형을 가정한다는 제한이 있음을 밝히며 Mann-Whitney U-statistic에 기반을 둔 새로운 비모수적인 접근방법을 제안하였다. 그러나 최근 다양한 연구(Pepe, Demler, Vickers, and Begg)에서 부분 F 검정 등을 통해 새롭게 추가된 지표가 통계적으로 유의함에도 DeLong의 방법은 내포 관계에 있는 두 모형의 AUC를 비교할 때 낮은 검정력(power)를 갖게 되어 너무 보수적(conservative)이라는 의견이 존재한다.

그리하여 본 논문에서는 기존에 널리 사용되는 위 방법과 최근 제안된 방법에 관하여 다양한 표본 수와 두 변수의 상관관계를 고려한 시뮬레이션을 통해 AUC를 비교하고, 이분형 자료에 대한 내포 관계에 있는 두 모형의 성능을 비교하기 위해 적절한 통계적 방법론이 무엇인지 고찰해보고자 한다.

1.2 연구 내용 및 방법

본 연구에서는 내포 관계에 있는 두 모형의 AUC 차이를 검정하기 위해 기존에 널리 사용되는 Hanley와 McNeil의 방법(1983), DeLong의 방법(1988), 그리고 최근 제안된 정확한 붓스트랩(Bootstrap) 분산 추정에 기반을 둔 Bandos의 방법(2007), 일반화 추측량(generalized pivotal quantity) 개념을 이용한 Li의 방법(2008), MRC(maximum rank correlation) 추정량에 기반을 둔 Heller의 방법(2017)에 관하여 다양한 표본 수와 두 변수의 상관관계를 고려한 시뮬레이션을 통해 위 다섯 가지 방법에서 제 1종 오류(Type 1 error)와 검정력(power)를 비교 및 평가한다.

1.3 논문의 구성

제 1장에서는 연구의 배경 및 목적과 연구 내용 및 방법을 소개한다. 제 2장에서는 대응표본에서 내포 관계에 있는 두 모형의 AUC 차이를 검정하는 다섯 가지 방법에 대한 이론적 배경과 개념을 정리한다. 제 3장에서는 기존에 알려진 지표와 새롭게 발견된 지표 사이의 상관관계가 두 모형의 AUC 차이에 미치는 영향에 대해 알아보고, 위 다섯 가지 방법에 대하여 모의실험을 통해 다양한 상황에서 계산되는 AUC를 비교 및 평가한다. 제 4장에서는 실제 자료에 대한 설명과 분석 결과를 제시하고, 마지막으로 제 5장에서는 본 연구를 통한 결론 및 고찰을 논의한다.

제 2장 이론적 배경

2.1에서는 AUC의 개념을 소개하고자 한다. 2.2에서는 대응표본에서 내포 관계에 있는 두 모형의 AUC 차이를 검정하는 다섯 가지 방법을 간략하게 정리한다.

2.1 AUC의 개념

질병 유무 등의 이분형 결과에 대한 모형을 구축한 뒤 모든 가능한 cut-off point(c)에 상응하는 민감도(sensitivity)와 특이도(specificity)를 계산한다. 좌표평면 가로축에 (1-특이도)를, 세로축에 민감도를 (0,0), (0,1), (1,0), (1,1) 범위 안에 표시하면 추정된 모형의 ROC 곡선을 얻을 수 있다. ROC 곡선 하단의 면적을 AUC라고 하며, 1에 가까울수록, 혹은 ROC 곡선이 (0,1)에 가까울수록 좋은 모형이라고 평가한다. AUC(θ)는 모형에서 추정된 질환군($Y=1$)과 비질환군($Y=0$)의 선형 예측변수(linear predictor)를 각각 $\beta^T \mathbf{X}_1$ 와 $\beta^T \mathbf{X}_2$ 라고 할 때 $\theta = P(\beta^T \mathbf{X}_1 > \beta^T \mathbf{X}_2)$ 로 정의된다.

이미 알려진 지표로만 이루어진 모형과 새로운 지표가 추가된 모형의 AUC를 각각 계산하고, 두 AUC의 차이를 통해 새로운 지표가 추가됨으로써 진단의 예측도가 얼마나 향상되었는지 판단한다. 다음으로 2.2에서 두 AUC의 차이 $\delta(=\theta_1 - \theta_2)$ 에 대해 검정하는 통계적 방법론을 알아본다.

2.2 AUC의 비교

기존에 알려진 지표를 \mathbf{X} , 새롭게 추가된 지표를 \mathbf{Z} 라고 할 때 AUC 차이를 검증하는 방법에 사용되는 자료 구조는 표 1과 같다. 질환군은 n 명, 비질환군은 m 명으로 정의한다.

표 1. 자료 구조

	기존에 알려진 지표	새롭게 추가된 지표
질환군 ($Y=1$)	X_1	Z_1
	\vdots	\vdots
	X_n	Z_n
비질환군 ($Y=0$)	X_{n+1}	Z_{n+1}
	\vdots	\vdots
	X_{n+m}	Z_{n+m}

본 연구는 대응표본에서 내포 관계에 있는 AUC를 비교하는 것이므로, 새로운 지표가 포함된 모형에서의 선형 예측변수를 $\eta_1 = \beta\mathbf{X} + \gamma\mathbf{Z}$ 라 하고, 기존에 알려진 지표로만 이루어진 모형에서의 선형 예측변수를 $\eta_2 = \beta^0\mathbf{X}$ 로 정의한다.

2.2.1 Hanley & McNeil's method (1983)

Hanley와 McNeil의 방법에서 두 모형 각각의 AUC는 비모수적으로 계산되며, 그 차이는 AUC의 평균과 모형 간 질환군과 비질환군 각각의 상관관계를 고려하여 추정한다. η_{ij} 는 i 번째 대상자의 j 모형에서의 선형 예측변수를 의미한다.

$$\hat{\theta}_j = \frac{1}{nm} \sum_{i=1}^{n+m} \sum_{k=1}^{n+m} I[Y_i > Y_k] \left[I(\eta_{ij} > \eta_{kj}) + \frac{1}{2} I(\eta_{ij} = \eta_{kj}) \right], \quad j = 1, 2$$

Q_1 을 무작위로 선택된 질환군 2명의 선형 예측변수가 무작위로 선택된 비 질환군 1명보다 모두 클 확률, Q_2 를 무작위로 선택된 질환군 1명의 선형 예측변수가 무작위로 선택된 비질환군 2명보다 클 확률이라고 정의하면 추정된 $\hat{\theta}$ 의 표준오차는 다음과 같이 계산된다.

$$Q_1 = \theta \div (2 - \theta)$$

$$Q_2 = 2\theta^2 \div (1 + \theta)$$

$$SE(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1 - \hat{\theta}) + (n - 1)(\hat{Q}_1 - \hat{\theta}^2) + (m - 1)(\hat{Q}_2 - \hat{\theta}^2)}{nm}}$$

대응표본이므로 발생하는 두 AUC의 상관계수 r 은 두 모형에서 각각 추정된 질환군의 선형 예측변수 간 Kendall τ 와 비질환군의 선형 예측변수 간 Kendall τ 의 평균, 그리고 두 AUC의 평균에 의해 결정되는데 Hanley와 McNeil의 논문에서 제시된 표는 두 AUC의 평균이 최소 0.70 이상일 때로 현실적인 제한점이 있다. AUC의 차이는 표준정규분포에 근사하여 $H_0: \delta = 0$ 에 대해 검정할 수 있으며, 아래 검정 통계량이 통계적으로 유의할 때 새로운 지표가 진단의 예측력을 향상시킨다고 할 수 있다.

$$Z = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{SE^2(\hat{\theta}_1) + SE^2(\hat{\theta}_2) - 2rSE(\hat{\theta}_1)SE(\hat{\theta}_2)}}$$

$100(1 - \alpha)\%$ 신뢰구간은 $\hat{\delta} \pm Z_{1 - \alpha/2} \sqrt{SE^2(\hat{\theta}_1) + SE^2(\hat{\theta}_2) - 2rSE(\hat{\theta}_1)SE(\hat{\theta}_2)}$ 이다.

2.2.2 DeLong's method (1988)

DeLong의 방법은 Hanley와 McNeil의 방법이 AUC의 분산을 추정함에 있어 Gaussian 모형을 가정한다는 제한점을 밝히며, Mann-Whitney statistic에 기반을 둔 새로운 비모수적인 접근방법을 제안하였다. 두 모형의 AUC는 2.2.1과 동일하게 비모수적인 방법으로 추정하며, Generalized U-statistic의 성질을 이용해 $\hat{\theta}$ 에 대한 분산을 다음과 같이 추정한다.

$$V_{10}^j(X_i) = \frac{1}{m} \sum_{k=1}^m \left[I(\eta_{ij} > \eta_{kj}) + \frac{1}{2} I(\eta_{ij} = \eta_{kj}) \right], \quad j = 1, 2,$$

$$V_{01}^j(Y_k) = \frac{1}{n} \sum_{i=1}^n \left[I(\eta_{ij} > \eta_{kj}) + \frac{1}{2} I(\eta_{ij} = \eta_{kj}) \right], \quad j = 1, 2,$$

$$s_{10}^{j,j'} = \frac{1}{n-1} \sum_{i=1}^n [(V_{10}^j(X_i) - \hat{\theta}_j)(V_{10}^{j'}(X_i) - \hat{\theta}_{j'})], \quad j, j' = 1, 2,$$

$$s_{01}^{j,j'} = \frac{1}{m-1} \sum_{k=1}^m [(V_{01}^j(Y_k) - \hat{\theta}_j)(V_{01}^{j'}(Y_k) - \hat{\theta}_{j'})], \quad j, j' = 1, 2,$$

$$\widehat{Var}(\hat{\theta}_1 - \hat{\theta}_2) = \frac{1}{n}(s_{10}^{1,1} + s_{10}^{2,2} - 2s_{10}^{1,2}) + \frac{1}{m}(s_{01}^{1,1} + s_{01}^{2,2} - 2s_{01}^{1,2})$$

최종적으로 추정된 $100(1-\alpha)\%$ 신뢰구간 $\hat{\delta} \pm Z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\theta}_1 - \hat{\theta}_2)}$ 에 '0'을 포함하는지 확인하여 $H_0: \delta = 0$ 에 대하여 검정하고, 통계적으로 유의할 때 새로운 지표가 추가됨으로써 진단의 예측력이 향상되었다고 할 수 있다.

