



## 2-표본 불량률(%) 검정

### 개요

2 비율 검정은 두 비율이 유의하게 다른지 확인하는 데 사용됩니다. 품질 분석에서 이 검정은 종종 제품 또는 서비스가 불량 또는 정상으로 분류될 때 불량품의 비율(%)이 2개의 독립적인 공정에서 수집한 표본에서 유의한 차이가 나는지 확인하기 위해 사용됩니다.

Minitab 보조 도구에는 2-표본 불량률(%) 검정이 포함되어 있습니다. 검정을 위해 수집되는 데이터는 2개의 독립 표본에 각각 포함되어 있고 이항 랜덤 변수의 관측치라고 가정되는 불량품 개수입니다. 보조 도구에서는 정확한 방법을 사용하여 가설 검정 결과를 계산합니다. 그러므로 실제 제1종 오류율은 검정에 대해 지정된 유의 수준(알파)에 가까울 것이며, 더 이상 조사할 필요가 없습니다. 그러나 보조 도구에서는 정규 근사 방법을 사용하여 불량률(%)의 차이에 대한 신뢰 구간(CI)과 정규 근사 검정의 이론적 검정력 함수를 계산하여 검정력 및 표본 크기 분석을 실시합니다. 이들 방법은 근사 방법이기 때문에 정확성을 평가해야 했습니다.

본 문서에서는 근사 신뢰 구간이 어떤 조건에서 정확한지 조사했습니다. 근사 방법의 이론적 검정력을 정확 검정의 실제 검정력과 비교하여 2-표본 불량률(%) 검정의 검정력과 표본 크기를 산정하기 위해 사용된 방법에 대해서도 조사합니다. 마지막으로, 자동으로 수행되고 보조 도구의 보고서 카드에 표시되는 다음 데이터 검사에 대해 조사하고 해당 검사가 분석 결과에 어떻게 영향을 미치는지 설명합니다.

- CI의 유효성
- 표본 크기

2-표본 불량률(%) 검정은 또한 다른 가정에 따라 달라집니다. 자세한 내용은 부록 A를 참조하십시오.

# 2-표본 불량률(%) 검정 방법

## 신뢰 구간의 정확도

보조 도구에서는 Fisher의 정확 검정을 사용하여 두 표본의 불량률(%)의 차이가 유의한지 평가하지만, 차이의 신뢰 구간은 정규 근사 방법에 기초합니다. 대부분의 통계 교과서에 나오는 일반 규칙에 따르면 이 근사 신뢰 구간은 각 표본에서 관측된 불량 개수와 비불량 개수가 5개 이상일 경우 정확합니다.

### 목적

당사는 정규 근사에 기초한 신뢰 구간이 어떤 조건에서 정확한지 평가하고자 했습니다. 구체적으로는 각 표본의 불량 및 비불량 개수와 관련된 일반 규칙이 근사 신뢰 구간의 정확도에 어떤 영향을 미치는지 확인하고자 했습니다.

### 방법

두 비율 간의 차이에 대한 신뢰 구간을 계산하는 데 사용되는 공식과 해당 구간의 정확성을 보장하기 위한 일반 규칙에 대한 설명은 부록 D를 참조하십시오. 또한 조사 과정에서 개발된 덜 엄격하게 수정된 규칙에 대해서도 설명합니다.

다양한 조건에서 근사 신뢰 구간의 정확도를 평가하기 위한 시뮬레이션을 수행했습니다. 시뮬레이션을 수행하기 위해 여러 Bernoulli 모집단에서 다양한 크기의 랜덤 표본 쌍을 추출했습니다. Bernoulli 모집단의 각 유형에 대해, Bernoulli 표본 반복실험 10,000쌍의 두 비율 간 차이에 대한 근사 신뢰 구간을 각각 계산했습니다. 그런 다음 10,000개 구간 중에서 두 비율 간에 실제 차이가 있는 구간의 비율(모의 포함 확률이라고 함)을 계산했습니다. 근사 구간이 정확하면 모의 포함 확률이 목표 포함 확률 0.95와 가까울 것입니다. 근사 구간의 정확도를 각 표본에 요구되는 최소 불량 및 비불량 개수에 대한 원래의 규칙 및 수정된 규칙과 연관지어 평가하기 위해, 10,000개 표본 쌍 중에서 각 규칙이 충족되는 비율(%)도 계산했습니다. 자세한 내용은 부록 D를 참조하십시오.

### 결과

두 비율의 차이에 대한 근사 신뢰 구간은 일반적으로 표본이 충분히 클 때, 즉 각 표본에서 관측된 불량 개수와 비불량 개수가 5개 이상일 때 정확합니다. 그러므로 이 규칙을 보고서 카드의 CI 유효성 검사에 채용했습니다. 이 규칙은 일반적으로 효과가 있지만, 경우에 따라서는 지나치게 보수적일 수 있고 두 비율이 0 또는 1에 가까울 경우 다소 관대할 수 있습니다. 자세한 내용은 데이터 검사 부분과 부록 D를 참조하십시오.

# 이론적 검정력 함수의 성능

보조 도구에서는 Fisher 검정을 사용하여 두 Bernoulli 모집단 비율(두 표본의 불량률(%))을 비교합니다. 그러나 이 정확 검정의 검정력 함수를 도출하기는 쉽지 않기 때문에 해당 정규 근사 검정의 이론적 검정력 함수를 사용하여 근사 검정력 함수를 도출해야 합니다.

## 목적

당사는 보조 도구에서 정규 근사 검정에 기초한 이론적 검정력 함수를 사용하여 2-표본 불량률(%) 검정의 검정력과 표본 크기 요건을 평가하는 것이 적절한지 판단하고자 했습니다. 그러기 위해서는 이 이론적 검정력 함수가 Fisher 정확 검정의 실제 검정력을 정확히 나타내는지 평가해야 했습니다.

## 방법

p-값 계산을 포함한 Fisher의 정확 검정 방법에 대한 자세한 설명은 부록 B를 참조하십시오. 정규 근사 검정에 기초한 이론적 검정력 함수는 부록 C에 정의되어 있습니다. 이 같은 정의를 토대로, 당사는 Fisher의 정확 검정을 두 표본의 불량률(%) 차이를 분석하는데 사용할 때 해당 검정의 (당사에서 모의 검정력 수준이라고 부르는) 실제 검정력 수준을 추정하기 위한 시뮬레이션을 수행했습니다.

