Minitab

4 귀하를 개선하는 새로운 방법 단일 모집단의 추정 Minitab 의 비율

기초 통계학의 일반적인 문제는 모집단에서 특정 관심 특성을 가진 개인의 비율을 추정하는 것입니다. 예를 들어, 품질 엔지니어는 특정 날짜에 대량 생산된 단위의 대량 배치에서 결점의 비율을 추정할 수 있습니다. 의학 과학자는 특정 병원체에 대한 예방 접종을 받았지만 그럼에도 불구하고 관련 질병을 경험한 일부 지역 사회의 개인의 비율을 조사하기를 원할 수 있습니다. 캠페인 관리자는 등록된 유권자 중 자신의 후보에게 투표할 의향이 있는 유권자의 비율에 관심이 있을 수 있습니다.

이 문제에 대해 가장 잘 알려진 구간 추정 방법은 Wald 신뢰 구간(CI)과 ClopperPearson 정확(1934) CI 라고 하는 교과서적인 정규 근사 방법입니다. 한편, Wald CI 는 CI 의 실제 신뢰 수준(또는 커버리지 확률)이 목표 명목 수준보다 훨씬 낮으며, 특히 실제 비율이 0 또는 1 에 가까울 때 매우 자유롭습니다(그림 1 참조). 반면, 정확한 Clopper-Pearson CI 는 CI 의 실제 신뢰 수준(또는 커버리지 확률)이 목표 명목 수준보다 훨씬 높다는 점에서 지나치게 보수적입니다. 이 두 가지 방법 모두 더 이상 실제 적용에 사용되어서는 안 됩니다(Agresti-Coull, 1998; Brown et al., 2001)을 참조하십시오.

그러나 최근 몇 년 동안 그들은 더 나은 중간 커버리지 확률을 가진 더 나은 CI 방법을 개발하는 데 중요한 역할을 했습니다. 예를 들어, Agresti-Coull 근사 CI 는 Wald CI 에 대한 조정입니다. Blaker (2000, 2001) 정확한 CI 는 반복 수치 알고리즘에서 Clopper-Pearson 신뢰 한계를 시작 추정치로 사용합니다. Minitab 은 새롭게 개선된 이러한 방법을 염두에 두고 단일 모집단 비율을 추정하기 위한 통계 도구를 업데이트하여 수정된 Blaker CI 및 검정 방법, Wilson/점수 CI 및 검정 방법(연속성 보정 포함 및 제외), Agresti-Coull CI 및 검정 방법의 4 가지 방법을 포함했습니다. 또한 Minitab 에서는 이러한 각 방법에 대해 CI 와 검정이 일관된 결과를 산출하도록 합니다.