2.2.3 Bandos' method (2007)

일반적인 붓스트랩(Bootstrap) 방법은 Monte-Carlo 알고리즘을 기반으로 복원추출을 허용하여 동일한 표본 수의 새로운 붓스트랩 표본을 B 개 추출하고, B 개의 데이터를 기반으로 로버스트(robust)하게 모수를 추정한다. 하지만 Monte-Carlo 추정은 상당히 많은 컴퓨터 작업량을 요구하며, 붓스트랩 표본을 추출하는 과정에서 표집편향(sampling bias)이 발생할 수 있다. 또한, 표본이 적을 때 모형이 수렴하지 않는 문제가 발생하여 특정 조합의 붓스트랩 표본만이 분석에 사용될 수 있다는 제한점이 존재한다.

이로 인해 Bandos와 그 연구진은 2.2.1과 동일하게 비모수적인 방법으로 AUC를 추정하고, 그에 대한 정확한 붓스트랩 분산(exact bootstrap variance)을 추정하는 공식을 제안하였다.

$$w_{ik} = I[Y_i > Y_k] \left(\left[I(\eta_{i1} > \eta_{k1}) + \frac{1}{2} I(\eta_{i1} = \eta_{k1}) \right] - \left[I(\eta_{i2} > \eta_{k2}) + \frac{1}{2} I(\eta_{i2} = \eta_{k2}) \right] \right)$$

$$\bar{w}_{i \cdot} = \frac{1}{m} \sum_{k=1}^m w_{ik}, \quad \bar{w}_{\cdot k} = \frac{1}{n} \sum_{i=1}^n w_{ik}$$

$$\bar{w}_{\cdot \cdot} = \hat{\delta} = \hat{\theta}_1 - \hat{\theta}_2$$

$$\begin{aligned} \text{Var}(\hat{\theta}_1 - \hat{\theta}_2) &= \frac{1}{n^2} \sum_{i=1}^n (\bar{w}_{i \cdot} - \bar{w}_{\cdot \cdot})^2 + \frac{1}{m^2} \sum_{k=1}^m (\bar{w}_{\cdot k} - \bar{w}_{\cdot \cdot})^2 \\ &\quad + \frac{1}{n^2 m^2} \sum_{i=1}^n \sum_{k=1}^m (w_{ij} - \bar{w}_{i \cdot} - \bar{w}_{\cdot k} + \bar{w}_{\cdot \cdot})^2 \end{aligned}$$

최종적으로 추정된 $100(1-\alpha)\%$ 신뢰구간 $\hat{\delta} \pm Z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\theta}_1 - \hat{\theta}_2)}$ 에 '0'을 포함하는지 확인하여 $H_0: \delta = 0$ 에 대하여 검정하고, 통계적으로 유의할 때 새로운 지표가 추가됨으로써 진단의 예측력이 향상되었다고 할 수 있다.

2.2.4 Li's method (2008)

Li와 그 연구진은 \mathbf{X} 와 \mathbf{Y} 가 이변량 정규분포를 따를 때 j 번째 진단검사의 AUC를 다음과 같이 계산한다.

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2\left(\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right)$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N_2\left(\boldsymbol{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}, \boldsymbol{\Psi} = \begin{bmatrix} \psi_1^2 & \psi_{12} \\ \psi_{12} & \psi_2^2 \end{bmatrix}\right)$$

$$\theta_j = \frac{\mu_j - \nu_j}{\sqrt{\sigma_j^2 + \psi_j^2}}$$

대응표본에서의 AUC 차이에 대한 정확한 신뢰구간을 추정하기 위하여 Li와 그 연구진은 일반화 추측량(generalized pivotal quantities, GPQ) 개념을 이용한 방법을 제안하였다. \mathbf{W} 를 미지의 모수 $\zeta = (\theta, \eta)$ 를 따르는 확률변수이고, θ 와 η 를 각각 관심 있는 모수와 장애 모수(nuisance parameter) 벡터라고 가정한다. 이때 \mathbf{W} 를 확률변수 \mathbf{W} 의 임의표본이고, \mathbf{w} 를 \mathbf{W} 의 관측값이라고 하자. 또한, $R = R(\mathbf{W}; \mathbf{w}, \zeta)$ 이 \mathbf{W} 와 \mathbf{w} , ζ 의 함수일 때 아래 두 가지 조건을 만족하면 GPQ라고 한다.

- (1) R 의 분포는 미지의 모수에 의존하지 않는다.
- (2) R 의 관측값인 $r = R(\mathbf{w}; \mathbf{w}, \zeta)$ 은 오직 (\mathbf{w}, θ) 의 함수로, 장애 모수 η 에 의존하지 않는다.

여기서 관측된 $r = \theta$ 일 때 GPQ는 FGQP(fiducial generalized pivotal quantity)라고 한다. 이를 기반으로 θ 의 양측 $100(1-\alpha)\%$ GCI(generalized confidence interval)는 $(R_{\alpha/2}, R_{1-\alpha/2})$ 이며, $R_{\alpha/2}$ 와 $R_{1-\alpha/2}$ 는 각각 분포 R 의

100($\alpha/2$)와 100(1- $\alpha/2$)번째 분위수이다. R 의 분위수는 Monte-Carlo 알고리즘을 통하여 추정 가능하다.

위의 이변량 정규분포 \mathbf{X}, \mathbf{Y} 의 임의표본을 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 과 $\mathbf{y}_1, \dots, \mathbf{y}_m$ 라고 하면 자료의 표본공분산행렬은 아래와 같이 구할 수 있다.

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \right) = \left(\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}, \frac{1}{n-1} \begin{bmatrix} SSX_1 & SSX_{12} \\ SSX_{12} & SSX_2 \end{bmatrix} \right)$$

$$(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\Psi}}) = \left(\frac{1}{m} \sum_{i=1}^m \mathbf{y}_i, \frac{1}{m-1} \sum_{i=1}^m (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \right) = \left(\begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix}, \frac{1}{m-1} \begin{bmatrix} SSY_1 & SSY_{12} \\ SSY_{12} & SSY_2 \end{bmatrix} \right)$$

Δ^T 는 $(\boldsymbol{\mu} - \boldsymbol{\nu})^T$, β 와 γ 는 두 모형에서의 회귀계수, $\sigma_{2|1}^2$ 과 $\psi_{2|1}^2$ 은 오차분산을 나타낸다고 할 때, 표본공분산행렬을 이용하여 모수 $(\Delta^T, \beta, \gamma, \sigma_{2|1}^2, \psi_{2|1}^2)$ 의 추정량을 아래와 같이 정의한다.

$$\begin{aligned} & (\mathbf{D}^T, B, G, SSX_{2|1}, SSY_{2|1}) \\ &= \left(\hat{\boldsymbol{\mu}}^T - \hat{\boldsymbol{\nu}}^T, \frac{SSX_{12}}{SSX_1}, \frac{SSY_{12}}{SSY_1}, SSX_2 - \frac{SSX_{12}^2}{SSX_1}, SSY_2 - \frac{SSY_{12}^2}{SSY_1} \right) \end{aligned}$$

AUC의 차이에 대한 100(1- α)% 신뢰구간을 추정하기 위해 Monte-Carlo 알고리즘을 이용하여 아래 다섯 단계를 반복 수행한다.

Step 1 : Monte-Carlo 알고리즘을 반복할 횟수 K 를 설정하고, $1 \leq k \leq K$ 을 만족하는 k 단계에서 각각 $n-1, n-2, m-1, m-2$ 의 자유도를 갖는 서로 독립인 카이제곱분포로부터 난수 $U_{1,k}, U_{2|1,k}, V_{1,k}, V_{2|1,k}$ 와 표준정규분포로부터 난수 $Z_{B,k}, Z_{G,k}$ 를 생성한다.