시뮬레이션을 수행하기 위해 여러 Bernoulli 모집단에서 다양한 크기의 랜덤 표본 쌍을 추출했습니다. Bernoulli 모집단의 각 범주별로 Fisher의 정확 검정을 각 10,000 쌍의 표본 반복실험에 대해 수행했습니다. 각 표본 크기에 대해, 주어진 차이를 탐지하는 검정의 모의 검정력을 검정이 유의했던 표본 10,000 쌍의 분수로 계산했습니다. 비교를 위해, 정규 근사 검정을 토대로 해당 이론적 검정력 수준도 계산했습니다. 근사가 효과적이면 이론적 검정력 수준 및 모의 검정력 수준이 유사할 것입니다. 자세한 내용은 부록 E를 참조하십시오.

## 결과

시뮬레이션에서는 정규 근사 검정의 이론적 검정력 함수가 일반적으로 Fisher 정확 검정의 모의 검정력 함수와 거의 같은 것으로 나타났습니다. 따라서 보조 도구에서는 정규 근사 검정의 이론적 검정력 함수를 사용하여 Fisher의 정확 검정을 수행할 때 실제로 중요한 차이를 탐지하기 위해 필요한 표본 크기를 추정합니다.

# 데이터 검사

## CI의 유효성

2-표본 불량률(%) 검정에서는 정확 검정을 사용하여 불량률(%)의 차이를 평가하기 때문에 각 표본의 불량 및 비불량 개수가 정확도에 큰 영향을 미치지 않습니다. 그러나 불량률(%) 간의 차이에 대한 신뢰 구간은 정규 근사에 기초합니다. 각 표본의 불량 및 비불량 개수가 증가하면 근사 신뢰 구간의 정확도도 높아집니다(부록 D 참조).

### 목적



당사는 표본의 불량 개수와 비불량 개수가 불량률(%)의 차이에 대한 근사 신뢰 구간이 정확하도록 보장하기에 충분한지 판단하고자 했습니다.

### 방법

대부분의 통계 교과서에 나오는 일반적인 규칙을 사용했습니다. 각 표본에 포함된 불량 및 비불량 개수가 각각 5개 이상이면 2-표본 불량률(%) 검정의 근사 신뢰 구간은 정확합니다. 자세한 내용은 위의 2-표본 불량률(%) 검정 방법 부분을 참조하십시오.

### 결과

2-표본 불량률(%) 검정 방법 부분에 요약되어 있는 시뮬레이션에서 볼 수 있듯이, 신뢰 구간의 정확도는 각 표본의 최소 불량 및 비불량 개수에 따라 달라집니다. 그러므로 보조 도구의 보고서 카드에는 두 불량률(%) 간의 차이에 대한 신뢰 구간의 정확도를 평가하는 데 도움이 되는 다음과 같은 상태가 표시됩니다.

상태	조건
	두 표본의 불량 및 비불량 개수가 모두 5개 이상임. 차이에 대한 신뢰 구간은 정확할 것임.
	표본에 있는 하나 이상의 불량 또는 비불량 개수가 5개 미만임. 차이에 대한 신뢰 구간이 정확하지 않을 수 있음.

## 표본 크기

일반적으로, 통계적 가설 검정은 "차이가 없다"는 귀무 가설을 기각하기 위한 증거를 수집하기 위해 수행합니다. 표본이 너무 작은 경우 검정의 검정력이 실제로 존재하는 차이를 탐지하기에 적절하지 않아 제2종 오류를 초래할 수 있습니다. 따라서 실제로 중요한 차이를 높은 확률로 탐지할 만큼 큰 표본 크기를 사용하도록 하는 것이 중요합니다.

### 목적

데이터에서 귀무 가설을 기각하기 위한 증거를 충분히 얻을 수 없는 경우, 표본 크기가 검정에서 탐지하고자 하는 실제 차이를 높은 확률로 탐지하기에 충분히 크지 확인하고자 합니다. 표본 크기 계획의 목적은 중요한 차이를 높은 확률로 탐지하기에 충분한 크기의

표본을 사용하도록 보장하는 것이지만, 표본이 너무 커서 무의미한 차이가 통계적으로 중요하다고 인식될 확률이 높아져서도 안 됩니다.






## 방법

2-표본 불량률(%) 검정의 검정력 및 표본 크기 분석은 Fisher 정확 검정의 실제 검정력에 대한 우수한 추정치를 제공하는 정규 근사 검정의 이론적 검정력 함수에 기초합니다(2-표본 불량률(%) 검정 방법 부분의 '이론적 검정력 함수의 성능'에 요약된 시뮬레이션 결과 참조). 이론적 검정력 함수는 불량률(%)과 결합 표본의 전체 불량률(%)의 목표 차이의 함수로 나타낼 수 있습니다.

## 결과

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

검정력 및 표본 크기를 확인할 때 2-표본 불량률(%) 검정에 대한 보조 도구의 보고서 카드는 결과 해석을 지원하기 위해 다음과 같은 상태를 표시합니다.

상태	조건
	검정에서 불량률(%)의 차이를 찾으므로 검정력은 문제가 되지 않음. 또는 검정력이 충분함. 검정에서 불량률(%)의 차이가 발견되지 않았지만, 표본 크기는 주어진 차이를 90% 이상의 확률로 탐지하기에 충분함(검정력 $\geq .90$ ).
	검정력이 충분할 수 있음. 검정에서 불량률(%)의 차이가 발견되지 않았지만, 표본 크기는 주어진 차이를 80%에서 90%의 확률로 탐지하기에 충분함( $.80 \leq$ 검정력 $< .90$ ). 90% 검정력을 달성하기 위해 필요한 표본 크기가 보고됨.
	검정력이 충분하지 않을 수 있음. 검정에서 불량률(%)의 차이가 발견되지 않았고 표본 크기는 주어진 차이를 60%에서 80%의 확률로 탐지하기에 충분함( $.60 \leq$ 검정력 $< .80$ ). 80% 검정력과 90% 검정력을 얻는 데 필요한 표본 크기가 보고됨.
	검정력이 충분하지 않음. 검정에서 불량률(%)의 차이가 발견되지 않았고 표본 크기는 주어진 차이를 60% 이상의 확률로 탐지할 수 있을 만큼 크지 않음(검정력 $< .60$ ). 80% 검정력과 90% 검정력을 얻는 데 필요한 표본 크기가 보고됨.
	검정에서 불량률(%) 간의 차이가 발견되지 않음. 탐지할 실제 차이가 지정되지 않음. 데이터에 따라 보고서에는 표본 크기와 알파를 기준으로 80% 및 90% 확률로 탐지할 수 있는 차이가 표시될 수 있음.

# 참고 문헌

Arnold, S.F. (1990). *Mathematical statistics*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Casella, G., & Berger, R.L. (1990). *Statistical inference*. Pacific Grove, CA: Wadsworth, Inc.