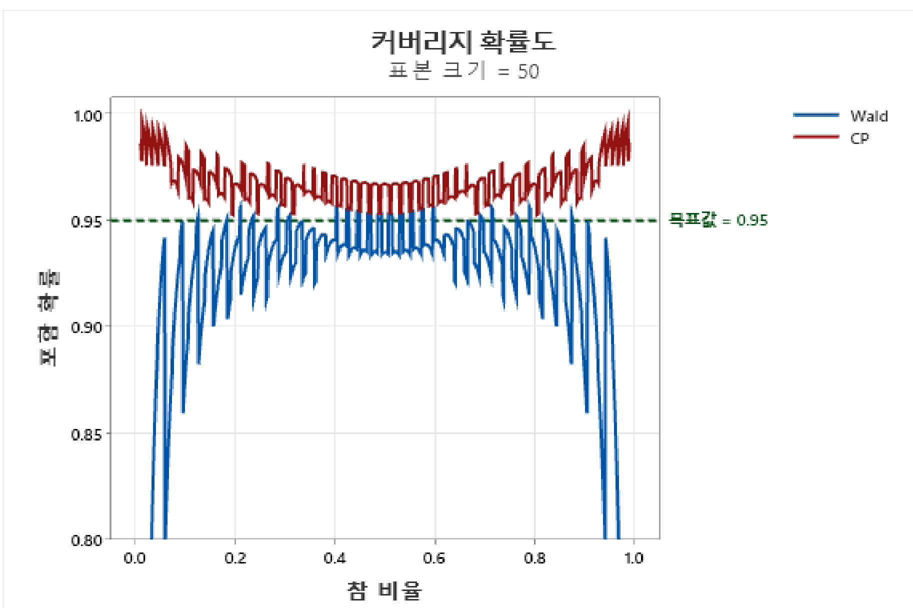

그림 1 : 표본 크기가 50 일 때 Wald CI 및 Clopper-Pearson(CP) CI 에 대한 커버리지 확률을 실제 비율의 함수로 비교합니다. 이 그래프는 Wald CI 와 Clopper Pearson CI 가 각각 과도하게 진보적이고 보수적이며, 특히 실제 비율이 0 또는 1 에 가까울 때 더욱 그러하다는 것을 보여줍니다. 실제 비율이 구간 (0, 1)에 균일하게 분포되어 있다고 가정하면 크기가 50 인 표본을 기반으로 하는 평균 커버리지 확률은 Wald CI 및 Clopper-Pearson CI 에 대해 각각 0.901 과 0.969 입니다.

4 가지 새로운 추정 방법을 만나보세요

4 개의 새로운 방법은 1 개의 정확한 CI 및 수정된 Blaker 방법이라고 하는 테스트 방법과 Wilson/score(Wilson) 방법, Yates 의 연속성 보정을 사용한 Wilson/score(Wilson CC) 방법 및 Agresti-Coull(AC) 방법을 포함한 3 개의 근사 CI 및 테스트 방법으로 구성됩니다. 이 문맥에서 정확한 방법은 일부 형태의 일반 근사 절차를 사용하여 얻은 근사 방법과 달리 방법의 유도에 근사가 사용되지 않음을 의미합니다.

1. 조정된 블레이커 방법

Klaschka 와 Reiczigel (2021)에 의해 조정된 Blaker 방법은 Blaker (2000, 2001)의 정확한 CI 및 테스트 방법을 수정한 것입니다. 이 수정은 원래 Blaker 알고리즘의 계산 집약적 특성과 CI 와 테스트 결과 간의 간헐적인 불일치를 해결합니다. 원래 Blaker CI 와 마찬가지로 결과 조정된 CI 는 정확하고 중첩되어 있으며 Clopper-Pearson CI 에 포함되어 있습니다. 그 결과, 수정된 Blaker CI 는 Clopper-Pearson CI 보다 덜 보수적입니다. 신뢰 수준이 높은 CI 에는 항상 신뢰 수준이 낮은 CI 가 포함된다는 점에서 CI 가 중첩됩니다. 예를 들어, 양측 95%(수정) Blaker CI 에는 항상 대응하는 양측 90% CI 가 포함됩니다. 중첩성은 이항과 같은 이산 분포에서 파생된 정확한 CI 방법의 매력적인 속성입니다. 예를 들어, Clopper-Pearson CI 는 중첩되어 있습니다. 그러나 반드시 중첩되지 않을 수 있는 정확한 CI 메서드가 있습니다. 예를 들어, 소위 Blyth-Still-Casella CI (Blyth and Still, 1983; Casella, 1986)는 가장 짧은 정확한 CI 가 보장되지만 중첩되지는 않습니다. Crow (1956) CI 도 중첩되지 않습니다. Blaker 또는 Adjusted Blaker 방법을 기반으로 하는 CI 계산은 수치 알고리즘이 필요하기 때문에 앞서 언급한 기존 CI 방법보다 더 복잡합니다. 그러나 현재 컴퓨팅 기술의 혁신으로 인해 더 나은 결과를 산출하는 복잡한 알고리즘을 구현하는 것을 더 이상 주저해서는 안 됩니다. 그림 2 는 Clopper-Pearson CI 에 비해 조정된 Blaker CI 의 개선 사항을 보여줍니다.