Step 2 : 추정된 모수들의 관측값인 $(\mathbf{d}^T, b, g, s s x_1, s s y_1, s s x_{2|1}, s s y_{2|1})$ 을 이용하여 $(\sigma_1^2, \sigma_{2|1}^2, \beta)$ 와 $(\psi_1^2, \psi_{2|1}^2, \gamma)$ 의 GPQ를 다음과 같이 계산한다.

$$(R_{\sigma_1^2}, R_{\sigma_{2|1}^2}, R_{\beta}) = \left(\frac{s s x_1}{U_1}, \frac{s s x_{2|1}}{U_{2|1}}, b - Z_{\beta} \sqrt{\frac{1}{U_{2|1}} \frac{s s x_{2|1}}{s s x_1}} \right)$$

$$(R_{\psi_1^2}, R_{\psi_{2|1}^2}, R_{\gamma}) = \left(\frac{s s y_1}{V_1}, \frac{s s y_{2|1}}{V_{2|1}}, g - Z_{\gamma} \sqrt{\frac{1}{V_{2|1}} \frac{s s y_{2|1}}{s s y_1}} \right)$$

Step 3 : Step 2의 두 식에 모수가 포함되어 있지 않으므로 앞서 언급한 GPQ의 두 가지 가정을 만족함을 확인할 수 있다. 마찬가지로 $(\sigma_2^2, \sigma_{12})$ 와 (ψ_2^2, ψ_{12}) 의 GPQ를 다음과 같이 계산한다.

$$(R_{\sigma_2^2}, R_{\sigma_{12}}) = (R_{\beta}^2 R_{\sigma_1^2} + R_{\sigma_{2|1}^2}, R_{\beta} R_{\sigma_1^2})$$

$$(R_{\psi_2^2}, R_{\psi_{12}}) = (R_{\gamma}^2 R_{\psi_1^2} + R_{\psi_{2|1}^2}, R_{\gamma} R_{\psi_1^2})$$

Step 4 : 그러므로 표본공분산행렬의 Σ 와 Ψ 의 GPQ를 각각 $R_{\Sigma} = \begin{bmatrix} R_{\sigma_1^2} & R_{\sigma_{12}} \\ R_{\sigma_{12}} & R_{\sigma_2^2} \end{bmatrix}$

와 $R_{\Psi} = \begin{bmatrix} R_{\psi_1^2} & R_{\psi_{12}} \\ R_{\psi_{12}} & R_{\psi_2^2} \end{bmatrix}$ 로 얻을 수 있으며, 표준이변량정규분포로부터 난수 Z_D 를

생성하여 Δ 의 GPQ를 다음과 같이 계산한다.

$$\begin{aligned} R_{\Delta} &= \begin{bmatrix} R_{\Delta_1} \\ R_{\Delta_2} \end{bmatrix} = \begin{bmatrix} R_{\mu_1 - \nu_1} \\ R_{\mu_2 - \nu_2} \end{bmatrix} \\ &= \mathbf{d} - \left(\frac{R_{\Sigma}}{n} + \frac{R_{\Psi}}{m} \right)^{1/2} Z_D \end{aligned}$$

마찬가지로 R_{Δ} 에 모수가 포함되어 있지 않으므로 δ 의 GPQ를 만족함을 확인할 수 있다. 나아가 표준정규분포의 누적분포함수를 이용하여 θ_1 과 θ_2 , δ 의 GPQ를 얻을 수 있다.

$$R_{\theta_j} = \Phi\left(\frac{R_{\mu_j - \nu_j}}{\sqrt{R_{\sigma_j^2} + R_{\psi_j^2}}}\right) \quad \text{for } j = 1, 2,$$

$$R_{\delta} = R_{\theta_1} - R_{\theta_2}$$

Step 5 : Step 1 ~ 4를 K 번 반복하여 K 개 R_{δ} 의 $100(\alpha/2)\%$ 와 $100(1-\alpha/2)\%$ 분위수를 두 AUC 차이의 $100(1-\alpha)\%$ 신뢰구간의 하한과 상한으로 설정한다.

최종적으로 추정된 $100(1-\alpha)\%$ 신뢰구간에 '0'을 포함하는지 확인하여 $H_0: \delta = 0$ 에 대하여 검정하고, 통계적으로 유의할 때 새로운 지표가 추가됨으로써 진단의 예측력이 향상되었다고 할 수 있다.

2.2.5 Heller's method (2017)

Heller와 그 연구진은 이분형 결과에 대한 모형의 회귀계수를 추정함에 있어, 일반적인 최대가능도(maximum likelihood; ML) 방법이 아닌 최대순위상관(maximum rank correlation; MRC) 방법을 이용한다. MRC 추정량은 Han(1987)이 제안한 추정 방법으로 가능도함수(likelihood function)를 최대화하는 대신 목적함수(objective function)가 되는 AUC를 최대화한다.

$$\operatorname{argmax}_{(\beta, \gamma)} \frac{1}{nm} \sum_i^{n+m} \sum_k^{n+m} I[Y_i > Y_k] I[\eta_i > \eta_k]$$

추정된 회귀계수는 Taylor 전개에서 1차 미분항을 '0'이 되게 하여 AUC 차이에 대한 점근분포(asymptotic distribution)를 이끌어낼 수 있으며, 모의실험에서 AUC의 차이가 음수가 되어 해석이 난해해지는 것을 방지한다.

AUC 차이의 분포를 추정하기 위하여 Taylor 전개를 이용하는데 $I(\cdot)$ 함수의 비연속성으로 인하여 미분이 불가능하다. 이로 인해 $I(\cdot)$ 함수를 표준정규 분포의 누적분포함수에 근사시켜 smoothed empirical AUC인 $\tilde{\theta}_j$ 를 얻는다.

$$I(\eta_{ij} > \eta_{kj}) \approx \Phi\left(\frac{\eta_{ij} - \eta_{kj}}{h_N}\right), \quad j = 1, 2$$

h_N 는 표본 수 $N(=n+m)$ 이 증가함에 따라 '0'으로 수렴하는 bandwidth로, ω^2 을 새로운 지표가 추가된 모형에서 추정된 선형 예측변수의 분산이라고 할 때 $h_N = \hat{\omega}(n+m)^{-0.333}$ 로 구할 수 있다.

(1) AUC의 증가가 발생하지 않을 경우

기준에 알려진 지표 \mathbf{X} 의 회귀계수를 $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_p)$, 새로운 지표 \mathbf{Z} 의 회귀계수를 $\boldsymbol{\gamma}^T = (\gamma_1, \dots, \gamma_q)$ 라고 할 때, 새로운 모형의 AUC와 그 Hessian 행렬(D)은 다음과 같다.

$$\theta_1 = \frac{1}{nm} \sum_{i=1}^N \sum_{k=1}^N I[Y_i > Y_k] \Phi\left[\frac{(\boldsymbol{\beta}^T X_i + \boldsymbol{\gamma}^T Z_i) - (\boldsymbol{\beta}^T X_k + \boldsymbol{\gamma}^T Z_k)}{h_N}\right]$$

$$D = \begin{bmatrix} D_{\beta\beta} & D_{\beta\gamma} \\ D_{\gamma\beta} & D_{\gamma\gamma} \end{bmatrix}, \quad D^{-1} = \begin{bmatrix} D^{\beta\beta} & D^{\beta\gamma} \\ D^{\gamma\beta} & D^{\gamma\gamma} \end{bmatrix}$$

Taylor 전개를 통해 AUC의 차이를 정리하면 $2N[\hat{\theta}_1 - \hat{\theta}_2]$ 가 asymptotic하게 아래와 같이 근사하며, 표본 수가 커질 때 자유도 1을 따르는 독립적인 χ^2 확률변수의 가중 합에 근사한다.

$$2N[\hat{\theta}_1 - \hat{\theta}_2] \approx N(\gamma - \hat{\gamma})^T [-D^{\gamma}]^{-1} (\gamma - \hat{\gamma})$$

$$P(2N[\hat{\theta}_1 - \hat{\theta}_2] \leq u) = P\left(\sum_{j=1}^q \lambda_j \chi_j^2 \leq u\right)$$

여기서 가중치 λ_j 는 MRC 방법으로 추정된 새로운 지표의 회귀계수 $\hat{\gamma}_{MRC}$ 의 asymptotic 분산과 $[D^{\gamma}]^{-1}$ 행렬 곱의 고유값을 나타낸다.

(2) AUC의 증가가 발생하는 경우

$\sqrt{N}[\theta_1 - \theta_2 - \delta]$ 을 자유도가 2인 two-sample U-statistic 형태로 변환하여, AUC 차이에 대하여 asymptotic normality을 통해 분산을 계산할 수 있다.

$$Var(\delta) = \frac{1}{n} \sigma_1^2 + \frac{1}{m} \sigma_2^2,$$

$$\hat{\sigma}_1^2 = \frac{1}{nm(m-1)} \sum_{i=1}^N \sum_{k=1, l=1, l \neq k}^N I(Y_i = 1) I(Y_k = 0) I(Y_l = 0) (e_{ik} - \hat{\delta})(e_{il} - \hat{\delta})$$

$$\hat{\sigma}_2^2 = \frac{1}{nm(n-1)} \sum_{i=1}^N \sum_{k=1, l=1, l \neq i}^N I(Y_i = 1) I(Y_k = 0) I(Y_l = 1) (e_{ik} - \hat{\delta})(e_{lk} - \hat{\delta})$$

여기서 e_{ik} 는 i 번째 대상자와 k 번째 대상자 각각의 새로운 모형에서 추정된 선형 예측변수의 차이에 bandwidth h_N 를 부여하고 표준정규분포의 누적분포 함수에 근사시킨 값과, 마찬가지로 기존 모형에서 추정된 선형 예측변수의 차이에 bandwidth h_N 을 부여하고 표준정규분포의 누적분포함수에 근사시킨 값의 차이를 나타낸다.

하지만 $\hat{\delta}$ 와 asymptotic하게 추정된 분산 $\widehat{Var}(\hat{\delta})$ 사이에 선형 관계가 강하게 발생하여 분산 안정화를 위해 $\hat{\tau} = \sqrt{\hat{\delta}}$ 로 재모수화 한다. Delta's method에 의하여 추정된 $\hat{\tau}$ 의 분산은 $\widehat{Var}(\hat{\tau}) = \frac{\widehat{Var}(\hat{\delta})}{4\hat{\delta}}$ 으로, 다시 AUC의 차이로 돌아와 δ 대한 $100(1-\alpha)\%$ 신뢰구간은 $[\hat{\tau} \pm Z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\tau})}]^2$ 으로 계산된다. (단, 하한의 제곱근이 음수인 경우 하한은 0으로 설정한다.)

제 3장 모의실험

제3장에서는 2.2에서 소개한 다섯 가지 방법에 대하여 다양한 표본 수와 기존에 알려진 지표와 새롭게 발견된 지표 사이의 상관관계를 고려한 모의실험을 수행한다. 새로운 생체지표가 발견되어 기존 지표로만 이루어진 모형에 추가되었을 때 AUC의 차이를 가장 잘 반영하는 통계적 방법론이 무엇인지 제 1종 오류(Type 1 error)와 검정력(power)를 통하여 비교 및 평가한다.