# 부록 A: 2-표본 불량률(%) 검정에 대한 추가 가정

2-표본 불량률(%) 검정은 다음 가정에 기초합니다.

- 각 표본의 데이터는  $n$ 개의 고유 품목으로 구성되며, 각 품목은 불량 또는 비불량으로 분류됩니다.
- 품목이 불량일 확률은 표본에 있는 각 품목마다 동일합니다.
- 품목이 불량일 우도는 다른 품목의 불량 여부에 영향을 받지 않습니다.

위 가정은 보조 도구의 보고서 카드의 데이터 검사에서 확인할 수 없습니다. 이 검정에는 원시 데이터가 아닌 요약 데이터가 입력되기 때문입니다.

# 부록 B: Fisher의 정확 검정

Bernoulli 분포에서 추출한 2개의 독립 랜덤 표본  $X_1, \dots, X_{n_1}$  과  $Y_1, \dots, Y_{n_2}$  를 관측한다고 가정합니다. 여기서,

$$p_1 = \Pr(X_i = 1) = 1 - \Pr(X_i = 0) \text{이고 } p_2 = \Pr(Y_j = 1) = 1 - \Pr(Y_j = 0) \text{입니다.}$$

다음 항목에서는 비율  $\delta = p_1 - p_2$  간의 차이에 대해 유추하는 절차에 대해 설명합니다.

## 공식 B1: Fisher의 정확 검정과 p-값

Fisher의 정확 검정에 대한 설명은 Arnold(1994)에서 확인할 수 있습니다. 당사는 이 검정에 대한 간략한 설명을 제공합니다.

$V$ 는 첫 표본의 성공 개수고  $v = n_1 \hat{p}_1$ 은 실험을 실시할 때 첫 표본에서 관측된 성공 개수입니다. 또한  $W$ 는 두 표본의 총 성공 개수고  $w = n_1 \hat{p}_1 + n_2 \hat{p}_2$ 은 실험을 실시할 때 관측된 성공 개수입니다.  $\hat{p}_1$ 과  $\hat{p}_2$ 는  $p_1$ 과  $p_2$ 의 표본 점 추정치입니다.

$\delta = p_1 - p_2 = 0$ 이라는 귀무 가설에 따라 주어진  $W$ 에 대한  $V$ 의 조건 분포는 확률 집단 함수가 다음과 같은 초기하학적 분포입니다.

$$f(v|w) = \frac{\binom{n_1}{v} \binom{n_2}{w-v}}{\binom{n_1+n_2}{w}}$$

$F(v|w)$ 는 분포의 c.d.f입니다. 그러면 단측 및 양측 검정에 대한 p-값은 다음과 같습니다.

- $\delta < 0$  또는 이와 동등한  $p_1 < p_2$ 에 대해 검정할 때  
p-값은  $F(v|w)$ 으로 계산됩니다. 여기서  $v$ 는 관측된  $V$  값이거나 첫 표본에서 관측된 성공의 수이고  $w$ 는 관측된  $W$  값이거나 두 표본에서 모두 관측된 성공의 수입니다.
- $\delta > 0$  또는 이와 동등한  $p_1 > p_2$ 에 대해 검정할 때  
p-값은  $1 - F(v-1|w)$ 으로 계산됩니다. 여기서  $v$ 는 관측된  $V$  값이거나 첫 표본에서 관측된 성공의 수이고  $w$ 는 관측된  $W$  값이거나 두 표본에서 모두 관측된 성공의 수입니다.
- $\delta \neq 0$  또는 이와 동등한  $p_1 \neq p_2$ 에 대해 검정할 때  
p-값은 다음 알고리즘에 따라 계산됩니다. 여기서  $m$ 은 위에서 설명한 초기하학적 분포의 모드입니다.
  - $v < m$ 이면 p-값이  $1 - F(y-1|w) + F(v|w)$ 로 계산됩니다. 여기서  $v$ 와  $w$ 는 위 정의를 따르며,  $y = \min\{k \geq m: f(k|w) \leq f(v|w)\}$ 입니다.
  - $v = m$ 이면 p-값은 1.0입니다.
  - $v > m$ 이면 p-값이  $1 - F(v-1|w) + F(y|w)$ 로 계산됩니다. 여기서  $v$ 와  $w$ 는 위 정의를 따르며,  $y = \max\{k \leq m: f(k|w) \leq f(v|w)\}$ 입니다.



# 부록 C: 이론적 검정력 함수

비율 2개(더 정확하게는 불량률(%) 2개)를 비교하기 위해, 부록 B에 설명된 대로 Fisher의 정확 검정을 사용합니다. 이 검정의 이론적 검정력 함수를 도출하기는 너무 복잡하기 때문에 근사 검정력 함수를 사용합니다. 더 구체적으로는 두 비율에 대한 잘 알려진 정규 근사 검정의 검정력 함수를 사용하여 Fisher의 정확 검정의 근사 검정력을 구합니다.

양측 검정에 대한 정규 근사의 검정력 함수는 다음과 같습니다.

$$\pi(n_1, n_2, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

설명:  $\delta = p_1 - p_2$ ,

$$se = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

그리고  $p_c = (n_1p_1 + n_2p_2)/(n_1 + n_2)$ .

$p_1 = p_2$ 을  $p_1 > p_2$ 에 대해 검정할 때 검정력 함수는 다음과 같습니다.

$$\pi(n_1, n_2, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

$p_1 = p_2$ 을  $p_1 < p_2$ 에 대해 검정할 때 검정력 함수는 다음과 같습니다.

$$\pi(n_1, n_2, \delta) = \Phi\left(\frac{-\delta - z_{\alpha}\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{se}\right)$$

# 부록 D: 근사 신뢰 구간

## 공식 D1: 두 비율의 차이에 대한 근사 신뢰 구간 계산

정규 근사에 기초한  $\delta = p_1 - p_2$ 에 대한 점근  $100(1 - \alpha)\%$  신뢰 구간은 다음과 같습니다.

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}$$

이 근사 신뢰 구간의 신뢰도를 평가하는 잘 알려진 일반 규칙은  $n_1\hat{p}_1 \geq 5$ ,  $n_1(1 - \hat{p}_1) \geq 5$ ,  $n_2\hat{p}_2 \geq 5$  및  $n_2(1 - \hat{p}_2) \geq 5$ 입니다. 다시 말해 신뢰 구간은 각 표본에서 관측된 성공 및 실패 수가 5개 이상일 경우에 정확합니다.

