

표준 편차 검정(표본 2개 이상)

개요

Minitab 보조 도구에는 독립 표본을 비교하여 표본의 변동성에 큰 차이가 있는지 판단하기 위한 분석이 2개 있습니다. 2-표본 표준 편차 검정은 표본 2개의 표준 편차를 비교하며, 표준 편차 검정은 2개보다 많은 표본의 표준 편차를 비교합니다. 이 문서에서는 $k = 2$ 인 k -표본 설계를 2-표본 설계라고 하고, $k > 2$ 인 k -표본 설계를 다중 표본 설계라고 합니다. 일반적으로, 이 두 가지 설계 유형은 별도로 조사됩니다(부록 A 참조).

표준 편차는 분산의 제곱근이므로, 표준 편차를 비교하는 가설 검정은 분산을 비교하는 가설 검정과 동일합니다. 모집단 2개 이상의 분산을 비교하기 위해 개발된 여러 통계분석 방법이 있습니다. 이런 검정 중에서 Levene/Brown-Forsythe 검정은 가장 강력하고 널리 사용되는 것 중 하나입니다. 그러나 Levene/Brown-Forsythe 검정의 검정력 성능은 2-표본 설계에서 해당 검정의 제1종 오류 속성보다 만족스럽지 않습니다. Pan(1999)에서는 정규 모집단을 포함한 일부 모집단의 경우 2-표본 설계의 검정력 상한이 표준 편차 간 차이의 크기에 관계 없이 1에 훨씬 못 미칠 수 있는 것으로 나타났습니다. 다시 말해, 이런 데이터 유형에 대해서는 표준 편차의 차이가 아무리 커도 서로 차이가 없다는 검정 결과가 나올 가능성이 더 큼니다. 이런 이유 때문에 보조 도구에서는 2-표본 표준 편차 검정에 Bonett 검정이라는 새로운 검정을 사용합니다. 보조 도구는 여러 표본 설계가 있는 표준 편차 검정에 다중 비교(MC) 절차를 사용합니다.

Layard(1978)의 두 표본 분산 동일성 검정을 수정한 버전인 Bonett(2006) 검정은 표본이 작은 검정의 성능을 개선합니다. Banga & Fox(2013A)에서는 Bonett 검정과 연관된 신뢰 구간을 도출하고 이것이 Levene/Brown-Forsythe 검정과 연관된 신뢰 구간만큼 정확하고 대부분의 분포에 대해 더 정밀한 것으로 나타났습니다. 또한 Banga & Fox(2013A)에서는 Bonett 검정이 Levene/Brown-Forsythe 검정만큼 강력하며, 대부분의 분포에 대해 더 검정력이 강하다고 결론지었습니다.

다중 비교(MC) 절차에는 각 표준 편차 쌍의 비교 구간에 기초한 여러 표본의 표준 편차(또는 분산)의 동질성 또는 동일성에 대한 전체적인 검정이 포함됩니다. 비교 구간의 한 쌍 이상이 겹치지 않을 경우에만 MC 검정이 유의하도록 비교 구간이 도출됩니다. Banga & Fox(2013B)에서는 MC 검정의 제1종 및 제2종 오류 속성이 대부분의 분포에 대해 Levene/Brown-Forsythe 검정과 유사한 것으로 나타났습니다. MC 검정의 중요한 장점 중 하나는 비교 구간의 그래픽 디스플레이로, 이는 표준 편차가 서로 다른 표본을 식별하는 데 효과적인 시각적 도구를 제공합니다. 설계에 표본이 2개뿐이면 MC 검정은 Bonett 검정과 동일합니다.

이 문서에서는 Bonett 검정과 MC 검정의 유효성을 서로 다른 데이터 분포와 표준 크기에 대해 평가합니다. 또한 대표본 근사 방법에 기초한 Bonett 검정에 사용되는 검정력과 표본 크기 분석에 대해 알아봅니다. 당사는 이러한 요인을 토대로 보조 도구가 데이터에 대해 자동으로 수행하고 보고서 카드에 표시하는 다음과 같은 검사를 개발했습니다.

- 비정상 데이터
- 정규성
- 검정의 유효성
- 표본 크기(2-표본 표준 편차 검정에만 해당)

표준 편차 검정 방법

Bonett 검정과 MC 검정의 유효성

Conover 외(1981)는 등분산 검정 비교 연구에서 Levene/Brown-Forsythe 검정이 제1종 및 제2종 오류율을 기준으로 가장 우수한 검정 중 하나임을 확인했습니다. 그 후로 2-표본 및 다중 표본 설계의 등분산 검정을 위한 기타 여러 가지 방법이 제안되었습니다(Pan, 1999; Shoemaker, 2003; Bonett, 2006). 예를 들어 Pan에서는 Levene/Brown-Forsythe 검정이 강력하고 해석이 간단하지만 표본을 같은 모집단(정규 모집단 포함)에서 추출했을 때 2개의 표준 편차 사이에서 중요한 차이를 탐지할 수 있을 만큼 검정력이 충분하지 않은 것으로 나타났습니다. 보조 도구는 이 중요한 한계로 인해 2-표본 표준 편차 검정에 Bonett 검정을 사용합니다(부록 A 또는 Banga & Fox, 2013A 참조). 보조 도구는 표본이 2개보다 많은 표준 편차 검정에 대해 MC 검정이 유의할 때 표준 편차가 서로 다른 표본을 식별하는 그래픽 디스플레이를 제공하는 비교 구간이 있는 MC 절차를 사용합니다(부록 A와 Banga & Fox, 2013B 참조).

목적

첫째, 당사는 모집단 2개의 표준 편차를 비교할 때 Bonett 검정이 얼마나 효과적인지 평가하고자 했습니다. 둘째, 모집단이 2개보다 많은 표준 편차를 비교할 때 MC 검정이 얼마나 효과적인지 평가하고자 했습니다. 구체적으로, 분포 유형이 서로 다른 다양한 크기의 표본에 해당 검정을 사용할 때 검정의 유효성을 평가하고자 했습니다.

방법

Bonett 검정과 MC 검정에 사용된 통계분석 방법은 부록 A에 정의되어 있습니다. 검정의 유효성을 평가하기 위해 검정의 제1종 오류율이 다른 조건에서도 목표 유의 수준(알파)과 가깝게 유지되는지 조사해야 했습니다. 이를 위해 독립 표본 2개의 표준 편차를 비교할 때 Bonett 검정의 유효성을 평가하기 위한 일련의 시뮬레이션을 수행하고 여러(k) 독립 표본의 표준 편차를 비교할 때 MC 검정의 유효성을 평가하기 위한 일련의 다른 시뮬레이션을 수행했습니다($k > 2$).

균형 및 불균형 설계를 모두 사용하여 여러 분포에서 추출한 다양한 크기의 여러(k) 무작위 표본을 1만 쌍 생성했습니다. 그런 다음, 각 실험에서 2개 표본의 표준 편차를 비교하기 위한 양측 Bonett 검정을 수행하거나 k개 표본의 표준 편차를 비교하기 위한 MC 검정을

수행했습니다. 사용된 목표 유의 수준은 $\alpha = 0.05$ 입니다. 10,000회의 반복실험 중에서 (실제 표준 편차가 동일할 때) 검정에 의해 귀무 가설이 기각된 횟수를 집계하고, 이 비율(모의 유의 수준이라고 함)을 목표 유의 수준과 비교했습니다. 검정이 효과적이면 실제 제1종 오류율을 나타내는 모의 유의 수준이

목표 유의 수준과 매우 유사할 것입니다. 2-표본 및 k-표본 시뮬레이션에 사용된 방법에 대한 자세한 내용은 부록 B를 참조하십시오.

결과

2-표본 비교에서 Bonett 검정의 모의 제1종 오류율은 표본 크기가 중간 이상일 때 설계의 균형 또는 불균형 여부에 관계 없이 목표 유의 수준과 유사했습니다. 그러나 극도로 치우친 모집단에서 작은 표본을 추출했을 때 Bonett 검정은 대체로 보수적이었으며, 제1종 오류율이 목표 유의 수준(즉, 목표 제1종 오류율)보다 약간 낮았습니다.

다중 표본 비교에서 MC 검정의 제1종 오류율은 표본 크기가 중간 이상일 때 분포에 관계 없이, 그리고 설계의 균형 또는 불균형 여부에 관계 없이 목표 유의 수준과 유사했습니다. 그러나 표본이 작고 극도로 치우친 경우에는 검정이 대체로 덜 보수적이었으며, 설계의 표본 수가 많을 때 제1종 오류율이 목표 유의 수준보다 높았습니다.

