이 문서는 Minitab 통계 소프트웨어의 보조 도구에서 사용되는 방법과 데이터 검사를 개발하기 위해 Minitab 통계 학자들이 실시한 연구에 대해 설명하는 전체 백서 중 하나입니다.

카이-제곱 검정

개요

현장에서 품질 전문가들은 연속형 데이터를 수집하는 것이 불가능하거나 편리하지 않을 때 공정을 평가하기 위해 범주형 데이터를 수집합니다. 예를 들어, 제품은 불량/양호와 같은 두 범주 또는 우수, 양호, 적정, 불량 등 세 개 이상의 범주로 분류할 수 있습니다. 또 한 가지 예로는 송장의 지불 기한 경과 일 수를 15일 이하, 16 - 30일, 31 - 45일, 45일 초과 등의 범주로 분류하는 재무 부서를 들 수 있습니다. 그 결과, 관심 변수는 각 범주에 포함되는 항목의 수입니다.

카이-제곱 검정은 유연성이 높아서 범주형 데이터를 포함한 여러 응용 프로그램에 사용됩니다. 보조 도구에서는 카이-제곱 검정을 다음과 같은 용도로 사용합니다.

- 다항 분포의 적합도 검정

  이 검정은 데이터가 과거와 같은 분포를 따르는지 여부를 확인하기 위해 사용할 수 있습니다. 분포는 다항 분포로서 항목들이 각 결과 범주에 포함되는 일련의 과거 백분율 또는 목표 백분율로 정의됩니다. 카이-제곱 검정은 백분율이 각각의 과거 또는 목표 백분율과 유의하게 다른지 여부를 공동으로 검정합니다.

- 3개 이상의 그룹에 대한 불량률(%)의 동일성 검정

  이 검정은 여러 그룹의 불량률(%) 간에 차이가 있는지 여부를 확인하기 위해 사용할 수 있습니다. 각 그룹은 다른 조작자에 의해, 다른 공장에서, 또는 다른 시간에 생산된 제품과 같이 관심 특성별로 다릅니다. 카이-제곱 검정은 한 그룹의 불량률(%)이 다른 그룹의 불량률(%)과 유의하게 다른지 여부를 공동으로 검정합니다.

- 두 범주형 변수 간의 연관성 검정

이 검정은 범주형 결과 변수(Y)가 다른 범주형 예측 변수(X)와 관계되거나 연관되어 있는지 여부를 확인하기 위해 사용할 수 있습니다. 카이-제곱 검정은 결과 변수와 예측 변수 간에 연관성이 있는지 여부를 공동으로 검정합니다. 보조 도구에서는 두 개 이상의 고유 값(두 개 이상의 표본)이 포함된 예측 변수(X)를 사용하여 카이-제곱 연관성 검정을 수행할 수 있습니다.

카이-제곱 검정 통계량에 대한 자세한 내용은 부록 A를 참조하십시오.

가설 검정이 포함된 방법의 경우, 검정에 대한 가정이 충족되는지, 검정의 검정력이 적절한지, 데이터 분석에 사용되는 근사가 유효한 결과를 생성하는지 확인하는 것이 좋습니다. 카이-제곱 검정의 경우 데이터 수집에 가정이 포함되며 데이터 검사에서는 이를 확인하지 않습니다.

Minitab에서는 근사 방법의 검정력 및 유효성에 주목합니다. 보조 도구에서는 이러한 근사 방법을 사용하여 데이터에 대해 다음과 같은 검사를 수행하고 보고서 카드에 결과를 표시합니다.

- 표본 크기
- 검정의 유효성
- 구간의 유효성

이 백서에서는 이러한 데이터 검사가 카이-제곱 검정과 실제로 어떤 관계가 있는지 조사하고 보조 도구에서 데이터 검사를 위해 어떤 지침을 설정했는 지에 대해 설명합니다.

데이터 검사

표본 크기

일반적으로, 통계적 가설 검정을 수행하는 기본 목적은 "차이 없음"의 귀무 가설을 기각하기 위한 증거를 수집하는 것입니다. 표본이 너무 작은 경우 검정의 검정력이 실제로 존재하는 불량률(%) 간의 차이를 탐지하기에 적절하지 않아 제2종 오류를 초래할 수 있습니다. 따라서 실제적으로 중요한 차이를 높은 확률로 탐지하기에 충분히 큰 표본 크기를 유지하는 것이 중요합니다.

표본 크기 데이터 검사는 검정의 검정력을 바탕으로 합니다. 이 계산에서는 사용자가 실제 모집단 모수와 귀무 가설에서의 값 간의 의미 있는 차이를 지정해야 합니다. 카이-제곱 적합도 및 카이-제곱 연관성 검정을 위해 이 실제 차이를 결정하고 표현하기가 어려웠기 때문에 보조 도구에서는 세 개 이상의 표본을 사용하는 카이-제곱 불량률(%) 검정을 위한 표본 크기만 검사합니다.

목적

데이터가 귀무 가설을 기각하기에 충분한 증거를 제공하지 않는 경우 Minitab에서는 표본 크기가 검정에서 실제적으로 중요한 차이를 높은 확률로 탐지하기에 충분히 큰지 여부를 확인하고자 합니다. 표본 크기 계획의 목적이 중요한 차이를 높은 확률로 탐지하기에 표본 크기를 충분히 크게 유지하는 것이지만, 의미없는 차이가 높은 확률로 통계적으로 유의하게 될 정도로 크지는 않아야 합니다.

방법

검정력과 표본 크기 분석은 부록 B에 표시된 공식을 바탕으로 합니다.

결과

데이터가 귀무 가설을 기각하기 위한 충분한 증거를 제공하지 않고 사용자가 실제 차이를 지정하지 않는 경우, 보조 도구에서는 표본 크기를 바탕으로 80% 및 90% 확률로 탐지할 수 있는 실제 차이를 계산합니다. 또한 사용자가 특히 실제적으로 중요한 차이를 제공하는 경우 보조 도구에서는 80% 및 90%의 확률로 차이를 탐지하는 표본 크기를 계산합니다.