참고: 이 항목과 다음 항목에서는 신뢰 구간에 대한 규칙을 가장 일반적인 형식인 각 표본의 성공 수와 실패 수를 기준으로 표현합니다. 성공은 관심 대상 이벤트고 실패는 관심 대상 이벤트의 여집합입니다. 그러므로 2-표본 불량률(%) 검정이라는 구체적인 맥락에서 "성공"의 수는 불량 개수와 같고 "실패" 수는 비불량 개수와 같습니다.

## 공식 D2: 근사 신뢰 구간에 대한 규칙

정규 근사에 기초한 신뢰 구간에 사용되는 일반 규칙에 따르면,  $n_1\hat{p}_1 \geq 5$ ,  $n_1(1 - \hat{p}_1) \geq 5$ ,  $n_2\hat{p}_2 \geq 5$  및  $n_2(1 - \hat{p}_2) \geq 5$ 인 경우 신뢰 구간이 정확하다고 규정됩니다. 즉, 구간의 실제 신뢰 수준은 각 표본에 성공(불량)이 5개 이상, 그리고 실패(비불량)가 5개 이상 포함되어 있을 경우 목표 신뢰 수준과 같거나 대략적으로 같습니다.

실제로는 실제 비율을 알 수 없기 때문에 규칙은 성공과 실패의 실제 비율이 아닌 추정 비율을 기준으로 표현됩니다. 그러나 실제 비율이 가정되었거나 알려진 이론적인 설정에서는 실제 비율을 사용하여 규칙을 직접 표현할 수 있습니다. 이런 경우에는 실제로 예상되는 성공의 수와 예상되는 실패의 수  $n_1p_1$ ,  $n_2p_2$ ,  $n_1(1 - p_1)$  및  $n_2(1 - p_2)$ 가 비율의 차이에 대한 신뢰 구간의 실제 포함 확률에 영향을 미치는 방법을 직접 평가할 수 있습니다.

실제 포함 확률은 성공 확률이  $p_1$  및  $p_2$ 인 2개의 Bernoulli 모집단에서 크기가  $n_1$  및  $n_2$ 인 표본의 쌍을 다수 추출하여 계산할 수 있습니다. 다음, 실제 포함 확률을 두 비율의 실제 차이를 포함하는 신뢰 구간을 생성하는 표본 쌍의 상대 빈도로 계산합니다.  $n_1p_1 \geq 5$ ,  $n_2p_2 \geq 5$ ,  $n_1(1 - p_1) \geq 5$  및  $n_2(1 - p_2) \geq 5$ 일 때 실제 포함 확률이 정확하면 대수의 강법칙에 따라  $n_1\hat{p}_1 \geq 5$ ,  $n_1(1 - \hat{p}_1) \geq 5$ ,  $n_2\hat{p}_2 \geq 5$  및  $n_2(1 - \hat{p}_2) \geq 5$ 일 때 포함 확률이 정확합니다. 따라서 이 규칙이 유효할 경우 실제 및 목표 신뢰 구간이 가까우면 두 Bernoulli 모집단에서 추출한 표본의 쌍 중 상당한 비율이  $n_1\hat{p}_1 \geq 5$ ,  $n_1(1 - \hat{p}_1) \geq 5$ ,  $n_2\hat{p}_2 \geq 5$  및  $n_2(1 - \hat{p}_2) \geq 5$ 일 것이라 기대할 수 있습니다. 이어지는 시뮬레이션에서는 이 규칙을 규칙 1이라고 지칭합니다.

또한 이 조사 과정에서는 많은 경우  $n_1p_1 \geq 5$ 고  $n_2p_2 \geq 5$ 이거나  $n_1(1 - p_1) \geq 5$ 고  $n_2(1 - p_2) \geq 5$ 면 구간의 모의 포함 확률이 목표 범위에 가까운 것으로 나타났습니다. 이에 따라  $n_1\hat{p}_1 \geq 5$ 고  $n_2\hat{p}_2 \geq 5$ 이거나  $n_1(1 - \hat{p}_1) \geq 5$ 고  $n_2(1 - \hat{p}_2) \geq 5$ 면 근사 신뢰 구간이 정확하다고 규정하는 보다 유연한 대체 규칙이 도출되었습니다. 이어지는 시뮬레이션에서는 이 수정된 규칙을 규칙 2고 지칭합니다.

# 시뮬레이션 D1: 근사 신뢰 구간의 정확도 평가

두 비율의 차이에 대한 근사 신뢰 구간이 정확한 조건을 평가하기 위한 시뮬레이션을 수행했습니다. 특히 구간의 정확도를 다음과 같은 일반 규칙과 연관지어서 조사했습니다.

규칙 1(기본)  $n_1p_1 \geq 5, n_2p_2 \geq 5, n_1(1-p_1) \geq 5$  및  $n_2(1-p_2) \geq 5$

규칙 2(수정)  $n_1\hat{p}_1 \geq 5$  및  $n_2\hat{p}_2 \geq 5$  또는  $n_1(1-\hat{p}_1) \geq 5$  및  $n_2(1-\hat{p}_2) \geq 5$

각 실험에서는 다음과 같은 비율에 의해 정의되는 Bernoulli 모집단 쌍에서 10,000쌍의 표본을 추출했습니다.

- A 비율:  $p_1$  과  $p_2$ 가 모두 1.0(또는 0)에 가까움. 시뮬레이션에서는 이 Bernoulli 모집단의 쌍을 나타내기 위해  $p_1 = 0.8$ 과  $p_2 = 0.9$ 를 사용했습니다.
- B 비율:  $p_1$  과  $p_2$ 가 모두 0.5에 가까움. 시뮬레이션에서는 이 Bernoulli 모집단의 쌍을 나타내기 위해  $p_1 = 0.4$ 과  $p_2 = 0.55$ 을 사용했습니다.
- C 비율:  $p_1$  은 0.5에 가깝고  $p_2$ 는 1.0에 가까움. 시뮬레이션에서는 이 Bernoulli 모집단의 쌍을 나타내기 위해  $p_1 = 0.4$ 와  $p_2 = 0.9$ 를 사용했습니다.

위의 비율 분류는 근사 신뢰 구간이 도출된 이항 분포에 대한 DeMoivre-Laplace 정규 근사에 기초한 것입니다. 이 정규 근사는 Bernoulli 표본이 10보다 크고 성공 확률이 0.5에 가까울 때 정확하다고 알려져 있습니다. 성공 확률이 0 또는 1에 가까우면 일반적으로 더 큰 Bernoulli 표본이 필요합니다.

두 쌍의 표본 크기는 모두 단일 값  $n$ 으로 고정되었습니다. 여기서  $n = 10, 15, 20, 30, \dots, 100$ 입니다. 두 규칙은 모두 표본 크기와 성공 비율로 통제될 수 있는 관측된 성공 및 실패 수에 의존하기 때문에, 연구는 일반성 손실이 없는 균형 설계( $n_1 = n_2 = n$ )로 제한되었습니다.