당사의 연구 결과는 Banga & Fox(2013A) 및 (2013B)의 결과와 일치했습니다. 당사는 Bonett 검정과 MC 검정이 최소 표본의 크기가 20개 이상일 때 효과적이라는 결론을 내렸습니다. 따라서 보조 도구 보고서 카드의 검정 유효성 검사에 이 최소 표본 크기 요건을 사용합니다(데이터 검사 항목 참조).

비교 구간

2개 이상의 표준 편차를 비교하기 위한 검정이 통계적으로 유의하여 하나 이상의 표준 편차가 나머지와 다른 것으로 나타나면, 다음 분석 단계에서는 어떤 표본이 통계적으로 다른지 확인합니다. 이 비교 작업을 직관적으로 수행하기 위해 각 표본과 연관된 신뢰 구간을 그래프로 표시하고 구간이 겹치지 않는 표본을 식별합니다. 그러나 개별 신뢰 구간은 비교에 사용할 수 없기 때문에 그래프를 근거로 내린 결론은 검정 결과와 일치하지 않을 수 있습니다.

목적

당사는 전체적인 검정이 유의할 때 전체적인 분산 동질성 검정과 분산이 서로 다른 표본을 식별하는 방법으로 모두 사용할 수 있는 개별 비교 구간을 계산하는 방법을 개발하고자 했습니다. MC 절차의 중요한 요구 사항 중 하나는 한 쌍 이상의 비교 구간이 겹치지 않음에

따라 2개 이상의 표본의 표준 편차가 다른 것으로 나타날 경우에만 전체 검정이 유의하다는 것입니다.

방법

여러 표준 편차를 비교하는 데 사용되는 MC 절차는 다중 쌍 비교에서 유래되었습니다. 각 표본 쌍은 Bonett(2006)의 2개 모집단 표준 편차 동일성 검정을 사용하여 비교합니다. 쌍 비교에는 Nayakama(2009)에 나와 있는 대표본 근사에 기초한 다중성 수정이 사용됩니다. 널리 사용되는 Bonferroni 수정보다 대표본 근사가 더 선호되는 이유는 표본 수가 증가할수록 Bonferroni 수정은 점점 더 보수적이 되기 때문입니다. 마지막으로, 비교 구간은 Hochberg 외(1982)의 최근사 절차에 기초한 쌍 비교에서 비롯되었습니다. 자세한 내용은 부록 A를 참조하십시오.

결과

MC 절차는 2개 이상의 비교 구간이 겹치지 않을 경우에만 전체 표준 편차 동일성 검정이 유의하다는 요구 사항을 충족합니다. 전체 검정이 유의하지 않으면 모든 비교 구간이 겹쳐야 합니다.

보조 도구는 요약 보고서의 표준 편차 비교 차트에 비교 구간을 표시합니다. 그리고 이 그래프 옆에 전체 표준 편차 동일성 검정인 MC 검정의 p-값을 표시합니다. 표준 편차 검정이 통계적으로 유의하면 하나 이상의 다른 구간과 겹치지 않는 비교 구간이 모두 빨간색으로 표시됩니다. 표준 편차 검정이 통계적으로 유의하지 않으면 빨간색으로 표시되는 구간이 없습니다.

이론적 검정력 성능(2-표본 설계에만 해당)

Bonett 및 MC 검정의 이론적 검정력 함수는 표본 크기를 계획하는 데 필요합니다. 2-표본 설계에 대해 검정의 근사 이론적 검정력 함수는 대표본 이론 방법을 사용하여 도출할 수 있습니다. 이 함수는 대표본 근사 방법의 결과로 얻어지므로, 검정을 실시할 때 정규 및 비정규 분포에서 생성된 작은 표본을 사용하여 함수의 속성을 평가해야 했습니다. 그러나 2개보다 많은 집단의 표준 편차를 비교할 경우 MC 검정의 이론적 검정력 함수를 쉽게 얻을 수 없습니다.

목적

당사는 보조 도구에서 대표본 근사에 기초한 이론적 검정력 함수를 사용하여 2-표본 표준 편차 검정에 대한 검정력 및 표본 크기 요구 사항을 평가할 수 있는지 확인하고자 했습니다. 이를 위해서는 Bonett 검정을 정규 및 비정규 분포를 포함한 여러 분포 유형에서 얻어진

데이터에 대해 수행할 때 근사 이론적 검정력 함수가 실제로 얻어지는 검정력을 정확히 나타내는지 평가해야 했습니다.

방법

2-표본 설계에 대한 Bonett 검정의 근사 이론적 검정력 함수는 부록 C에 나와 있습니다.

당사는 Bonett 검정을 사용하여 (당사가 모의 검정력 수준이라고 부르는) 실제 검정력 수준을 추정하기 위한 시뮬레이션을 수행했습니다. 첫째, 정규 및 비정규 분포를 포함한 여러 분포에서 추출한 여러 크기의 무작위 표본 쌍을 생성했습니다. 각 분포에 대해 Bonett 검정을 1만 쌍의 표본 반복실험에 대해 각각 수행했습니다. 각 표본 크기 쌍에 대해 주어진 차이를 탐지하는 검정의 모의 검정력을 검정이 유의한 표본 1만 쌍의 분수로 계산했습니다. 비교를 위해 검정의 근사 이론적 검정력 함수를 사용하여 해당 검정력 수준도 계산했습니다. 근사가 효과적이라면 이론적 및 모의 검정력 수준이 유사할 것입니다. 자세한 내용은 부록 D를 참조하십시오.

결과

시뮬레이션에 따르면 Bonett 검정의 이론적 및 모의 검정력 함수는 표본 크기가 작을 때 대부분의 분포에서 거의 같고 최소 표본 크기가 20개에 도달할 때 더 유사한 것으로 나타났습니다. 꼬리의 크기가 중간 이하인 대칭 및 근대칭 분포에서 이론적 검정력 수준은 모의(실제) 검정력 수준보다 약간 더 높습니다. 그러나 치우친 분포와 꼬리가 두꺼운 분포에서는 모의(실제) 검정력 수준보다 더 작습니다. 자세한 내용은 부록 D를 참조하십시오. 전체적으로, 당사의 연구 결과에서는 이론적 검정력 함수가 표본 크기를 계획하는 좋은 기초가 되는 것으로 나타났습니다.

데이터 검사

비정상 데이터

비정상 데이터는 특이치라고도 하는 극도로 크거나 작은 데이터 값입니다. 비정상 데이터는 분석 결과에 중대한 영향을 미칠 수 있으며, 특히 표본이 작을 때 통계적으로 유의한 결과를 찾을 확률에 영향을 미칠 수 있습니다. 비정상 데이터는 데이터 수집에 문제가 있음을 나타내거나, 연구하는 프로세스의 비정상적인 동작에 기인할 수 있습니다. 그러므로 이런 데이터 점은 종종 조사할만한 가치가 있으며, 가능하면 수정해야 합니다. 시뮬레이션 연구에서는 데이터에 특이치가 있을 때 Bonett 검정과 MC 검정이 보수적인 것으로 나타났습니다(부록 B 참조). 검정의 실제 유의 수준은 특히 작은 표본을 사용하여 분석을 수행할 때 목표 수준보다 훨씬 낮습니다.

목적

당사는 전체 표본에 비해 매우 크거나 매우 작고 분석 결과에 영향을 미칠 수 있는 데이터 값이 있는지 확인하는 방법을 개발하고자 했습니다.


방법


당사는 상자 그림에서 특이치를 식별하기 위해 사용되는 Hoaglin, Iglewicz & Tukey(1986)가 설명한 방법을 토대로 비정상 데이터를 검사하는 방법을 개발했습니다.

결과

보조 도구는 분포의 하위 또는 상위 사분위수를 벗어난 사분위간 범위의 1.5배 이상인 데이터 점을 비정상 데이터로 구분합니다. 하위 및 상위 사분위수는 데이터의 25 및 75 백분위수에 해당됩니다. 사분위간 범위는 두 사분위수의 차이입니다. 이 방법을 사용하면 특정 특이치를 각각 탐지할 수 있기 때문에 특이치가 여러 개일 때도 효과적입니다.

비정상 데이터가 있는지 검사할 때 보조 도구는 다음과 같은 상태 지표를 보고서 카드에 표시합니다.

상태	조건
	비정상적인 데이터 점 없음.

상태	조건
	하나 이상의 데이터 점이 비정상이고 결과에 중대한 영향을 미칠 수 있음.

정규성

정규성 가정에 따라 도출되는 대부분의 등분산 검정과 달리 표준 편차 동일성에 대한 Bonett 검정과 MC 검정은 특정 데이터 분포에 대해 가정하지 않습니다.

