



일원 분산 분석

개요

일원 분산 분석은 그룹 3개 이상의 평균을 비교하여 서로 유의하게 다른지 판단하기 위해 사용합니다. 또 하나의 중요한 기능은 특정 그룹 간의 차이를 추정하는 것입니다.

일원 분산 분석에서 그룹 간의 차이를 탐지하는 가장 흔한 방법은 모든 표본의 모집단이 같지만 알 수 없는 표준 편차를 공유한다는 가정에 기초한 F-검정입니다. 실제로는 표본의 표준 편차가 서로 다른 경우가 많은 것으로 나타났습니다. 따라서 F-검정 대신 다른 표준 편차를 처리할 수 있는 Welch 방법을 조사하고자 했습니다. 표준 편차가 같지 않은 표본에 대해 설명하는 여러 비교를 계산하는 방법도 개발하고자 했습니다. 이 방법을 사용하여 서로 다른 그룹을 쉽게 구분하는 방법을 제공하는 개별 구간을 그래프로 표시할 수 있습니다.

이 문서에서는 다음에 대한 Minitab 보조 도구 일원 분산 분석 절차에 사용된 방법을 개발한 방법에 대해 설명합니다.

- Welch 검정
- 다중 비교 구간

또한 비정상 데이터 유무, 검정의 표본 크기와 검정력, 그리고 데이터의 정규성을 포함하여 일원 분산 분석 결과의 유효성에 영향을 미칠 수 있는 다양한 조건에 대해서도 조사했습니다. 보조 도구에서는 이러한 조건에 따라 데이터에 대해 다음과 같은 검사를 자동으로 수행하고 검사 결과를 보고서 카드에 표시합니다.

- 비정상 데이터
- 표본 크기
- 데이터 정규성

이 문서에서는 이러한 조건이 실제로 일원 분산 분석과 어떤 관계가 있는지 조사하고 보조 도구에서 해당 조건을 확인하기 위한 가이드라인을 정한 방법에 대해 설명합니다.

일원 분산 분석 방법

F-검정과 Welch 검정의 비교

일원 분산 분석에 널리 사용되는 F-검정은 같지만 알 수 없는 표준 편차(σ)를 모든 그룹이 공유한다는 가정에 기초합니다. 실제로 이 가정은 참인 경우가 드물며, 이로 인해 제1종 오류율을 관리하는 데 문제가 발생합니다. 제1종 오류는 귀무 가설을 잘못 기각할 (표본이 다르지 않은데 유의하게 다르다고 결론 지을) 확률입니다. 표본의 표준 편차가 서로 다르면 검정이 잘못된 결론에 도달할 가능성이 더 많습니다. 이 문제를 해결하기 위해 F-검정에 대한 대안으로 Welch 검정이 개발되었습니다(Welch, 1951).

목적

당사는 보조 도구의 일원 분산 분석 절차에 F-검정과 Welch 검정 중 무엇을 사용할 것인지 결정하고자 했습니다. 이를 위해 F-검정과 Welch 검정의 실제 검정 결과가 검정에 대한 목표 유의 수준(알파 또는 제1종 오류율)과 얼마나 일치하는지, 즉 검정이 주어진 표본 크기와 표본 표준 편차에서 귀무 가설을 잘못 기각하는 횟수가 목표값보다 많은지 적은지 평가해야 했습니다.

방법

F-검정과 Welch 검정을 비교하기 위해 표본 수, 표본 크기 및 표본 표준 편차를 바꿔가면서 여러 시뮬레이션을 수행했습니다. 각 조건에 대해 F-검정과 Welch 방법을 모두 사용하여 분산 분석 검정을 10,000회 수행했습니다. 표본의 평균이 같아서 각 검정에 대해 귀무 가설이 참이 되도록 랜덤 데이터를 생성했습니다. 그런 다음, 목표 유의 수준 0.05와 0.01을 사용하여 검정을 수행했습니다. 10,000회의 검정 중에서 F-검정과 Welch 검정에 의해 귀무 가설이 실제로 기각된 횟수를 집계하고 이 비율을 목표 유의 수준과 비교했습니다. 검정이 효과적이면 추정 제1종 오류가 목표 유의 수준과 매우 유사합니다.

결과

모든 검정 조건에서 Welch 방법이 F-검정 이상으로 효과적인 것으로 나타났습니다. 예를 들어 Welch 검정을 사용하여 표본 5개를 비교할 때 제1종 오류율은 0.0460과 0.0540 사이로, 목표 유의 수준 0.05와 매우 가까웠습니다. 이 결과는 Welch 방법의 제1종 오류율이 표본의 표본 크기와 표준 편차가 다를 때도 목표값과 일치함을 나타냅니다.

반면에 F-검정의 제1종 오류율은 0.0273과 0.2277 사이입니다. 다음과 같은 조건에서는 F-검정이 특히 효과적이지 못했습니다.

- 가장 큰 표본의 표준 편차가 가장 클 경우에도 제1종 오류율은 0.05 미만으로 하락했습니다. 이 조건으로 인해 검정이 더 보수적이 되며, 단순히 표본 크기를 늘리는 것은 표본의 표준 편차가 같지 않을 때 효과적인 해결책이 되지 않음을 보여줍니다.
- 표본 크기가 같지만 표준 편차가 다르면 제1종 오류율이 0.05보다 높았습니다. 표준 편차가 더 큰 표본이 나머지 표본보다 크기가 작을 경우에도 오류율이 0.05보다 높았습니다. 특히 더 작은 표본의 표준 편차가 더 크면 F-검정에 의해 귀무 가설이 잘못 기각될 위험이 크게 증가합니다.

시물레이션 방법 및 결과에 대한 자세한 내용은 부록 A를 참조하십시오.

Welch 방법은 표본의 표준 편차 및 크기가 같지 않을 때 효과적이었기 때문에 당사는 Welch 방법을 보조 도구의 일원 분산 분석 절차로 사용합니다.

비교 구간

분산 분석 검정이 통계적으로 유의하여 표본 중 하나 이상의 평균이 나머지 표본과 다른 것으로 나타나면 다음 분석 단계에서는 어떤 표본이 통계적으로 다른지 확인합니다. 이를 비교하는 직관적인 방법으로, 각 표본과 연관된 신뢰 구간을 그래프로 표시하고 구간이 겹치지 않는 표본을 식별합니다. 그러나 개별 신뢰 구간은 비교에 사용되지 않기 때문에 그래프를 근거로 내린 결론은 검정 결과와 일치하지 않을 수 있습니다. 표준 편차가 같은 표본의 다중 비교 방법에 대해서는 출판물에서 다뤄진 적이 있지만, 이 방법을 확대하여 표준 편차가 다른 표본에 대해 설명하는 데도 사용할 수 있어야 했습니다.

목적

당사는 여러 표본을 비교하는 데 사용할 수 있고 검정 결과와도 최대한 일치하는 개별 비교 구간을 계산하는 방법을 개발하고자 했습니다. 또한 나머지 표본과 통계적으로 다른 표본이 무엇인지 확인하는 시각적인 방법도 제공하고자 했습니다.

방법

기본적인 여러 다중 비교 방법(Hsu 1996)은 각 평균 쌍 간의 차이에 대한 구간을 제공하면서 다중 비교 시 증가하는 오류를 관리합니다. 표본 크기가 같고 표준 편차가 같다고 가정되는 특별한 경우에는 각 평균에 대한 개별 구간을 모든 쌍의 차이에 대한 구간과 정확히 일치하도록 표시할 수 있습니다. 표본 크기가 다르고 표준 편차가 같다고 가정되는 경우에 대해 Hochberg, Weiss & Hart(1982)는 Tukey-Kramer의 다중 비교 방법에 기초한 쌍 간 차이에 대한 구간과 대략적으로 같은 개별 구간을 개발했습니다. 보조 도구에서는 이와 동일한 방식을 표준 편차가 같다고 가정하지 않는 Games-Howell 다중 비교 방법에 적용합니다. Minitab 16 릴리스의 보조 도구에서 사용된 방식은 Games-Howell 방식과 개념이 유사했지만, Games-Howell 방식을 직접 응용한 것은 아니었습니다. 자세한 내용은 부록 B를 참조하십시오.

결과

보조 도구는 일원 분산 분석 요약 보고서의 평균 비교 차트에 비교 구간을 표시합니다. 분산 분석 검정이 통계적으로 유의하면, 하나 이상의 다른 구간과 겹치지 않는 비교 구간이 모두 빨간색으로 표시됩니다. 검정과 비교 구간이 불일치할 수 있지만, 두 가지 방법 모두 귀무 가설이 참일 때 기각할 확률이 같기 때문에 이런 결과가 나오는 경우는 드뭅니다. 분산 분석 검정이 유의하지만 모든 구간이 겹치면 겹치는 부분이 가장 작은 쌍이 빨간색으로 표시됩니다. 분산 분석 검정이 통계적으로 유의하지 않으면 일부 구간이 겹치지 않는 경우에도 빨간색으로 표시되는 구간이 없습니다.

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. 따라서 이런 데이터 점은 조사할만한 가치가 많으며, 가능하면 수정해야 합니다.

목적

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



방법

Minitab에서는 Hoaglin, Iglewicz & Tukey(1986)가 설명한 상자 그림에서 특이치를 식별하는 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 25번째 및 75번째 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 확인할 때 보조 도구 보고서 카드에는 다음과 같은 상태가 표시됩니다.

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

표본 크기

검정력은 실제로 존재하는 유의한 효과 또는 차이를 찾을 수 있는 확률을 나타내므로, 모든 가설 검정에서 중요한 속성입니다. 검정력은 귀무 가설을 기각하고 대립 가설을 인정할 확률입니다. 검정의 검정력을 높일 수 있는 가장 쉬운 방법 중 하나는 표본 크기를 늘리는 것입니다. 보조 도구에서는 검정력이 낮은 검정에 대해 지정된 차이를 찾기 위해 필요한 표본 크기를 표시합니다. 차이가 지정되지 않으면 적절한 검정력으로 탐지할 수 있는 차이를 보고합니다. 보조 도구에서는 정확한 검정력 공식이 없는 Welch 방법을 사용하기 때문에 이 정보를 제공하기 위해 검정력을 계산할 방법을 개발해야 했습니다.

목적

검정력 계산 방법을 개발하기 위해서는 두 가지 문제를 해결해야 했습니다. 첫째, 보조 도구에서는 사용자에게 전체 평균 집합을 입력할 것을 요구하지 않고 실질적인 영향이 있는 평균의 차이만 입력할 것을 요구합니다. 모든 주어진 차이에 대해 해당 차이를 생성할 수 있는 가능한 평균 구성의 수는 무한합니다. 따라서 가능한 모든 평균 구성에 대해 검정력을 계산할 수 없기 때문에 검정력을 계산할 때 사용할 평균을 결정할 합리적인 방법을 개발해야 했습니다. 둘째, 보조 도구에서는 같은 표본 크기나 표준 편차가 필요하지 않은 Welch 방법을 사용하기 때문에 검정력을 계산할 방법을 개발해야 했습니다.

방법

가능한 평균 구성의 수가 무한하다는 문제를 해결하기 위해 Minitab의 표준 일원 분산 분석 절차에 사용되는 방식에 기초한 방법을 개발했습니다(통계분석 > 분산 분석 > 일원 분산 분석). 두 평균만 지정된 양만큼 차이가 나고 나머지 평균이 같은 (평균의 가중 평균으로 설정된) 경우에 초점을 맞췄습니다. 이 방식에서는 (3개 이상이 아닌) 2개의 평균만 전체 평균과 다르다고 가정하기 때문에 보수적인 추정 검정력이 얻어집니다. 그러나 표본의 크기 또는 표준 편차가 다를 수 있기 때문에 검정력 계산은 다르다고 가정되는 2개의 평균에 따라 달라집니다.