두 모집단 비율(모의 신뢰 수준이라고 함)의 차이에 대한 신뢰 구간의 실제 신뢰 수준을 추정하기 위해 10,000개의 구간 중에서 두 비율의 실제 차이를 포함하는 구간의 비율을 계산했습니다. 각 실험의 목표 포함 확률은 0.95였습니다. 그 밖에 10,000개 표본 중에서 두 규칙에 따른 조건이 충족되는 표본의 비율(%)도 산정했습니다.

참고: 일부 소규모 표본의 경우 비율 간 차이의 추정 표준 오류는 0이었습니다. 해당 표본은 "퇴화"로 간주되어 실험에서 제외되었습니다. 그 결과 표본 반복실험의 수가 10,000보다 약간 작은 경우가 간혹 있었습니다.

결과는 표 1-11에 나와 있으며, 아래 그림 1에 그래프로 표시되어 있습니다.

표 1  $n=10$ 일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 10							
범주	비율(p)	np	n(1-p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	$p_1$	0.80	8.00	2.00	0.907	0.0	99.1
	$p_2$	0.90	9.00	1.00			

n = 10							
범주		비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)
B	$p_1$	0.40	4.00	6.00	0.928	4.4	63.0
	$p_2$	0.55	5.50	4.50			
C	$p_1$	0.45	4.50	5.50	0.919	0.0	48.3
	$p_2$	0.90	9.00	1.00			

표 2 n=15일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 15							
범주		비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)
A	$p_1$	0.80	12.00	3.00	0.938	0.2	100.0
	$p_2$	0.90	13.50	1.50			
B	$p_1$	0.40	6.00	9.00	0.914	65.0	97.3
	$p_2$	0.55	8.25	6.75			
C	$p_1$	0.45	6.75	8.25	0.93	1.2	86.9
	$p_2$	0.90	13.50	1.50			

표 3 n=20일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 20							
범주		비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)
A	$p_1$	0.80	16.00	4.00	0.942	1.5	100.0
	$p_2$	0.90	18.00	2.00			
B	$p_1$	0.40	8.00	12.00	0.943	92.8	99.9
	$p_2$	0.55	11.00	9.00			
C	$p_1$	0.45	9.00	11.00	0.934	4.1	98.2
	$p_2$	0.90	18.00	2.00			

표 4 n=30일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 30							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	p <sub>1</sub>	0.80	24.00	6.00	0.941	4.3	100.0
	p <sub>2</sub>	0.90	27.00	3.00			
B	p <sub>1</sub>	0.40	12.00	18.00	0.944	99.7	100.0
	p <sub>2</sub>	0.55	16.50	13.50			
C	p <sub>1</sub>	0.45	13.50	16.50	0.938	7.2	100.0
	p <sub>2</sub>	0.90	27.00	3.00			

표 5 n=40일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 40							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	p <sub>1</sub>	0.80	32.00	8.00	0.941	35.1	100.0
	p <sub>2</sub>	0.90	36.00	4.00			
B	p <sub>1</sub>	0.40	16.00	24.00	0.945	100.0	100.0
	p <sub>2</sub>	0.55	22.00	18.00			
C	p <sub>1</sub>	0.45	18.00	22.00	0.945	37.7	100.0
	p <sub>2</sub>	0.90	36.00	4.00			

표 6 n=50일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 50							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	p <sub>1</sub>	0.80	40.00	10.00	0.942	36.4	100.0
	p <sub>2</sub>	0.90	45.00	5.00			

n = 50							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
B	p <sub>1</sub>	0.40	20.00	30.00	0.944	100.0	100.0
	p <sub>2</sub>	0.55	27.50	22.50			
C	p <sub>1</sub>	0.45	22.50	27.50	0.935	38.3	100.0
	p <sub>2</sub>	0.90	45.00	5.00			

표 7 n=60일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 60							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	p <sub>1</sub>	0.80	48.00	12.00	0.947	72.8	100.0
	p <sub>2</sub>	0.90	54.00	6.00			
B	p <sub>1</sub>	0.40	24.00	36.00	0.947	100.0	100.0
	p <sub>2</sub>	0.55	33.00	27.00			
C	p <sub>1</sub>	0.45	27.00	33.00	0.949	73.1	100.0
	p <sub>2</sub>	0.90	54.00	6.00			

표 8 n=70일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 70							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	p <sub>1</sub>	0.80	56.00	14.00	0.939	71.7	100.0
	p <sub>2</sub>	0.90	63.00	7.00			
B	p <sub>1</sub>	0.40	28.00	42.00	0.945	100.0	100.0
	p <sub>2</sub>	0.55	38.50	31.50			
C	p <sub>1</sub>	0.45	31.50	38.50	0.944	71.8	100.0
	p <sub>2</sub>	0.90	63.00	7.00			

표 9 n=80일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 80							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	p <sub>1</sub>	0.80	64.00	16.00	0.947	91.3	100.0
	p <sub>2</sub>	0.90	72.00	8.00			
B	p <sub>1</sub>	0.40	32.00	48.00	0.947	100.0	100.0
	p <sub>2</sub>	0.55	44.00	36.00			
C	p <sub>1</sub>	0.45	36.00	44.00	0.948	91.3	100.0
	p <sub>2</sub>	0.90	72.00	8.00			

표 10 n=90일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 90							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	p <sub>1</sub>	0.80	72.00	18.00	0.947	95.18	100.0
	p <sub>2</sub>	0.90	81.00	9.00			
B	p <sub>1</sub>	0.40	36.00	54.00	0.951	100.0	100.0
	p <sub>2</sub>	0.55	49.50	40.50			
C	p <sub>1</sub>	0.45	40.50	49.50	0.945	95.2	100.0
	p <sub>2</sub>	0.90	81.00	9.00			

표 11 n=100일 때 규칙 1과 규칙 2를 충족하는 표본의 모의 포함 확률과 백분율. 목표 포함 확률은 0.95입니다.

n = 100							
범주	비율(p)	np	n(1 - p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)	
A	p <sub>1</sub>	0.80	80.00	20.00	0.952	97.7	100.0
	p <sub>2</sub>	0.90	90.00	10.00			

n = 100							
범주		비율(p)	np	n(1-p)	포함 확률	규칙 1을 충족하는 표본 비율(%)	규칙 2를 충족하는 표본 비율(%)
B	$p_1$	0.40	40.00	60.00	0.945	100.0	100.0
	$p_2$	0.55	55.00	45.00			
C	$p_1$	0.45	45.00	55.00	0.948	97.7	100.0
	$p_2$	0.90	90.00	10.00			