3.1 모의실험 설계

본 논문에서는 모의실험을 위해 표 1의 자료구조를 바탕으로 다음과 같은 가상의 자료를 생성하였다. $(\mathbf{X}_1, \mathbf{Z}_1)^T$ 는 질환군($Y=1$) n 명에 대한 두 변수를 나타내고, $(\mathbf{X}_2, \mathbf{Z}_2)^T$ 는 비질환군($Y=0$) m 명에 대한 두 변수를 나타낸다.

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{Z}_1 \end{bmatrix} \sim N_2\left(\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right)$$

$$\begin{bmatrix} \mathbf{X}_2 \\ \mathbf{Z}_2 \end{bmatrix} \sim N_2\left(\boldsymbol{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}, \boldsymbol{\Psi} = \begin{bmatrix} \psi_1^2 & \psi_{12} \\ \psi_{12} & \psi_2^2 \end{bmatrix}\right)$$

\mathbf{X}_1 과 \mathbf{X}_2 을 결합하여 기존에 알려진 지표 \mathbf{X} 으로, \mathbf{Z}_1 과 \mathbf{Z}_2 를 결합하여 새롭게 추가된 지표 \mathbf{Z} 라 정의한다. 질환군의 평균에 대한 모수 $\mu_1 = 2$, μ_2 는 0과 1.8, 비질환군의 평균에 대한 모수 $(\nu_1, \nu_2) = (0, 0)$ 으로 설정하여 새로운 지표가 모형에 추가되었을 때 실제 AUC의 차이가 없을 경우와 차이가 있을 경우를 고려한다. 모든 대상자의 분산 $\sigma_1^2 = \sigma_2^2 = \psi_1^2 = \psi_2^2 = 1$ 로 고정한다.

3.2 두 지표의 상관관계에 따른 AUC

3.1에서 생성된 자료를 이용하여 구축된 두 로지스틱 회귀모형은 아래와 같으며, AUC의 차이를 계산하기 위하여 질환군과 비질환군 각각에서의 선형 예측변수를 정의할 수 있다.

$$- \text{Full model} : \log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \alpha + \beta\mathbf{X} + \gamma\mathbf{Z} \Rightarrow \begin{cases} Y=1 : \alpha + \beta\mathbf{X}_1 + \gamma\mathbf{Z}_1 \\ Y=0 : \alpha + \beta\mathbf{X}_2 + \gamma\mathbf{Z}_2 \end{cases}$$

$$- \text{Reduced model} : \log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \alpha^0 + \beta^0\mathbf{X} \Rightarrow \begin{cases} Y=1 : \alpha^0 + \beta^0\mathbf{X}_1 \\ Y=0 : \alpha^0 + \beta^0\mathbf{X}_2 \end{cases}$$

$$\therefore \zeta_1 = \begin{bmatrix} \alpha + \beta\mathbf{X}_1 + \gamma\mathbf{Z}_1 \\ \alpha^0 + \beta^0\mathbf{X}_1 \end{bmatrix}, \quad \zeta_2 = \begin{bmatrix} \alpha + \beta\mathbf{X}_2 + \gamma\mathbf{Z}_2 \\ \alpha^0 + \beta^0\mathbf{X}_2 \end{bmatrix}$$

두 모형에서의 질환군의 선형 예측변수 ζ_1 와 비질환군의 선형 예측변수 ζ_2 의 평균 벡터 (μ', ν')와 공분산 행렬 (Σ', Ψ')은 다음과 같다.

$$\mu_1' = \alpha + \beta\mu_1 + \gamma\mu_2, \quad \mu_2' = \alpha^0 + \beta^0\mu_1$$

$$\nu_1' = \alpha + \beta\nu_1 + \gamma\nu_2, \quad \nu_2' = \alpha^0 + \beta^0\nu_1$$

$$\sigma_1^{2'} = V(\alpha + \beta\mathbf{X}_1 + \gamma\mathbf{Z}_1) = \beta^2 V(\mathbf{X}_1) + \gamma^2 V(\mathbf{Z}_1) + 2Cov(\beta\mathbf{X}_1, \gamma\mathbf{Z}_1) = \beta^2\sigma_1^2 + \gamma^2\sigma_2^2 + 2\beta\gamma\sigma_{12}$$

$$\psi_1^{2'} = V(\alpha + \beta\mathbf{X}_2 + \gamma\mathbf{Z}_2) = \beta^2 V(\mathbf{X}_2) + \gamma^2 V(\mathbf{Z}_2) + 2Cov(\beta\mathbf{X}_2, \gamma\mathbf{Z}_2) = \beta^2\psi_1^2 + \gamma^2\psi_2^2 + 2\beta\gamma\psi_{12}$$

$$\sigma_2^{2'} = V(\alpha^0 + \beta^0\mathbf{X}_1) = \beta^{02} V(\mathbf{X}_1) = \beta^{02}\sigma_1^2, \quad \psi_2^{2'} = V(\alpha^0 + \beta^0\mathbf{X}_2) = \beta^{02} V(\mathbf{X}_2) = \beta^{02}\psi_1^2$$

그러므로 새로운 지표가 추가되었을 때와 기존 지표만을 이용했을 때 각각의 AUC는 다음과 같이 정의할 수 있다.

$$\theta_1 = \Phi \left[\frac{\beta(\mu_1 - \nu_1) + \gamma(\mu_2 - \nu_2)}{\sqrt{\beta^2(\sigma_1^2 + \psi_1^2) + \gamma^2(\sigma_2^2 + \psi_2^2) + 2\beta\gamma(\sigma_{12} + \psi_{12})}} \right]$$

$$\theta_2 = \Phi \left[\frac{\beta^0(\mu_1 - \nu_1)}{\sqrt{\beta^{0^2}(\sigma_1^2 + \psi_1^2)}} \right]$$

3.1에서 모든 대상자의 분산은 1로 고정하였고, 생성된 질환군의 기준에 알려진 지표 \mathbf{X} 의 평균 모수가 2이므로 β^0 과 β 는 양수일 것으로 예상 가능하다. 다시 위에서 정의된 θ_1 와 θ_2 은 다음과 같이 정리할 수 있다.

$$\theta_1 = \Phi \left[\frac{\beta(\mu_1 - \nu_1) + \gamma(\mu_2 - \nu_2)}{\sqrt{2\beta^2 + 2\gamma^2 + 2\beta\gamma(\sigma_{12} + \psi_{12})}} \right]$$

$$\theta_2 = \Phi \left[\frac{2\beta^0}{\sqrt{2\beta^{0^2}}} \right] = \Phi[\sqrt{2}] = 0.9214$$

그러므로 모의실험을 통해 질환군과 비질환군에서 각각 생성한 두 변수의 상관관계가 실제 AUC 차이에 어떠한 영향을 미치는지 고찰하여 본다. 아래 모의실험은 각 경우에 대하여 10만 번 반복 시행하여 얻은 회귀계수와 두 모형에서의 AUC, 두 AUC의 차이(δ)의 평균을 나타낸다. 질환군에서 기준에 알려진 지표 \mathbf{X} 의 평균 모수는 2로 고정되어 있으므로, 새롭게 발견된 지표 \mathbf{Z} 의 평균 모수가 0일 때에는 두 AUC의 차이가 없을 것으로, 2일 때에는 두 AUC의 차이가 있을 것으로 예상된다. 질환군과 비질환군은 각각 20명, 100명으로 고정하였다.

(1) AUC의 차이가 없고 독립인 경우 : $(\mu_1, \mu_2, \nu_1, \nu_2, \sigma_{12}, \psi_{12}) = (2, 0, 0, 0, 0, 0)$

$$\theta_1 = \Phi \left[\frac{2\beta}{\sqrt{2(\beta^2 + \gamma^2)}} \right]$$

질환군에서 \mathbf{Z} 의 평균 모수 $\mu_2 = 0$ 이므로 $\gamma \approx 0$ 일 것으로 예상된다.

표 2. AUC의 차이가 없고 상관관계가 없는 경우

σ_{12}	ψ_{12}	α	β	γ	α^0	β^0	θ_1	θ_2	δ
0	0	-3.923	2.2076	0.0007	-3.8306	2.1556	0.9243	0.9213	0.003

(2) AUC의 차이가 없고 상관관계가 있는 경우 : $(\mu_1, \mu_2, \nu_1, \nu_2) = (2, 0, 0, 0)$

$$\theta_1 = \Phi \left[\frac{2\beta}{\sqrt{2\beta^2 + 2\gamma^2 + 2\beta\gamma(\sigma_{12} + \psi_{12})}} \right]$$