이 문제를 해결하기 위해 최선 및 최악의 경우를 나타내는 평균 두 쌍을 식별합니다. 최악의 경우는 표본 분산에 비해 표본 크기가 비교적 작고 검정력이 최소화될 때 발생하며, 최선의 경우는 표본 분산에 비해 표본 크기가 비교적 크고 검정력이 극대화될 때 발생합니다. 모든 검정력 계산에서는 정확히 2개의 평균이 평균의 전체 가중 평균과 다르다는 가정에서 검정력을 최소화하고 극대화하는 두 가지의 극단적인 경우를 고려합니다.

검정력 계산을 개발하기 위해 Kulinskaya 외 (2003)에 제시된 방법이 사용되었습니다. 시뮬레이션에서 얻은 검정력 계산과 평균 구성의 문제를 해결하기 위해 개발한 방법과 Kulinskaya 외 (2003)에 제시된 방법을 비교했습니다. 검정력이 평균 구성에 따라 어떻게 달라지는지 더 분명하게 보여주는 다른 검정력 근사도 조사했습니다. 검정력 계산에 대한 자세한 내용은 부록 C를 참조하십시오.

결과






세 가지 방법을 비교한 결과 Kulinskaya 방법이 우수한 검정력 근사를 제공하고 당사의 구성 평균 처리 방법이 적절한 것으로 나타났습니다.

데이터로부터 귀무 가설을 반박하는 충분한 증거를 얻을 수 없으면 보조 도구는 주어진 표본 크기에 대해 80% 및 90% 확률로 탐지할 수 있는 실제 차이를 계산합니다. 또한 실제 차이를 지정하면 보조 도구가 해당 차이에 대해 최소 및 최대 검정력 값을 계산합니다.

검정력 값이 90% 미만이면 보조 도구는 지정된 차이와 관측된 표본 표준 편차를 토대로 표본 크기를 계산합니다. 표본 크기를 통해 최소 및 최대 검정력 값이 모두 90% 이상이 되는 결과를 생성하기 위해 지정된 차이가 변동성이 가장 큰 두 평균 사이에 있다고 가정합니다.

사용자가 차이를 지정하지 않으면 보조 도구는 검정력 값 범위의 최대값이 60%인 최대 차이를 찾습니다. 이 값은 검정력 보고서에서 60% 검정력에 해당되는 빨간색 막대와 노란색 막대 사이의 경계에 표시됩니다. 검정력 값 범위의 최소값이 90%인 가장 작은 차이도 찾습니다. 이 값은 검정력 보고서에서 90% 검정력에 해당되는 노란색 막대와 녹색 막대 사이의 경계에 표시됩니다.

검정력과 표본 크기를 확인할 때 보조 도구의 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	조건
	데이터에서 평균 간에 차이가 있다는 결론을 내릴 만한 충분한 증거를 얻을 수 없음. 차이가 지정되지 않음.
	검정에서 평균의 차이가 탐지되므로 검정력은 문제가 되지 않음. 또는 검정력이 충분함. 검정에서 평균의 차이를 찾지 못했지만 표본 크기는 주어진 차이를 90% 이상의 확률로 탐지하기에 충분함.
	검정력이 충분할 수 있음. 검정에서 평균의 차이를 찾지 못했지만 표본 크기는 주어진 차이를 80% - 90%의 확률로 탐지하기에 충분함. 90% 검정력을 달성하기 위해 필요한 표본 크기가 보고됨.
	검정력이 충분하지 않을 수 있음. 검정에서 평균의 차이를 찾지 못했지만 표본 크기는 주어진 차이를 60% - 80%의 확률로 탐지하기에 충분함. 80% 검정력과 90% 검정력을 달성하는 데 필요한 표본 크기가 보고됨.
	검정력이 충분하지 않음. 검정에서 평균의 차이를 찾지 못했으며 표본 크기는 주어진 차이를 60% 이상의 확률로 탐지하기에 충분하지 않음. 80% 검정력과 90% 검정력을 달성하는 데 필요한 표본 크기가 보고됨.

정규성

많은 통계분석 방법에서는 흔히 데이터가 정규 분포를 따른다고 가정합니다. 데이터가 정규 분포를 따르지 않더라도 정규성 가정에 기초한 방법이 매우 효과적일 수 있습니다. 이 내용은 표본 평균의 분포가 근사 정규 분포를 따르고 표본 크기가 커질수록 근사 정규 분포가 거의 정규 분포가 된다는 중심 극한 정리에 의해 부분적으로 설명됩니다.

목적

당사의 목표는 합당한 수준으로 양호한 정규 분포의 근사를 제공하려면 표본이 얼마나 커야 하는지 확인하는 것이었습니다. 당사는 Welch 검정과 비교 구간을 다양한 비정규 분포를 따르고 크기가 중간 이하인 표본을 사용하여 조사하고자 했습니다. 그리고 Welch 방법과 비교 구간의 실제 검정 결과가 검정에 대해 선택된 유의 수준(알파 또는 제1종 오류율)과 얼마나 일치하는지, 즉 검정이 서로 다른 표본 크기 및 수준 수와 비정규 분포가 제공될 경우 귀무 가설을 예상보다 더 많이 잘못 기각하는지 확인하고자 했습니다.

방법



제1종 오류를 평가하기 위해 표본 수, 표본 크기 및 데이터 분포를 바꿔가면서 여러 시뮬레이션을 수행했습니다. 시뮬레이션에는 정규 분포에서 크게 이탈한 치우친 분포와 꼬리가 두꺼운 분포가 포함되었습니다. 크기와 표준 편차는 각 검정의 표본 간에 일정했습니다.

각 조건에 대해 Welch 방법과 비교 구간을 사용하여 분산 분석 검정을 10,000번 수행했습니다. 표본의 평균이 같아서 각 검정에 대해 귀무 가설이 참이 되도록 랜덤 데이터를 생성했습니다. 그런 다음, 목표 유의 수준 0.05를 사용하여 검정을 수행했습니다. 10,000회의 검정 중에서 검정에 의해 귀무 가설이 실제로 기각된 횟수를 집계하고 이 비율을 목표 유의 수준에 비교했습니다. 비교 구간에 대해서는 10,000회 검정 중에서 구간이 차이를 하나 이상 나타낸 횟수를 집계했습니다. 검정이 효과적이면 제1종 오류율이 목표 유의 수준과 매우 유사할 것입니다.

결과

각 검정과 비교 구간은 전반적으로 표본 크기가 최소 10개 또는 15개인 모든 조건에서 매우 효과적입니다. 수준이 9개 이하인 검정에서 대부분의 경우 결과는 모두 표본 크기가 10개일 때 목표 유의 수준의 3% 이내였고 표본 크기가 15개일 때 2% 이내였습니다. 수준이 10개 이하인 검정에서 대부분의 경우 결과는 표본 크기가 15개일 때 3% 이내였고 표본 크기가 20개일 때 2% 이내였습니다. 자세한 내용은 부록 D를 참조하십시오.

검정은 표본이 상대적으로 작을 때 효과적이기 때문에 보조 도구는 데이터의 정규성을 검사하지 않습니다. 대신, 보조 도구는 표본 크기를 확인하고 수준이 2-9개일 경우 표본이 15개 미만이고 수준이 10-12개일 경우 표본이 20개 미만을 나타냅니다. 이런 결과를 토대로 보조 도구의 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	조건
	표본 크기가 15개 또는 20개 이상이므로 정규성은 문제가 되지 않음.
	일부 표본 크기가 15개 또는 20개 미만이므로 정규성이 문제가 될 수 있음.

참고 문헌

- Dunnett, C. W. (1980). Pairwise Multiple Comparisons in the Unequal Variance Case. *Journal of the American Statistical Association*, 75, 796-800.
- Hoaglin, D. C., Iglewicz, B. & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81, 991-999.
- Hochberg, Y., Weiss G. & Hart, S. (1982). On graphical procedures for multiple comparisons. *Journal of the American Statistical Association*, 77, 767-772.
- Hsu, J. (1996). *Multiple comparisons: Theory and methods*. Boca Raton, FL: Chapman & Hall.
- Kulinskaya, E., Staudte, R. G. & Gao, H. (2003). Power approximations in testing for unequal means in a One-Way ANOVA weighted for unequal variances, *Communication in Statistics*, 32 (12), 2353-2371.
- Welch, B.L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34, 28-35
- Welch, B.L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika* 38, 330-336.

부록 A: F-검정과 Welch 검정의 비교

F-검정은 표준 편차가 같다는 가정을 위반할 때 제1종 오류율이 높아지는 결과를 초래할 수 있습니다. Welch 검정은 이런 문제를 방지하기 위해 만들어졌습니다.

Welch 검정

k 모집단에서 크기가 n_1, \dots, n_k 인 랜덤 표본을 관측합니다. μ_1, \dots, μ_k 는 모집단 평균을 나타내고 $\sigma_1^2, \dots, \sigma_k^2$ 은 모집단 분산을 나타냅니다. $\bar{x}_1, \dots, \bar{x}_k$ 는 표본 평균을 나타내고 s_1^2, \dots, s_k^2 은 표본 분산을 나타냅니다. 이 때 다음 가설을 검정하고자 합니다.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

$$\text{일부 } i, j \text{에 대해 } H_1: \mu_i \neq \mu_j.$$

k 평균의 동일성을 검정하는 Welch 검정에서는 다음 통계를

$$W^* = \frac{\sum_{j=1}^k w_j (\bar{x}_j - \hat{\mu})^2 / (k-1)}{1 + [2(k-2)/(k^2-1)] \sum_{j=1}^k h_j}$$

F(k - 1, f) 분포에 비교합니다. 여기서,

$$w_j = \frac{n_j}{s_j^2},$$

$$W = \sum_{j=1}^k w_j,$$

$$\hat{\mu} = \frac{\sum_{j=1}^k w_j \bar{x}_j}{W},$$

$$h_j = \frac{(1 - w_j/W)^2}{n_j - 1} \text{ 및}$$

$$f = \frac{k^2 - 1}{3 \sum_{j=1}^k h_j} \text{입니다.}$$

Welch 검정은 $W^* \geq F_{k-1, f, 1-\alpha}$ (확률 α 로 초과되는 F 분포의 백분위수)일 경우 귀무 가설을 기각합니다.

동일하지 않은 표준 편차

다음 항목에서는 표준 편차가 같다는 가정 위반 시 F-검정의 민감도에 대해 설명하고 이 검정을 Welch 검정과 비교합니다.

아래 결과는 $N(0, \sigma^2)$ 를 따르는 5개의 표본을 사용한 일원 분산 분석의 결과입니다. 각 행은 F-검정과 Welch 검정을 사용한 10,000번에 걸친 시뮬레이션에 기초하고 있습니다. 5번째 표본의 표준 편차를 다른 표본보다 2배 및 4배로 높임으로써 표준 편차에 대한 두 가지 조건을 검정했습니다. 표본 크기에 대한 세 가지 조건, 즉 표본 크기가 같고, 5번째 표본이 나머지보다 크고, 5번째 표본이 나머지보다 작은 경우를 검정했습니다.

표 1 목표 유의 수준 $\alpha = 0.05$ 일 때 모의 F-검정 및 Welch 검정의 제1종 오류율

표준 편차 ($\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5$)	표본 크기 (n_1, n_2, n_3, n_4, n_5)	F-검정	Welch 검정
1, 1, 1, 1, 2	10, 10, 10, 10, 20	.0273	.0524
1, 1, 1, 1, 2	20, 20, 20, 20, 20	.0678	.0462
1, 1, 1, 1, 2	20, 20, 20, 20, 10	.1258	.0540
1, 1, 1, 1, 4	10, 10, 10, 10, 20	.0312	.0460
1, 1, 1, 1, 4	20, 20, 20, 20, 20	.1065	.0533
1, 1, 1, 1, 4	20, 20, 20, 20, 10	.2277	.0503

표본 크기가 같을 때(2번 및 5번 행) F-검정이 귀무 가설을 잘못 기각할 확률은 목표값인 0.05보다 크고 표본 편차가 더 많이 다를수록 확률이 증가합니다. 표준 편차가 가장 큰 표본의 크기가 감소하면 문제가 더욱 악화됩니다. 반면에 표준 편차가 가장 큰 표본의 크기를 늘리면 기각 확률이 감소합니다. 그러나 표본 크기를 너무 많이 늘리면 기각 확률이 너무 작아져서 귀무 가설 하에서 검정이 필요 이상으로 보수적이 될 뿐만 아니라 대립 가설 하에서도 검정의 검정력에 부정적인 영향을 미칩니다. 이런 결과를 모든 경우 목표 유의 수준 0.05와 가깝게 일치하는 Welch 검정과 비교하십시오.