검정력과 표본 크기를 확인할 때 세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정의 보조 도구 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	조건
	검정에서 불량률(%) 간의 차이가 탐지되므로 검정력은 문제가 되지 않음. 또는 검정력이 충분함. 검정에서 불량률(%) 간의 차이를 찾지 못했지만 표본이 주어진 차이를 90% 이상의 확률로 탐지하기에 충분함.
	검정력이 충분할 수 있음. 검정에서 불량률(%) 간의 차이를 찾지 못했지만 표본이 주어진 차이를 80% - 90%의 확률로 탐지하기에 충분함. 90% 검정력에 도달하는 데 필요한 표본 크기가 보고됨.
	검정력이 충분하지 않을 수 있음. 검정에서 불량률(%) 간의 차이를 찾지 못했으며 표본이 주어진 차이를 60% - 80%의 확률로 탐지하기에 충분함. 80% 검정력 및 90% 검정력에 도달하는 데 필요한 표본 크기가 보고됨.
	검정력이 충분하지 않음(< 60%). 검정에서 불량률(%) 간의 차이를 찾지 못함. 80% 검정력 및 90% 검정력에 도달하는 데 필요한 표본 크기가 보고됨.
	검정에서 불량률(%) 간의 차이를 찾지 못함. 불량률(%) 간의 실제 차이를 지정하지 않았으므로 보고서에는 표본 크기와 알파를 바탕으로 80% - 90% 확률로 탐지할 수 있는 차이가 표시됨.

검정의 유효성

χ^2 검정 통계량은 근사적으로만 카이-제곱 분포를 따릅니다. 표본 크기가 클수록 근사가 개선됩니다. 이 항목에서는 정확한 결과를 얻기 위해 필요한 최소 표본 크기를 결정하는 데 사용되는 근사를 평가합니다.

셀의 기대 카운트가 작은 것이 제1종 오류율(알파)에 미치는 영향을 조사하여 검정 통계량에 대한 카이-제곱 근사를 평가합니다. 검정의 유효성을 평가하기 위해 제1종 오류를 사용함으로써 다음과 같은 사항을 보장하는 규칙을 개발했습니다.

- 귀무 가설이 참인데 기각할 확률이 작고 원하는 제1종 오류율에 가깝습니다.
- p-값을 정확하게 계산하는 데 중요한 정규 분포의 꼬리를 적절히 근사시킬 수 있습니다.

Minitab에서는 일반적인 기준을 사용하여 셀의 기대 카운트가 5보다 작거나 같은 셀을 기대 카운트가 작은 셀로 정의했습니다.

Minitab에서는 귀무 가설 하의 비율을 정의하기 위해 비균등 비율 모형과 균등 비율 모형, 두 개의 모형을 개발했습니다. 자세한 내용은 부록 C를 참조하십시오. 두 가지 모형은 모두 이 백서에서 이후에 언급되는 시뮬레이션에 사용됩니다. 모형은 각 카이-제곱 검정에 사용되는데 한 가지 경우만 예외입니다. 비균등 비율 모형은 세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정에 적용되지 않습니다.

검정의 유효성 데이터 검사는 보조 도구의 모든 카이-제곱 검정에 적용됩니다. 각 데이터 검사는 아래 설명되어 있습니다.

카이-제곱 적합도 검정

목적

Minitab에서는 셀의 기대 카운트가 작은 정도 및 빈도가 제1종 오류율에 미치는 영향을 조사하여 검정 통계량에 대한 카이-제곱 근사를 평가했습니다.

방법

비율이 비균등 비율 모형 또는 균등 비율 모형으로 설명된 다항 분포에서 크기 n의 표본이 추출되었습니다(부록 C 참조). 각 조건에 대해 목표 유의 수준 0.05에서 10000번의 카이-제곱 적합도 검정을 수행했습니다. 각 검정에 대해 실제 제1종 오류를 $\frac{\text{기각된 검정 수}}{\text{반복실험 횟수}(10000)}$로 계산했습니다. 허용 가능한 제1종 오류의 범위를 [0.03 - 0.07]로 정의하고, 제1종 오류율이 이 범위 내에 있는 최소 표본 크기를 기록했습니다.

결과

시뮬레이션 결과를 보면 셀의 목표 카운트가 작은 비율이 50%이하인 경우, 셀의 목표 카운트가 1.25보다 작으면 p-값이 올바르지 않을 수 있습니다. 또한 셀의 목표 카운트가 작은 비율이 50%보다 큰 경우, 셀의 목표 카운트가 2.5보다 작으면 p-값이 올바르지 않을 수 있습니다. 자세한 내용은 부록 D를 참조하십시오.

카이-제곱 적합도 검정의 유효성을 확인하는 경우 보조 도구의 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	조건
	작은 목표 셀 카운트의 백분율이 50%보다 작거나 같은 경우 최소 목표 셀 카운트는 1.25보다 크거나 같습니다. 또는 작은 목표 셀 카운트의 백분율이 50%보다 큰 경우 최소 목표 셀 카운트는 2.5보다 크거나 같습니다. 표본이 충분한 목표 카운트를 얻기에 충분히 큽니다. 검정에 대한 p-값은 정확해야 합니다.
	위의 조건이 유지되지 않는 경우.

연관성에 대한 카이-제곱 검정

목적

Minitab에서는 셀의 기대 카운트가 작은 정도 및 빈도가 제1종 오류율에 미치는 영향을 조사하여 검정 통계량에 대한 카이-제곱 근사를 평가했습니다.

방법

비율이 비균등 비율 모형 또는 균등 비율 모형으로 정의된 다항 분포에서 크기 n_i의 표본이 추출되었습니다(부록 C 참조). 단순성 때문에 Minitab에서는 $n_i = n\ \forall i$을 선택했습니다. 각 조건에 대해 목표 유의 수준 0.05에서 10000번의 카이-제곱 연관성 검정을 수행했습니다. 각 검정에 대해 실제 제1종 오류율을 $\frac{\text{기각된 검정 수}}{\text{반복실험 횟수}(10000)}$으로 계산했습니다. 허용 가능한 제1종 오류의 범위를 [0.03 - 0.07]로 정의하고, 제1종 오류율이 이 범위 내에 있는 최소 표본 크기를 기록했습니다.