표 3. AUC의 차이가 없고 상관관계가 있는 경우

σ_{12}	ψ_{12}	α	β	γ	α^0	β^0	θ_1	θ_2	δ
0.1	0.1	-3.948	2.2315	-0.2225	-3.831	2.156	0.9253	0.9213	0.004
0.1	0.3	-4.0272	2.3377	-0.5436	-3.8311	2.1561	0.9284	0.9213	0.0071
0.1	0.5	-4.1753	2.5659	-0.9464	-3.8313	2.1563	0.9334	0.9213	0.0121
0.1	0.7	-4.4309	3.0399	-1.5667	-3.8305	2.156	0.9404	0.9213	0.0191
0.3	0.1	-4.0221	2.2817	-0.3848	-3.8314	2.1563	0.9283	0.9213	0.007
0.3	0.3	-4.1711	2.4442	-0.7328	-3.8316	2.1564	0.9336	0.9213	0.0123
0.3	0.5	-4.4303	2.7637	-1.1954	-3.8315	2.1564	0.9411	0.9213	0.0198
0.3	0.7	-4.8284	3.3871	-1.9391	-3.8296	2.1554	0.9508	0.9213	0.0295
0.5	0.1	-4.1594	2.3691	-0.5737	-3.8317	2.1565	0.9333	0.9213	0.012
0.5	0.3	-4.4079	2.6072	-0.9695	-3.8321	2.1568	0.941	0.9213	0.0197
0.5	0.5	-4.7951	3.0339	-1.5165	-3.8313	2.1563	0.951	0.9213	0.0297
0.5	0.7	-5.2818	3.7648	-2.3746	-3.8275	2.154	0.9633	0.9212	0.0421
0.7	0.1	-4.4081	2.5198	0.8215	-3.8319	2.1566	0.9406	0.9213	0.0193
0.7	0.3	-4.8252	2.8745	-1.309	-3.832	2.1567	0.9507	0.9213	0.0294
0.7	0.5	-5.5112	3.5169	-2.0356	-3.8306	2.1559	0.9632	0.9213	0.0419
0.7	0.7	-6.8463	4.9235	-3.4427	-3.822	2.1501	0.9774	0.9210	0.0564

(3) AUC의 차이가 있고 독립인 경우 : $(\mu_1, \mu_2, \nu_1, \nu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0, 0, 0, 0)$

$$\theta_1 = \Phi \left[\frac{2\beta + 1.8\gamma}{\sqrt{2(\beta^2 + \gamma^2)}} \right]$$

질환군에서 Var_2 의 평균 모수 $\mu_2 \neq 0$ 이므로 $\beta_2 \approx 0$ 이 아닐 것으로 예상된다.

표 4. AUC의 차이가 있고 상관관계가 없는 경우

σ_{12}	ψ_{12}	α	β	γ	α^0	β^0	θ_1	θ_2	δ
0	0	-6.0225	2.3288	2.0993	-3.8236	2.1511	0.973	0.9211	0.0519

(4) AUC의 차이가 있고 상관관계가 있는 경우 : $(\mu_1, \mu_2, \nu_1, \nu_2) = (2, 1.8, 0, 0)$

$$\theta_1 = \Phi \left[\frac{2\beta + 1.8\gamma}{\sqrt{2\beta^2 + 2\gamma^2 + 2\beta\gamma(\sigma_{12} + \psi_{12})}} \right]$$

표 5. AUC의 차이가 있고 상관관계가 있는 경우

σ_{12}	ψ_{12}	α	β	γ	α^0	β^0	θ_1	θ_2	δ
0.1	0.1	-5.7425	2.1826	1.9227	-3.8266	2.1531	0.967	0.9212	0.0458
0.1	0.3	-5.3362	1.956	1.6592	-3.8295	2.155	0.961	0.9213	0.0397
0.1	0.5	-5.0106	1.7938	1.4441	-3.8306	2.1558	0.955	0.9213	0.0337
0.1	0.7	-4.7492	1.6864	1.2526	-3.8314	2.1564	0.9492	0.9213	0.0279
0.3	0.1	-5.1505	1.9653	1.6957	-3.8283	2.1542	0.9607	0.9213	0.0394
0.3	0.3	-4.9758	1.8388	1.5151	-3.8302	2.1555	0.9548	0.9213	0.0335
0.3	0.5	-4.7181	1.7115	1.3173	-3.8316	2.1565	0.9489	0.9213	0.0276
0.3	0.7	-4.4982	1.6336	1.1272	-3.8323	2.157	0.9433	0.9213	0.022
0.5	0.1	-4.9597	1.9075	1.6095	-3.8292	2.1549	0.9544	0.9213	0.0331
0.5	0.3	-4.6922	1.7505	1.3937	-3.8313	2.1563	0.9485	0.9213	0.0272
0.5	0.5	-4.4736	1.6487	1.2021	-3.8317	2.1566	0.9428	0.9213	0.0215
0.5	0.7	-4.2938	1.6071	1.0051	-3.8321	2.1569	0.9376	0.9213	0.0163
0.7	0.1	-4.682	1.8113	1.4913	-3.83	2.1554	0.9479	0.9213	0.0266
0.7	0.3	-4.4565	1.6808	1.2866	-3.8316	2.1565	0.9422	0.9213	0.0209
0.7	0.5	-4.2726	1.6082	1.0946	-3.8321	2.1569	0.9367	0.9213	0.0154
0.7	0.7	-4.126	1.6132	0.8745	-3.8326	2.1573	0.9318	0.9213	0.0105

위 (1)-(4)의 모의실험을 통해 질환군과 비질환군에서 각각 두 변수의 상관관계가 강하게 존재할수록, 두 AUC의 차이가 없을 것이라고 예상한 경우 ($\mu_1 = 2, \mu_2 = 0$) 오히려 차이가 발생하였으며 두 AUC의 차이가 있을 것이라고 예상한 경우($\mu_1 = 2, \mu_2 = 1.8$) 오히려 차이가 감소하였다. 그러므로 두 AUC의 차이가 없다고 할 수 없는 (2)번을 제외한 상황에 대하여 모의실험을 통해 다양한 표본 수를 설정하고 제2장에서 소개한 5가지 방법들에 대해 비교한다.

3.3 자료의 생성

본 연구의 모의실험에서는 3.1에서 설정한 모수를 기반으로 생성된 가상의 자료를 이용하며, 2.2에서 정리한 방법에 대하여 모두 1,000번 자료를 새롭게 생성하고 내포 관계에 있는 두 모형의 AUC 차이에 대한 검정을 수행한다.

질환군 n 명과 비질환군 m 명에 대한 표본 수는 (10,10), (10,30), (10,50), (10,100), (10,200), (30,30), (30,50), (30,100), (30,200), (50,50), (50,100), (50,200), (100,100), (100,200), (200,200)으로 다양하게 설정한다.

$(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 0, 0, 0)$ 인 경우, 새로운 지표가 추가된다 하더라도 실제 AUC의 증가는 없을 것이므로 귀무가설 $H_0: \delta = 0$ 를 기각할 제 1종 오류 (Type I error)를 계산하고, $(\mu_1, \mu_2) = (2, 1.8)$ 인 경우, 새로운 지표가 추가됨으로써 실제 AUC가 증가하였을 것이므로 귀무가설을 기각할 검정력(power)를 계산한다. 이때 질환군과 비질환군에서의 두 변수의 상관계수 σ_{12} 와 ψ_{12} 는 각각 (0,0), (0.1, 0.1), (0.1, 0.3), (0.1, 0.5), (0.3, 0.1), (0.3, 0.3), (0.3, 0.5), (0.5, 0.1), (0.5, 0.3), (0.5, 0.5)로 한다.

Heller의 방법에 대하여 최대순위상관(MRC) 방법을 통해 회귀계수를 추정하고, 붓스트랩에 기반을 둔 분산을 계산하여 신뢰구간을 얻는다.

3.4 모의실험 결과

모의실험의 결과는 제 1종 오류와 검정력, 95% 신뢰구간 폭의 평균인 기대 길이(expected length)로 평가한다. <표 6>은 질환군에서 새롭게 추가된 지표의 평균 모수 $\mu_2 = 0$ 으로, 변수 간 상관관계는 '0'으로 설정하여 실제 AUC의 차이가 없는 경우 제 1종 오류(%)를 계산한 결과이다. Heller 방법을 제외한 나머지 4가지 방법에서 모두 실제 AUC 차이를 잘 반영하고 있음을 알 수 있으나 제 1종 오류가 대부분 '0'에 가까워 검정력을 통하여 너무 보수적이지 않은지 확인할 필요가 있다.

<표 7>부터 <표 16>까지는 질환군에서 새롭게 추가된 지표의 평균 모수 $\mu_2 = 1.8$ 로, 질환군과 비질환군 각각에서의 변수 간 상관관계는 0, 0.1, 0.3, 0.5로 설정하여, 실제 AUC의 차이가 있는 경우의 검정력(%)을 계산한 결과이다. Hanley와 McNeil의 방법은 고정된 AUC와 상관계수의 평균을 이용하기에 다른 방법들에 비하여 검정력이 많이 낮고 신뢰구간의 폭 또한 넓다. 다음으로 DeLong 방법의 검정력은 최근 소개된 Li나 Heller의 방법에 비하여 항상 낮다. 이 결과는 최근 다양한 연구에서 언급된 바와 같이 내포 관계에 있는 두 모형의 실제 AUC 차이가 있음에도 불구하고 이를 반영하지 못하는, 즉 보수적임을 확인할 수 있다. AUC 차이에 대한 정확한 붓스트랩 분산을 계산한 Bandos의 방법은 일반적인 Monte-Carlo를 이용한 결과에 비해 낮은 검정력을 갖지만 DeLong 방법보다 약간 높거나 거의 비슷하다. (Monte-Carlo 결과는 제시하지 않음.) Heller의 방법은 표본 수와 관계없이 항상 다른 방법들에 비해 높은 검정력을 가지며 신뢰구간의 폭 또한 작다.

마지막으로 동일한 표본 수일 때, 질환군과 비질환군 각각에서 두 변수 간의 상관관계가 강해질수록 검정력이 낮아진다. 이는 3.2에서 고찰한 바와 같이 변수 간의 상관관계가 강해질수록 실제 AUC의 차이가 오히려 감소되기 때문이다. 또한, 표본 수가 커지면 자연스럽게 검정력도 증가함을 확인할 수 있다.