다음, $k = 7$ 표본의 경우에 대해 시뮬레이션을 실시했습니다. 표의 각 행에는 10,000번 시뮬레이션한 F-검정이 요약되어 있습니다. 표준 편차와 표본 크기를 다양하게 변경했습니다. 목표 유의 수준은 $\alpha = 0.05$ 와 $\alpha = 0.01$ 입니다. 위와 마찬가지로 목표 값과 매우 큰 차이가 있을 수 있습니다. 변동성이 더 클 때 더 작은 표본 크기를 사용하면 제1종 오류 발생 확률이 매우 높아지는 한편, 더 큰 표본을 사용하면 검정이 극도로 보수적이 될 수 있습니다. 결과는 아래 표 2에 나와 있습니다.

표 2 표본이 7개인 F-검정의 제1종 오류율

표준 편차 ($\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7$)	표본 크기 ($n_1, n_2, n_3, n_4, n_5, n_6, n_7$)	목표값 $\alpha = 0.05$	목표값 $\alpha = 0.01$
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	21, 21, 21, 21, 22, 22, 12	0.0795	0.0233
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	20, 21, 21, 21, 21, 24, 12	0.0785	0.0226
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	20, 21, 21, 21, 21, 21, 15	0.0712	0.0199
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	20, 20, 20, 21, 21, 23, 15	0.0719	0.0172
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	20, 20, 20, 20, 21, 21, 18	0.0632	0.0166
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	20, 20, 20, 20, 20, 20, 20	0.0576	0.0138
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	18, 19, 19, 20, 20, 20, 24	0.0474	0.0133
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	18, 18, 18, 18, 18, 18, 32	0.0314	0.0057

표준 편차 ($\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7$)	표본 크기 ($n_1, n_2, n_3, n_4, n_5, n_6, n_7$)	목표값 $\alpha = 0.05$	목표값 $\alpha = 0.01$
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	15, 18, 18, 19, 20, 20, 30	0.0400	0.0085
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	12, 18, 18, 18, 19, 19, 36	0.0288	0.0064
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	15, 15, 15, 15, 15, 15, 50	0.0163	0.0025
1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.9	12, 12, 12, 12, 12, 12, 68	0.0052	0.0002
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	21, 21, 21, 21, 22, 22, 12	0.1097	0.0436
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	20, 21, 21, 21, 21, 24, 12	0.1119	0.0452
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	20, 21, 21, 21, 21, 21, 15	0.0996	0.0376
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	20, 20, 20, 21, 21, 23, 15	0.0657	0.0345
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	20, 20, 20, 20, 21, 21, 18	0.0779	0.0283
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	20, 20, 20, 20, 20, 20, 20	0.0737	0.0264
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	18, 19, 19, 20, 20, 20, 24	0.0604	0.0204
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	18, 18, 18, 18, 18, 18, 32	0.0368	0.0122
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	15, 18, 18, 19, 20, 20, 30	0.0390	0.0117
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	12, 18, 18, 18, 19, 19, 36	0.0232	0.0046
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	15, 15, 15, 15, 15, 15, 50	0.0124	0.0026
1.75, 1.75, 1.75, 1.75, 1.75, 1.75, 3.5	12, 12, 12, 12, 12, 12, 68	0.0027	0.0004
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	21, 21, 21, 21, 22, 22, 12	0.134	0.0630
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	20, 21, 21, 21, 21, 24, 12	0.1329	0.0654
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	20, 21, 21, 21, 21, 21, 15	0.1101	0.0484
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	20, 20, 20, 21, 21, 23, 15	0.1121	0.0495
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	20, 20, 20, 20, 21, 21, 18	0.0876	0.0374
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	20, 20, 20, 20, 20, 20, 20	0.0808	0.0317
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	18, 19, 19, 20, 20, 20, 24	0.0606	0.0243

표준 편차 ($\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7$)	표본 크기 ($n_1, n_2, n_3, n_4, n_5, n_6, n_7$)	목표값 $\alpha = 0.05$	목표값 $\alpha = 0.01$
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	18, 18, 18, 18, 18, 18, 32	0.0356	0.0119
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	15, 18, 18, 19, 20, 20, 30	0.0412	0.0134
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	12, 18, 18, 18, 19, 19, 36	0.0261	0.0068
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	15, 15, 15, 15, 15, 15, 50	0.0100	0.0023
1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 1.68333, 3.9	12, 12, 12, 12, 12, 12, 68	0.0017	0.0003
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	21, 21, 21, 21, 22, 22, 12	0.1773	0.1006
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	20, 21, 21, 21, 21, 24, 12	0.1811	0.1040
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	20, 21, 21, 21, 21, 21, 15	0.1445	0.0760
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	20, 20, 20, 21, 21, 23, 15	0.1448	0.0786
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	20, 20, 20, 20, 21, 21, 18	0.1164	0.0572
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	20, 20, 20, 20, 20, 20, 20	0.1020	0.0503
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	18, 19, 19, 20, 20, 20, 24	0.0834	0.0369
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	18, 18, 18, 18, 18, 18, 32	0.0425	0.0159
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	15, 18, 18, 19, 20, 20, 30	0.0463	0.0168
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	12, 18, 18, 18, 19, 19, 36	0.0305	0.0103
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	15, 15, 15, 15, 15, 15, 50	0.0082	0.0021
1.55, 1.55, 1.55, 1.55, 1.55, 1.55, 4.7	12, 12, 12, 12, 12, 12, 68	0.0013	0.0001

부록 B: 비교 구간

평균 비교 차트를 사용하여 모집단 평균 간 차이의 통계적인 유의성을 평가할 수 있습니다.

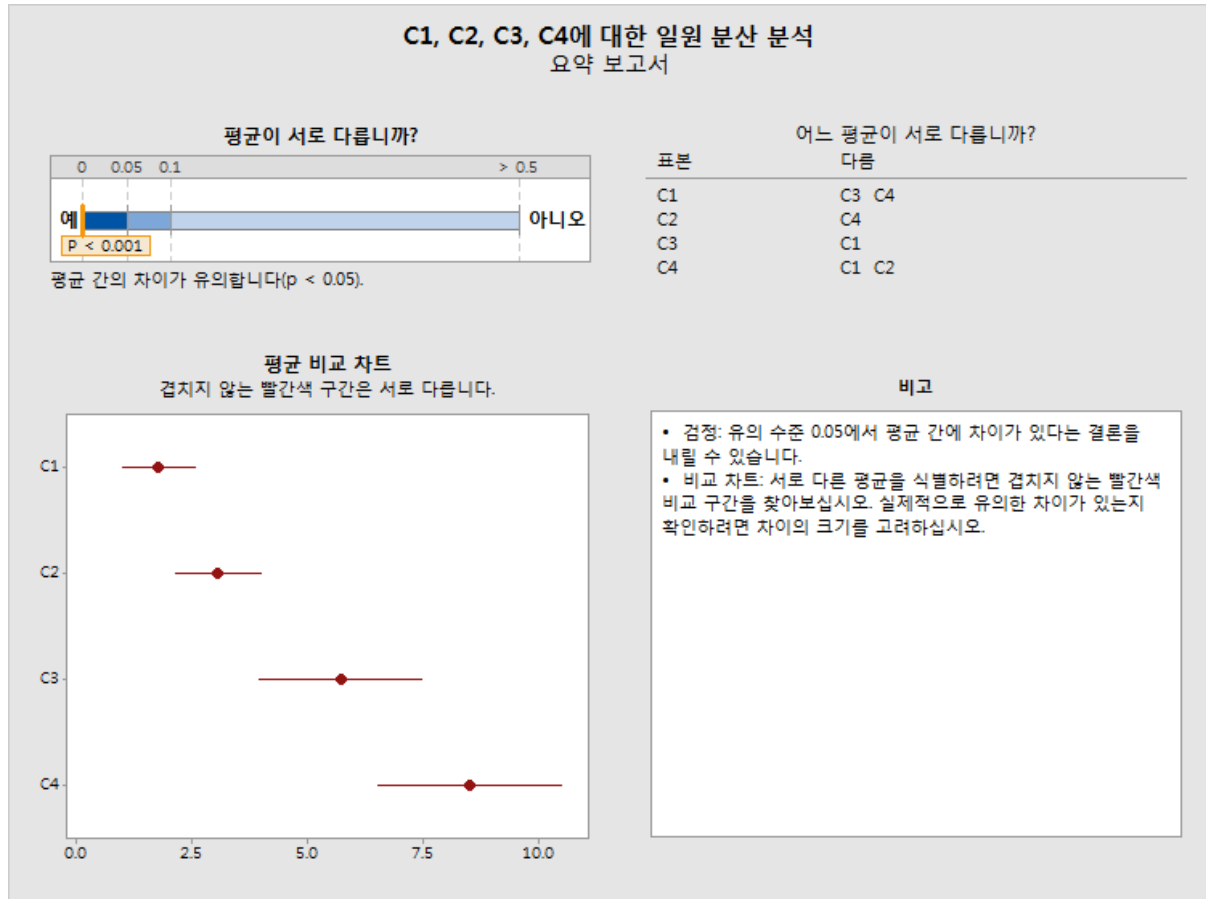
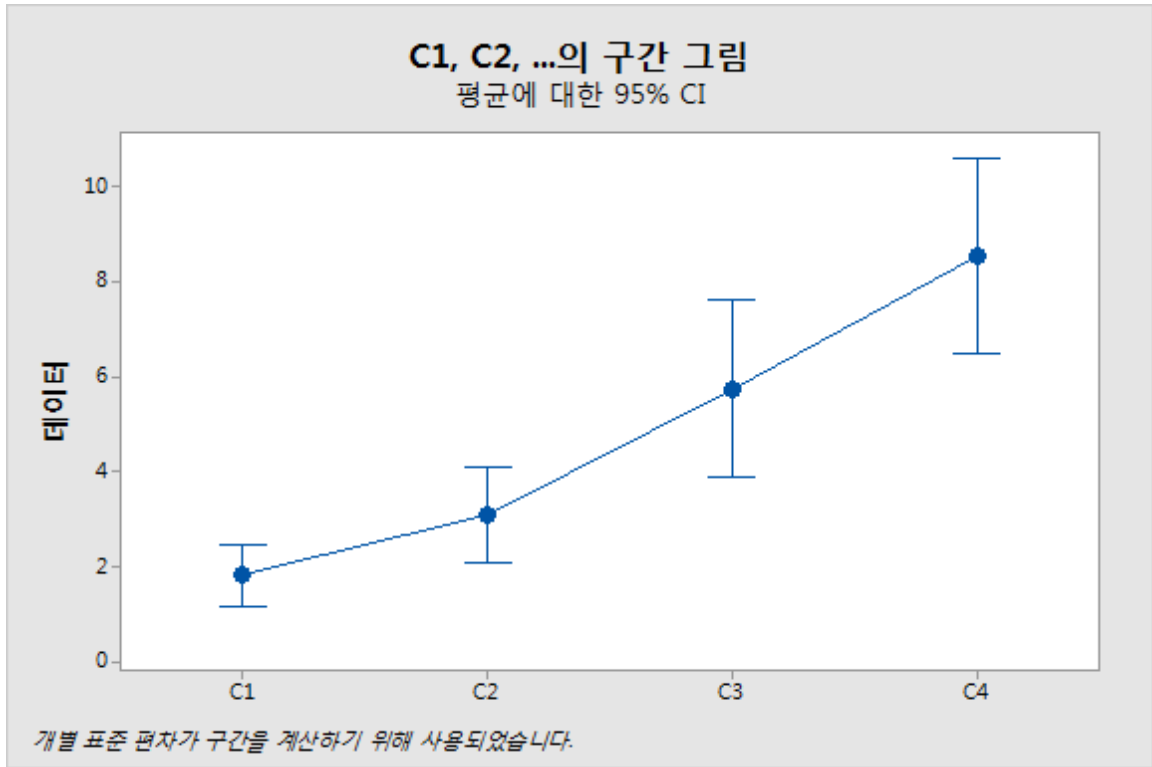


그림 1 보조 도구 일원 분산 분석 요약 보고서의 평균 비교 차트

Minitab의 표준 일원 분산 분석 절차의 결과에서도 일련의 유사한 구간이 나타납니다(통계분석 > 분산 분석 > 일원 분산 분석).