목적

Bonett 검정과 MC 검정은 대표본 근사 방식에 기초하고 있지만, 당사는 각 검정이 작은 표본의 정규 및 비정규 데이터에도 효과적인지 확인하고자 했습니다. 또한 데이터의 정규성이 표준 편차 검정의 결과와 어떤 관계가 있는 지도 사용자에게 알리고자 했습니다.

방법


검정의 유효성을 서로 다른 조건에서 평가하기 위해 Bonett 검정 및 MC 검정의 제1종 오류율을 다양한 표본 크기의 정규 및 비정규 데이터를 사용하여 조사하기 위한 시뮬레이션을 수행했습니다. 자세한 내용은 표준 편차 검정 방법 항목과 부록 B를 참조하십시오.

결과


시뮬레이션에서는 표본이 충분히 클 경우(최소 표본 크기 ≥ 20) 데이터의 분포가 Bonett 검정 또는 MC 검정의 제1종 오류 속성에 중대한 영향을 미치지 않는 것으로 나타났습니다. 검정의 제1종 오류율은 정규 및 비정규 데이터의 경우 모두 목표 오류율과 일관되게 유사합니다.

보조 도구는 제1종 오류율에 대한 이 같은 결과를 토대로 정규성에 대한 정보를 보고서 카드에 표시합니다.

2-표본 설계에 대해 보조 도구는 다음 지표를 표시합니다.

상태	조건
	이 분석에는 Bonett 검정이 사용됩니다. 이 검정은 표본이 충분히 클 경우 정규 및 비정규 데이터에 대해 모두 우수한 성능을 나타냅니다.

다중 표본 설계에 대해 보조 도구는 다음 지표를 표시합니다.

상태	조건
	이 분석에는 다중 비교 검정이 사용됩니다. 이 검정은 표본이 충분히 클 경우 정규 및 비정규 데이터에 대해 모두 우수한 성능을 나타냅니다.

검정의 유효성

표준 편차 검정 방법 항목에서는 2-표본 및 다중 (k) 비교의 경우 모두 Bonett 검정과 MC 검정의 제1종 오류율이 표본 크기가 중간 이상일 때 균형 및 불균형 설계의 정규 및 비정규 데이터에 대해 목표 오류율과 유사한 것으로 나타났습니다. 그러나 표본이 작을 때는 Bonett 및 MC 검정이 대체로 효과적이지 못했습니다.

목적



당사는 사용자의 데이터를 토대로 2-표본 및 다중 (k) 표본에 대한 표준 편차 검정 결과의 유효성을 평가하기 위한 규칙을 적용하고자 했습니다.

방법

검정의 유효성을 서로 다른 조건에서 평가하기 위해 앞의 표준 편차 검정 방법 항목에서 설명한 대로 Bonett 검정 및 MC 검정의 제1종 오류율을 다양한 데이터 분포, 표본 수 및 표본 크기를 사용하여 조사하기 위한 시뮬레이션을 수행했습니다. 자세한 내용은 부록 B를 참조하십시오.

결과

Bonett 검정과 MC 검정은 최소 표본의 크기가 20개 이상일 때 효과적입니다. 따라서 보조 도구는 표준 편차 검정의 유효성을 평가하기 위해 다음과 같은 상태 지표를 보고서 카드에 표시합니다.

상태	조건
	표본 크기가 20개 이상이므로 p-값이 정확할 것임.
	일부 표본 크기가 20개 미만이므로 p-값이 정확하지 않을 수 있음. 표본 크기를 20개 이상으로 늘려 보십시오.

표본 크기(2-표본 표준 편차 검정에만 해당)

일반적으로, 통계적 가설 검정은 "차이가 없는" 귀무 가설을 기각하기 위한 증거를 수집하기 위해 수행합니다. 표본이 너무 작으면 검정의 검정력이 실제로 존재하는 차이를 탐지하는 데 충분하지 않을 수 있으며, 이는 제2종 오류로 이어집니다. 따라서 실제로 중요한 차이를 높은 확률로 탐지할 만큼 큰 표본 크기를 사용하는 것이 중요합니다.

목적

데이터에서 귀무 가설을 기각하기 위한 증거를 충분히 얻을 수 없는 경우에는 표본 크기가 검정에서 탐지하고자 하는 실제 차이를 높은 확률로 탐지하기에 충분히 큰지 확인하고자 했습니다. 표본 크기 계획의 목적은 중요한 차이를 높은 확률로 탐지하기에 충분히 큰 표본 크기를 사용하기 위한 것이지만, 표본이 너무 커서 무의미한 차이가 통계적으로 중요하다고 인식될 확률이 높아져서도 안 됩니다.


방법





2-표본 표준 편차 검정의 검정력 및 표본 크기 분석은 일반적으로 검정의 실제 검정력 함수에 대한 좋은 추정치를 제공하는 Bonett 검정의 근사 검정력 함수에 기초합니다(방법 항목의 '이론적 검정력 함수의 성능'에 요약된 시뮬레이션 결과 참조).

결과

귀무 가설을 반박하는 충분한 증거를 데이터로부터 얻을 수 없으면, 보조 도구는 Bonett 검정의 근사 검정력 함수를 사용하여 80% 및 90% 확률로 탐지할 수 있는 실제 차이를 주어진 표본 크기에 대해 계산합니다. 또한 탐지하고자 하는 특정 실제 차이를 제공할 경우 보조 도구는 정규 근사 검정의 검정력 함수를 사용하여 80% 및 90%의 차이 탐지 확률을 얻기 위해 필요한 표본 크기를 계산합니다.

2-표본 표준 편차 검정에 대한 보조 도구 보고서 카드는 결과 해석을 지원하기 위해 검정력 및 표본 크기를 확인할 때 다음과 같은 상태 지표를 표시합니다.

상태	조건
	검정에서 표준 편차의 차이가 탐지되므로 검정력은 문제가 되지 않음. 또는 검정력이 충분함. 검정에서는 표준 편차의 차이가 발견되지 않았지만, 표본 크기는 주어진 차이를 90% 이상의 확률로 탐지하는 데 충분함.

상태	조건
	<p>검정력이 충분할 수 있음. 검정에서는 표준 편차의 차이가 발견되지 않았지만, 표본 크기는 주어진 차이를 80% - 90% 확률로 탐지하는 데 충분함. 90% 검정력을 달성하기 위해 필요한 표본 크기가 보고됨.</p>
	<p>검정력이 충분하지 않을 수 있음. 검정에서 표준 편차의 차이를 찾지 못했지만, 표본 크기는 주어진 차이를 60% - 80% 확률로 탐지하는 데 충분함. 80% 검정력과 90% 검정력을 달성하는 데 필요한 표본 크기가 보고됨.</p>
	<p>검정력이 충분하지 않음. 검정에서 표준 편차의 차이를 찾지 못했으며, 표본 크기는 주어진 차이를 60% 이상의 확률로 탐지하는 데 충분하지 않음. 80% 검정력과 90% 검정력을 달성하는 데 필요한 표본 크기가 보고됨.</p>
	<p>검정에서 표준 편차의 차이를 찾지 못함. 탐지할 실제 차이를 지정하지 않았으므로 보고서에는 표본 크기와 알파를 기준으로 80% 및 90% 확률로 탐지할 수 있는 차이가 표시됨.</p>

참고 문헌

- Arnold, S.F. (1990). *Mathematical Statistics*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Banga, S.J. & Fox, G.D. (2013A). On Bonett's Robust Confidence Interval for a Ratio of Standard Deviations. *백서, Minitab Inc.*
- Banga, S.J. & Fox, G.D. (2013B) A graphical multiple comparison procedure for several standard deviations. *백서, Minitab Inc.*
- Bonett, D.G. (2006). Robust confidence interval for a ratio of standard deviations. *Applied Psychological Measurements*, 30, 432-439.
- Brown, M.B., & Forsythe, A.B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69, 364-367.
- Conover, W.J., Johnson, M.E., & Johnson, M.M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, 23, 351-361.
- Gastwirth, J. L. (1982). Statistical properties of a measure of tax assessment uniformity. *Journal of Statistical Planning and Inference*, 6, 1-12.
- Hochberg, Y., Weiss G., & Hart, S. (1982). On graphical procedures for multiple comparisons. *Journal of the American Statistical Association*, 77, 767-772.
- Layard, M.W.J. (1973). Robust large-sample tests for homogeneity of variances. *Journal of the American Statistical Association*, 68, 195-198.
- Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Probability and statistics* (278-292). Stanford University Press, Palo Alto, California.
- Nakayama, M.K. (2009). Asymptotically valid single-stage multiple-comparison procedures. *Journal of Statistical Planning and Inference*, 139, 1348-1356.
- Pan, G. (1999) On a Levene type test for equality of two variances. *Journal of Statistical Computation and Simulation*, 63, 59-71.
- Shoemaker, L. H. (2003). Fixing the F test for equal variances. *The American Statistician*, 57 (2), 105-114.