표 6. 실제 AUC 차이가 없는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 0, 0, 0)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질환군	비질환군	Type 1 error (%)	Length	Type 1 error (%)	Length	Type 1 error (%)	Length	Type 1 error (%)	Length	Type 1 error (%)	Length
10	10	0	0.182	0	0.106	0	0.114	0	0.112	7.2	0.082
10	30	0	0.151	0	0.06	0	0.064	0.1	0.066	5.4	0.059
10	50	0	0.145	0	0.048	0	0.051	0.2	0.055	7.8	0.058
10	100	0	0.137	0.1	0.037	0	0.038	0.3	0.048	5.2	0.053
10	200	0	0.136	0.2	0.031	0.2	0.032	0.8	0.043	7.6	0.049
30	30	0	0.085	0	0.031	0	0.034	0	0.026	3.2	0.033
30	50	0	0.081	0	0.024	0	0.026	0.1	0.021	5.2	0.029
30	100	0	0.078	0	0.018	0	0.019	0	0.016	5.6	0.022
30	200	0	0.076	0	0.014	0	0.015	0.3	0.014	9.6	0.019
50	50	0	0.064	0.1	0.019	0	0.02	0	0.015	4	0.022
50	100	0	0.06	0	0.013	0	0.014	0	0.011	7.4	0.016
50	200	0	0.059	0.2	0.01	0.1	0.011	0	0.009	9.4	0.013
100	100	0	0.044	0.1	0.009	0	0.009	0	0.007	5	0.011
100	200	0	0.042	0.1	0.006	0	0.007	0.1	0.005	9.4	0.008
200	200	0	0.031	0	0.004	0	0.005	0	0.004	5.8	0.005

표 7. 실제 AUC 차이가 있으면서 자료의 상관관계가 없는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0, 0)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질환군	비질환군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	1.6	0.255	0.9	0.215	1.1	0.22	14.9	0.248	59.4	0.168
10	30	0.8	0.223	6.3	0.154	6.5	0.155	33.6	0.191	69.6	0.142
10	50	1.1	0.217	11.9	0.14	12.5	0.14	41.4	0.182	77.6	0.134
10	100	1.4	0.21	21.5	0.126	24	0.123	44.1	0.172	87	0.121
10	200	0.8	0.21	29.9	0.121	35	0.116	48.9	0.169	94.6	0.118
30	30	16.5	0.133	36.5	0.108	34.4	0.11	79	0.112	93.8	0.099
30	50	17	0.13	59.8	0.096	57.2	0.097	91.7	0.097	98.2	0.09
30	100	19.2	0.128	79.8	0.085	78.8	0.085	94.2	0.086	100	0.081
30	200	17.5	0.126	89.3	0.078	89.3	0.078	96.8	0.083	100	0.076
50	50	44.4	0.104	82.4	0.082	80.6	0.083	98.5	0.082	99.6	0.078
50	100	46.7	0.1	96.1	0.07	95.7	0.071	99.6	0.069	100	0.068
50	200	51.3	0.099	98.5	0.064	98.5	0.064	99.8	0.063	100	0.061
100	100	91.9	0.073	99.9	0.057	99.9	0.058	100	0.056	100	0.055
100	200	95.7	0.072	100	0.05	100	0.05	100	0.048	100	0.048
200	200	100	0.052	100	0.041	100	0.041	100	0.039	100	0.039

표 8. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.1, 0.1)$

표본 수		Hanley & McNeil		DeLong		Bando		Li		Heller	
질 환 군	비 질 환 군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	1.2	0.249	0.9	0.205	0.9	0.21	9.7	0.237	49	0.163
10	30	0.5	0.22	5.1	0.147	5	0.149	26	0.176	61.6	0.132
10	50	0.7	0.212	8	0.132	8.6	0.132	31.7	0.169	71.2	0.127
10	100	0.6	0.205	17.3	0.118	19	0.115	36.6	0.162	79.8	0.119
10	200	0.6	0.207	24.1	0.114	27.5	0.11	40	0.16	91.6	0.111
30	30	10	0.131	26.8	0.102	25.2	0.104	67.5	0.104	87.4	0.094
30	50	9.6	0.128	48.4	0.091	45.1	0.092	80.4	0.092	94.6	0.085
30	100	9.8	0.126	69.2	0.08	68.1	0.081	88.8	0.081	99	0.077
30	200	8.5	0.124	81.3	0.073	81.5	0.073	91.1	0.078	100	0.071
50	50	29.4	0.102	70.6	0.078	67.8	0.079	92.4	0.077	99	0.074
50	100	29.2	0.098	90.3	0.066	89.3	0.067	98.1	0.065	100	0.064
50	200	32.5	0.098	95.5	0.06	95.4	0.06	99	0.059	100	0.058
100	100	77.8	0.072	99.1	0.054	99.1	0.054	99.9	0.052	100	0.052
100	200	86.4	0.071	100	0.047	100	0.047	100	0.045	100	0.045
200	200	99.9	0.051	100	0.038	100	0.038	100	0.037	100	0.037

표 9. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.1, 0.3)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질 환 군	비 질 환 군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	1.1	0.241	1	0.191	1	0.197	6.2	0.223	40	0.151
10	30	0.5	0.213	3	0.139	3	0.141	16.2	0.171	48.6	0.129
10	50	0.2	0.204	5.8	0.124	5.8	0.124	21.3	0.165	54.6	0.127
10	100	0.5	0.198	10.8	0.111	12.8	0.109	27.7	0.157	67.4	0.112
10	200	0.2	0.201	14.4	0.108	17	0.105	30.3	0.156	81.2	0.105
30	30	5.7	0.128	18.6	0.096	16.9	0.098	50.4	0.097	75.4	0.088
30	50	4.6	0.126	33.8	0.086	31.6	0.087	67	0.084	90.8	0.081
30	100	4.8	0.123	51.1	0.076	49.4	0.077	77.6	0.078	97.4	0.073
30	200	3.9	0.121	66.4	0.07	66.7	0.07	82.3	0.075	99	0.068
50	50	17.2	0.1	55	0.073	53.2	0.074	82.4	0.071	97.2	0.069
50	100	15.4	0.096	78.5	0.062	77	0.063	93.1	0.061	99.8	0.06
50	200	17.8	0.095	89.8	0.057	89.8	0.057	96.6	0.057	100	0.055
100	100	57.8	0.071	97.4	0.05	97.2	0.051	99.4	0.048	100	0.049
100	200	66.8	0.069	99.7	0.044	99.7	0.045	100	0.042	100	0.042
200	200	98.2	0.05	100	0.036	100	0.036	100	0.034	100	0.034

표 10. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.1, 0.5)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질 환 군	비 질 환 군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	1	0.234	1.1	0.182	1.1	0.188	4.6	0.209	35.8	0.152
10	30	0.2	0.205	1.6	0.131	1.7	0.133	12.6	0.162	37.6	0.121
10	50	0.2	0.197	3.4	0.116	3.8	0.116	12.9	0.151	44.6	0.115
10	100	0.4	0.192	6.5	0.105	7.5	0.103	16.6	0.153	52.8	0.11
10	200	0.1	0.194	9	0.103	10	0.1	20.8	0.151	65	0.101
30	30	2.5	0.124	11	0.089	9.7	0.091	37.1	0.09	63.8	0.081
30	50	2.7	0.122	20.8	0.08	19.2	0.082	49.4	0.079	81.8	0.076
30	100	2.5	0.119	33.1	0.072	31	0.072	59.4	0.074	92	0.069
30	200	2	0.118	43.8	0.067	43.8	0.067	69.2	0.072	97.4	0.065
50	50	8.6	0.097	39.9	0.067	36.7	0.068	69	0.066	92.4	0.064
50	100	6.6	0.093	58.1	0.058	55.6	0.059	82.6	0.057	97.4	0.056
50	200	7.1	0.093	75.1	0.054	74.6	0.054	91	0.054	99.4	0.052
100	100	34.1	0.069	90.8	0.046	89.6	0.047	98.3	0.044	99.8	0.045
100	200	38.6	0.067	98.2	0.042	97.8	0.042	99.9	0.04	100	0.04
200	200	86.4	0.049	100	0.033	100	0.033	100	0.031	100	0.032

표 11. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.3, 0.1)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질 환 군	비 질 환 군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	1.2	0.243	1	0.194	1.1	0.199	7	0.221	40.8	0.153
10	30	0.1	0.215	3.8	0.137	3.4	0.139	18.7	0.164	54.2	0.123
10	50	0.1	0.207	7.5	0.121	7.4	0.122	21.3	0.152	63.2	0.12
10	100	0.2	0.201	14.7	0.107	15.9	0.105	30.9	0.146	73	0.107
10	200	0.1	0.202	21.9	0.103	24.5	0.1	33.1	0.143	86	0.1
30	30	5	0.128	18.7	0.096	16.5	0.098	50.7	0.097	73.6	0.087
30	50	4.9	0.126	35.3	0.085	32.9	0.086	65.5	0.084	90.2	0.079
30	100	4.3	0.123	58.2	0.074	56.3	0.074	77.6	0.074	97.8	0.07
30	200	3.5	0.122	71.9	0.067	72	0.067	82.7	0.07	99.8	0.064
50	50	17.4	0.1	55.3	0.073	52.3	0.074	82.9	0.071	96.6	0.07
50	100	14.6	0.096	81.3	0.061	78.9	0.062	93.3	0.059	100	0.059
50	200	16.6	0.096	90.2	0.055	90.1	0.055	96.3	0.054	100	0.053
100	100	57.8	0.071	97.4	0.05	97.2	0.051	99.5	0.048	100	0.049
100	200	66.8	0.069	99.7	0.043	99.6	0.044	100	0.041	100	0.041
200	200	97.5	0.05	100	0.036	100	0.036	100	0.034	100	0.034