그러나 위 구간은 단순히 평균의 개별적인 신뢰 구간일 뿐입니다. 분산 분석 검정(F 또는 Welch)에서 일부 평균이 다르다는 결론이 나오면 자연스럽게 겹치지 않는 구간을 찾고 어떤 평균이 다른지에 대한 결론을 내리고자 하는 경향이 있습니다. 이러한 비공식적인 개별 신뢰 구간 분석은 종종 합리적인 결론으로 이어지지만, 분산 분석 검정과 같은 방법으로 오류 확률을 관리하지 않습니다. 모집단 수에 따라 구간은 차이가 있다는 결론에 도달할 확률이 검정보다 훨씬 더 크거나 작을 수 있습니다. 그 결과 두 가지 방법은 서로 일치하지 않는 결론에 쉽게 도달할 수 있습니다. 비교 차트는 다중 비교 시 Welch 검정 결과와 더 일관되게 일치하도록 만들어졌지만 항상 완전한 일관성을 얻을 수 있는 것은 아닙니다.

Minitab(통계분석 > 분산 분석 > 일원 분산 분석)의 Tukey-Kramer 및 Games-Howell 비교와 같은 다중 비교 방법을 사용하면 개별 평균 간의 차이에 대한 통계적으로 타당한 결론을 도출할 수 있습니다. 이 두 가지 방법은 각 평균 쌍의 차이에 대한 구간을 제공하는 쌍 비교 방법입니다. 모든 구간에 추정하는 차이가 동시에 포함될 확률은 이상입니다. Tukey-Kramer 방법은 등분산 가정에 의존하는 한편, Games-Howell 방법은 등분산을 요구하지 않습니다. 동일 평균 귀무 가설이 참이면 모든 차이가 0이고 Games-Howell 구간에 0이 하나도 포함되지 않을 확률은 최대 α 입니다. 따라서 구간을 사용하여 유의 수준이 인 가설 검정을 수행할 수 있습니다. 당사는 Games-Howell 구간을 보조 도구에서 비교 차트 구간을 도출하는 시작점으로 사용합니다.

모든 차이 $\mu_i - \mu_j (1 \leq i < j \leq k)$ 에 대한 구간 집합 $[L_{ij}, U_{ij}]$ 에 대해 같은 정보를 제공하는 개별 평균 $\mu_i (1 \leq i \leq k)$ 에 대한 구간 집합 $[L_i, U_i]$ 를 찾고자 합니다. 이를 위해 차이 d 가 $\mu_i \in [L_i, U_i]$ 및 $\mu_j \in [L_j, U_j]$ 가 존재하고 $\mu_i - \mu_j = d$ 일 경우에만 구간 $[L_{ij}, U_{ij}]$ 안에 있어야 합니다. 구간의 끝점은 등식에 의해 연관되어야 합니다.

$$U_i - L_j = U_{ij} \text{ 과}$$

$$L_i - U_j = L_{ij}.$$

k = 2일 경우에는 차이가 하나뿐이지만 개별 구간이 2개이므로 정확한 비교 구간을 얻을 수 있습니다. 실제로 이 조건을 충족하는 구간 너비는 상당히 유연합니다. k = 3일 경우 차이와 개별 구간이 각각 3개이므로 조건을 충족하는 것도 가능하지만, 이제는 구간 너비를 유연하게 설정할 수 없습니다. k = 4일 경우 차이가 6개지만 개별 구간은 4개뿐입니다. 비교 구간은 같은 정보를 더 적은 구간을 사용하여 전달해야 합니다. 일반적으로 k ≥ 4일 경우 차이가 개별 평균보다 많으므로 차이가 있는 구간에 같은 너비 등의 추가적인 조건을 지정하지 않는 한 정확한 해결 방법이 없습니다.

Tukey-Kramer 구간은 표본 크기가 같을 경우에만 너비가 같습니다. 같은 너비는 등분산 가정의 결과이기도 합니다. Games-Howell 구간은 등분산을 가정하지 않으므로 너비가 같지 않습니다. 보조 도구에서는 근사 방법에 의존하여 비교 구간을 정의해야 합니다.

$\mu_i - \mu_j$ 에 대한 Games-Howell 구간은 다음과 같습니다.

$$\bar{x}_i - \bar{x}_j \pm |q^*(k, \hat{\nu}_{ij})| \sqrt{s_i^2/n_i + s_j^2/n_j}$$

여기서 $q^*(k, \hat{\nu}_{ij})$ 는 비교하는 평균의 수인 k와 쌍(i, j)과 연관된 자유도인 ν_{ij} 에 좌우되는 표준화 범위 분포의 해당 백분위수입니다.

$$\hat{\nu}_{ij} = \frac{\left(\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}\right)^2}{\left(\frac{s_i^2}{n_i}\right)^2 \frac{1}{n_i - 1} + \left(\frac{s_j^2}{n_j}\right)^2 \frac{1}{n_j - 1}}.$$

Hochberg, Weiss & Hart(1982)는 다음을 사용하여 이런 쌍 비교와 대략적으로 동일한 개별 구간을 얻었습니다.

$$\bar{x}_i \pm |q^*(k, \nu)| s_p X_i.$$

X_i 값은 다음을 최소화하도록 선택됩니다.

$$\sum \sum_{i \neq j} (X_i + X_j - a_{ij})^2,$$

설명:

$$a_{ij} = \sqrt{1/n_i + 1/n_j}.$$

다음 형식의 Games-Howell 비교에서 구간을 도출하여 이 방식을 이분산에 맞게 수정합니다.

$$\bar{x}_i \pm d_i.$$

d_i 값은 다음을 최소화하도록 선택됩니다.

$$\sum \sum_{i \neq j} (d_i + d_j - b_{ij})^2,$$

설명:

$$b_{ij} = |q^*(k, \hat{\nu}_{ij})| \sqrt{s_i^2/n_i + s_j^2/n_j}.$$

해는 다음과 같습니다.

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

아래 그래프에서는 두 가지 방법, 즉 현재 사용하는 Games-Howell에 기반을 둔 방법과 Minitab 16 릴리스에서 사용하던 자유도 평균 계산에 기초한 방법을 사용하여 Welch 검정 시뮬레이션 결과를 비교 구간 결과와 비교합니다. 수직 축은 10,000번의 시뮬레이션 중에서 Welch 검정에 의해 귀무 가설이 잘못 기각되거나 일부 비교 구간이 겹치지 않은 횟수의 비율입니다. 각 예에서 목표 알파는 $\alpha = 0.05$ 입니다. 각 시뮬레이션은 표준 편차와 표본 크기가 같지 않은 다양한 경우를 포함합니다. 수평 축의 각 위치는 서로 다른 경우를 나타냅니다.