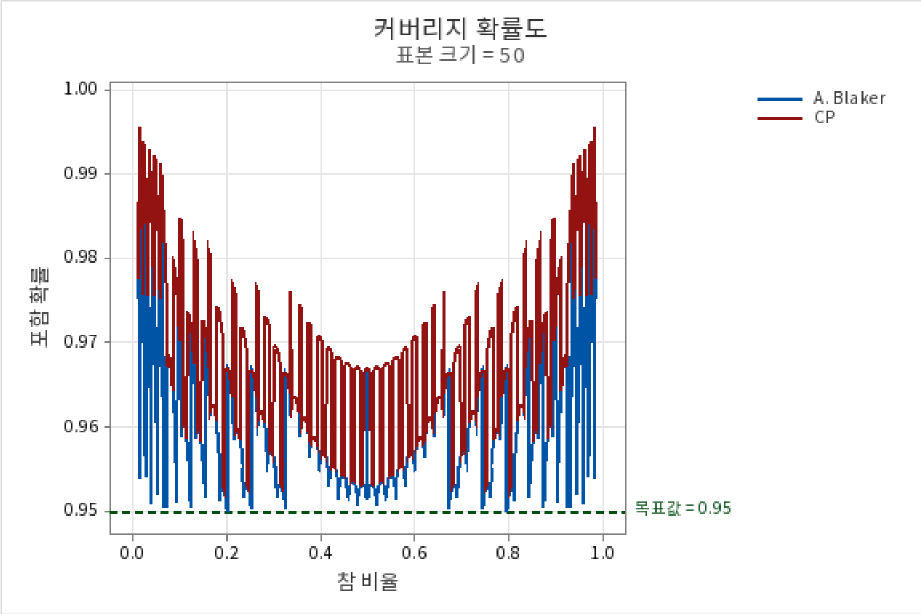

그림 2 : 표본 크기가 50 일 때 실제 비율의 함수로 수정된 Blaker(A. Blaker) CI 와 Clopper-Pearson(CP) CI 에 대한 커버리지 확률을 비교합니다. 그래프는 Clopper-Pearson CI 의 커버리지 확률이 적어도 수정된 Blaker CI 의 커버리지 확률임을 나타냅니다. 이는 조정된 Blaker CI 가 Clopper-Pearson CI 에 포함되어 있다는 사실과 일치합니다. 크기가 50 인 주어진 표본에 대해 평균 커버리지

확률(실제 비율이 단위 구간에 균일하게 분포되어 있다고 가정)은 수정된 Blaker CI 및 Clopper-Pearson CI 에 대해 각각 0.960 과 0.969 입니다.

2. Wilson 및 Wilson CC 방법

Wilson (1927) CI 방법은 검정 통계량의 분모에 대한 고전적인 표준 오차 $\sqrt{\hat{p}(1-\hat{p})/n}$과 달리 null 표준 오차 $\sqrt{p_0 (1-p_0)/n}$을 사용하는 검정인 점수 검정에 해당하는 CI 로 파생됩니다. 이러한 이유로 Wilson/score CI 라고도 합니다. 실제 커버리지 확률은 목표 명목 수준보다 작거나 클 수 있지만 실제 비율이 0 또는 1 에 가까운 경우를 제외하고는 가깝게 유지됩니다(그림 3 참조). Yates 의 연속성 수정을 사용하여 Wilson/score CI 를 보수적으로 조정할 수 있습니다. Minitab 에서는 CI 버전(Yates 의 연속성 수정 포함 및 제외)과 일치 가설 검정을 모두 제공합니다.