부록 A: Bonnett 검정 및 다중 비교

검정 방법

Bonnett 방법(2-표본 설계) 또는 다중 비교(MC) 절차(다중 표본 설계)를 사용하여 표준 편차 또는 분산에 대한 추론을 도출하기 위한 기본 가정은 다음과 같이 설명할 수 있습니다.

$X_{11}, \dots, X_{1n_1}, \dots, X_{k1}, \dots, X_{kn_k}$ 가 k 개($k \geq 2$)의 독립 무작위 표본이고, 각 표본이 평균 μ_i 와 분산 σ_i^2 을 알 수 없고 각각에 대해 $i = 1, \dots, k$ 인 분포에서 추출된 경우를 생각해 보십시오. 표본의 모집단 분포의 공통 유한 첨도를 $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$ 이라고 가정합니다. 이 가정은 이론을 도출하는 데 중요하지만, 대부분 표본이 충분히 큰 실제 용도에서는 중요하지 않습니다(Banga & Fox, 2013A).

방법 A1: Bonnett의 두 분산 동일성 검정

Bonnett 검정은 2개의 분산 또는 표준 편차를 비교하는 2-표본 설계에만 적용됩니다. 이 검정은 Layard(1978)의 2-표본 설계 등분산 검정을 수정한 것입니다. 두 표본 설계 분산의 동일성을 검정하는 유의 수준이 α 인 양측 Bonnett 검정은 다음과 같은 경우에만 null 동일성 가설을 기각합니다.

$$|\ln(c S_1^2 / S_2^2)| > z_{\alpha/2} \sqrt{\frac{\hat{\gamma}_P - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P - g_2}{n_2 - 1}}$$

설명:

S_i 는 표본 i 의 표본 표준 편차입니다

$$g_i = (n_i - 3) / n_i, i = 1, 2$$

$z_{\alpha/2}$ 는 표준 정규 분포의 상위 $\alpha/2$ 백분위수를 말합니다.

$\hat{\gamma}_P$ 는 다음 식으로 표현되는 합동 첨도 추정자입니다.

$$\hat{\gamma}_P = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (X_{1j} - m_1)^4 + \sum_{j=1}^{n_2} (X_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]^2}$$

합동 첨도 추정자 식에서 m_i 는 표본 i 의 절사 평균이며, 절사 비율은 $1/[2(n_i - 4)^{1/2}]$ 입니다.

위에서 상수 c 는 불균형 설계에서 불균등 꼬리 오류 확률의 효과를 줄이기 위한 소규모 표본 조정으로 포함됩니다. 이 상수의 공식은 $c = c_1 / c_2$ 입니다.

$$c_i = \frac{n_i}{n_i - z_{\alpha/2}}, i = 1, 2$$

설계가 균형이면(즉 $n_1 = n_2$ 이면) 검정의 p-값을 다음과 같이 구합니다.

$$P = 2 \Pr(Z > z)$$

여기서 Z 는 표준 정규 분포의 형태로 분포된 무작위 변수이며, z 는 주어진 데이터에 기초한 다음 통계의 관측값입니다. 이 통계는 다음과 같습니다.

$$Z = \frac{\ln(C S_1^2/S_2^2)}{se}$$

설명

$$se = \sqrt{\frac{\hat{y}_P - g_1}{n_1 - 1} + \frac{\hat{y}_P - g_2}{n_2 - 1}}$$

반면에, 설계가 불균형일 경우에는 검정의 p-값을 다음과 같이 구합니다.

$$P = 2\min(\alpha_L, \alpha_U)$$

여기서 $\alpha_L = \Pr(Z > z_L)$ 및 $\alpha_U = \Pr(Z > z_U)$ 입니다. 변수 z_L 는 다음 함수의 최소 제곱근이고

$$L(z, S_1, S_2, n_1, n_2) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} - z se + \ln \frac{S_1^2}{S_2^2} - \ln \rho_0^2, z < \min(n_1, n_2)$$

z_U 는 $L(z, S_2, S_1, n_2, n_1)$ 함수의 최소 제곱근입니다.

방법 A2: 다중 비교 검정과 비교 구간

$k(k \geq 2)$ 개의 독립 그룹 또는 표본이 있다고 가정합니다. 당사의 목표는 k 개 구간 중 2개 이상의 구간이 겹치지 않을 경우에만 표준 편차 동일성 검정이 유의하도록 모집단 표준 편차에 대해 k 구간 시스템을 구성하는 것이었습니다. 이런 구간을 비교 구간이라고 합니다. 이 비교 방법은 Tukey-Kramer가 처음 개발하고 이후에 Hochberg 외(1982)가 일반화한 일원 분산 분석 모형의 평균을 다중 비교하는 절차와 유사합니다.

표준 편차 2개 비교

2-표본 설계에서는 Bonett 검정과 연관된 표준 편차 비율의 신뢰 구간을 직접 계산하여 표준 편차 간 차이의 크기를 평가할 수 있습니다(Banga & Fox, 2013A). 실제로 당사는 이 방법을 Minitab 17 릴리스의 통계분석 > 기초 통계 > 두 표본 분산에 사용합니다. 그러나 보조 도구에서 표준 편차 비율의 신뢰 구간보다 더 해석하기 쉬운 비교 구간을 제공하고자 했습니다. 이를 위해 방법 A1에서 설명한 Bonett 절차를 사용하여 두 표본의 비교 구간을 정했습니다.

표본이 2개일 때는 Bonett 등분산 검정과 연관된 다음 합격 구간에 0이 없을 경우에만 Bonett의 등분산 검정이 유의합니다.

$$\ln(c_1 S_1^2) - \ln(c_2 S_2^2) \pm z_{\alpha/2} \sqrt{\frac{\hat{\gamma}_P - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P - g_2}{n_2 - 1}}$$

여기서 합동 척도 추정치 $\hat{\gamma}_P$ 와 $g_i, i = 1, 2$ 는 이전에 주어진 것과 같습니다.

이 구간을 토대로, 2개 비교 구간이 겹치지 않을 경우에만 분산 또는 표준 편차 동일성 검정이 유의하도록 다음 2개의 비교 구간을 도출합니다. 두 구간은 다음과 같습니다.

$$\left[S_i \sqrt{C_i \exp(-z_{\alpha/2} V_i)}, S_i \sqrt{C_i \exp(z_{\alpha/2} V_i)} \right], i = 1, 2$$

설명

$$V_i = \frac{\sqrt{\frac{\hat{\gamma}_P - g_i}{n_i - 1}}}{\sqrt{\frac{\hat{\gamma}_P - g_i}{n_i - 1} + \frac{\hat{\gamma}_P - g_j}{n_j - 1}}} \sqrt{\frac{\hat{\gamma}_P - g_i}{n_i - 1} + \frac{\hat{\gamma}_P - g_j}{n_j - 1}}, i = 1, 2; j = 1, 2; i \neq j$$

두 구간을 표준 편차 동일성 검정 절차로 사용하는 것은 Bonett의 표준 편차 동일성 검정과 동일합니다. 구체적으로, 두 구간은 Bonett의 표준 편차 동일성 검정이 유의할 경우에만 겹치지 않습니다. 그러나 두 구간은 표준 편차의 신뢰 구간이 아니라 표준 편차의 다중 비교에만 해당됩니다. 같은 이유로, Hochberg 외는 평균을 비교하는 데 사용되는 이와 유사한 구간을 불확실성 구간이라고 지칭합니다. 당사에서는 해당 구간을 비교 구간이라고 합니다.

비교 구간 절차는 Bonett의 표준 편차 동일성 검정과 동일하기 때문에 비교 구간과 연관된 p-값은 앞에서 설명한 Bonett의 두 표본 표준 편차 동일성 검정의 p-값과 동일합니다.