결과

셀의 최소 기대 카운트는 X 값의 수 및 기대 카운트가 작은 셀의 비율에 따라 달라집니다.

- 비균등 비율 모형의 경우 X 값의 수가 각각 (2 또는 3) 및 (4, 5 또는 6)일 때 셀의 기대 카운트가 작은 비율이 50% 이하이면 셀의 최소 기대 카운트는 2 이하 및 1 이하입니다. 또한 X 값의 수가 각각 (2 또는 3) 및 (4, 5 또는 6)일 때 셀의 기대

카운트가 작은 비율이 50%보다 크면 셀의 최소 기대 카운트가 3 이하 및 1.5 이하입니다.

- 균등 비율 모형의 경우 X 값의 수가 (2 또는 3)일 때 셀의 최소 기대 카운트는 2 이하이며, X 값의 수가 (4, 5 또는 6)일 때 셀의 최소 기대 카운트는 1.5 이하입니다.

자세한 내용은 부록 E를 참조하십시오.

카이-제곱 연관성 검정의 유효성을 확인하는 경우 보조 도구의 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	X 변수 값의 수	조건
	2 또는 3	작은 기대 셀 카운트(5보다 작거나 같음)의 백분율이 50%보다 작거나 같은 경우 최소 기대 셀 카운트는 2보다 크거나 같습니다. 작은 기대 셀 카운트(5보다 작거나 같음)의 백분율이 50%보다 큰 경우 최소 기대 셀 카운트는 3보다 크거나 같습니다.
	4, 5 또는 6	작은 기대 셀 카운트(5보다 작거나 같음)의 백분율이 50%보다 작거나 같은 경우 최소 기대 셀 카운트는 1보다 크거나 같습니다. 작은 기대 셀 카운트(5보다 작거나 같음)의 백분율이 50%보다 큰 경우 최소 기대 셀 카운트는 2(편의상 1.5를 2로 반올림함)보다 크거나 같습니다.
	모든 경우	위의 조건이 유지되지 않는 경우

세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정

목적

Minitab에서는 셀의 기대 카운트가 작은 정도 및 빈도가 제1종 오류율에 미치는 영향을 조사하여 검정 통계량에 대한 카이-제곱 근사를 평가했습니다.

방법

Minitab에서는 모형 p = $p_i = p_j \; \forall i, j$을 정의했으며, 여기서 p = 0.001, 0.005, 0.01, 0.025 및 0.25입니다. 위에 설명된 p_i의 값을 갖는 이항 분포에서 크기 n_i의 표본을 추출했습니다. 단순성 때문에 $n_i = n \; \forall i$을 선택했습니다. 각 조건에 대해 목표 유의 수준 0.05에서

10000번의 카이-제곱 불량률(%) 검정을 수행했습니다. 각 검정에 대해 실제 제1종 오류를 $\frac{\text{기각된 검정 수}}{\text{반복실험 횟수}(10000)}$로 정의했습니다. 허용 가능한 제1종 오류의 범위를 [0.03 - 0.07]로 정의하고, 제1종 오류율이 이 범위 내에 있는 최소 표본 크기를 기록했습니다.

결과

X 값이 3 - 6개일 때 최소 기대 불량품 및 비불량품 수가 1.5보다 크거나 같으면 검정의 제1종 오류율이 구간 [0.03, 0.07]에 포함됩니다. X 값이 7 - 12개일 때 최소 기대 불량품 및 비불량품의 수가 1보다 크거나 같으면 검정의 제1종 오류율이 구간 [0.03, 0.07]에 포함됩니다.

자세한 내용은 부록 F를 참조하십시오.

세 개 이상의 표본에 대한 카이-제곱 검정의 유효성을 확인하는 경우 보조 도구의 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	X 값의 수	조건
	3~6	최소 기대 불량품 및 비불량품의 수가 1.5보다 크거나 같습니다.
	7 - 12	최소 기대 불량품 및 비불량품의 수가 1보다 크거나 같습니다.
	모든 경우	위의 조건이 유지되지 않는 경우

구간의 유효성

세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정과 카이-제곱 적합도 검정의 비교 구간은 정규 근사를 바탕으로 합니다. 또한 카이-제곱 적합도 검정의 개별 신뢰 구간도 정규 근사를 바탕으로 합니다. 이 섹션에서는 정규 근사의 유효성을 평가합니다. 대부분의 통계 교과서에 있는 일반 규칙에 따라 관측된 카운트가 5 이상인 경우에는 근사 신뢰 구간이 정확합니다.

구간 유효성 데이터 검사는 세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정과 카이-제곱 적합도 검정에 적용됩니다.

세 개 이상의 표본에 대한 카이-제곱 불량률(%)

목적

Minitab에서는 각 표본에서 관측된 최소 불량품 및 비불량품의 수에 대한 일반 규칙을 평가하여 근사 신뢰 구간이 정확한지 확인하고자 했습니다.

방법

먼저 비교 차트에 사용되는 구간을 정의합니다. 구간 내에 약 α의 전체 오류률이 포함되도록 끝점을 정의하고, 구간이 모두 겹치지 않으면, 모 불량률(%)이 서로 다르다는 것을 나타냅니다. 사용된 공식은 부록 G를 참조하십시오.

비교 구간은 비교 신뢰 구간 쌍을 바탕으로 합니다. 자세한 내용은 일원 분산 분석에 대한 보조 도구 백서의 비교 구간 항목을 참조하십시오. 각 쌍 $(p_i - p_j)$에 대해 정규 근사 신뢰 구간을 사용한 후 Bonferroni 다중 비교 절차를 사용하여 실험 전반의 오류율을 관리합니다. 따라서 쌍체 비교 절차의 구간 중 하나의 유효성만 평가하면 정규 근사의 비교 구간에 대한 영향을 파악할 수 있습니다.