표 12. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.3, 0.3)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질 환 군	비 질 환 군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	0.8	0.233	0.8	0.181	0.9	0.187	2.9	0.208	33.2	0.148
10	30	0.1	0.206	2.1	0.128	2.1	0.13	11	0.154	43	0.119
10	50	0.1	0.199	4.4	0.113	4.5	0.114	16.2	0.144	48.4	0.115
10	100	0.2	0.194	9.6	0.1	10.5	0.099	18.6	0.139	60.2	0.101
10	200	0.1	0.196	13.1	0.097	14.7	0.095	21.6	0.137	76.2	0.094
30	30	2.5	0.124	10.4	0.089	9.4	0.091	37.1	0.089	63.2	0.081
30	50	2.1	0.122	23	0.079	21.6	0.081	49.4	0.078	83	0.074
30	100	2.2	0.12	40.2	0.069	38.2	0.07	61.7	0.07	93	0.066
30	200	1	0.118	56.3	0.063	55.7	0.063	70.5	0.067	97.8	0.061
50	50	7.8	0.098	40	0.067	37.8	0.068	67.7	0.066	92.8	0.064
50	100	6.3	0.094	64.3	0.057	61.9	0.058	83.3	0.055	98.4	0.055
50	200	6.2	0.093	80	0.052	79.2	0.052	90.4	0.051	99.8	0.05
100	100	33	0.069	91.8	0.046	90.5	0.047	98.5	0.045	100	0.045
100	200	38.4	0.067	98.6	0.04	98.4	0.041	99.7	0.038	100	0.039
200	200	86.8	0.049	100	0.033	100	0.033	100	0.031	100	0.032

표 13. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.3, 0.5)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질 환 군	비 질 환 군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	0.6	0.225	0.9	0.17	1	0.176	3	0.186	28.6	0.139
10	30	0.1	0.197	0.8	0.118	1	0.121	7.5	0.143	29.6	0.11
10	50	0.1	0.19	2.3	0.105	2.3	0.106	9.8	0.136	36.6	0.105
10	100	0	0.185	5.3	0.093	5.6	0.092	11.7	0.131	46.2	0.099
10	200	0.1	0.188	8.4	0.092	8.7	0.089	13.1	0.131	59.2	0.091
30	30	1.5	0.12	6.1	0.081	5.1	0.083	24.5	0.082	49.6	0.074
30	50	0.7	0.117	12.5	0.073	10.9	0.074	32.3	0.072	67	0.069
30	100	0.7	0.115	23.6	0.064	22.6	0.065	40.8	0.065	81.4	0.062
30	200	0.5	0.113	33.9	0.059	33.6	0.059	51.5	0.063	92.8	0.057
50	50	3	0.094	25.4	0.061	23	0.062	50.2	0.059	81.2	0.058
50	100	2.7	0.091	41.6	0.052	40.4	0.053	65.9	0.051	94.4	0.051
50	200	1.7	0.089	61.4	0.048	60.6	0.048	75.7	0.047	98.6	0.046
100	100	14.4	0.067	76.6	0.042	73.8	0.042	91.8	0.04	98.6	0.041
100	200	17.1	0.065	92.2	0.037	91.8	0.037	98.4	0.035	100	0.036
200	200	61.3	0.047	99.9	0.03	99.9	0.03	100	0.028	100	0.029

표 14. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.5, 0.1)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질환군	비질환군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	0.7	0.235	0.5	0.182	0.7	0.189	6.3	0.209	33	0.148
10	30	0	0.208	2.6	0.125	2.2	0.128	13.4	0.142	46.8	0.115
10	50	0	0.202	5.8	0.109	5.7	0.111	18.2	0.135	55	0.108
10	100	0	0.195	13.7	0.094	13.8	0.094	23	0.126	71.2	0.093
10	200	0	0.197	21.2	0.09	23.7	0.088	26.9	0.123	82	0.086
30	30	2.1	0.124	10.9	0.089	9.3	0.091	38.4	0.09	64.8	0.08
30	50	2.3	0.123	25.8	0.078	22.9	0.079	48.1	0.076	82.6	0.073
30	100	1.4	0.12	48.1	0.066	46.1	0.067	65	0.065	94	0.063
30	200	0.8	0.119	63.2	0.059	62.4	0.059	73.3	0.061	98	0.056
50	50	8	0.097	39.7	0.067	37	0.068	67.4	0.065	90.6	0.064
50	100	6.3	0.094	68.1	0.055	66.4	0.056	84.3	0.053	99	0.053
50	200	6.2	0.093	84.2	0.049	84.1	0.049	91.1	0.047	100	0.046
100	100	33.7	0.069	90.7	0.046	89.5	0.047	98.6	0.045	100	0.045
100	200	38.4	0.067	98.1	0.039	98	0.039	99.7	0.037	100	0.037
200	200	88.1	0.049	100	0.033	100	0.033	100	0.031	100	0.032

표 15. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.5, 0.3)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질 환 군	비 질 환 군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	0.6	0.224	0.5	0.168	0.6	0.175	4.1	0.186	28.8	0.141
10	30	0	0.199	1	0.115	1	0.119	8.1	0.131	37	0.105
10	50	0	0.192	3.6	0.1	3.3	0.102	11.8	0.127	37.8	0.097
10	100	0	0.187	8.2	0.087	8.5	0.087	15.5	0.121	53.8	0.089
10	200	0	0.19	12.1	0.085	13.2	0.083	16.3	0.117	70.6	0.082
30	30	1.1	0.12	5.3	0.081	4.8	0.084	24.9	0.081	47.8	0.072
30	50	0.7	0.118	14.4	0.071	11.7	0.073	33.9	0.069	68.2	0.067
30	100	0.6	0.117	30.8	0.061	28.7	0.062	45	0.061	84	0.058
30	200	0.3	0.115	45.4	0.055	45.1	0.055	56.5	0.058	93.8	0.052
50	50	3	0.094	25.4	0.061	22.8	0.062	50	0.059	80.8	0.058
50	100	2.6	0.091	48.2	0.051	45.4	0.052	67.5	0.049	95.6	0.049
50	200	1.5	0.09	68.2	0.045	67.4	0.046	77.7	0.044	98.4	0.043
100	100	14.9	0.067	76	0.042	73.9	0.042	92.7	0.04	98.8	0.041
100	200	16.4	0.065	93.6	0.036	92.9	0.036	98.2	0.034	100	0.034
200	200	60.9	0.047	99.4	0.03	99.4	0.03	100	0.028	100	0.029

표 16. 실제 AUC 차이가 있으면서 자료의 상관관계가 있는 경우 : $(\mu_1, \mu_2, \sigma_{12}, \psi_{12}) = (2, 1.8, 0.5, 0.5)$

표본 수		Hanley & McNeil		DeLong		Bandos		Li		Heller	
질 환 군	비 질 환 군	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length	Power (%)	Length
10	10	0.3	0.216	0.5	0.155	0.5	0.162	1.6	0.175	19.4	0.122
10	30	0	0.19	0.5	0.105	0.6	0.109	5.8	0.122	27	0.102
10	50	0.1	0.183	1.1	0.092	1.4	0.093	7.3	0.116	25.6	0.095
10	100	0	0.177	4.3	0.08	4.2	0.08	7.8	0.113	38.6	0.086
10	200	0	0.18	6.8	0.079	7.5	0.077	10.1	0.11	50.8	0.078
30	30	0.4	0.115	3.1	0.072	2.5	0.075	13	0.072	32.4	0.065
30	50	0.4	0.113	7.1	0.064	5.7	0.066	18.4	0.062	51.8	0.061
30	100	0.2	0.111	16.5	0.055	15.3	0.056	26.2	0.055	68.6	0.053
30	200	0.1	0.109	24.8	0.05	24.4	0.051	35	0.053	81.6	0.048
50	50	0.8	0.089	14.1	0.054	12.2	0.056	30.7	0.052	62.4	0.052
50	100	0.8	0.086	25.1	0.046	23.6	0.046	44.3	0.044	85.2	0.044
50	200	0.4	0.085	43.5	0.041	42.1	0.041	53.2	0.04	94.6	0.039
100	100	4.5	0.064	54.2	0.037	52.1	0.038	74.7	0.035	95	0.036
100	200	3.6	0.062	78.8	0.032	77	0.032	90.2	0.03	99.4	0.031
200	200	24.3	0.045	96.1	0.026	95.7	0.026	99.1	0.025	100	0.025

제 4장 실제 자료 적용

4.1 자료 설명

자료는 2008년 5월부터 2008년 12월까지 총 8개월 동안 신촌 세브란스 병원 영상의학과에서 수집되었다. 자료의 목적은 초음파 검사 결과를 통하여 갑상선 결절(thyroid nodule)을 구분하고 악성 위험률을 예측하기 위함이다.

참고 문헌의 주제에 부합하지 않지만, 본 연구의 목적을 위하여 갑상선 결절을 종속변수로, 연속형 변수인 연령과 병변(lesion)의 크기를 독립변수로 설정한다. 총 병변은 3,084개로 악성(malignant)은 839개, 양성(benign)은 2,245개이다. 이 중에서 악성의 비율(27%)을 유지하여 무작위로 800개의 병변만을 선택한다.

병변의 크기를 기존에 알려진 지표로, 연령을 새롭게 추가된 지표로 간주하여 내포 관계인 두 모형을 구축하고 AUC의 차이가 존재하는지 2장에서 소개한 5가지 방법을 통해 비교 및 평가한다.

4.2 자료 분석 결과

로지스틱 회귀분석 결과, 새롭게 추가된 지표로 간주한 연령의 추정된 회귀계수는 -0.0252 ($SE=0.0073$, $p\text{-value}<0.001$)로 유의수준 $\alpha=0.05$ 하에서 통계적으로 유의하다. 그러므로 실제 AUC가 증가하였을 것이라고 기대할 수 있다.

2장에서 소개한 5가지 방법을 통해 계산한 두 AUC 차이의 95% 신뢰구간과 폭은 다음과 같다.