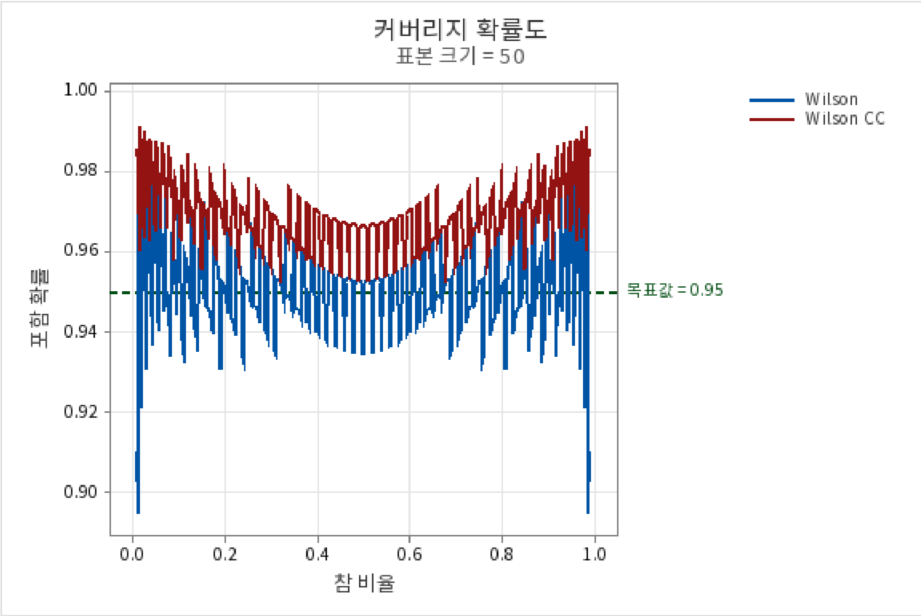

그림 3 : 표본 크기가 50 일 때 Wilson/점수(Wilson) CI 와 Yates 의 연속성 보정(Wilson CC)을 사용한 Wilson/점수에 대한 커버리지 확률을 실제 비율의 함수로 비교합니다. 그래프는 실제 비율의 크기에 따라 Wilson CC CI 가 항상 보수적인 반면 Wilson CI 는 보수적이고 진보적이라는 것을 보여줍니다. 특히, Wilson CI 는 실제 비율이 0 또는 1 에 매우 가까울 때 너무 자유로워지는 경향이 있습니다. 크기가 50 인 주어진 표본의 경우 Wilson CI 및 Wilson CC CI 에 대한 평균 커버리지 확률은 각각 0.952 와 0.969 입니다.

3. Agresti-Coull 방법

Agresti-Coull CI 는 지나치게 자유로운 고전적 Wald CI 의 조정에서 얻어집니다. 결과 CI 는 Wilson CI 와 유사한 커버리지 속성을 갖지만 일반적으로 조금 더 보수적입니다. 또한 두 가지 유형의 CI 는 동일한 중간점을 갖지만 Wilson CI 는 항상 Agresti-Coull CI 에 포함됩니다. 그림 4 에서 볼 수 있듯이 실제 비율이 보통일 때 기본적으로 동일한 커버리지 확률을 갖습니다. 그러나 Agresti Coull CI 는 일반적으로 실제 비율이 0 또는 1 에 가까울 때 덜 자유롭습니다. 그림 4 에서 볼 수 있듯이 크기가 50 인 표본의 경우 실제 비율이 0 또는 1 에 가까워질수록 Agresti-Coull CI 가 보수적으로 변합니다. Agresti-Coull CI 의 또 다른 매력은 Wald CI 에서 물려받은 구현의 단순성입니다. 또한 특히 신뢰 수준이 95%일 때 가르치고 기억하기 쉽습니다. 이 신뢰 수준의 경우 일반적으로 Wald CI 를 도출하기 위해 조정한 것에 대한 각서로 "2 개의 성공과 2 개의 실패를 더하는" CI 방법이라고 합니다.