결과

정규 근사의 유효성을 평가하기 위해서는 근사가 불량률(%) 간의 차이에 대한 한 구간에 어떤 영향을 미치는지만 조사하면 됩니다. 따라서 2-표본 불량률(%)에 대해 개발된 일반 규칙을 간단히 사용할 수 있습니다. 자세한 내용은 2-표본 불량률(%) 검정에 대한 보조 도구 백서의 2-표본 불량률(%) 검정 방법 항목을 참조하십시오. 2-표본 불량률(%) 검정의 시뮬레이션 결과에 따르면 일반적으로 불량률(%) 간의 차이에 대한 근사 신뢰 구간의 정확도는 표본이 충분히 클 때, 즉 각 표본에서 관측된 불량품의 수 및 비불량품의 수가 5 이상일 때 신뢰할 수 있습니다.

세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정의 구간의 유효성을 확인하는 경우 보조 도구의 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	조건
	모든 표본에 5개 이상의 불량품 및 5개 이상의 비불량품이 있습니다. 비교 구간은 정확해야 합니다.
	위의 조건이 유지되지 않는 경우

카이-제곱 적합도 검정

목적

Minitab에서는 각 표본에서 관측된 최소 불량품 및 비불량품의 수에 대한 일반 규칙을 평가하여 근사 신뢰 구간이 정확한지 확인하고자 했습니다.

방법

보조 도구의 카이-제곱 적합도 검정에는 비교 및 개별 신뢰 구간이 포함됩니다. Minitab에서는 비율에 대한 표준 정규 근사 구간을 활용하고 Bonferroni 수정(Goodman, 1965)을 사용하여 다중 구간을 수정합니다. 따라서 Bonferroni 동시 구간은 다음과 같이 계산됩니다.

$$p_{i\text{하한}} = p_i - Z_{\alpha/2k}\sqrt{\frac{\mathrm{p_i}(1-\mathrm{p_i})}{\mathrm{N}}}$$

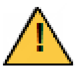
$$p_{i\text{상한}} = p_i + Z_{\alpha/2k}\sqrt{\frac{\mathrm{p_i}(1-\mathrm{p_i})}{\mathrm{N}}}$$

구간 내에 약 ∝의 전체 오류율이 포함되도록 끝점을 정의하고, 구간이 모두 목표 비율 값을 포함하지 않으면, 실제 비율은 목표 비율과 다르다는 것을 나타냅니다. 개별 구간은 Bonferroni 구간과 같은 양식을 활용하지만 $Z_{\alpha/2}$을 사용하여 다중 구간을 수정합니다.

결과

위에 설명된 두 가지 방법은 모두 보조 도구의 2-표본 불량률(%) 검정에 정의된 것과 유사한 방법론을 따릅니다. 따라서 해당 검정을 위해 개발된 정규 근사의 유효성에 대해 유사한 규칙을 사용할 수 있습니다. 자세한 내용은 2-표본 불량률(%) 검정에 대한 보조 도구 백서의 2-표본 불량률(%) 검정 방법 항목을 참조하십시오. 이 백서에서는 표본 카운트가 5보다 작을 때 비교 구간 및 개별 신뢰 구간이 정확하지 않을 수 있다는 결론을 내렸습니다.

카이-제곱 적합도 검정의 구간에 대한 유효성을 확인하는 경우 보조 도구의 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	조건
	모든 표본 카운트가 5 이상입니다. 구간이 정확해야 합니다.
	표본 카운트가 5보다 작습니다.

참고 문헌

Agresti, A. (1996). An introduction to categorical data analysis. New York, NY: Wiley.

Read, T. & Cressie, N. (1988). Goodness-of-fit statistics for discrete multivariate data. New York, NY: Springer-Verlag.

Fienberg, S. (1980). The analysis of cross-classified categorical data. Cambridge, MA: MIT Press.

Goodman, L. (1965). On simultaneous confidence intervals for multinomial proportions. *Technometrics, 7*, 247-254.

부록 A: 카이-제곱 검정 통계량

보조 도구에서는 다음과 같은 형식의 카이-제곱 검정 통계량을 사용합니다.

$$x^2 = \sum_{ij} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

여기서,

O_{ij} = 관측된 카운트(아래 표에 정의되어 있음)

사례	$\boldsymbol{O_{ij}}$
다항 분포의 적합도 검정	i^{th} 결과의 관측된 카운트는 O_{i1}로 정의됩니다.
3개 이상의 불량률(%)의 동일성 검정	i^{th} 표본의 관측된 불량품 및 비불량품 수는 각각 O_{i1} 및 O_{i2} 로 정의됩니다.
두 범주형 변수 간의 연관성 검정	X 변수의 i^{th} 값 및 Y 변수의 j^{th} 값에 대한 관측된 카운트는 O_{ij}로 정의됩니다.

E_{ij} = 기대 카운트(아래 표에 정의되어 있음)

사례	$\boldsymbol{E_{ij}}$
다항 분포의 적합도 검정	E_{i1} = np_i $i = 1, \dots, k$ (k = 결과의 수) n = 표본 크기 p_i = 과거 비율 $\sum_i p_i = 1$
3개 이상의 불량률(%)의 동일성 검정	E_{i1} = $n_i p$ (불량품의 경우) E_{i2} = $n_i(1-p)$ (비불량품의 경우) $i = 1, \dots, k$ (k = 표본 개수) n_i = i^{th} 표본 크기 p = 전체 불량률

사례	$\boldsymbol{E_{ij}}$
두 범주형 변수 간의 연관성 검정	E_{ij} = $\frac{(n_{i.}n_{.j})}{n_{..}}$ $i = 1, \dots, m$ (m = X 값의 수) $j = 1, \dots, k$ (k = Y 값의 수) $n_{i.}$ = X 변수의 i^{th} 값에 대한 전체 카운트 $n_{.j}$ = Y 변수의 j^{th} 값에 대한 전체 카운트 $n_{..}$ = 전체 표본 크기

부록 B: 세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정의 검정력

Minitab에서는 비중심 카이-제곱 분포를 사용하여 $p_i = p_j = p \, \forall i,j$인 검정의 검정력을 계산합니다. 비중심 모수는 n_i 및 $p_i \forall i$에 따라 달라집니다.