여러 표준 편차 비교

그룹 또는 표본이 2개보다 많은 경우 k 비교 구간을 모임 유의 수준이 α 인 $k(k-1)/2$ 쌍 동시 표준 편차 동일성 검정으로부터 도출합니다. 더 구체적으로, X_{i1}, \dots, X_{in_i} 및 X_{j1}, \dots, X_{jn_j} 가 표본 쌍 (i, j) 의 표본 데이터인 경우를 생각해 볼 수 있습니다. 특정 표본 쌍 (i, j) 에 대한 표준 편차 동일성 검정은 표본이 2개인 경우와 마찬가지로 다음 구간에 0이 없을 경우에만 α' 수준에서 유의합니다.

$$\ln(c_i S_i^2) - \ln(c_j S_j^2) \pm z_{\alpha'/2} \sqrt{\frac{\hat{\gamma}_{ij} - g_i}{n_i - 1} + \frac{\hat{\gamma}_{ij} - g_j}{n_j - 1}}$$

위에서 $\hat{\gamma}_{ij}$ 는 표본 쌍 (i, j) 에 기초한 합동 척도 추정치이며, 다음과 같이 계산됩니다.

$$\hat{\gamma}_{ij} = (n_i + n_j) \frac{\sum_{l=1}^{n_i} (X_{il} - m_i)^4 + \sum_{l=1}^{n_j} (X_{jl} - m_j)^4}{[(n_i - 1)S_i^2 + (n_j - 1)S_j^2]^2}$$

또한 이전에 정의한 것처럼 m_i 는 절사 비율이 $1/[2(n_i - 4)^{1/2}]$ 인 표본 i 의 절사 평균이며, 다음이 성립됩니다.

$$g_i = \frac{n_i - 3}{n_i}, g_j = \frac{n_j - 3}{n_j}, c_i = \frac{n_i}{n_i - z_{\alpha/2}}, c_j = \frac{n_j}{n_j - z_{\alpha/2}}$$

동시 쌍 검정의 수는 $k(k - 1)/2$ 이므로, 실제 모임 오류율이 목표 유의 수준 α' 에 가깝도록 α 수준을 선택해야 합니다. 가능한 조정 중 하나는 Bonferroni의 근사값에 기초한 조정입니다. 그러나 Bonferroni 수정은 설계의 표본 수가 증가할수록 점점 보수적이 된다고 널리 알려져 있습니다. 더 나은 방법은 Nakayama(2008)가 제시한 정규 근사에 기초한 방법입니다. 이 방법을 사용하면 $z_{\alpha'/2}$ 가 $q_{\alpha,k}/\sqrt{2}$ 로 대체될 뿐입니다. 여기서 $q_{\alpha,k}$ 는 k 개의 독립적이고 동일하게 분포된 표준 정규 무작위 변수 범위의 상위 α 점입니다. 즉, 다음과 같습니다.

$$\Pr\left(\max_{1 \leq i < j \leq k} |Z_i - Z_j| \leq q_{\alpha,k}\right) = 1 - \alpha$$

여기서 Z_1, \dots, Z_k 는 독립적이고 동일하게 분포된 표준 정규 무작위 변수입니다.

또한 Hochberg 외(1982)의 방법과 유사한 위에서 설명한 표본 쌍 절차에 가장 근접한 절차를 사용하면 특정 표본 쌍 (i, j) 에 대해 다음 공식이 성립될 경우에만 표준 편차 동일성의 귀무 가설을 기각합니다.

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > q_{\alpha,k} (V_i + V_j) / \sqrt{2}$$

여기서 V_i 는

$$\sum_{i \neq j} (V_i + V_j - b_{ij})^2$$

수량을 사용하여 다음과 같이 최소화하기 위해 선택됩니다.

$$b_{ij} = \sqrt{\frac{\hat{\gamma}_{ij} - g_i}{n_i - 1} + \frac{\hat{\gamma}_{ij} - g_j}{n_j - 1}}$$

Hochberg 외(1982)에 설명된 이 문제에 대한 해결책은 다음을 선택하는 것입니다.

$$V_i = \frac{(k - 1) \sum_{j \neq i} b_{ij} - \sum_{1 \leq j < l \leq k} b_{jl}}{(k - 1)(k - 2)}$$

논리적으로, 근사 절차에 기초한 검정은 다음 k 개 구간 중 한 쌍 이상이 겹치지 않을 경우에만 유의합니다.

$$\left[S_i \sqrt{C_i \exp(-q_{\alpha,k} V_i / \sqrt{2})}, S_i \sqrt{C_i \exp(q_{\alpha,k} V_i / \sqrt{2})} \right], i = 1, \dots, k$$

MC 검정과 연관된 전체 p-값을 계산하기 위해 P_{ij} 가 표본 쌍 (i, j) 와 연관된 p-값이라고 하겠습니다. 그러면 다중 비교 검정과 연관된 전체 p-값은 논리적으로 다음과 같습니다.

$$P = \min\{P_{ij}, 1 \leq i < j \leq k\}$$

P_{ij} 를 계산하기 위해 다음을 사용하여 방법 A1에서 주어진 2-표본 설계의 알고리즘을 실행합니다.

$$se = V_i + V_j$$

여기서 V_i 는 위에서 주어진 것과 같습니다.

더 구체적으로 $n_i \neq n_j$ 일 경우 다음과 같습니다.

$$P_{ij} = \min(\alpha_L, \alpha_U)$$

여기서 $\alpha_L = \Pr(Q > z_L\sqrt{2})$, $\alpha_U = \Pr(Q > z_U\sqrt{2})$, 변수 z_L 는 함수 $L(z, S_i, S_j, n_i, n_j)$ 의 최소 제곱근이고, 변수 z_U 는 함수 $L(z, S_j, S_i, n_j, n_i)$ 의 최소 제곱근이며, Q 는 범위 분포가 이전에 정의된 것과 같은 무작위 변수입니다.

$n_i = n_j$ 면 $P_{ij} = \Pr(Q > |z_o|\sqrt{2})$ 입니다. 설명:

$$z_o = \frac{\ln S_i^2 - \ln S_j^2}{V_i + V_j}$$

부록 B: Bonett 검정 및 다중 비교 검정의 유효성

시뮬레이션 B1: Bonett 검정의 유효성(2-표본 모형, 균형 및 불균형 설계)

당사는 속성이 서로 다른 분포에서 추출한 크기가 중간 이하인 무작위 표본의 쌍을 생성했습니다. 포함된 분포는 다음과 같습니다.

- 표준 정규 분포($N(0,1)$)
- 대칭 및 얇은 꼬리 분포. 균등 분포($U(0,1)$)와 두 매개 변수를 모두 3으로 설정한 베타 분포($B(3,3)$) 포함.
- 대칭 및 두꺼운 꼬리 분포. 자유도가 5-10인 t 분포($t(5), t(10)$)와 위치가 0이고 척도가 1인 Laplace 분포(Lpl) 포함.
- 치우친 분포와 두꺼운 꼬리 분포. 척도가 1인 지수 분포(Exp)와 자유도가 5-10인 카이 제곱 분포($Chi(5), Chi(10)$) 포함.
- 왼쪽으로 치우친 분포와 두꺼운 꼬리 분포. 구체적으로, 매개 변수가 각각 8과 1로 설정된 베타 분포($B(8,1)$).

또한 특이치의 직접적인 영향을 평가하기 위해 다음과 같이 정의되는 오염된 정규 분포로부터 표본의 쌍을 생성했습니다.

$$CN(p, \sigma) = pN(0,1) + (1 - p)N(0, \sigma)$$

여기서 p 는 혼합 모수이고 $1 - p$ 는 (특이치의 비율과 같은) 오염 비율입니다. 연구에 사용할 오염된 정규 분포 2개를 선택했습니다. 하나는 모집단의 10%가 특이치인 $CN(0.9,3)$ 였고, 다른 하나는 모집단의 20%가 특이치인 $CN(0.8,3)$ 였습니다. 이 두 분포는 대칭이며, 특이치로 인해 꼬리가 깎입니다.

각 분포에서 추출한 각 표본 쌍에 대해 목표 유의 수준이 $\alpha = 0.05$ 인 양측 Bonett 검정을 수행했습니다. 각각의 경우에 모의 유의 수준은 1만 쌍의 표본 반복실험을 기준으로 한 것이고 5%의 목표 유의 수준을 사용했기 때문에 시뮬레이션 오류는 $\sqrt{0.95(0.05)/10,000} = 0.2\%$ 였습니다.

시뮬레이션 결과는 아래 표 1에 요약되어 있습니다.

표 1 균형 및 불균형 2-표본 설계에서 양측 Bonett 검정의 모의 유의 수준. 목표 유의 수준은 0.05입니다.