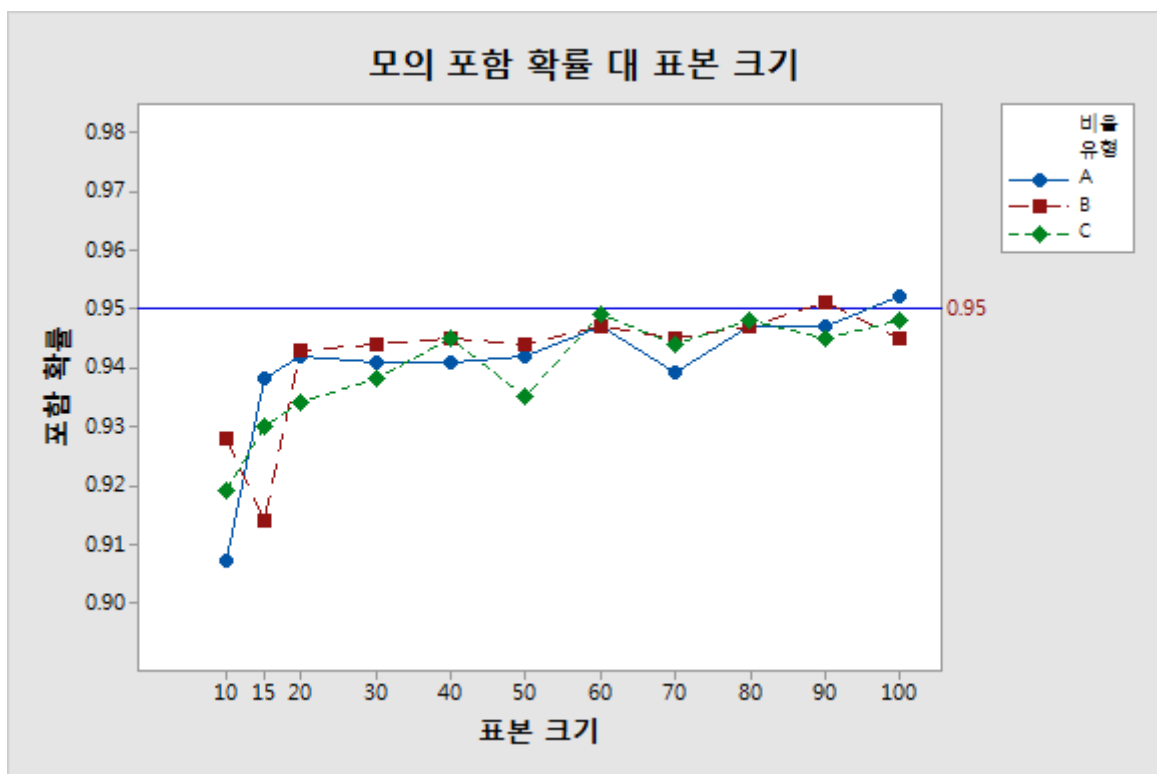


그림 1 각 Bernoulli 모집단 범주별로 표본 크기에 대해 표시된 모의 포함 확률.

표 1-11과 그림 1에 요약된 결과를 보면 (두 비율이 모두 0.5에 가까운) Bernoulli 모집단 범주 B에서 추출된 표본이 일반적으로 더 안정적이고 목표 범위 0.95에 더 가까운 모의 포함 확률을 생성함을 알 수 있습니다. 이 범주에서 두 모집단 모두의 기대 성공 및 실패 수는 표본이 작을 때도 다른 범주보다 더 큼니다.

반면에 (두 비율이 모두 1.0에 가까운) 범주 A 또는 (한 비율이 1.0에 가깝고 다른 하나는 0에 가까운) 범주 C의 Bernoulli 모집단 쌍에서 추출한 표본의 모의 포함 확률은 기대되는 성공의 수(np) 또는 기대되는 실패의 수(n(1-p))가 충분히 큰 경우를 제외하고 표본이 더 작을 때 목표에서 더 많이 벗어납니다.

예를 들어  $n = 15$ 일 때 범주 A의 Bernoulli 모집단에서 추출한 표본에 대해 생각해 보십시오. 각 모집단에 대해 기대되는 성공의 수는 12.0 및 13.5고 기대되는 실패의 수는 3.0 및 1.5입니다. 기대되는 실패 수는 두 모집단에서 모두 5보다 작지만, 모의 포함 확률은



약 0.94입니다. 이 같은 결과로 인해, 각 표본에 대해 기대되는 성공의 수 또는 기대되는 실패의 수 중 하나만 5 이상이면 된다는 규칙 2를 만들게 되었습니다.

규칙 1과 규칙 2가 신뢰 구간에 대한 근사를 효과적으로 평가할 수 있는 정도를 보다 완전하게 평가하기 위해, 규칙 1을 충족하는 표본의 백분율과 규칙 2를 충족하는 표본의 백분율을 실험에서 시뮬레이션된 포함 확률에 대해 표시한 그래프를 작성했습니다. 그래프는 그림 2에 나와 있습니다.