여기서,

n_i = i^{th} 표본의 표본 크기

각 p_i는 비율 차이 = δ에서 계산된 대체 비율을 나타냅니다. 이 부록의 다음 항목, 대체 비율 계산을 참조하십시오.

Minitab에서는 카이-제곱 분포의 비중심 모수를 다음과 같이 계산합니다.

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

여기서,

$$O_{i1} = n_i p_i$$

$$O_{i2} = n_i (1 - p_i)$$

그리고 검정의 검정력을 다음과 같이 계산합니다.

$$\text{Prob}(X \geq x_{1-\alpha} \mid \chi^2)$$

여기서,

X = 비중심 모수가 χ^2인 비중심 카이-제곱 분포의 랜덤 변수

$x_{1-\alpha}$ = 중심 카이-제곱 분포에 대해 $1 - \alpha$에서 평가된 역 cdf

대체 비율 계산

Minitab에서는 대체 비율을 다음과 같이 정의했습니다.

$$p_i = p_c + \frac{n_j}{n_i + n_j} \delta$$

$$p_j = p_c - \frac{n_i}{n_i + n_j} \delta$$

$$p_m = p_c \forall m \neq i,j$$

$$0 < \delta < 1$$

여기서,

$$p_c = \frac{1}{N_T}\sum_{i=1}^{k} n_i \widehat{p_i}$$

$\widehat{p_i}$ = i^{th} 표본에 대한 표본 불량품 비율

NT = 총 관측치 수

n_i = i^{th} 표본의 표본 크기

일부 차이 δ의 경우 $p_i > 1$ 또는 $p_j < 0$입니다. 따라서 Minitab에서는 다음과 같은 규칙을 개발합니다.

p_j < 0인 경우
$$p_i = \delta$$
$$p_j = 0$$
$$p_m = \frac{\delta}{2}\ \forall m \neq i, j$$

p_i > 1인 경우
$$p_i = 1$$
$$p_j = 1 - \delta$$
$$p_m = 1 - \frac{\delta}{2}\ \forall m \neq i, j$$

두 개의 가장 작은 n_i 값을 사용하면 검정력이 최소가 되고 두 개의 가장 큰 n_i 값을 사용하면 검정력이 최대가 됩니다.

부록 C: 비균등 비율 모형 및 균등 비율 모형

비균등 비율 모형

Minitab에서는 Read and Cressie(1988)를 따라 귀무 가설 하의 비율 집합을 다음과 같이 정의합니다.

k - 1 가까이의 δ 을 선택하고(여기서 k = 각 표본에 대한 비율의 수) 작은 p_i의 집합을

$i = 1, \ldots, r$에 대한 $p_i = \frac{(1 - \frac{\delta}{k-1})}{k}$

으로 나머지 p_i를

$i = r + 1, \ldots, k$에 대한 $p_i = \frac{(1 - \sum_{i=1}^{r} p_i)}{(k-r)}$으로 정의합니다.

시뮬레이션에서 사용한 δ 값은 표 1에 나열되어 있습니다.

표 1 작은 p_i을 생성하는 시뮬레이션에 사용된 δ

k	δ	$p_{i=1,\ldots,r}$
3	1.95	0.008
4	2.95	0.004
5	3.90	0.005
6	4.90	0.003

각 k에 대해 작은 p_i's. 의 집합 크기를 변경하기 위해 r = 1,..., k - 1로 바꾸었습니다. 예를 들어, k = 3의 경우 표 2에 설명된 다음과 같은 두 모형을 얻었습니다.

표 2 k = 3인 비균등 비율 모형에 대한 p_i의 값

r	p1	p2	p3
1	0.008	0.496	0.496

r	p1	p2	p3
2	0.008	0.008	0.984

균등 비율 모형

셀의 기대 카운트가 작은 비율이 100%인 모형을 얻기 위해 다음과 같이 정의된 균등 비율 모형을 사용합니다.

$$p_i = \frac{1}{k} \forall\, i$$

표본 크기가 매우 작은 상태에서 이 모형을 사용하면 모든 셀의 기대 카운트가 작은 것으로 간주됩니다. 균등 비율 모형을 사용하는 경우, 셀의 기대 카운트를 작게 하려면 표본 크기가 매우 작아야 하지만, 실제로는 이러한 일이 발생할 가능성이 없습니다.

부록 D: 카이-제곱 적합도 검정의 유효성

비균등 비율 모형에서, 셀의 기대 카운트가 작은 %와 제1종 오류율을 [0.03 - 0.07] 구간 내에 있게 하는 최소 기대 카운트를 그래프로 표시했습니다(그림 1 참고).