분포	n_1, n_2	모의 수준	분포	n_1, n_2	모의 수준
N(0,1)	10, 10	0.038	기대	10, 10	0.052
	20, 10	0.043		20, 10	0.051
	20, 20	0.045		20, 20	0.049
	30, 10	0.044		30, 10	0.044
	30, 20	0.046		30, 20	0.042
	25, 25	0.048		25, 25	0.043
	30, 30	0.048		30, 30	0.042
	40, 40	0.051		40, 40	0.042
	50, 50	0.047		50, 50	0.039
t(5)	10, 10	0.044	Chi(5)	10, 10	0.040
	20, 10	0.042		20, 10	0.043
	20, 20	0.046		20, 20	0.040
	30, 10	0.041		30, 10	0.039
	30, 20	0.046		30, 20	0.043
	25, 25	0.048		25, 25	0.042
	30, 30	0.043		30, 30	0.043
	40, 40	0.046		40, 40	0.040
	50, 50	0.050		50, 50	0.039

분포	n_1, n_2	모의 수준	분포	n_1, n_2	모의 수준
t(10)	10, 10	0.041	Chi(10)	10, 10	0.044
	20, 10	0.040		20, 10	0.042
	20, 20	0.045		20, 20	0.041
	30, 10	0.046		30, 10	0.043
	30, 20	0.045		30, 20	0.045
	25, 25	0.046		25, 25	0.046
	30, 30	0.048		30, 30	0.038
	40, 40	0.045		40, 40	0.042
	50, 50	0.051		50, 50	0.049
	Lpl	10, 10		0.054	B(8,1)
20, 10		0.056	20, 10	0.045	
20, 20		0.055	20, 20	0.048	
30, 10		0.057	30, 10	0.042	
30, 20		0.058	30, 20	0.047	
25, 25		0.057	25, 25	0.041	
30, 30		0.053	30, 30	0.040	
40, 40		0.047	40, 40	0.042	
50, 50		0.048	50, 50	0.038	

분포	n_1, n_2	모의 수준	분포	n_1, n_2	모의 수준
B(3,3)	10, 10	0.032	CN(0.9,3)	10, 10	0.024
	20, 10	0.037		20, 10	0.022
	20, 20	0.042		20, 20	0.018
	30, 10	0.039		30, 10	0.019
	30, 20	0.038		30, 20	0.020
	25, 25	0.039		25, 25	0.019
	30, 30	0.041		30, 30	0.015
	40, 40	0.044		40, 40	0.020
	50, 50	0.046		50, 50	0.017
	U(0,1)	10, 10		0.030	CN(0.8,3)
20, 10		0.032	20, 10	0.019	
20, 20		0.031	20, 20	0.020	
30, 10		0.034	30, 10	0.017	
30, 20		0.034	30, 20	0.020	
25, 25		0.034	25, 25	0.021	
30, 30		0.037	30, 30	0.017	
40, 40		0.043	40, 40	0.023	
50, 50		0.043	50, 50	0.020	

표 1에서 볼 수 있듯이, 표본 크기가 더 작으면 Bonett 검정의 모의 유의 수준은 꼬리 크기가 중간 이하인 대칭 또는 근대칭 분포의 목표 유의 수준(0.05)보다 낮습니다. 반면에, 많이 치우친 분포에서 소규모 표본을 추출하면 모의 수준이 목표 수준보다 약간 더 큰 경향이 있습니다.

표본 크기가 크거나 약간 크면 모의 유의 수준은 모든 분포에서 목표 수준과 비슷합니다. 실제로 검정 성능은 지수 분포와 베타(8,1) 분포처럼 많이 치우친 분포에도 합당한 수준입니다.

또한 특이치는 큰 표본보다 작은 표본에 더 많은 영향을 미치는 것으로 나타납니다. 오염된 정규 분포의 모의 유의 수준은 두 표본의 최소 크기가 20개에 도달할 때 약 0.020에서 안정화되었습니다.

두 표본의 최소 크기가 20개일 때는 모의 유의 수준이 평평한 균등 분포와 오염된 정규 분포를 제외하고 계속 구간 [0.038, 0.058]에 포함되었습니다. 모의 유의 수준이 0.040이면 0.05의 목표 수준에 비해 약간 보수적이지만, 이 제1종 오류율은 대부분의 실제 용도에서 허용될 수 있는 수준입니다. 그러므로 두 표본의 최소 크기가 20개 이상이면 Bonett 검정이 유효하다는 결론을 내릴 수 있습니다.

시뮬레이션 B2: MC 검정의 유효성(다중 표본 모형)

I부: 균형 설계

당사는 균형 설계를 사용한 다중 표본 모형에 MC 검정이 얼마나 효과적인지 조사하기 위한 시뮬레이션을 수행했습니다. 이를 위해, 앞의 시뮬레이션 B1에 나열된 일련의 분포를 사용하여 동일한 분포로부터 같은 크기의 표본을 k 개 생성했습니다. 설계의 표본 수는 $k = 3$, $k = 4$ 및 $k = 6$ 이 되도록 선택했으며, 각 실험의 k 표본 크기를 10, 15, 20, 25, 50 및 100으로 고정했습니다.

각 설계의 동일한 표본에 대해 목표 유의 수준이 $\alpha = 0.05$ 인 양측 MC 검정을 수행했습니다. 각각의 경우에 모의 유의 수준은 1만 쌍의 표본 반복실험을 기준으로 한 것이었고 5%의 목표 유의 수준이 사용되었기 때문에 시뮬레이션 오류는 $\sqrt{0.95(0.05)/10,000} = 0.2\%$ 였습니다.

시뮬레이션 결과는 아래의 표 2a와 2b에 요약되어 있습니다.

표 2a 균형 다중 표본 설계에서 양측 다중 비교 검정의 모의 유의 수준. 검정의 목표 유의 수준은 0.05입니다.

분포	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	모의 수준	n_i	모의 수준	n_i	모의 수준
N(0,1)	10	0.038	10	0.038	10	0.036
	15	0.040	15	0.041	15	0.039

분포	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	모의 수준	n_i	모의 수준	n_i	모의 수준
	20	0.039	20	0.040	20	0.041
	25	0.045	25	0.047	25	0.047
	50	0.046	50	0.046	50	0.052
	100	0.049	100	0.049	100	0.052
t(5)	10	0.042	10	0.044	10	0.042
	15	0.041	15	0.044	15	0.046
	20	0.043	20	0.045	20	0.045
	25	0.046	25	0.048	25	0.046
	50	0.040	50	0.039	50	0.038
	100	0.038	100	0.040	100	0.040
T(10)	10	0.033	10	0.037	10	0.038
	15	0.040	15	0.042	15	0.041
	20	0.042	20	0.043	20	0.043
	25	0.041	25	0.042	25	0.045
	50	0.047	50	0.044	50	0.047
	100	0.048	100	0.046	100	0.047
Lpl	10	0.056	10	0.063	10	0.071
	15	0.056	15	0.061	15	0.063
	20	0.054	20	0.058	20	0.059
	25	0.051	25	0.056	25	0.58
	50	0.045	50	0.051	50	0.049

분포	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	모의 수준	n_i	모의 수준	n_i	모의 수준
B(3,3)	100	0.044	100	0.047	100	0.050
	10	0.031	10	0.031	10	0.031
	15	0.037	15	0.036	15	0.034
	20	0.035	20	0.036	20	0.037
	25	0.039	25	0.038	25	0.040
	50	0.044	50	0.044	50	0.044
U(0,1)	100	0.044	100	0.046	100	0.043
	10	0.029	10	0.025	10	0.023
	15	0.026	15	0.027	15	0.026
	20	0.028	20	0.030	20	0.028
	25	0.034	25	0.033	25	0.032
	50	0.041	50	0.036	50	0.036
기대	100	0.048	100	0.047	100	0.045
	10	0.063	10	0.073	10	0.076
	15	0.056	15	0.058	15	0.064
	20	0.051	20	0.053	20	0.057
	25	0.043	25	0.045	25	0.050
	50	0.033	50	0.037	50	0.038
	100	0.033	100	0.035	100	0.035

표 2b 균형 다중 표본 설계에서 양측 다중 비교 검정의 모의 유의 수준. 검정의 목표 유의 수준은 0.05입니다.