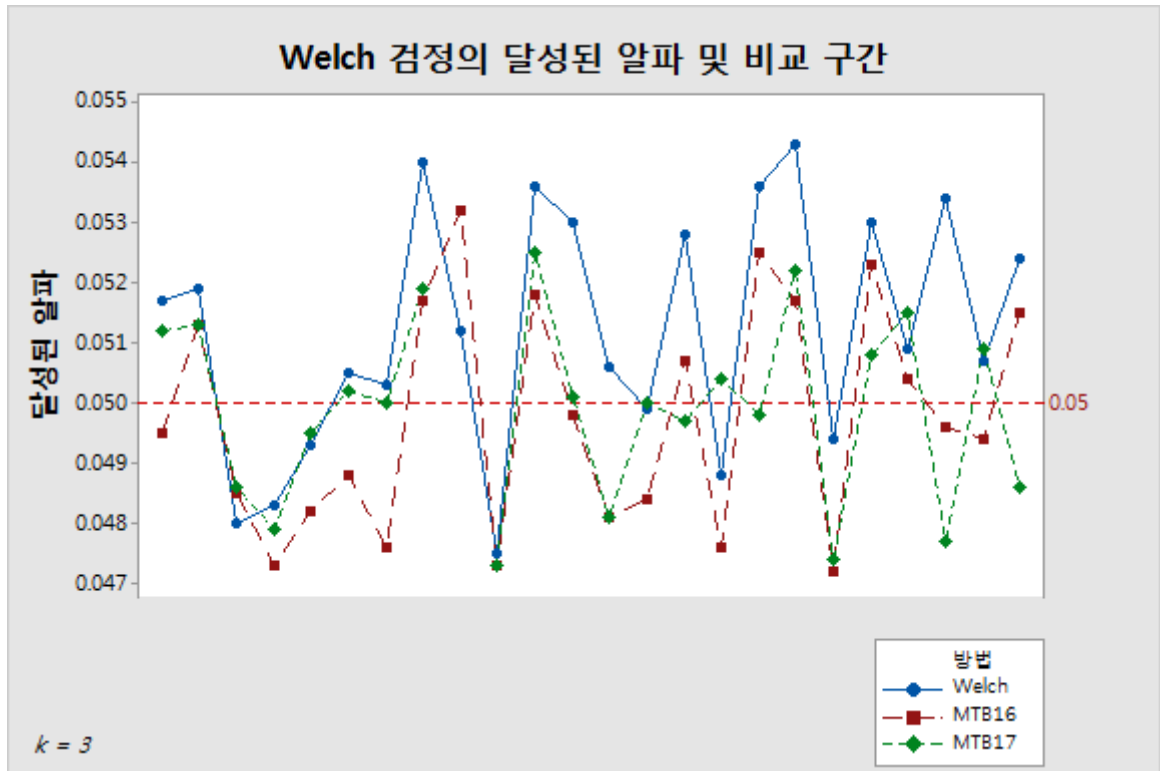


그림 2 3개 표본에 대해 비교 구간을 계산하는 두 가지 방법과 비교한 Welch 검정

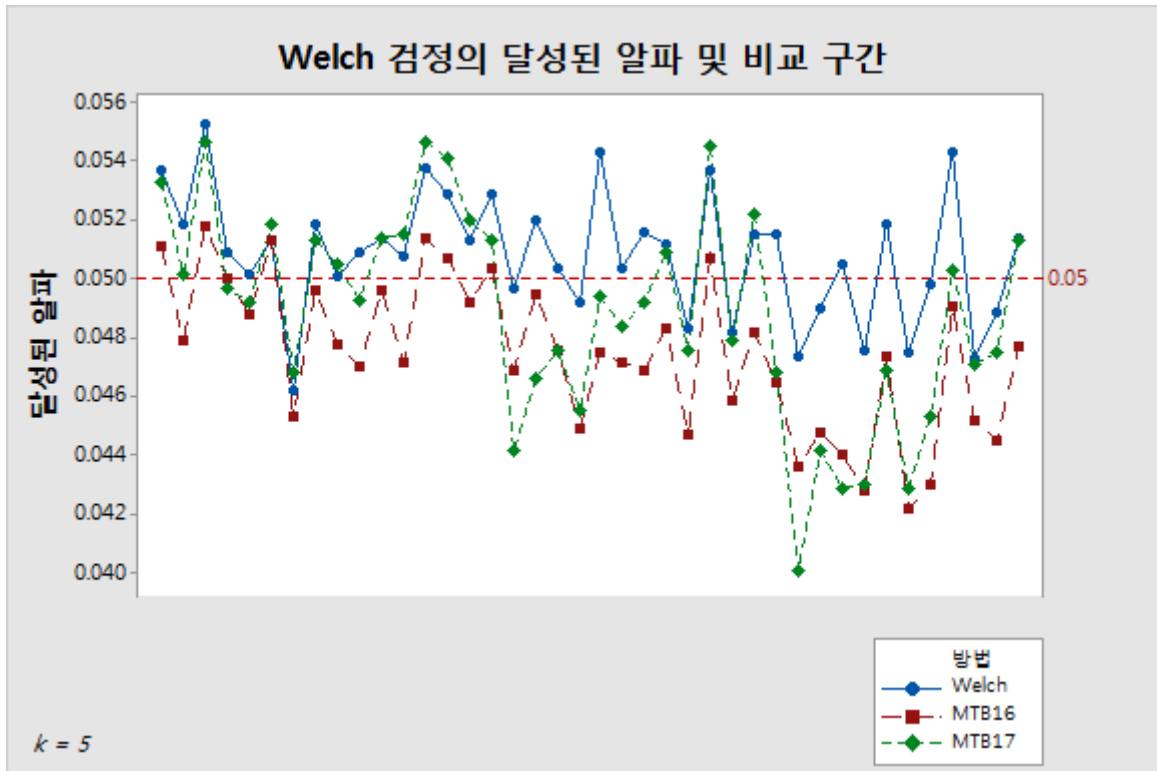


그림 3 5개 표본에 대해 비교 구간을 계산하는 두 가지 방법과 비교한 Welch 검정

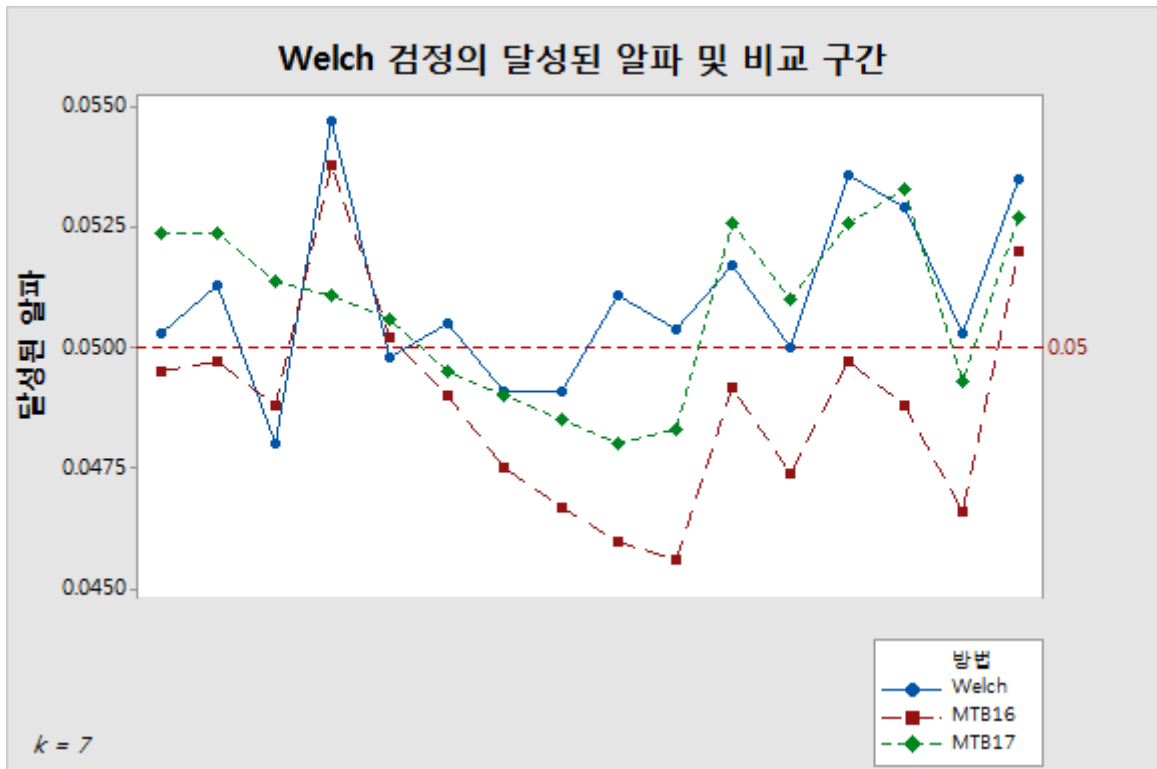


그림 4 7개 표본에 대해 비교 구간을 계산하는 두 가지 방법과 비교한 Welch 검정

이러한 결과는 목표값 0.05 주위의 좁은 범위에 시뮬레이트된 알파 값을 표시합니다. 또한 Minitab 릴리스 17에서 구현된 Games-Howell에 기반을 둔 방법을 사용한 결과는 Minitab 16 릴리스에 사용된 방법보다 Welch 검정의 결과와 더 가깝게 일치한다고 할 수 있습니다.

구간의 포함 확률이 다른 표준 편차에 민감할 수 있다는 증거가 있습니다. 그러나 민감도는 F-검정보다 훨씬 덜 극단적입니다. 아래 그래프에서는 $k = 5$ 인 경우의 해당 의존성을 나타냅니다.



그림 5 다른 표준 편차를 사용한 시뮬레이션 결과

가설 검정과 비교 구간 함께 사용

드물지만 가설 검정과 비교에서 귀무 가설 기각에 대해 일치하지 않는 결과가 나타나는 경우가 있습니다. 검정에서 귀무 가설이 기각되더라도 비교 구간이 모두 겹칠 수 있습니다. 반대로, 검정에서 귀무 가설을 기각하지 못해도 겹치지 않는 구간이 있을 수 있습니다. 그러나 두 가지 방법 모두 참인 귀무 가설을 기각할 확률이 같기 때문에 이렇게 불일치하는 결과가 나오는 경우는 드뭅니다.

이런 경우에는 먼저 검정 결과를 고려하고 검정이 유의할 경우 비교를 사용하여 추가로 조사합니다. 검정에서 귀무 가설이 유의 수준 α 로 기각되면, 하나 이상의 다른 구간과 겹치지 않는 비교 구간이 빨간색으로 표시됩니다. 이는 해당 그룹 평균이 적어도 다른 하나와 다르다는 시각적 단서로 사용됩니다. 모든 구간이 겹치는 경우에도 검정이 유의하면 겹치는 부분이 가장 작은 쌍이 빨간색으로 표시되어 "가장 확률이 높은" 차이를 나타냅니다(아래 그림 6 참조). 이는 특히 겹치는 부분이 매우 적은 다른 쌍이 있을 경우 다소 임의적인 선택입니다. 그러나 차이의 한계가 0보다 더 가까운 다른 쌍은 없습니다.

C1, C2, C3에 대한 일원 분산 분석 요약 보고서

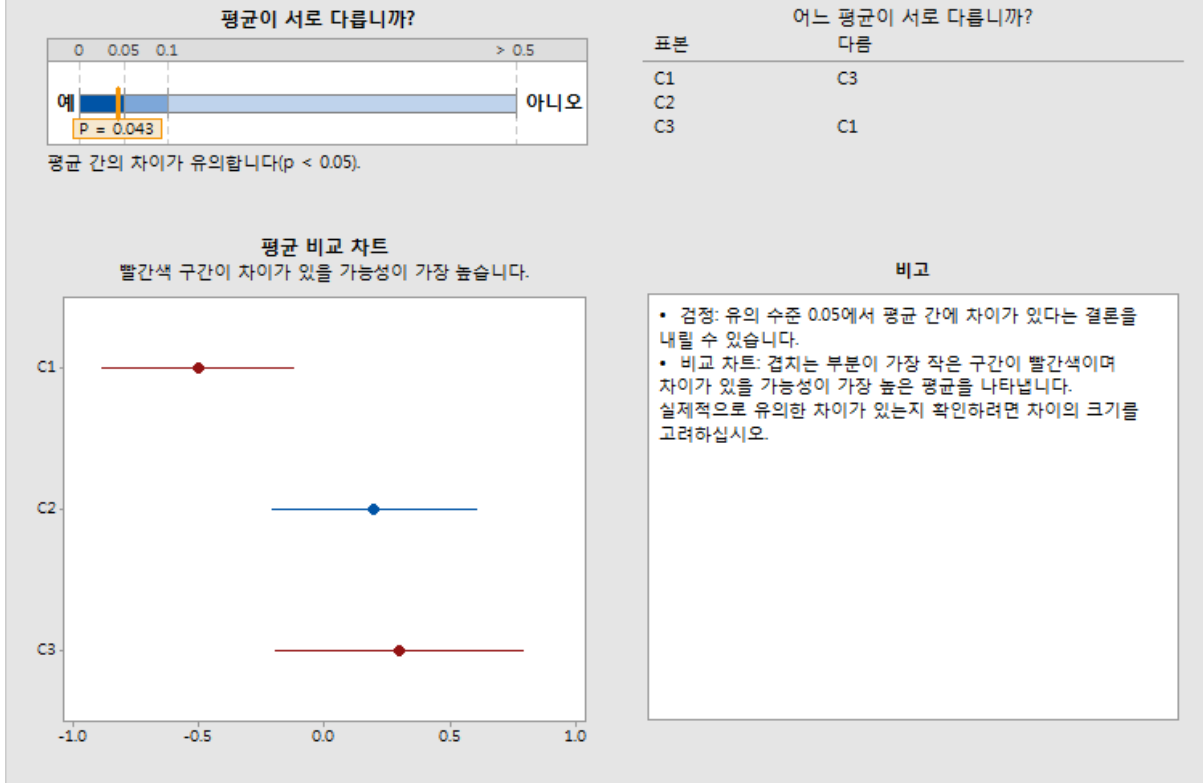
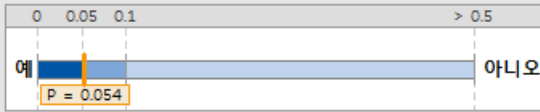


그림 6 구간이 표본 간에 겹치는 경우에도 빨간색으로 표시된 유의한 검정

검정에서 귀무 가설을 기각하지 못할 경우, 겹치지 않는 구간이 있어도 빨간색으로 표시되는 구간이 없습니다(아래 그림 7 참조). 이런 구간은 평균 간에 차이가 있음을 암시하지만, 귀무 가설을 기각하지 못하는 것은 귀무 가설이 참이라는 결론을 내리는 것과는 다릅니다. 단지 관측된 차이가 확률을 원인으로 배제할 만큼 크지 않음을 나타낼 뿐입니다. 또한 이런 상황에서는 겹치지 않는 구간 사이의 간격이 일반적으로 매우 작으므로, 매우 작은 차이가 여전히 구간과 일치하고 실질적인 영향이 있는 차이가 있음을 나타내는 것이 아닐 수 있음을 언급해야 합니다.

C1, C2, C3에 대한 일원 분산 분석 요약 보고서

평균이 서로 다른지?



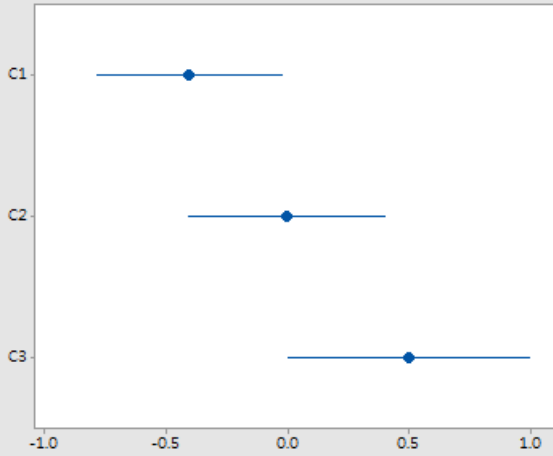
평균 간의 차이가 유의하지 않습니다($p > 0.05$).