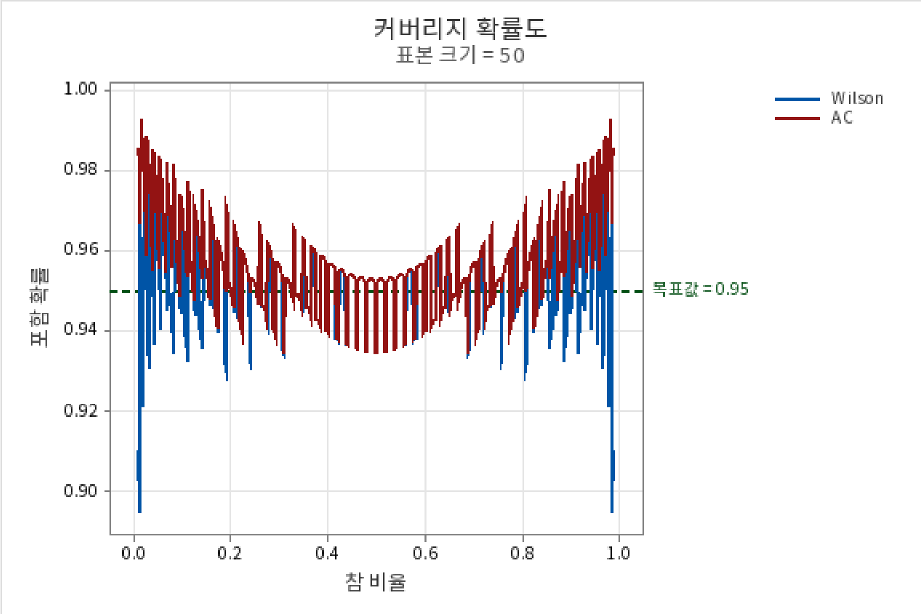

그림 4 : 표본 크기가 50 일 때 Wilson/score(Wilson) CI 및 Agresti-Coull(AC) CI 에 대한 커버리지 확률을 실제 비율의 함수로 비교합니다. 그래프는 두 방법이 구간(0.3, 0.75)에서 실제 비율의 중간 값에 대해 본질적으로 동일한 커버리지 확률을 산출한다는 것을 나타냅니다. 그러나 실제 비율이 0 또는 1 에 가까운 값의 경우 Wilson CI 는 진보적이지만 Agresti-Coull CI 는 보수적입니다. 이는 Wilson CI 가 Agresti-Coull CI 에 포함되어 있다는 사실과 일치합니다. 크기가 50 인 주어진 표본에 대해 Wilson CI 및 Agresti-Coull CI 에 대한 평균 커버리지 확률은 각각 0.952 와 0.958 입니다.

몇 가지 간단한 예시

대량 생산 공장의 품질 엔지니어가 특정 날짜에 1465 개의 대량 생산 제품 중 무작위 표본을 선택했습니다. 1465 개 품목에 대한 독립적인 테스트 결과, 53 개 품목에 결함이 있는 것으로 밝혀졌습니다. 엔지니어는 특정 날짜에 생산된 불량품의 비율이 2.5%와 유의하게 다른지 알고 싶어 합니다.

이 문제는 통계적 가설 검정 문제로 다루어지지만, 응용 통계학자들은 검정의 p-값과 함께 점 추정치와 신뢰 구간을 분석 결과에 포함시키는 것이 점점 더 권장되고 있습니다. Minitab 은 특히 기초 통계 모듈에서 가능한 한 이 규칙을 따릅니다. 예를 들어, Minitab 을 사용하면 수정된 Blaker 방법을 기반으로 한 위 질문에 대한 분석 결과는 다음과 같습니다.

단일 비율에 대한 검정 및 CI

방법

p 사건 비율

Method Adjusted Blaker 의 정확한 방법

기술 통계량

n 사건 표본 p p 에 대한 95% CI

1465 53 0.036177 (0.027353 , 0.046822)

검정

귀무 가설: p= 0.025

대립가설 H_1: p ≠ 0.025

P-값

0.009

4 가지 방법 각각에 대해 유사한 출력을 생성할 수 있습니다. 다음 표에 각 방법의 특징과 장점이 요약되어 있습니다.

방법	95%	일치 검정의 p-값
블레이커를 조정했습니다.	(2.74%,4.68%)	0.009
윌슨과 예이츠의 정정	(2.75%,4.74%)	0.008
윌슨/스코어	(2.78%,4.70%)	0.006
아그레스티-쿨	(2.77%,4.74%)	0.007

이 예에서는 모든 p-값이 0.05 보다 작기 때문에 0.05 의 유의 수준에서 불량률(%)이 2.5%와 다르다는 동일한 결론을 얻을 수 있습니다. 모든 방법에 대한 신뢰 구간과 해당 p-값은 부분적으로 비슷한데, 그 이유는 표본의 크기가 매우 크기 때문입니다. 또한 각 방법에 대한 CI 는 귀무 가설 검정의 p-값과 일치하는 귀무 가설 비율 값(2.5%)을 포함하지 않습니다.

위의 예에서 품질 엔지니어가 50 개의 항목만 테스트하고 2 개의 항목에 결함이 있음을 발견했다고 가정합니다. 또한 엔지니어가 불량품의 비율이 1.0%와 유의하게 다른지 알고 싶다고 가정합니다. 다음 표에 각 방법의 특징과 장점이 요약되어 있습니다.