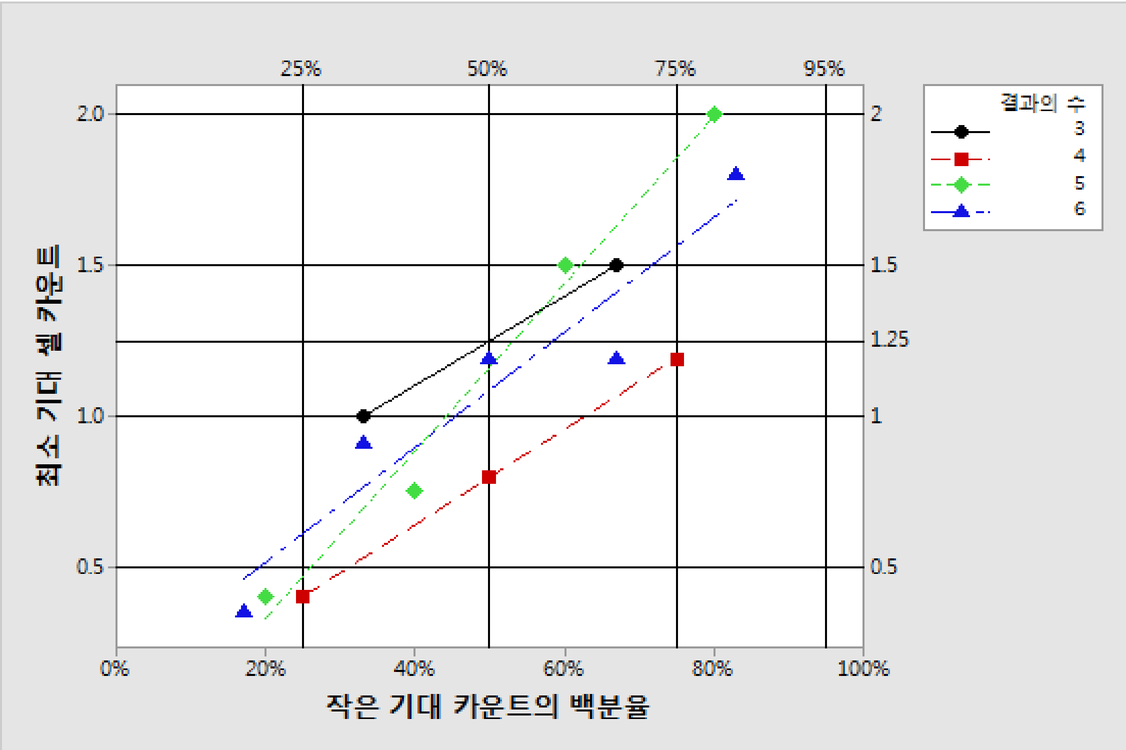


그림 1 구간 [0.03, 0.07]의 제1종 오류율을 얻는 데 필요한 최소 기대 셀 카운트 대 작은 기대 셀 카운트의 백분율.

그림 1에서 작은 기대 셀 카운트의 백분율이 50%보다 작은 경우 최소 기대 셀 카운트는 1.25보다 작거나 같습니다. 모든 기대 셀 카운트는 2보다 작거나 같습니다. 이러한 시뮬레이션 결과를 바탕으로 한 보조 도구 보고서 카드에 사용되는 규칙은 보수적입니다.

그런 다음 귀무 분포를 정의하기 위해 균등 비율 모형을 사용하여 동일한 시뮬레이션을 수행했습니다. 표 4에는 균등 비율 모형을 사용한 시뮬레이션의 결과가 요약되어 있습니다.

표 4 구간 [0.03, 0.07]의 제1종 오류율을 얻기 위한 최소 기대 셀 카운트

k	최소 기대 셀 카운트
3	2.5
4	1.25
5	1
6	1.4

위에 표시된 대로 균등 비율 모형에서는 셀 카운트의 100%가 작습니다. 표 4를 보면 최소 기대 셀 카운트가 모두 2.5보다 작거나 같아 보조 도구 보고서 카드에 사용되는 규칙을 지원한다는 것을 알 수 있습니다.

부록 E: 카이-제곱 연관성 검정의 유효성

비균등 비율 모형에서 각 X 값의 수에 대한 작은 기대 셀 카운트(%)에 대해 구간 [0.03, 0.07]의 제1종 오류율을 얻는 데 필요한 최소 기대 셀 카운트를 그래프로 표시했습니다(그림 2 참고).

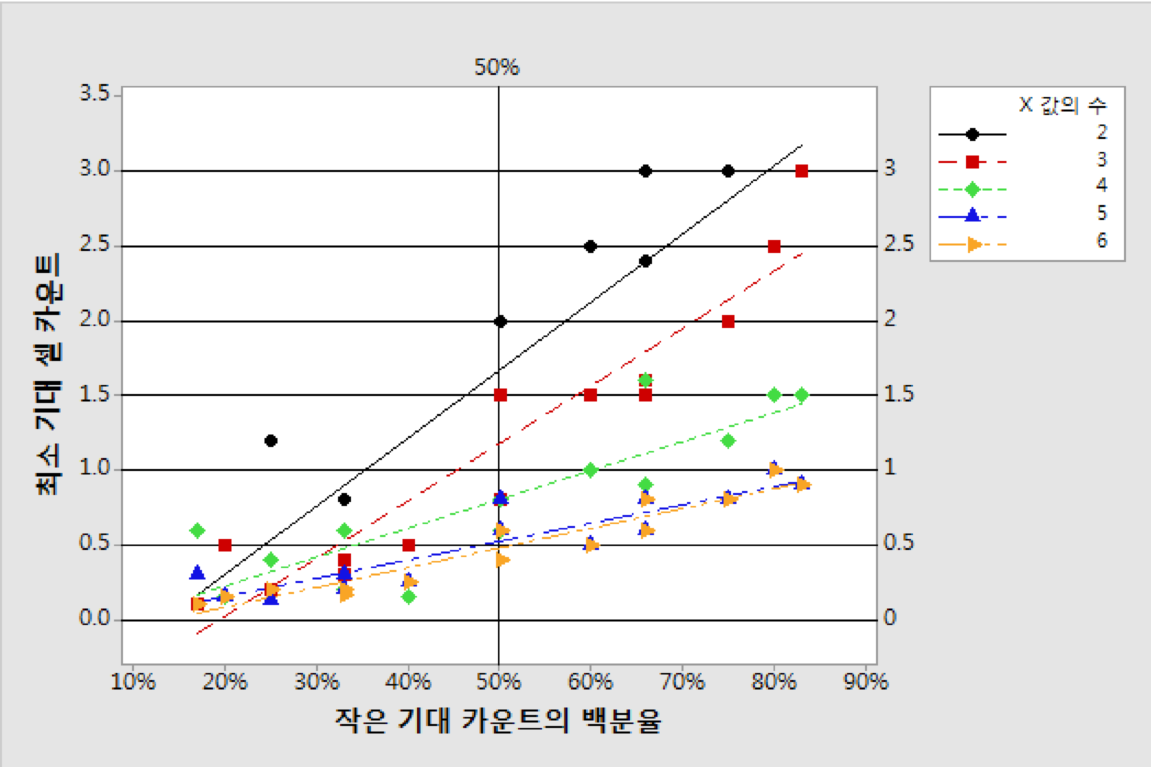


그림 2 구간 [0.03, 0.07]의 제1종 오류율을 얻는 데 필요한 최소 기대 셀 카운트 대 작은 기대 셀 카운트의 백분율.