표 17. 실제 자료 분석 결과

95% CI	Hanley	DeLong	Bandos	Li	Heller
LCI	-0.0282	-0.0135	-0.0135	0.0006	0.0018
UCI	0.0434	0.0286	0.0286	0.0274	0.0274
Length	0.0716	0.0421	0.0421	0.0268	0.0256

Wald 검정 결과 추가된 변수가 통계적으로 유의함에도 Hanley와 McNeil의 방법과 DeLong 방법은 AUC의 차이가 없다는 결과를 낸다. 정확한 붓스트랩 분산을 계산한 Bandos 방법은 DeLong 방법과 동일한 결과를 나타내며, 반면 Li의 방법과 Heller의 방법은 두 AUC에 차이가 있다는 결과를 낸다. 제3장의 모의실험을 통해 확인한 바와 같이 통계적으로 유의한 변수가 추가되었음에도 DeLong의 방법이 최근 소개된 방법들에 비해 보수적임을 알 수 있다.

제 5장 결론 및 고찰

이미 질병 발생에 대한 예측 모형 구축은 활발하게 진행되어 왔으며, 기존에 알려진 많은 임상지표와 생체지표가 확립되어 있다. 최근 들어 유전체 연구를 통해 새롭게 발견되는 생체지표나 기술의 발달로 새롭게 이루어지는 수술이나 약물 요법 등 기존에 확립된 지표 이외에도 활용 가능한 지표는 다양하다. 조금이라도 질병 발생을 예측할 수 있는 변수라면, 발견되는 즉시 모형에 추가하는 것이 현명하다. 하지만 이런 통계학적 측면과 대비되게 임상적 측면도 고려해야 한다. 환자에게 위해가 가해지거나 검사 절차가 복잡하고 비싸다면 아주 뛰어난 예측 변수가 아닌 이상 필요시에만 활용하는 편이 더 이성적일 것이다.

새롭게 발견한 지표를 추가하여 모형을 구축한 결과, Wald 혹은 가능도비 검정을 통해 종속변수에 크게 영향을 미치는 변수를 발견하였음에도 불구하고 계산된 AUC는 기존 모형에 비해 크게 증가하지 않는 결과는 빈번하다. 또한, 최근 다양한 연구에서 내포 관계에 있는 두 모형의 AUC 차이를 검정하기 위해 가장 널리 사용됐던 DeLong의 방법이 너무 보수적이라고 보고되고 있다.

질환군과 비질환군의 평균 모수가 같다면, 즉 같은 분포를 이루고 있다면 구축한 모형의 성능은 뛰어나지 않을 것이며, 반면 평균 모수에 차이가 있다면 모형은 질병 여부를 잘 구분할 것이다. 마찬가지로 새롭게 추가된 지표에서 질환군과 비질환군의 평균 모수가 동일할 때 기존 모형과 비교하여 AUC 차이는 없을 것이며, 평균 모수에 차이가 있을 때 AUC는 증가할 것이다. 하지만 3.2를 통해 각 군에서 두 변수 사이의 상관관계에 따라 새로운 모형에서의 성능이 달라짐을 확인하였다. 상관관계가 강하게 존재할수록 평균 모수가 동일할 때, 즉 AUC의 차이가 없을 것이라 예상한 경우 오히려 새로운 모형에서의 AUC가 증가하였으며, 평균 모수가 다를 때 AUC가 예상만큼 증가하지 않았다. 3장의 모의실험을 통하여 다양한 표본 수에서 여러 상관관계를 통해 동일한 결과를 확인할 수 있었다. 그러므로 새로운 지표가 발견되었을 때 단순히 위치 모수뿐만 아니라 기존에 확립된 지표들과의 상관관계도 모형의 성능에 영향을 미친다는 점을 인지해야 한다.

본 연구에서는 내포 관계에 있는 두 모형의 AUC 차이를 검정하기 위하여 질환군과 비질환군의 상관관계를 이용한 Hanley와 McNeil의 방법, 비모수적 접근을 활용한 DeLong의 방법, 정확한 붓스트랩 분산 계산에 기반을 둔 Bandos의 방법, GPQ 개념을 이용한 Li, 그리고 MRC 추정량을 통한 Heller의 방법을 비교하였다.

먼저 Hanley와 McNeil의 방법은 다른 방법들에 비해 분산이 크며, 두 AUC의 평균과 변수 간 상관관계를 통하여 고정된 상관계수를 사용한다는 제한점이 있다. 또한, 적어도 두 모형의 AUC 평균이 0.70 이상이 되어야 하기에 현실적으로 적용하기에 어려움이 존재한다. 두 번째로 DeLong의 방법은 최근

소개된 Li나 Heller의 방법에 비하여 항상 낮은 검정력을 가진다. 모의실험과 실제 자료를 통하여, 실제 모형의 성능을 증가시킬 수 있는 변수가 추가되었음에도 이를 반영하지 못한다. 세 번째로 정확한 붓스트랩 분산을 계산하는 Bandos의 방법 또한 내포 관계에 있는 두 모형의 AUC 차이를 검정하기에는 부적절하다. 본 논문에 제시하지 않았지만, 오히려 Monte-Carlo 알고리즘에 기반을 둔 붓스트랩 방법으로 얻은 검정력이 더 높다.

가장 최근에 제안된 Heller의 방법은 모형의 회귀계수를 추정함에 있어 일반적인 최대가능도(ML) 방법이 아닌 최대순위상관(MRC) 방법을 사용하여 두 모형에서의 AUC 차이를 asymptotic 분포로 근사시킨다. 그러나 논문에서 제시된 분산 추정 수식에 다소 오류가 있어 보이며 시뮬레이션에서 동일한 결과를 얻지 못하였다. 이로 인해 시뮬레이션에서는 Heller 방법의 핵심인 MRC 추정량을 이용하되 붓스트랩을 통하여 AUC 차이에 대해 검정하였다. 최종적으로 2007년에 새롭게 제안된 Li의 방법이 AUC의 차이를 가장 잘 반영하며 전반적으로 신뢰구간의 간격도 좁아, 비교한 5가지 방법 중에서 가장 좋은 검정 방법임을 확인할 수 있다.

참고문헌

J. A. Hanley, B. J. McNeil. (1983). "A method of comparing the areas under receiver operating characteristic curves derived from the same cases". *Radiology*, 148, 839-843.

E. R. DeLong. (1988). "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach". *Biometrics*, 44, 837-845.

A. I. Bandos. (2007). "Exact bootstrap variances of the area under ROC curve". *Communications in Statistics-Theory and Methods*, 36, 2443-2461.

C. R. Li. (2008). "On the exact interval estimation for the difference in paired areas under the ROC curves". *Statistics in Medicine*, 27, 224-242.

G. Heller (2017). "Inference for the difference in the area under the ROC curve derived from nested binary regression models". *Biostatistics*, 18, 260-274

L. J. Wei, W. E. Johnson (1985). "Combining dependent tests with incomplete repeated measurements". *Biometrika*, 72, 359-364.

A. K. Han (1987). "Non-parametric analysis of a generalized regression model". *Journal of Econometrics*, 35, 303-316.

M. S. Pepe, Tianxi Cai, Gary Longton (2006). "Combining predictors for classification using the area under the receiver operating characteristic curve". *Biometrics*, 62, 221-229.

J. A. Hanley, B. J. McNeil (1982). "The meaning and use of the area under a receiver operating characteristic(ROC) curve". *Radiology*, 143, 29-36.

A. J. Vickers, A. M. Cronin, C. B. Begg (2011). "One statistical test is sufficient for assessing new predictive markers". *BMC Medical Research Methodology*, 11, 13.

O. V. Demler, M. J. Pencina, R. D'Agostino (2011). "Equivalence of improvement in area under ROC curve and linear discriminant analysis coefficient under assumption of normality". *Statistics in Medicine*, 30, 1410-1418.

O. V. Demler, M. J. Pencina, R. D'Agostino (2012). "Misuse of DeLong test to compare AUCs for nested models". *Statistics in Medicine*, 31, 2577-2587.

JY Kwak (2011). "Thyroid imaging reporting and data system for US features of nodules". *Radiology*, 260, 892-899.

ABSTRACT

Comparison of methods for testing the difference of two areas under the ROC curves derived from two nested models

Lee, Ho Jae
Dept. of Biostatistics and Computing
The Graduate School
Yonsei University

While most clinical features are well-established, there are various variables that can be utilized, such as new bio-markers and surgeries newly discovered. It is important with both statistical and clinical aspects to assess performance for nested models. The area under the receiver operating characteristic curve(AUC) is a widely used measure of model discrimination in a binary regression model.

Although novel markers are statistically significantly associated with the outcome, only trivial improvement in the difference in AUCs derived from two models is observed frequently. Recent researches have reported that DeLong's method is too conservative to compare two nested regression models. In this study, we assess the performance among conventional Hanley and McNeil's method(1983), DeLong's method(1988), Bandos' method(2007) calculating exact bootstrap variance, Li's method(2008) based on the generalized pivotal quantity(GPQ) theory, and Heller's method(2017)

computed using maximum rank correlation(MRC) methodology.

Hanley and McNeil's method is difficult to apply in practice, because it is computed by fixed correlation coefficients between two models and require the average of two AUCs be at least 0.70. DeLong's method fails to reflect the true AUC increment even if added variables produce small p-values or high odds ratios. Bandos' method is similar to DeLong's, and is not proper to test for nested models. Not included in this study, but the general bootstrap method based on Monte-Carlo algorithm shows better performance. Finally Li's method is the most appropriate test among above the five methods compared, because it reflects the difference in AUCs best and overall confidence intervals are narrow.

Keywords : Nested models, Difference in AUCs, generalized pivotal quantity, Maximum rank correlation, Bootstrap