어느 평균이 서로 다른지?
다름

표본	어느 평균이 서로 다른지? 다름
C1	
C2	아무 것도 식별되지 않음
C3	

평균 비교 차트

파란색은 유의한 차이가 없음을 나타냅니다.



비고

- 검정: 유의 수준 0.05에서 평균 간에 차이가 있다는 결론을 내릴 충분한 증거가 없습니다.
- 비교 차트: 파란색 구간은 평균이 유의하게 다르지 않음을 나타냅니다.

그림 7 검정 실패, 표본 간에 겹치는 부분이 없을 때도 빨간색으로 표시되는 구간 없음

부록 C: 표본 크기

일원 분산 분석에서 검정하려는 모수는 서로 다른 그룹 또는 모집단의 모집단 평균 $\mu_1, \mu_2, \dots, \mu_k$ 입니다. 모수가 같으면 귀무 가설이 충족됩니다. 평균 간에 차이가 있으면 대립 가설이 충족됩니다. 귀무 가설을 기각할 확률은 평균이 귀무 가설을 충족할 경우 α 보다 크지 않아야 합니다. 실제 확률은 분포의 표준 편차와 표본 크기에 따라 결정됩니다. 귀무 가설 이탈을 탐지하는 검정력은 표준 편차가 작거나 표본이 클수록 증가합니다.

비중심 F 분포를 사용하여 표준 편차가 같은 정규 분포 가정 하에서 F-검정의 검정력을 계산할 수 있습니다. 비중심 모수는 다음과 같습니다.

$$\theta_F = \sum_{i=1}^k n_i (\mu_i - \mu)^2 / \sigma^2$$

여기서 μ 는 평균의 가중 평균이고,

$$\mu = \sum_{i=1}^k n_i \mu_i / \sum_{i=1}^k n_i,$$

σ 는 상수라고 가정되는 표준 편차입니다. 다른 모든 조건이 같으면 검정력은 θ_F 와 함께 증가합니다. 이는 평균이 귀무 가설에서 더 멀리 이탈할수록 검정력이 증가하는 경우 정확히 해당됩니다.

F-검정과 달리 Welch 검정은 검정력을 간단히 계산하는 정확한 공식이 없습니다. 그러나 여기서는 합당한 수준으로 우수한 근사 공식 2개를 살펴볼 것입니다. 첫 근사 공식은 F-검정의 검정력과 유사한 방법으로 비중심 F-분포를 사용합니다. 비중심 모수는 여전히 다음 형식을 사용합니다.

$$\theta_W = \sum_{i=1}^k w_i (\mu_i - \mu)^2$$

여기서 μ 는 가중 평균입니다.

$$\mu = \sum_{i=1}^k w_i \mu_i / \sum_{j=1}^k w_j$$

그러나 가중치는 알려진 표준 편차 σ_i^2 에 대한 결과를 시뮬레이션하거나 표본 표준 편차 s_i^2 을 추정하는 경우에 따라 표준 편차와 표본 크기, 즉 $w_i = n_i / \sigma_i^2$ 또는 $w_i = n_i / s_i^2$ 에 따라 결정됩니다. 따라서 근사 검정력은 다음과 같이 계산됩니다.

$$P(F_{k-1, f, \theta_W} \geq F_{k-1, f, 1-\alpha})$$

여기서 분모 자유도는 다음과 같습니다.

$$f = \frac{k^2 - 1}{3 \sum_{i=1}^k (1 - w_i / \sum_{j=1}^k w_j) / (n_i - 1)}$$

아래에 나와 있는 것처럼 이 식은 시뮬레이션에서 관측된 검정력에 대한 합당한 수준으로 우수한 근사를 제공합니다. 보조 도구 메뉴에서 검정력을 계산하는 데는 다른 근사가 사용되지만, 위의 근사식은 좋은 통찰력을 부여하고 보조 도구 메뉴에서 검정력을 계산하는 데 사용하는 평균 구성을 선택하는 기초가 됩니다.

평균 구성

Minitab(통계분석 > 분산 분석 > 일원 분산 분석)에서 검정력과 표본 크기에 대해 사용되는 방식에 따라 보조 도구는 사용자에게 검정력을 평가할 평균의 전체 집합을 요청하지 않습니다. 대신, 실질적인 영향이 있는 평균 간 차이를 사용자에게 요청합니다. 제공된

차이에 대해 최대 및 최소 평균이 해당 양만큼 차이가 나는 가능한 평균 구성의 수는 무한합니다. 예를 들어 다음은 모두 5개 평균 집합에서 최대 차이가 10입니다.

$$\mu_1 = 0, \mu_2 = 5, \mu_3 = 5, \mu_4 = 5, \mu_5 = 10;$$

$$\mu_1 = 5, \mu_2 = 0, \mu_3 = 10, \mu_4 = 10, \mu_5 = 0;$$

$$\mu_1 = 0, \mu_2 = 10, \mu_3 = 0, \mu_4 = 0, \mu_5 = 0;$$

그리고 무한히 더 많이 있습니다.

여기서는 Minitab(통계분석 > 검정력과 표본 크기 > 일원 분산 분석)에서 검정력과 표본 크기에 사용되는 2개 평균을 제외한 모든 평균이 평균의 (가중) 평균과 일치하고 나머지 2개 평균이 지정된 양만큼 차이가 나는 경우를 선택하는 방식을 따릅니다. 그러나 분산과 표본 크기가 같지 않을 가능성으로 인해 비중심 모수(따라서 검정력)는 여전히 차이가 나는 것으로 가정되는 평균 2개에 따라 결정됩니다.

2개를 제외한 모든 평균이 전체 가중 평균 μ 와 같고 두 평균(예: $\mu_i > \mu_j$)이 서로 다르고 전체 평균과 다른 평균 구성 μ_1, \dots, μ_k 가 있다고 생각해 보십시오. $\Delta = \mu_i - \mu_j$ 는 두 평균의 차이를 나타냅니다. $\Delta_i = \mu_i - \mu$ 및 $\Delta_j = \mu - \mu_j$ 입니다. 따라서 $\Delta = \Delta_i + \Delta_j$ 입니다. 또한 μ 는 모든 k 개 평균의 가중 평균을 나타내고 ($k - 2$)개의 평균은 μ 와 같다고 가정되므로 다음이 성립합니다.

$$\mu = \left[\sum_{l \neq i, j} w_l \mu_l + w_i(\mu + \Delta_i) + w_j(\mu - \Delta_j) \right] / \sum_{l=1}^k w_l = \mu + (w_i \Delta_i - w_j \Delta_j) / \sum_{l=1}^k w_l.$$

따라서

$$w_i \Delta_i = w_j \Delta_j = w_j(\Delta - \Delta_i) \text{고,}$$

그러므로

$$\Delta_i = \frac{w_j}{w_i + w_j} \Delta$$

$$\Delta_j = \frac{w_i}{w_i + w_j} \Delta$$

이 특정한 평균 구성에 대해 Welch 검정과 관련된 비중심 모수를 다음과 같이 계산할 수 있습니다.

$$\begin{aligned} \theta_W &= w_i(\mu_i - \mu)^2 + w_j(\mu_j - \mu)^2 \\ &= \frac{w_i w_j^2 \Delta^2 + w_j w_i^2 \Delta^2}{(w_i + w_j)^2} = \frac{w_i w_j \Delta^2}{w_i + w_j} \end{aligned}$$

이 수량은 w_j 가 고정일 경우 w_i 에서 증가하고, 반대의 경우도 성립합니다. 그러므로 가중치가 가장 큰 쌍(i, j)에서 극대화되고 가중치가 가장 작은 쌍에서 최소화됩니다. 모든 검정력 계산에는 정확히 2개의 평균이 평균의 전체 가중 평균과 다르다는 가정 하에서 검정력을 최소화하고 극대화하는 두 가지의 극단적인 경우가 고려됩니다.

검정에 대해 차이를 지정하면 최대 및 최소 검정력 값이 해당 차이에 대해 평가됩니다. 해당 검정력의 범위는 보고서에 60% 이하의 검정력은 빨간색이고 90% 이상의 검정력은 녹색이고, 60%에서 90% 사이의 검정력은 노란색인 색상별 막대로 표시됩니다. 보고서 카드 결과는 이 색상별 척도에서 검정력 범위가 어디에 속하는 지에 따라 결정됩니다. 전체

범위가 빨간색에 속하면 모든 그룹 쌍의 검정력이 60% 이하이고 빨간색 아이콘이 보고서 카드에 표시되어 검정력이 불충분한 문제가 있음을 나타냅니다. 전체 범위가 녹색에 속하면 모든 그룹의 검정력이 90% 이상이고 보고서 카드에 녹색 아이콘이 표시되어 검정력이 충분한 조건임을 나타냅니다. 기타 모든 조건은 보고서 카드에 노란색 아이콘으로 표시되는 중간 상황으로 간주됩니다.

녹색 조건이 충족되지 않는 경우 보조 도구는 사용자가 지정한 차이와 관측된 표본 표준 편차에 따라 녹색 조건을 가져올 표본 크기를 계산합니다. 추정 검정력은 가중치 $w_i = n_i/s_i^2$ 을 통해 표본 크기에 따라 결정됩니다. 모든 표본의 표본 크기가 같다고 가정하면 최소 가중치 2개는 표본 표준 편차가 가장 큰 그룹 2개에 해당됩니다. 보조 도구는 지정된 차이가 변동성이 가장 큰 두 그룹 사이에 있을 경우 최소 90% 이상의 검정력을 부여하는 표본 크기를 찾습니다. 따라서 모든 그룹에 대해 이 크기 이상의 표본을 추출하면 전체 검정력 값의 범위가 녹색 조건을 충족하는 90% 이상이 되는 결과로 이어집니다.

사용자가 검정력 계산에 사용할 차이를 지정하지 않으면 보조 도구는 계산된 검정력 범위의 최대값이 60%가 되는 최대 차이를 찾습니다. 이 값은 60% 검정력에 해당되는 막대의 빨간색과 노란색 부분 사이의 경계에 표시됩니다. 계산된 검정력 범위의 최소값이 90%가 되는 최소 차이도 찾습니다. 이 값은 90% 검정력에 해당되는 막대의 노란색과 녹색 부분 사이의 경계에 표시됩니다.

검정력 계산

검정력은 Kulinskaya 외 (2003)로 인해 근사를 사용하여 계산됩니다.

정의:

$$\lambda = \sum_{i=1}^k w_i (\mu_i - \mu)^2,$$

$$A = \sum_{i=1}^k h_i,$$

$$B = \sum_{i=1}^k w_i (\mu_i - \mu)^2 (1 - w_i/W) / (n_i - 1),$$

$$D = \sum_{i=1}^k w_i^2 (\mu_i - \mu)^4 / (n_i - 1),$$

$$E = \sum_{i=1}^k w_i^3 (\mu_i - \mu)^6 / (n_i - 1)^2.$$

Welch 통계의 계산자 $\sum_{i=1}^k w_i (\bar{x}_i - \hat{\mu})^2$ 의 첫 누적률 3개는 다음과 같이 추정할 수 있습니다.

$$\kappa_1 = k - 1 + \lambda + 2A + 2B,$$

$$\kappa_2 = 2(k - 1 + 2\lambda + 7A + 14B + D),$$

$$\kappa_3 = 8(k - 1 + 3\lambda + 15A + 45B + 6D + 2E).$$

$F_{k-1, f, 1-\alpha}$ 는 $F(k-1, f)$ 분포의 $(1-\alpha)$ 분위수를 나타냅니다. $W^* \geq F_{k-1, f, 1-\alpha}$ 는 크기가 α 인 Welch 검정에서 귀무 가설을 기각하는 기준입니다.