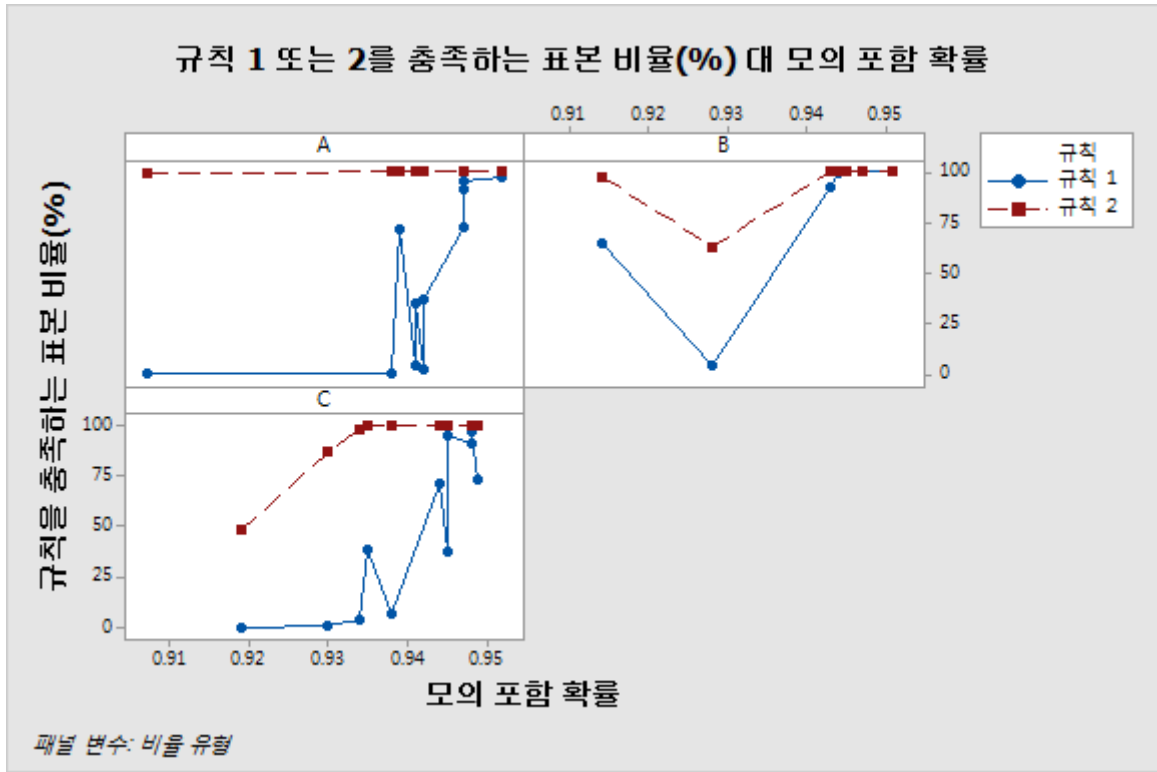


그림 2 각 Bernoulli 모집단 범주의 모의 포함 확률에 대해 표시한 규칙 1과 규칙 2를 충족하는 표본의 백분율.

그래프에는 모의 포함 확률이 목표 범위 .095에 근접할수록 각 규칙의 요건을 충족하는 표본의 백분율이 일반적으로 100%에 근접하는 것으로 나타나 있습니다. 범주 A 및 C의 Bernoulli 모집단에서 추출된 표본에 대해, 규칙 1은 표본이 작을 때 엄격합니다. 이는 모의 포함 확률이 목표값이 가까움에도 불구하고 규칙을 충족하는 표본의 비율(%)이 매우 낮다는 사실에 의해 증명됩니다. 예를 들어  $n = 20$ 이고 표본을 범주 A의 Bernoulli 모집단에서 추출했을 때 모의 포함 확률은 0.942입니다(표 3 참조). 그러나 규칙을 충족하는 표본의 비율(0.015)은 0에 가깝습니다(그림 2 참조). 그러므로 이런 경우에는 규칙이 너무 보수적일 수 있습니다.

반면에 규칙 2는 범주 A의 Bernoulli 모집단에서 추출한 작은 표본에 대해 덜 엄격합니다. 예를 들어 표 1에서 볼 수 있는 것과 같이  $n = 10$ 이고 표본을 범주 A의 Bernoulli 모집단에서 추출했을 때 모의 포함 확률은 규칙을 충족하는 표본의 0.907 및 99.1%입니다.

결론적으로 규칙 1은 표본이 작을 때 지나치게 보수적인 경향이 있습니다. 규칙 2는 덜 보수적이며, 표본 크기가 작을 때 선호될 수 있습니다. 그러나 규칙 1은 잘 알려져 있고 널리 인정됩니다. 규칙 2는 유망한 잠재력을 보이지만, 앞에서 본 바와 같이 경우에 따라 너무

개방적일 수 있습니다. 두 규칙을 합쳐서 각 규칙의 장점을 이용하는 방법도 가능하지만, 이 방법을 적용할 수 있으려면 추가 조사를 거쳐야 합니다.

# 부록 E: 실제 검정력과 이론적 검정력 비교

## 시뮬레이션 E1: Fisher의 정확 검정을 사용한 실제 검정력 추정

Fisher 정확 검정의 (모의 검정력 수준이라고 하는) 추정 실제 검정력 수준을 정규 근사 검정의 검정력 함수에 기초한 (근사 검정력 수준이라고 하는) 이론적 검정력 수준과 비교하기 위한 시뮬레이션을 설계했습니다. 각 실험에서는 Bernoulli 모집단 쌍에서 10,000쌍의 표본을 추출했습니다. 각 표본 쌍에 대해, 비율의 차이가  $p_1 - p_2 = -0.20$ 이 되도록 비율을 선택했습니다.

- A 비율:  $p_1$  과  $p_2$ 가 모두 1.0(또는 0)에 가까움. 시뮬레이션에서는 이 Bernoulli 모집단의 쌍을 나타내기 위해  $p_1 = 0.70$ 과  $p_2 = 0.90$ 를 사용했습니다.
- B 비율:  $p_1$  과  $p_2$ 가 모두 0.5에 가까움. 시뮬레이션에서는 이 Bernoulli 모집단의 쌍을 나타내기 위해  $p_1 = 0.40$ 과  $p_2 = 0.60$ 를 사용했습니다.
- C 비율:  $p_1$  은 0.5에 가깝고  $p_2$ 는 1.0에 가까움. 시뮬레이션에서는 이 Bernoulli 모집단의 쌍을 나타내기 위해  $p_1 = 0.55$ 과  $p_2 = 0.75$ 를 사용했습니다.

두 쌍의 표본 크기는 모두 단일 값  $n$ 으로 고정되었습니다. 여기서  $n = 10, 15, 20, 30, \dots, 100$ 입니다. 일반적으로 두 표본의 크기가 같다고 가정되기 때문에 연구를 균형 설계( $n_1 = n_2 = n$ )로 제한했습니다. 실제로 중요한 차이를 특정한 검정력으로 탐지하기 위해 필요한 공통 표본 크기를 계산했습니다.

각 시뮬레이션 결과를 토대로 Fisher 정확 검정의 실제 검정력을 추정하기 위해 표본 10,000쌍 중 목표 유의 수준  $\alpha = 0.05$ 에서 양측 검정이 유의했던 쌍의 비율을 분수로 계산했습니다. 다음, 비교를 위해 정규 근사 검정을 토대로 해당 이론적 검정력 수준도 계산했습니다. 결과는 아래의 표 12에 나와 있습니다.