분포	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	모의 수준	n_i	모의 수준	n_i	모의 수준
Chi(5)	10	0.040	10	0.046	10	0.048
	15	0.043	15	0.046	15	0.049
	20	0.040	20	0.040	20	0.042
	25	0.040	25	0.045	25	0.042
	50	0.037	50	0.038	50	0.040
	100	0.036	100	0.037	100	0.038
Chi(10)	10	0.042	10	0.045	10	0.045
	15	0.038	15	0.044	15	0.047
	20	0.036	20	0.039	20	0.040
	25	0.043	25	0.044	25	0.045
	50	0.041	50	0.040	50	0.042
	100	0.038	100	0.040	100	0.042
B(8,1)	10	0.058	10	0.060	10	0.066
	15	0.057	15	0.061	15	0.064
	20	0.049	20	0.051	20	0.055
	25	0.044	25	0.046	25	0.050
	50	0.037	50	0.037	50	0.039
	100	0.037	100	0.038	100	0.039

분포	$k = 3$ $n_1 = n_2 = n_3$		$k = 4$ $n_1 = n_2 = n_3 = n_4$		$k = 6$ $n_1 = n_2 = \dots = n_6$	
	n_i	모의 수준	n_i	모의 수준	n_i	모의 수준
CN(0.9,3)	10	0.020	10	0.018	10	0.016
	15	0.022	15	0.020	15	0.017
	20	0.014	20	0.012	20	0.008
	25	0.011	25	0.011	25	0.008
	50	0.009	50	0.007	50	0.006
	100	0.010	100	0.008	100	0.008
CN(0.8, 3)	10	0.017	10	0.015	10	0.011
	15	0.013	15	0.011	15	0.008
	20	0.012	20	0.012	20	0.009
	25	0.013	25	0.010	25	0.009
	50	0.011	50	0.011	50	0.009
	100	0.014	100	0.012	100	0.010

표 2a와 2b에서 볼 수 있듯이, 표본 크기가 작을 때 MC 검정은 균형 설계에서 대칭 및 근대칭 분포에 대해 대체로 보수적입니다. 반면에, 검정은 지수 및 베타(8, 1) 분포처럼 많이 치우친 분포에서 얻은 소규모 표본에 대해 개방적입니다. 그러나 표본 크기가 커지면 모의 유의 수준이 목표 유의 수준(0.05)에 근접합니다. 또한 표본 크기가 크지 않을 경우 표본 수는 표본 검정의 성능에 많은 영향을 미치지 않는 것으로 보입니다. 그러나 데이터가 특이치로 오염된 경우에는 검정 성능에 상당한 영향을 미칩니다. 데이터에 특이치가 있으면 검정이 일관되게, 그리고 지나치게 보수적입니다.

II부: 불균형 설계

당사는 불균형 설계에서 MC 검정 성능이 얼마나 효과적인지 조사하기 위한 시뮬레이션을 수행했습니다. 이를 위해, 앞의 시뮬레이션 B1에서 설명한 일련된 분포를 사용하여 동일한 분포로부터 3개 표본을 생성했습니다. 처음 실시한 실험에서 첫 두 표본의 크기는 $n_1 = n_2 = 10$ 이었고 세 번째 표본의 크기는 $n_3 = 15, 20, 25, 50, 100$ 이었습니다. 두 번째로 실시한

실험에서는 첫 두 표본의 크기가 $n_1 = n_2 = 15$ 이었고 세 번째 표본 집합의 크기가 $n_3 = 20, 25, 30, 50, 100$ 이었습니다. 세 번째로 실시한 실험에서는 최소 표본 크기를 20으로 설정했으며, 첫 두 표본의 크기는 $n_1 = n_2 = 20$ 이었고 세 번째 표본의 크기는 $n_3 = 25, 30, 40, 50, 100$ 이었습니다.

각 분포에서 추출한 동일한 표본 3개에 대해 목표 유의 수준이 $\alpha = 0.05$ 인 양측 MC 검정을 수행했습니다. 각각의 경우에 모의 유의 수준은 1만 쌍의 표본 반복실험을 기준으로 한 것이었고 5%의 목표 유의 수준을 사용했기 때문에 시뮬레이션 오류는 $\sqrt{0.95(0.05)/10,000} = 0.2\%$ 였습니다.

시뮬레이션 결과는 아래의 표 3a와 3b에 요약되어 있습니다.

표 3a 다중 표본 불균형 설계에서 다중 비교 검정의 모의 유의 수준. 검정의 목표 유의 수준은 0.05입니다.

분포	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	모의 수준	n_3	모의 수준	n_3	모의 수준
N(0,1)	15	0.032	20	0.040	25	0.045
	20	0.037	25	0.039	30	0.041
	25	0.038	30	0.037	40	0.043
	50	0.041	50	0.044	50	0.041
	100	0.042	100	0.042	100	0.044
t(5)	15	0.040	20	0.042	25	0.043
	20	0.036	25	0.040	30	0.037
	25	0.044	30	0.036	40	0.038
	50	0.033	50	0.036	50	0.035
	100	0.032	100	0.031	100	0.032

분포	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	모의 수준	n_3	모의 수준	n_3	모의 수준
t(10)	15	0.039	20	0.042	25	0.042
	20	0.038	25	0.041	30	0.040
	25	0.040	30	0.041	40	0.041
	50	0.037	50	0.043	50	0.042
	100	0.036	100	0.039	100	0.040
Lpl	15	0.059	20	0.060	25	0.054
	20	0.057	25	0.054	30	0.051
	25	0.056	30	0.051	40	0.050
	50	0.049	50	0.051	50	0.050
	100	0.048	100	0.047	100	0.046
B(3,3)	15	0.034	20	0.033	25	0.037
	20	0.031	25	0.035	30	0.039
	25	0.031	30	0.034	40	0.039
	50	0.036	50	0.039	50	0.038
	100	0.035	100	0.039	100	0.039
U(0,1)	15	0.027	20	0.030	25	0.032
	20	0.030	25	0.030	30	0.031
	25	0.028	30	0.032	40	0.036
	50	0.039	50	0.034	50	0.037
	100	0.042	100	0.038	100	0.042

분포	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	모의 수준	n_3	모의 수준	n_3	모의 수준
기대	15	0.061	20	0.053	25	0.042
	20	0.060	25	0.052	30	0.047
	25	0.054	30	0.049	40	0.043
	50	0.050	50	0.046	50	0.041
	100	0.044	100	0.040	100	0.040

표 3b 다중 표본 불균형 설계에서 MC 검정의 모의 유의 수준. 검정의 목표 유의 수준은 0.05입니다.

분포	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	모의 수준	n_3	모의 수준	n_3	모의 수준
Chi(5)	15	0.047	20	0.045	25	0.041
	20	0.043	25	0.042	30	0.039
	25	0.043	30	0.039	40	0.040
	50	0.039	50	0.037	50	0.040
	100	0.034	100	0.035	100	0.034
Chi(10)	15	0.043	20	0.042	25	0.042
	20	0.039	25	0.038	30	0.041
	25	0.040	30	0.041	40	0.038
	50	0.038	50	0.041	50	0.042
	100	0.035	100	0.034	100	0.035
B(8,1)	15	0.056	20	0.052	25	0.048
	20	0.054	25	0.046	30	0.044
	25	0.050	30	0.047	40	0.046

분포	$n_1 = n_2 = 10$		$n_1 = n_2 = 15$		$n_1 = n_2 = 20$	
	n_3	모의 수준	n_3	모의 수준	n_3	모의 수준
	50	0.046	50	0.043	50	0.043
	100	0.043	100	0.042	100	0.044
CN(0.9,3)	15	0.017	20	0.020	25	0.017
	20	0.020	25	0.019	30	0.012
	25	0.017	30	0.016	40	0.013
	50	0.019	50	0.016	50	0.012
	100	0.014	100	0.016	100	0.010
CN(0.8, 3)	15	0.012	20	0.013	25	0.013
	20	0.016	25	0.012	30	0.012
	25	0.014	30	0.010	40	0.010
	50	0.015	50	0.010	50	0.013
	100	0.012	100	0.011	100	0.010

표 3a와 3b에 나와 있는 모의 유의 수준은 앞에서 균형 설계 다중 표본에 대해 보고된 것과 일치합니다. 그러므로 MC 검정의 성능은 불균형 설계의 영향을 받지 않는 것으로 보입니다. 또한 최소 표본 크기가 20개 이상이면 모의 유의 수준은 오염된 데이터를 제외하고 목표 수준과 유사합니다.

결론적으로, 최소 표본 크기가 20개 이상일 때 MC 검정은 균형 및 불균형 설계에서 모두 다중 (k) 표본에 대해 효과적입니다. 그러나 표본이 더 작을 경우 검정은 대칭 및 근대칭 데이터에 대해 보수적이고 많이 치우친 데이터에 대해 개방적입니다.

부록 C: 이론적 검정력 함수

MC 검정의 정확한 이론적 검정력 함수는 제공되지 않습니다. 그러나 2-표본 설계에 대해서는 대표본 이론 방법에 기초한 근사 검정력 함수를 얻을 수 있습니다. 다중 표본 설계에 대해 유사한 근사를 도출하려면 더 많은 연구가 필요합니다.

그러나 2-표본 설계의 경우에는 대표본 이론 방법을 사용하여 Bonett 검정의 이론적 검정력 함수를 얻을 수 있습니다. 보다 구체적으로, 아래에 제시된 검정 통계 T 는 자유도가 1인 카이 제곱 분포로서 점근적으로 분포됩니다.

$$T = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\gamma - g_1}{n_1 - 1} + \frac{\gamma - g_2}{n_2 - 1}}$$

이 T 식에서 $\hat{\rho} = S_1/S_2$, $\rho = \sigma_1/\sigma_2$, $g_i = (n_i - 3)/n_i$ 및 γ 는 두 모집단의 알 수 없는 공통 첨도입니다.