다음과 같은 경우

$$q = (k - 1) \left[1 + \frac{2(k-2)A}{k^2 - 1} \right] F_{k-1, f, 1-\alpha},$$

$$b = \kappa_1 - 2\kappa_2^2 / \kappa_3,$$

$$c = \kappa_3 / (4\kappa_2) \text{ [참고: } c \text{의 식은 Kulinskaya 외 (2003)에 괄호 없이 나와 있습니다.]}$$

$$v = 8\kappa_2^3 / \kappa_3^2.$$

Welch 검정의 추정 근사 검정력은 다음과 같습니다.

$$P(\chi_v^2 \geq \frac{q-b}{c})$$

여기서 χ_v^2 은 자유도가 v 인 카이 제곱 무작위 변수입니다.

다음 결과는 10,000번에 걸친 시뮬레이션에 기초하여 근사 방법 2개의 검정력과 광범위한 예의 모의 검정력을 비교합니다.

표 3 모의 검정력과 비교한 두 근사 방법의 검정력 계산

예	알파	모의 검정력	비중심 F	Kulinskaya 외
μ 's: 0, 0, 0, -0.1724, 0.8276	0.10	0.1372	0.135702	0.135795
σ 's: 2, 2, 2, 2, 4	0.05	0.0739	0.072563	0.069512
n 's: 12, 12, 12, 12, 10	0.01	0.0195	0.016587	0.012538
μ 's: 0, 0, 0, -0.3448, 1.6552	0.10	0.2498	0.251064	0.257455
σ 's: 2, 2, 2, 2, 4	0.05	0.1574	0.153128	0.156215
n 's: 12, 12, 12, 12, 10	0.01	0.0541	0.045211	0.042195
μ 's: 0, 0, 0, -0.5172, 2.4828	0.10	0.4534	0.44557	0.453506
σ 's: 2, 2, 2, 2, 4	0.05	0.3211	0.311994	0.321575
n 's: 12, 12, 12, 12, 10	0.01	0.1273	0.121225	0.125065
μ 's: 0, 0, 0, -0.6896, 3.3104	0.10	0.662	0.671317	0.670296
σ 's: 2, 2, 2, 2, 4	0.05	0.5219	0.533819	0.538617
n 's: 12, 12, 12, 12, 10	0.01	0.2842	0.271316	0.282759
μ 's: 0, 0, 0, -0.8620, 4.1380	0.10	0.8417	0.852589	0.846697
σ 's: 2, 2, 2, 2, 4	0.05	0.7382	0.752173	0.746121
n 's: 12, 12, 12, 12, 10	0.01	0.4883	0.487601	0.49323
μ 's: 0, 0, 0, -1.0344, 4.9656	0.10	0.9429	0.952077	0.954929
σ 's: 2, 2, 2, 2, 4	0.05	0.8866	0.901485	0.897937
n 's: 12, 12, 12, 12, 10	0.01	0.691	0.711055	0.703379
μ 's: 0, 0, 0, 0, 0, -0.148148, 1.85185	0.10	0.2011	0.189392	0.200114
σ 's: 2, 2, 2, 2, 2, 2, 5	0.05	0.1201	0.108986	0.11742
n 's: 20, 20, 20, 20, 20, 20, 10	0.01	0.0385	0.028986	0.031456
μ 's: 0, 0, 0, 0, 0, -0.296296, 3.70370	0.10	0.4942	0.485917	0.500143
σ 's: 2, 2, 2, 2, 2, 2, 5	0.05	0.3677	0.351593	0.375296
n 's: 20, 20, 20, 20, 20, 20, 10	0.01	0.177	0.149041	0.177189

예	알파	모의 검정력	비중심 F	Kulinskaya 외
μ 's: 0, 0, 0, 0, 0, -0.444444, 5.55556 σ 's: 2, 2, 2, 2, 2, 2, 5 n 's: 20, 20, 20, 20, 20, 20, 10	0.10 0.05 0.01	0.8125 0.7131 0.4876	0.829702 0.727384 0.474291	0.819542 0.720807 0.49469
μ 's: 0, 0, 0, 0, 0, -0.592593, 7.40741 σ 's: 2, 2, 2, 2, 2, 2, 5 n 's: 20, 20, 20, 20, 20, 20, 10	0.10 0.05 0.01	0.9645 0.9286 0.7938	0.977211 0.949997 0.831174	0.984213 0.949239 0.814067
μ 's: 0, 0, 0, 0, 0, -0.740741, 9.25926 σ 's: 2, 2, 2, 2, 2, 2, 5 n 's: 20, 20, 20, 20, 20, 20, 10	0.10 0.05 0.01	0.9961 0.9895 0.9528	0.998947 0.996653 0.977536	1.00 1.00 0.98705
μ 's: 0, 0, 0, 0, 0, -0.888889, 11.1111 σ 's: 2, 2, 2, 2, 2, 2, 5 n 's: 20, 20, 20, 20, 20, 20, 10	0.10 0.05 0.01	0.9999 0.9995 0.9943	0.999985 0.999926 0.99891	1.00 1.00 1.00
μ 's: 0, 0, 0, 0, 0, -0.518519, 6.48148 σ 's: 2, 2, 2, 2, 2, 2, 5 n 's: 20, 20, 20, 20, 20, 20, 10	0.10 0.05 0.01	0.9059 0.8403 0.6511	0.929392 0.868721 0.67121	0.924696 0.85672 0.66652
μ 's: 0, 0, 0, 0, 0, -.5, .5 σ 's: 2, 2, 2, 2, 2, 2, 2 n 's: 12, 12, 12, 12, 12, 12, 12	0.10 0.05 0.01	0.187 0.1098 0.0315	0.186658 0.106600 0.027773	0.18329 0.100189 0.021332
μ 's: 0, 0, 0, 0, 0, -1, 1 σ 's: 2, 2, 2, 2, 2, 2, 2 n 's: 12, 12, 12, 12, 12, 12, 12	0.10 0.05 0.01	0.4734 0.3394 0.1378	0.474736 0.338655 0.137788	0.472469 0.33443 0.128693
μ 's: 0, 0, 0, 0, 0, -1.5, 1.5 σ 's: 2, 2, 2, 2, 2, 2, 2 n 's: 12, 12, 12, 12, 12, 12, 12	0.10 0.05 0.01	0.8228 0.7112 0.4391	0.817355 0.707319 0.441154	0.810181 0.698461 0.431868
μ 's: 0, 0, 0, 0, 0, -2, 2 σ 's: 2, 2, 2, 2, 2, 2, 2 n 's: 12, 12, 12, 12, 12, 12, 12	0.10 0.05 0.01	0.9691 0.9312 0.7817	0.973246 0.940585 0.799339	0.973319 0.936546 0.785099
μ 's: 0, 0, 0, 0, 0, -2.5, 2.5 σ 's: 2, 2, 2, 2, 2, 2, 2 n 's: 12, 12, 12, 12, 12, 12, 12	0.10 0.05 0.01	0.9984 0.9936 0.9587	0.998579 0.99533 0.967674	0.999763 0.997481 0.966249
μ 's: 0, 0, 0, 0, 0, -3, 3 σ 's: 2, 2, 2, 2, 2, 2, 2 n 's: 12, 12, 12, 12, 12, 12, 12	0.10 0.05 0.01	1.00 0.9997 0.9959	0.999975 0.99987 0.997927	1.00 1.00 0.99961

예	알파	모의 검정력	비중심 F	Kulinskaya 외
μ 's: 0, 0, 0, 0, 0, -3.5, 3.5 σ 's: 2, 2, 2, 2, 2, 2, 2 n's: 12, 12, 12, 12, 12, 12, 12	0.10 0.05 0.01	1.00 1.00 0.99998	1.00 1.00 0.99995	1.00 1.00 1.00
μ 's: 0, 0, 0, 0, 0, -1.75, 1.75 σ 's: 2, 2, 2, 2, 2, 2, 2 n's: 12, 12, 12, 12, 12, 12, 12	0.10 0.05 0.01	0.914 0.8418 0.619	0.921225 0.852755 0.633815	0.916652 0.843856 0.620704
μ 's: 0, -0.5, 0.5 σ 's: 2, 2, 2 n's: 12, 12, 12	0.10 0.05 0.01	0.2548 0.1549 0.0470	0.259249 0.160861 0.049045	0.257149 0.156251 0.042292
μ 's: 0, -1, 1 σ 's: 2, 2, 2 n's: 12, 12, 12	0.10 0.05 0.01	0.654 0.5205 0.2612	0.659073 0.522885 0.26355	0.654105 0.515816 0.252469
μ 's: 0, -1.5, 1.5 σ 's: 2, 2, 2 n's: 12, 12, 12	0.10 0.05 0.01	0.9364 0.8747 0.6614	0.935939 0.87562 0.664478	0.937768 0.872608 0.652563
μ 's: 0, -1.75, 1.75 σ 's: 2, 2, 2 n's: 12, 12, 12	0.10 0.05 0.01	0.981 0.9522 0.8251	0.981434 0.9561 0.830726	0.986815 0.959796 0.823624
μ 's: 0, -2, 2 σ 's: 2, 2, 2 n's: 12, 12, 12	0.10 0.05 0.01	0.9953 0.9878 0.9308	0.995969 0.988175 0.931922	0.999332 0.993705 0.933446
μ 's: 0, -2.5, 2.5 σ 's: 2, 2, 2 n's: 12, 12, 12	0.10 0.05 0.01	0.9999 0.9997 0.9949	0.999923 0.999634 0.994725	1.00 1.00 0.99909
μ 's: 0, -3, 3 σ 's: 2, 2, 2 n's: 12, 12, 12	0.10 0.05 0.01	1.00 1.00 0.9999	1.00 1.00 0.99985	1.00 1.00 1.00
μ 's: 0, -3.5, 3.5 σ 's: 2, 2, 2 n's: 12, 12, 12	0.10 0.05 0.01	1.00 1.00 0.9999	1.00 1.00 1.00	1.00 1.00 1.00
μ 's: 0, -0.142857, 0.857143 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.1452 0.0790 0.0223	0.143156 0.077699 0.018200	0.146824 0.077538 0.014338

예	알파	모의 검정력	비중심 F	Kulinskaya 외
μ 's: 0, -0.285714, 1.71429 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.2765 0.1787 0.0624	0.27424 0.170628 0.051588	0.286222 0.179469 0.050335
μ 's: 0, -0.428571, 2.57143 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.4861 0.3487 0.1467	0.476925 0.338626 0.132405	0.490018 0.355743 0.141352
μ 's: 0, -0.50000, 3 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.5846 0.4425 0.2107	0.588533 0.444491 0.19729	0.596795 0.460707 0.212798
μ 's: 0, -0.571429, 3.42857 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.6933 0.5631 0.3052	0.694684 0.555731 0.279131	0.696773 0.567129 0.299302
μ 's: 0, -0.714286, 4.28571 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.848 0.7402 0.4871	0.861469 0.759703 0.480052	0.859329 0.759762 0.497421
μ 's: 0, -0.857143, 5.14286 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.9434 0.8869 0.6649	0.952562 0.898817 0.687058	0.961913 0.902716 0.692591
μ 's: 0, -1, 6 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.9849 0.9609 0.8294	0.987981 0.967589 0.847436	0.999989 0.985049 0.853787
μ 's: 0, -1.14286, 6.85714 σ 's: 2, 2, 4 n's: 14, 12, 8	0.10 0.05 0.01	0.9976 0.989 0.9222	0.997776 0.99222 0.940972	1.00 1.00 0.96383
μ 's: 1, 2, 3 σ 's: 0.3, 2.4, 3.6 n's: 13, 19, 25	0.10 0.05 0.01	0.8838 0.7995 0.5632	0.882194 0.797869 0.556486	0.884649 0.802137 0.563208
μ 's: 1, 2, 3 σ 's: 2.77489, 2.77489, 2.77489 n's: 13, 19, 25	0.10 0.05 0.01	0.5649 0.4305 0.1994	0.566831 0.431302 0.201329	0.565141 0.428126 0.195734

위 결과는 각 근사와 시뮬레이션에 의해 추정된 검정력 값의 불일치가 표시된 아래 그래프에 요약되어 있습니다.