그림 2를 보면 최소 기대 셀 카운트가 X 값의 수 및 작은 기대 셀 카운트의 백분율에 따라 달라진다는 것을 알 수 있습니다.

그림 2를 보면 X 값의 수가 각각 2 또는 3 및 4, 5 또는 6일 때 작은 기대 셀 카운트(%)가 50%보다 작거나 같으면 최소 기대 셀 카운트가 2보다 작거나 같고 1보다 작거나 같다는 것을 알 수 있습니다. 또한 X 값의 수가 각각 2 또는 3 및 4, 5 또는 6일 때 작은 기대 셀 카운트의 백분율이 50%보다 크면 최소 기대 셀 카운트가 3보다 작거나 같거나 1.5보다 작거나 같습니다.

균등 비율 모형의 경우 X 값의 수(m) 및 Y 값의 수(k)에 대해 최소 기대 셀 카운트를 그래프로 표시했습니다(그림 3 참고).

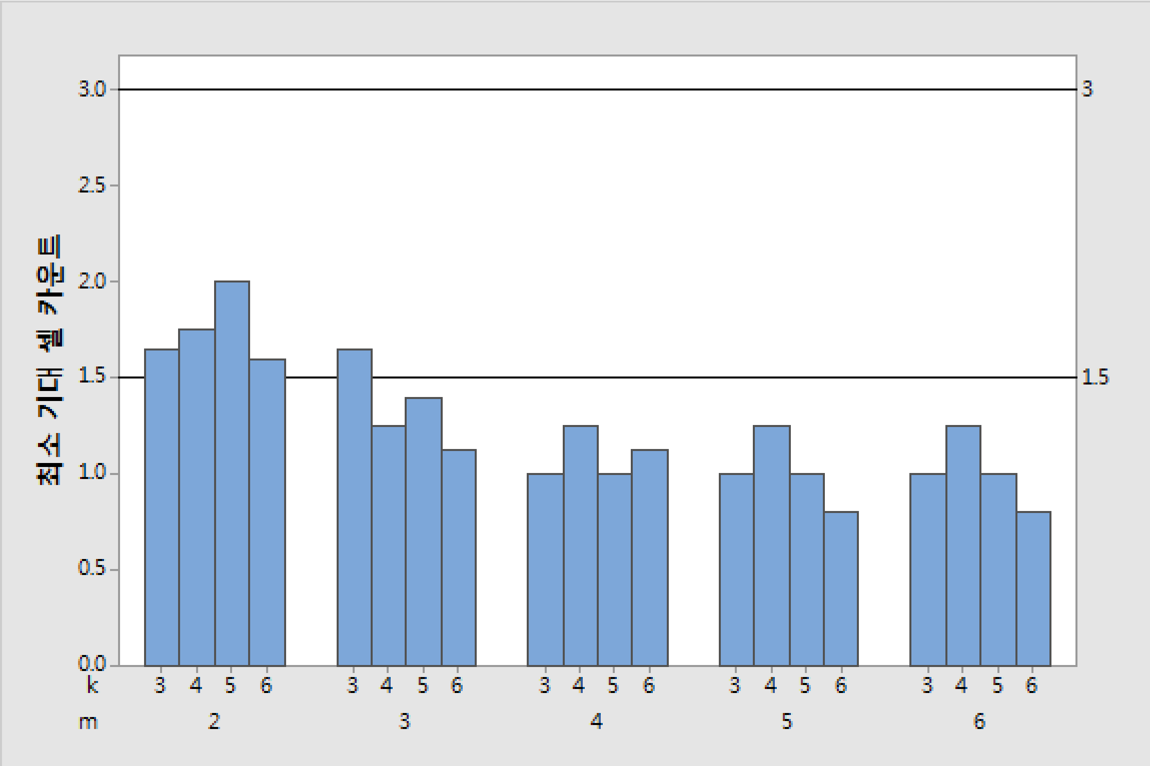


그림 3 구간 [0.03, 0.07]의 제1종 오류율을 얻는 데 필요한 최소 기대 셀 카운트 대 X 값(m) 및 Y 값(k)

그림 3을 보면 X 값의 수가 2 또는 3일 때 최소 기대 셀 카운트가 2보다 작거나 같고 X 값의 수가 4, 5 또는 6일 때 최소 기대 셀 카운트가 1.5보다 작거나 같다는 것을 알 수 있습니다. 시뮬레이션 결과를 바탕으로 한 보조 도구 보고서 카드의 규칙은 보수적입니다.

부록 F: 세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정의 유효성

각 p 및 각 m = 3, 4, 5,..., 12에 대해 최소 기대 셀 카운트를 그래프로 표시했습니다. 결과는 그림 4와 5에 표시되어 있습니다.