따라서 근사 유의 수준이 α 인 양측 Bonett 등분산 검정의 이론적 검정력 함수는 다음 식으로 나타낼 수 있습니다.

$$\pi(n_1, n_2, \rho) = 1 - \Phi\left(z_{\alpha/2} - \frac{\ln \rho^2}{se}\right) + \Phi\left(-z_{\alpha/2} - \frac{\ln \rho^2}{se}\right)$$

설명

$$se = \sqrt{\frac{\gamma - g_1}{n_1 - 1} + \frac{\gamma - g_2}{n_2 - 1}}$$

단측 검정의 경우 $\sigma_1 > \sigma_2$ 에 대해 검정할 때 근사 검정력 함수는 다음과 같습니다.

$$\pi(n_1, n_2, \rho) = 1 - \Phi\left(z_{\alpha/2} - \frac{\ln \rho^2}{se}\right)$$

그리고 $\sigma_1 < \sigma_2$ 에 대해 검정할 때는 근사 검정력 함수가 다음과 같습니다.

$$\pi(n_1, n_2, \rho) = \Phi\left(-z_{\alpha/2} - \frac{\ln \rho^2}{se}\right)$$

데이터 분석의 표본 크기 계획 단계 도중에 모집단의 공통 척도 γ 는 알 수 없습니다. 따라서 γ 의 계획 값을 얻으려면 조사자는 일반적으로 전문가의 의견이나 이전 실험 결과에 의존해야 합니다. 해당 정보를 얻을 수 없으면 소규모 시험 연구를 수행하여 대규모 시험 계획을 수립하는 것이 좋은 경우가 많습니다. 계획 값 γ 는 시험 연구에서 추출한 표본을 사용하여 다음 식으로 표현되는 합동 척도 형태로 얻어집니다.

$$\hat{\gamma}_P = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (X_{1j} - m_1)^4 + \sum_{j=1}^{n_2} (X_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]^2}$$

보조 도구 메뉴에서 계획 추정치 γ 는 제공된 사용자 데이터에 따라 소급적으로 얻어집니다.

부록 D: 이론적 및 모의 검정력 비교

시뮬레이션 D1: Bonett 검정의 모의(실제) 검정력

당사는 Bonett 검정의 모의 검정력 수준을 부록 C에서 도출한 근사 검정력 함수에 기초한 검정력 수준에 비교하기 위한 시뮬레이션을 수행했습니다.

이 시뮬레이션을 위해 앞에서 설명한 각 분포에 대해 10,000쌍의 표본을 생성했습니다(시뮬레이션 B1 참조). 선택된 표본 크기는 대체로 검정의 모의 유의 수준이 앞의 시뮬레이션 B1에서 얻은 결과를 기준으로 목표 유의 수준과 합당할 정도로 유사할 만큼 컸습니다.

표준 편차 비율 $\rho = \sigma_1/\sigma_2 = 1/2$ 에서 모의 검정력 수준을 평가하기 위해 모든 표본 쌍의 두 번째 표본에 상수 2를 곱했습니다. 그 결과, 주어진 분포와 주어진 표본 크기 n_1 및 n_2 에 대해 모의 검정력 수준이 양측 Bonett 검정이 유의했던 표본 반복실험 10,000쌍의 분수로 계산되었습니다. 검정의 목표 유의 수준은 $\alpha = 0.05$ 로 고정되었습니다. 비교를 위해 부록 C에서 도출한 근사 검정력 함수를 토대로 해당되는 이론적 검정력 수준을 계산했습니다.

결과는 아래 표 4에 나와 있습니다.

표 4 모의 검정력 수준과 양측 Bonett 검정의 근사 검정력 수준 비교. 목표 유의 수준은 0.05입니다.

분포	n_1, n_2	평가자 검정력	모의 검정력	분포	n_1, n_2	평가자 검정력	모의 검정력
N(0,1)	20, 10	0.627	0.527	기대	20, 10	0.222	0.227
	20, 20	0.83	0.765		20, 20	0.322	0.368
	20, 30	0.896	0.846		20, 30	0.377	0.434
	20, 40	0.925	0.886		20, 40	0.412	0.475
	30, 15	0.825	0.771		30, 15	0.32	0.307
	30, 30	0.954	0.925		30, 30	0.458	0.50
	30, 45	0.98	0.97		30, 45	0.531	0.579

분포	n_1, n_2	평가자 검정력	모의 검정력	분포	n_1, n_2	평가자 검정력	모의 검정력
	30, 60	0.989	0.984		30, 60	0.575	0.622
t(5)	20, 10	0.222	0.379	Chi(5)	20, 10	0.355	0.347
	20, 20	0.322	0.569		20, 20	0.517	0.53
	20, 30	0.377	0.637		20, 30	0.597	0.616
	20, 40	0.412	0.69		20, 40	0.644	0.661
	30, 15	0.32	0.545		30, 15	0.513	0.51
	30, 30	0.458	0.733		30, 30	0.701	0.711
	30, 45	0.531	0.795		30, 45	0.781	0.793
	30, 60	0.575	0.828		30, 60	0.823	0.833
t(10)	20, 10	0.476	0.45	Chi(10)	20, 10	0.454	0.414
	20, 20	0.673	0.673		20, 20	0.646	0.631
	20, 30	0.756	0.749		20, 30	0.73	0.717
	20, 40	0.80	0.803		20, 40	0.776	0.771
	30, 15	0.668	0.659		30, 15	0.641	0.618
	30, 30	0.85	0.852		30, 30	0.828	0.819
	30, 45	0.91	0.911		30, 45	0.892	0.882
	30, 60	0.936	0.937		30, 60	0.921	0.912
Lpl	20, 10	0.321	0.33	B(8,1)	20, 10	0.363	0.278
	20, 20	0.469	0.519		20, 20	0.528	0.463
	20, 30	0.545	0.585		20, 30	0.609	0.549
	20, 40	0.59	0.632		20, 40	0.655	0.60
	30, 15	0.466	0.475		30, 15	0.524	0.419

분포	n_1, n_2	평가자 검정력	모의 검정력	분포	n_1, n_2	평가자 검정력	모의 검정력
	30, 30	0.647	0.673		30, 30	0.713	0.634
	30, 45	0.729	0.758		30, 45	0.792	0.737
	30, 60	0.773	0.80		30, 60	0.833	0.777
B(3,3)	20, 10	0.777	0.628	CN(0.9,3)	20, 10	0.238	0.284
	20, 20	0.939	0.869		20, 20	0.346	0.452
	20, 30	0.973	0.936		20, 30	0.405	0.517
	20, 40	0.984	0.964		20, 40	0.442	0.561
	30, 15	0.935	0.871		30, 15	0.343	0.374
	30, 30	0.993	0.98		30, 30	0.491	0.598
	30, 45	0.998	0.995		30, 45	0.567	0.70
	30, 60	0.999	0.999		30, 60	0.612	0.719
U(0,1)	20, 10	0.916	0.74	CN(0.8,3)	20, 10	0.26	0.223
	20, 20	0.992	0.95		20, 20	0.379	0.396
	20, 30	0.998	0.985		20, 30	0.444	0.467
	20, 40	0.999	0.995		20, 40	0.484	0.52
	30, 15	0.991	0.941		30, 15	0.376	0.354
	30, 30	1.0	0.996		30, 30	0.535	0.549
	30, 45	1.0	1.0		30, 45	0.614	0.65
	30, 60	1.0	1.0		30, 60	0.661	0.706

결과는 근사 검정력 수준과 모의 검정력 수준이 대체로 서로 유사함을 보여줍니다. 표본 크기가 커질수록 더 유사해집니다. 꼬리의 크기가 중간 이하인 대칭 및 근대칭 분포에서 근사 검정력 수준은 대체로 모의 검정력 수준보다 약간 더 큼니다. 그러나 꼬리가 두꺼운 대칭

분포나 치우친 분포에서는 모의 검정력 수준보다 약간 더 작습니다. 두 검정력 함수의 차이는 표본을 자유도가 5인 t 분포에서 생성하는 경우를 제외하고 대체로 중요하지 않습니다.

전체적으로 최소 표본 크기가 20에 도달하면 근사 검정력 수준과 모의 검정력 수준이 매우 유사합니다. 따라서 근사 검정력 함수를 토대로 표본 크기를 계획할 수 있습니다.

© 2015, 2017 Minitab Inc. All rights reserved.

Minitab®, Quality. Analysis. Results.® and the Minitab® logo are all registered trademarks of Minitab, Inc., in the United States and other countries. See minitab.com/legal/trademarks for more information.