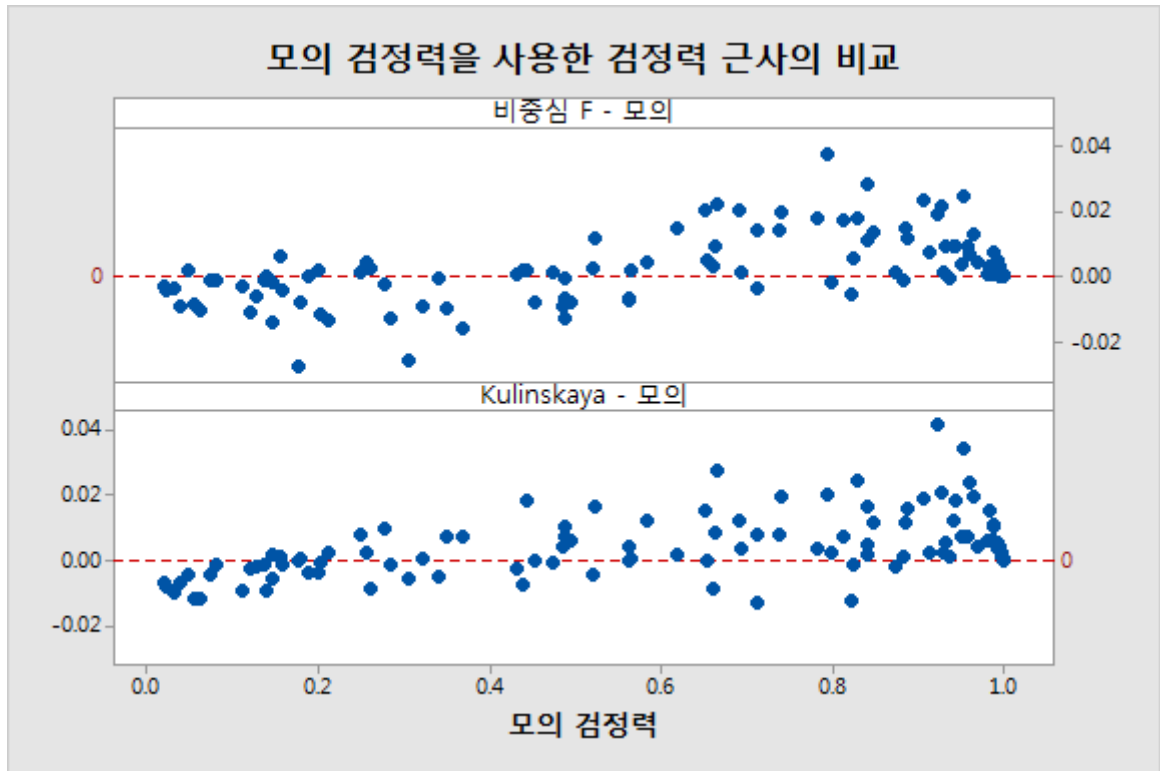


그림 8 두 검정력 근사와 시뮬레이션에 의해 추정된 검정력의 비교

부록 D: 정규성

다음 항목에는 여러 비정규 분포에서 추출한 중간 이하 크기의 표본을 사용하여 Welch 검정과 비교 구간의 성능을 조사한 시뮬레이션이 나와 있습니다.

아래 표에는 동일 평균 귀무 가설 하에서 다양한 분포 유형에 대해 실시한 시뮬레이션 결과가 요약되어 있습니다. 각 예의 표준 편차도 모두 같고 모든 표본의 크기가 같습니다. 표본 개수 $k = 3, 5$ 또는 7 입니다.

각 셀에는 10,000번에 걸친 시뮬레이션에 기초한 제1종 오류율의 추정치가 나와 있습니다. 목표 유의 수준(목표 α)은 0.05 입니다.

표 4 서로 다른 분포에 대해 같은 평균을 사용한 Welch 검정의 시뮬레이션 결과

분포	표본 크기 $n = 10$			표본 크기 $n = 15$		
	$k = 3$	$k = 5$	$k = 7$	$k = 3$	$k = 5$	$k = 7$
N(0,1)	0.0490	0.0486	0.0512	0.0534	0.0522	0.0550
T(3)	0.0371	0.0361	0.0348	0.0353	0.0385	0.0365
T(5)	0.0440	0.0425	0.0439	0.0435	0.0428	0.0428
Laplace(0,1)	0.0433	0.0354	0.0345	0.0445	0.0397	0.0407
균등(-1, 1)	0.0544	0.0640	0.0718	0.0517	0.0573	0.0585
베타(3, 3)	0.0504	0.0577	0.0622	0.0501	0.0538	0.0564
지수	0.0508	0.0621	0.0748	0.0483	0.0633	0.0779
카이-제곱(3)	0.0473	0.0579	0.0753	0.0499	0.0588	0.0703
카이-제곱(5)	0.0458	0.0594	0.0643	0.0504	0.0606	0.0679
카이-제곱(10)	0.0463	0.0510	0.0585	0.0463	0.0552	0.0567
베타(8, 1)	0.0500	0.0622	0.0775	0.0549	0.0653	0.0760

제1종 오류율은 모두 표본 크기가 10일 때도 목표 α 의 3% 이내였습니다. 그룹이 더 많고 분포가 정규 분포와 거리가 멀수록 차이가 더 많이 나는 경향이 있습니다. 표본 크기가 10일 때 합격 확률이 2%보다 높은 경우는 $k = 7$ 일 때뿐이었습니다. 이런 경우는 정규 분포보다 꼬리가 훨씬 더 짧은 균등 분포와 매우 치우친 지수, 카이-제곱(3) 및 베타(8, 1) 분포에서만 발생합니다. 표본 크기를 15로 늘리면 균등 분포에 대한 결과가 현저히 개선되지만 매우 치우친 분포 2개는 그렇지 않습니다.

비교 구간에 대해 유사한 시뮬레이션을 수행했습니다. 이 경우 시뮬레이트된 α 는 10,000번 중에 일부 구간이 겹치지 않는 시뮬레이션의 수입입니다. 목표 $\alpha = 0.05$ 입니다.

표 5 서로 다른 분포에 대해 같은 평균을 사용한 비교 구간 시뮬레이션 결과

분포	표본 크기 n = 10			표본 크기 n = 15		
	k = 3	k = 5	k = 7	k = 3	k = 5	k = 7
N(0,1)	0.0493	0.0494	0.0469	0.0538	0.0518	0.0561
t(3)	0.0378	0.0321	0.0254	0.0347	0.0343	0.0289
t(5)	0.0449	0.0399	0.0361	0.0447	0.0444	0.0412
Laplace(0,1)	0.0438	0.0305	0.0246	0.0456	0.0366	0.0348
균등(-1, 1)	0.0559	0.0605	0.0699	0.0534	0.0607	0.0590
베타(3, 3)	0.0515	0.0569	0.0615	0.0510	0.0553	0.0568
지수	0.0353	0.0254	0.0207	0.0346	0.0310	0.0275
카이-제곱(3)	0.0375	0.0305	0.0296	0.0384	0.0359	0.0339
카이-제곱(5)	0.0405	0.0390	0.0353	0.0417	0.0433	0.0416
카이-제곱(10)	0.0425	0.0428	0.0447	0.0435	0.0476	0.0464
베타(8, 1)	0.0381	0.0352	0.0287	0.0459	0.0428	0.0403

Welch 검정과 마찬가지로, 제1종 오류율은 모두 표본 크기가 10일 때도 목표 α 의 3% 이내였습니다. 표본이 더 많고 분포가 정규 분포와 거리가 멀수록 차이가 더 많이 나는 경향이 있습니다. 표본 크기가 10일 때 오류율은 때때로 k = 7일 경우 (그리고 k = 5일 경우 한 번) 2% 넘게 차이가 났습니다. 이런 경우는 자유도가 3인 꼬리가 매우 두꺼운 t 분포, Laplace 분포, 그리고 매우 치우친 지수 및 카이-제곱 (3) 분포에서 발생합니다. 표본 크기를 15로 늘리면 결과가 개선되어 t(3) 및 지수 분포만 모의 α 값이 목표값을 2% 이상 벗어납니다. Welch 검정의 결과와 달리 비교 구간의 경우 더 큰 차이는 보수적인 편입니다.

보조 도구의 일원 분산 분석에서는 최대 k = 12 표본까지 허용되므로 다음에는 표본이 7개보다 많은 경우의 결과를 고려합니다. 아래 표에는 k = 9 그룹에서 비정규 데이터에 Welch 검정을 사용한 제1종 오류율이 나와 있습니다. 이번에도 목표 $\alpha = 0.05$ 입니다.

표 6 표본이 9개인 서로 다른 분포에 대한 Welch 검정 시뮬레이션 결과

분포	k = 9
t(3)	0.0362
t(5)	0.0426
Laplace(0,1)	0.0402
균등(-1, 1)	0.0625
베타(3, 3)	0.0584

지수	0.0885
카이-제곱(3)	0.0774
카이-제곱(5)	0.0686
카이-제곱(10)	0.0581
베타(8, 1)	0.0863

매우 치우친 분포는 목표 α 에서 가장 크게 벗어납니다. 목표값에서 4% 이상 벗어나는 오류율은 없지만 지수 분포는 4% 가까이 벗어납니다. 보고서 카드에서는 표본 크기가 15면 모든 결과가 적어도 합당한 수준에서 목표 α 에 가깝기 때문에 비정규 데이터 문제 플래그를 표시하지 않는 데 충분하다고 간주됩니다 .

크기 $n = 15$ 인 표본에서는 $k = 12$ 표본에 도달하자 검정력이 좋지 못했습니다. 아래에서는 광범위한 표본 크기에 대해 극단적인 비정규 분포를 사용하여 실시한 Welch 검정의 시뮬레이션 결과를 고려합니다. 이는 표본 크기에 대한 합당한 기준을 도출하는 데 도움이 될 것입니다.

표 7 표본이 12개인 서로 다른 분포에 대한 Welch 검정 시뮬레이션 결과

n	T(3)	균등 분포	카이-제곱(5)
10	0.0397	0.0918	0.0792
15	0.0351	0.0695	0.0717
20	0.0362	0.0622	0.0671
30	0.0408	0.0573	0.0657

이런 분포에 대해 목표 α 에서 2% 약간 넘게 벗어나는 것을 받아들일 수 있는 경우 $n = 15$ 로 충분합니다. 차이를 2% 미만으로 유지하려면 표본 크기가 20이 되어야 합니다. 이제 보다 치우친 카이-제곱(3) 및 지수 분포에서 얻은 결과를 고려합니다.

표 8 표본이 12개인 카이-제곱 및 지수 분포에 대한 Welch 검정 시뮬레이션 결과

n	카이-제곱(3)	지수
10	0.1013	0.1064
15	0.0854	0.1079
20	0.0850	0.0951
30	0.0746	0.0829
40	0.0727	0.0735
50	0.0675	0.0694

이처럼 매우 치우친 분포는 더 큰 문제를 제기합니다. 목표 $\alpha = 0.05$ 에서 3%를 훨씬 넘는 차이를 받아들일 수 있을 경우 카이-제곱(3) 분포에 대해서도 $n = 15$ 가 충분하다고 간주할 수 있지만, 지수 분포의 경우에는 $n = 30$ 에 가까운 표본 개수가 필요할 것입니다. 특정 표본 크기의 기준은 약간 임의적이고 $n = 20$ 이면 광범위한 분포에 매우 효과적이고 극도로 치우친 분포에도 어느 정도 효과적이지만, 10-12개 표본의 경우 $n = 20$ 을 최소 권장 표본 크기로 사용합니다. 극도로 치우친 분포에 대해서도 차이를 매우 작게 유지해야 할 경우에는 더 큰 표본이 권장됩니다.