표 12 세 Bernoulli 모집단 범주의 근사 검정력 수준과 비교한 Fisher 정확 검정의 모의 검정력 수준. 목표 유의 수준은  $\alpha = 0.05$ 입니다.

n	A 비율		B 비율		C 비율	
	$p_1 = 0.70$ $p_2 = 0.90$		$p_1 = 0.40$ $p_2 = 0.60$		$p_1 = 0.55$ $p_2 = 0.75$	
	모의 검정력	근사 검정력	모의 검정력	근사 검정력	모의 검정력	근사 검정력
10	0.063	0.193	0.056	0.14	0.056	0.149
15	0.151	0.271	0.097	0.19	0.101	0.204
20	0.244	0.348	0.146	0.24	0.183	0.259

n	A 비율		B 비율		C 비율	
	$p_1 = 0.70$ $p_2 = 0.90$		$p_1 = 0.40$ $p_2 = 0.60$		$p_1 = 0.55$ $p_2 = 0.75$	
	모의 검정력	근사 검정력	모의 검정력	근사 검정력	모의 검정력	근사 검정력
30	0.37	0.49	0.256	0.338	0.272	0.366
40	0.534	0.612	0.371	0.431	0.381	0.466
50	0.641	0.711	0.477	0.516	0.491	0.556
60	0.726	0.789	0.536	0.593	0.56	0.635
70	0.814	0.849	0.61	0.661	0.649	0.703
80	0.87	0.893	0.66	0.72	0.716	0.76
90	0.907	0.925	0.716	0.77	0.772	0.808
100	0.939	0.948	0.792	0.812	0.812	0.848

표 12에는 세 Bernoulli 모집단 범주(A, B, C)에 대해 모두 근사 검정력이 모의 검정력보다 높은 경향이 있다는 결과가 나와 있습니다. 예를 들어 A 범주의 비례에 대해 절대 차이 - 0.20을 근사 검정력 수준 0.91로 탐지하기 위해 필요한 실제 표본 크기는 약 90입니다. 반대로, 근사 이론적 검정력 함수에 기초한 해당 표본 크기 추정치는 약 85입니다. 그러므로 근사 검정력 함수에 기초한 표본 크기 추정치는 일반적으로 주어진 검정력 수준을 얻기 위해 필요한 실제 표본 크기보다 약간 더 작습니다.

결과를 아래 그림 3에 나와 있는 검정력 곡선으로 표시하면 이 관계를 더 분명히 볼 수 있습니다.

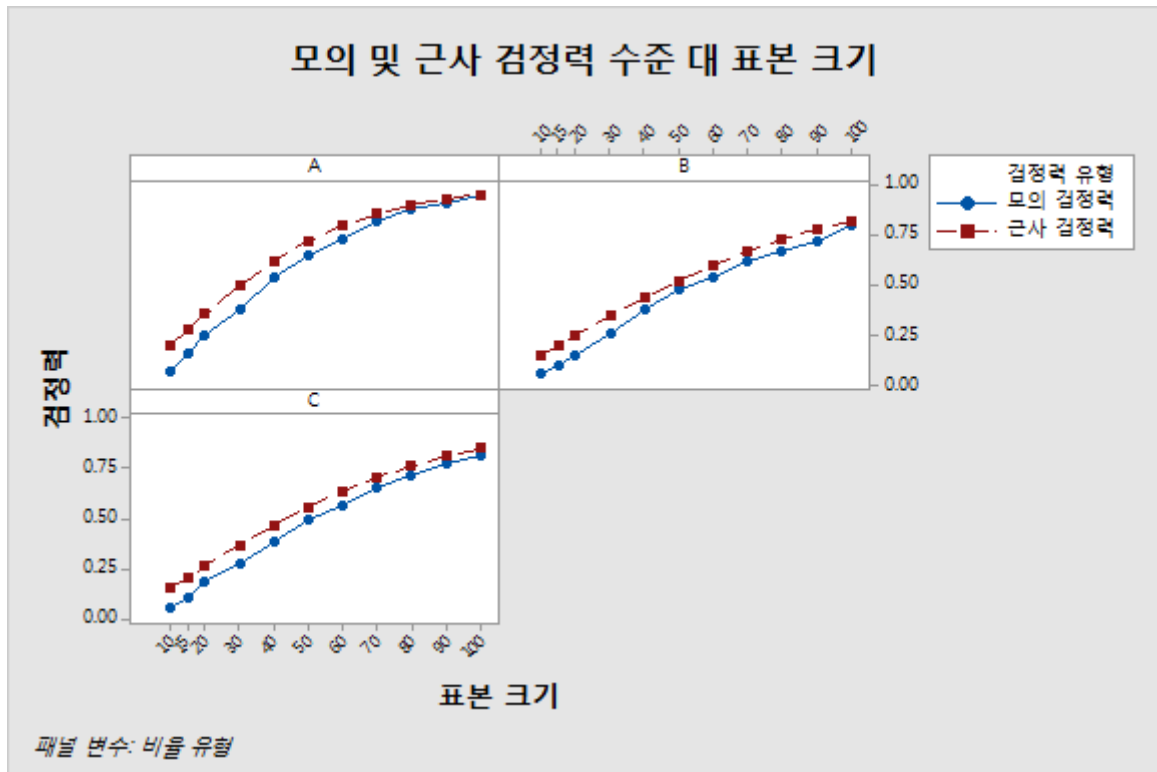


그림 3 두 비율의 비교를 위한 양측 검정의 모의 및 근사 검정력 수준 그래프. 검정력 수준은 각 Bernoulli 모집단 범주에 대한 별도의 패널에 표본 크기를 기준으로 표시되었습니다.

모의 검정력 곡선은 세 Bernoulli 모집단 범주(A, B, C)에 대해 모두 근사 검정력 곡선보다 낮지만, 곡선 간 차이의 크기는 표본이 추출된 Bernoulli 모집단의 실제 비율에 따라 달라집니다. 예를 들어 두 비율이 0.5에 가까우면(범주 B) 두 검정력 수준은 일반적으로 가깝습니다. 그러나 모집단 범주 A 및 C와 관련된 비율에 대한 검정력 곡선 2개의 차이는 작은 표본에서 더 두드러지게 나타납니다.

이런 결과는 정규 근사 검정의 이론적 검정력 함수가 일반적으로 Fisher 정확 검정의 모의 검정력 함수와 거의 같음을 나타냅니다. 그러므로 보조 도구에서는 Fisher의 정확 검정을 수행하기 전에 정규 근사 검정의 이론적 검정력 함수를 사용하여 표본 크기를 추정합니다. 그러나 근사 검정력 함수를 사용하여 계산한 표본 크기는 두 비율(불량률(%))의 차이를 탐지하는 일정한 검정력을 얻는 데 필요한 실제 표본 크기보다 약간 더 작을 수 있습니다.

© 2020 Minitab, LLC. All rights reserved. Minitab®, Minitab Workspace™, Companion by Minitab®, Salford Predictive Modeler®, SPM®, and the Minitab® logo are all registered trademarks of Minitab, LLC, in the United States and other countries. Additional trademarks of Minitab, LLC can be found at [www.minitab.com](http://www.minitab.com). All other marks referenced remain the property of their respective owners.