방법	95%	해당 검정의 p-값
블레이커를 조정했습니다.	(0.72%,13.35%)	0.089
윌슨과 예이츠의 정정	(0.70%,14.86%)	0.155
윌슨/스코어	(1.10%,13.46%)	0.033
아그레스티-쿨	(0.34%,14.22%)	0.124

이 경우 Wilson/score 방법만 0.05 의 유의 수준에서 불량품 %가 1.0%와 다르다는 유의한 결론을 산출합니다. 동일한 유의 수준에서 다른 모든 방법은 차이가 있다고 판단하기에 증거가 불충분하다는 반대 결론을 내립니다. 여러 방법 간 결과의 불일치는 대부분 표본의 크기가 적당하기 때문입니다. 평균적으로 이러한 방법의 커버리지 확률은 표본 크기가 증가함에 따라 명목 수준에 근접합니다(아래 그림 5 참조). 그러나 작거나 중간 정도의 표본 설계의 경우 각 방법과 관련된 평균 커버리지 확률의 차이가 더 두드러져 해당 CI 의 폭이 눈에 띄게 달라집니다. 그러나 중요한 질문은 어떤 결과를 상사에게 보고할 것인가 하는 것입니다. 이 질문에 대한 직접적인 답은 실제 비율의 크기 또는 응용 분야에 대한 사전 지식과 같은 요인에 따라 달라질 수 있기 때문에 더 나은 답은 없습니다. 다음 섹션에서는 몇 가지 대략적인 일반 지침을 제공합니다.

결론

그림 2, 3 및 4 는 4 개의 CI 방법인 Adjusted Blaker, Wilson, Wilson CC 및 Agresti-Coull 이 서로 다른 커버리지 확률 속성을 갖는다는 것을 보여줍니다. Wilson CC 가 가장 보수적이며 조정된 Blaker 가 그 뒤를 잇습니다. Agresti Coull 과 Wilson 은 실제 비율의 크기에 따라 종종 진보적이고 보수적입니다. 전반적으로 Wilson CI 방법은 4 가지 방법 중 가장 자유롭습니다. 또한 그림 5 는 평균적으로 4 가지 방법 모두 보수적이며 Wilson CC 가 가장 보수적이며 조정된 Blaker, Agresti-Coull 및 Wilson 방법이 그 뒤를 잇는다는 것을 나타냅니다. 주어진 문제에 대해 선택할 적절한 방법은 특정 응용 프로그램, 표본 크기 및 실제 비율의 크기에 대한 사전 지식을 사용할 수 있는지 여부에 따라 달라질 수 있습니다.

예를 들어, 규제 기관은 종종 소비자를 보호하기 위해 보수적인 방법에 관심이 있습니다. 그러나 너무 보수적인 방법은 엄격한 규제를 낳을 수 있고 너무 자유로운 방법은 느슨한 규제를 낳을 수 있습니다. 일반적으로 중간 정도에서 큰 표본 설계의 경우 수정된 Blaker 방법 또는 Willson CC 방법을 선택하는 것이 좋습니다. 작은 표본 설계의 경우 실제 비율 값에 대한 사전 지식이 있는 경우 Wilson 또는 Agresti-Coull 이 적합할 수 있습니다. 이러한 사전 지식은 종종 이전의 유사한 경험 또는 비율의 대략적인 추정치를 얻기 위해 특별히 고안된 파일럿 소규모 연구를 기반으로 합니다. 예를 들어, 결점 비율이 일반적으로 0 에 가까운 응용 분야의 품질 관리 영역에서 선택할 방법은 조사자가 보수적인 결과를 원하는지 여부에 따라 달라집니다. 보수적인 결과는 Agresti-Coull 방법을 기반으로 할 수 있는 반면 진보적인 결과는 Wilson/score 방법을 기반으로 할 수 있습니다. 마지막으로, 더 중요한 것은 유형 II 오류를 방지하거나 CI 의 너비를 제어하기 위해 표본 크기를 미리 계획하는 것이 좋습니다. Minitab 에는 "이를 위한 앱"도 있습니다. Minitab 에서는 가설 검정과 관련된 유형 II 오차를 제어하거나 CI 의 너비를 제어하기 위해 적절한 표본 크기를 결정하는 도구를 사용할 수 있습니다.