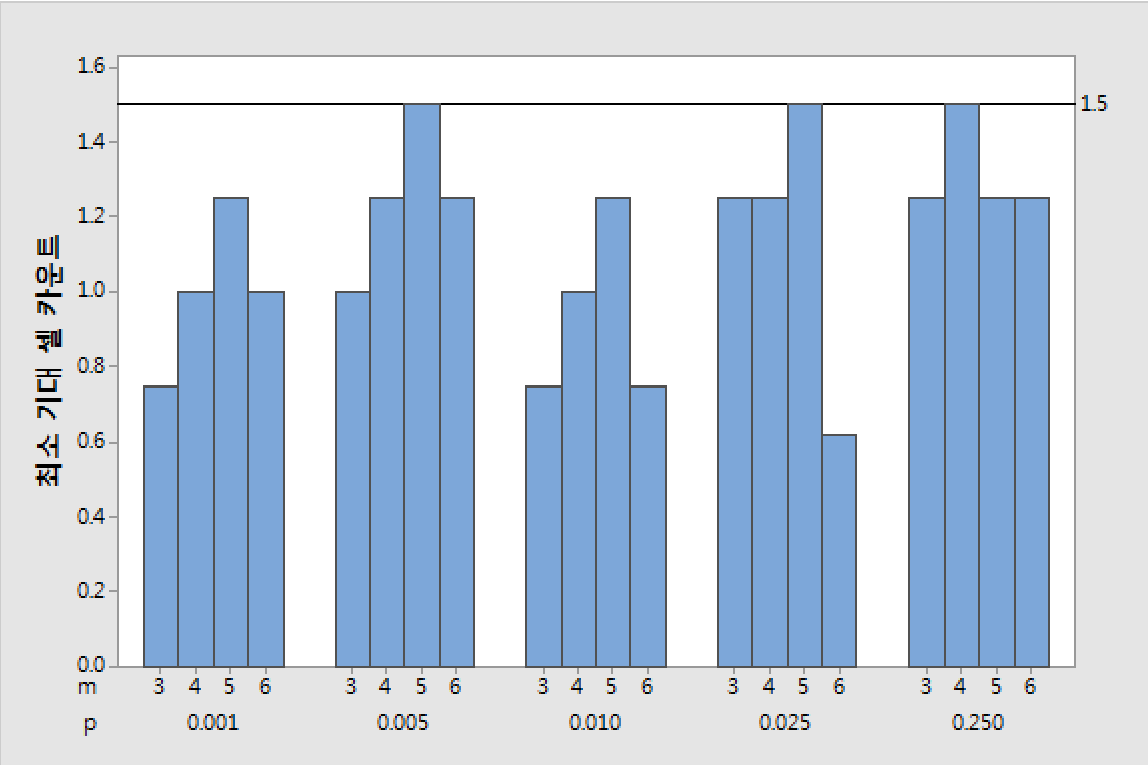


그림 4 구간 [0.03, 0.07]의 제1종 오류율을 얻는 데 필요한 최소 기대 셀 카운트 대 X 값의 수(m = 3 - 6)

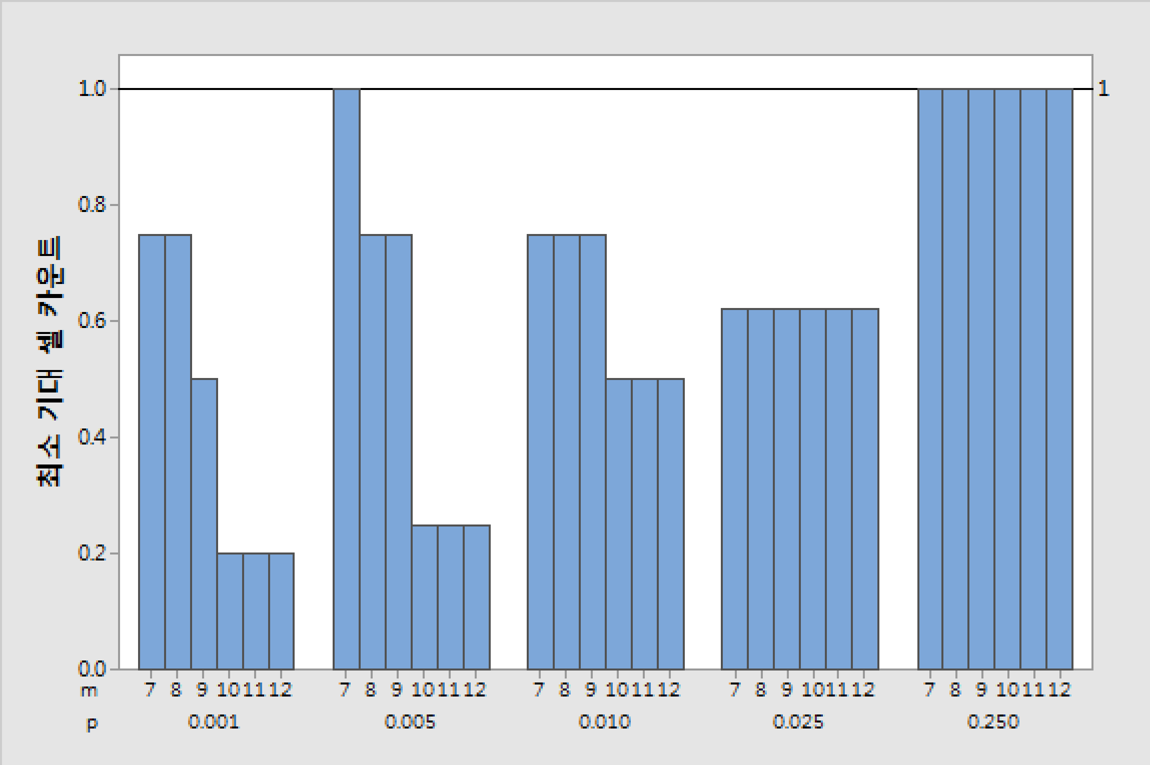


그림 5 구간 [0.03, 0.07]의 제1종 오류율을 얻는 데 필요한 최소 기대 셀 카운트 대 X 값의 수(m = 7 - 12)

X 값의 수가 3, 4, 5 또는 6인 경우 기대 셀 카운트가 1.5보다 크거나 같으면 검정의 제1종 오류율이 구간 [0.03, 0.07]에 포함됩니다. X 값의 수가 7, 8, 9,…, 12인 경우 기대 셀 카운트가 1보다 크거나 같으면 검정의 제1종 오류율이 구간 [0.03, 0.07]에 포함됩니다.

부록 G: 세 개 이상의 표본에 대한 카이-제곱 불량률(%) 검정의 비교 구간

p_i 의 하한과 상한은 다음과 같이 정의됩니다.

$$p_{i\text{하한}} = p_i - Z_{\alpha/c} X_i$$

$$p_{i\text{상한}} = p_i + Z_{\alpha/c} X_i$$

여기서,

c = 비교의 수 = k (k - 1) /2

여기서 k는 표본의 수입니다.

$Z_{\alpha/c}$ = (평균이 0이고 표준 편차가 1인 정규 분포의 1 - $\frac{\alpha}{2c}$) 백분위수)

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

여기서,

$$b_{ij} = \sqrt{\frac{p_i(1 - p_i)}{n_i} + \frac{p_j(1 - p_j)}{n_j}}$$