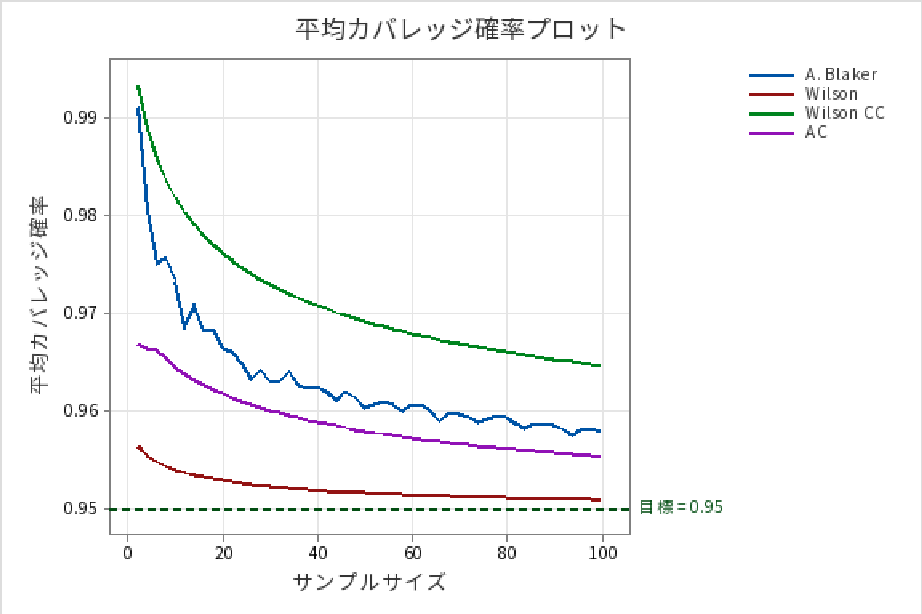

그림 5 : 표본 크기의 함수로 나타낸 4 가지 CI 방법 모두의 평균 커버리지 확률. 평균 커버리지 확률은 실제 비율이 단위 구간에 균일하게 분포되어 있다고 가정하여 계산됩니다. 평균 커버리지 곡선은 평균적으로 Yates 의 연속성 보정(Wilson CC) CI 방법을 사용한 Wilson/점수가 가장 보수적이며 수정된 Blaker(A. Blaker), Agresti-Coull(AC) 및 Wilson/score(Wilson) CI 방법이 그 뒤를 잇는다는 것을 보여줍니다. 평균 커버리지 확률 곡선은 표본 크기가 증가함에 따라 목표 공칭 커버리지 수준에 근접합니다. 또한 근사 방법(Wilson CC, Wilson, Agresti-Coull)에 대한 평균 커버리지 곡선은 매끄럽지만 정확히 조정된 Blaker 에 대한 커버리지 곡선은 공칭 커버리지에 가까워짐에 따라 약간의 흔들림 움직임이 있습니다. 이는 아마도 조정된 Blaker CI 방법이 더 개선될 수 있음을 나타냅니다.

기준

Agresti, A. and Coull, B. A. (1998). Approximate is better than “Exact” for interval Estimation of Binomial Proportion. The American Statistician 52, 119–125

Blaker, H. (2000). Confidence Curves and Improved Exact Confidence Intervals for Discrete Distributions. The Canadian Journal of Statistics, 28, 783–798

Blaker, H. (2001). Corrigenda: Confidence curves and improves exact confidence intervals for discrete distributions. The Canadian Journal of Statistics, 29, 681.

Blyth, C. R. and Still, H. A. (1983). Binomial Confidence Intervals. Journal of the American Statistical Association 78, 108–116.

Brown, L. D., Cai, T. and Das Gupta, A. (2001). Interval Estimation for a Binomial Proportion. Statistical Science 16, 101–133.

Casella, G., 1986. Refining binomial confidence intervals. Canad. J. Statist. 14, 113-129.

Clopper, C. J. and Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of Binomial. Biometrika 26, 404–413

Crow, E.L., 1956. Confidence intervals for a proportion. Biometrika 43, 423-435.

Klaschka, J. and Reiczigel, J. (2021). On matching confidence intervals and tests for some discrete distributions: methodological and computational aspects, Computational Statistics, Springer, vol. 36(3), 1775-1790.

Wilson E. B. (1927) Probable Inference, the Law of Successions and Statistical Inference. J. Amer. Statist. Assoc. 22, 209–21