

2-표본 t-검정

개요

2-표본 t-검정은 2개의 독립적인 집단이 서로 다른지 비교하기 위해 사용할 수 있습니다. 이 검정은 두 모집단이 모두 정규 분포를 따르고 분산이 같다는 가정 하에서 도출됩니다. 정규성 가정은 중요하지 않지만(Pearson, 1931; Barlett, 1935; Geary, 1947), 등분산 가정은 표본 크기가 크게 다른 경우에 중요합니다(Welch, 1937; Horsnell, 1953).

일부 실무자는 기존 2-표본 t 절차를 수행하기 전에 등분산을 평가하기 위한 예비 검정을 먼저 수행합니다. 그러나 이런 분산 검정에는 중요한 가정과 제한이 적용되기 때문에 이 방법에는 심각한 단점이 있습니다. 예를 들어 기존 F-검정 같은 여러 등분산 검정은 정규성이탈에 민감하게 반응합니다. 그 외에 Levene/Brown-Forsythe처럼 정규성 가정에 의존하지 않는 검정은 분산의 차이를 탐지하는 검정력이 낮습니다.

B.L. Welch는 2개의 독립적인 정규 분포의 분산이 다를 경우 두 분포의 평균을 비교하는 근사 방법을 개발했습니다(Welch, 1947). Welch의 수정된 t-검정은 등분산 가정에 따라 도출되지 않으므로 사용자는 등분산 검정을 먼저 실시하지 않고도 두 모집단의 평균을 비교할 수 있습니다.

이 문서에서는 Welch의 수정된 t 방법을 기존 2-표본 t 절차와 비교하고 어느 절차의 신뢰성이 가장 높은지 결정합니다. 또한 자동으로 수행되고 보조 도구 보고서 카드에 표시되는 다음 데이터 검사에 대해 설명하고 이런 검사가 분석 결과에 어떤 영향을 미치는지 설명합니다.

- 정규성
- 비정상 데이터
- 표본 크기

2-표본 t-검정 방법

기존 2-표본 t-검정과 Welch의 t-검정 비교

분산이 같은 2개의 정규 분포 모집단에서 얻어진 데이터에 대해 기존 2-표본 t-검정의 검정력은 Welch의 t-검정과 같거나 더 우수합니다. 정규성 가정은 기존 절차에 중요하지 않지만(Pearson, 1931; Barlett, 1935; Geary, 1947), 등분산 가정은 유효한 결과를 보장하는 데 중요합니다. 더 구체적으로 말하면 기존 절차는 표본 크기가 다를 때 표본 크기에 관계 없이 등분산 가정에 민감합니다(Welch, 1937; Horsnell, 1953). 그러나 실제로는 등분산 가정이 참인 경우가 드물며, 이로 인해 제1종 오류율이 더 높아질 수 있습니다. 그러므로 두 표본의 편차가 다를 때 기존 2-표본 t-검정을 사용하면 검정에서 잘못된 결과가 나올 가능성이 더 큼니다.

Welch의 t-검정은 분산이 같다고 가정하지 않아 표본 크기에 관계 없이 동일하지 않은 분산에 민감하게 반응하지 않으므로 기존 t-검정을 대체할 수 있는 실용적인 대안입니다. 그러나 Welch의 t-검정은 근사 기반이고 표본 크기가 작을 때 효과적이지 않을 수 있습니다. 당사는 Welch의 t-검정이나 기존 2-표본 t-검정이 보조 도구에서 사용할 수 있는 가장 신뢰할 수 있는 실용적인 검정인지 판단하고자 했습니다.

목적

당사는 시뮬레이션 연구와 이론적인 도출을 통해 Welch의 t-검정과 기존 2-표본 t-검정 중 어느 검정의 신뢰도가 더 높은지 파악하고자 했습니다. 보다 구체적으로는 다음을 조사하고자 했습니다.

- 데이터가 정규 분포를 따르고 분산이 같을 때 다양한 표본 크기에서 기존 2-표본 t-검정과 Welch t-검정 모두의 제1종 및 제2종 오류율
- 기존 2-표본 t-검정이 효과적이지 않은 분산이 다른 불균형 설계에 대한 Welch의 t-검정의 제1종 및 제2종 오류율

방법

시뮬레이션에서는 다음 세 가지 부분에 초점을 맞췄습니다.

- 정규성, 비정규성, 등분산, 이분산, 균형 및 불균형 설계가 포함된 다양한 모형 가정 하에서 기존 2-표본 t-검정과 Welch t-검정의 모의 검정 결과를 비교했습니다. 자세한 내용은 부록 A를 참조하십시오.
- Welch t-검정의 검정력 함수를 도출하고 기존 2-표본 t-검정의 검정력 함수와 비교했습니다. 자세한 내용은 부록 B를 참조하십시오.
- 비정규성이 Welch t-검정의 이론적 검정력 함수에 미치는 영향을 조사했습니다.

결과

기존 2-표본 t 모형에 대한 가정이 참이면 Welch의 t-검정은 소규모 불균형 설계를 제외하고 기존 2-표본 t-검정과 검정력이 유사합니다. 그러나 기존 2-표본 t-검정은 등분산 가정에 대한 민감도가 높기 때문에 소규모 불균형 설계에서 검정력이 낮을 수도 있습니다. 또한 실제 설정에서는 두 모집단의 분산이 정확히 동일함을 증명하기가

어렵습니다. 따라서 Welch의 t-검정과 비교한 기존 2-표본 검정의 이론적인 우월성은 실용적인 가치가 거의 또는 전혀 없습니다. 이런 이유로 보조 도구에서는 Welch의 t-검정을 사용하여 두 모집단의 평균을 비교합니다. 자세한 시뮬레이션 결과는 부록 A, B, C를 참조하십시오.

데이터 검사

정규성

보조 도구에서 독립적인 두 모집단의 평균을 비교하기 위해 사용하는 Welch의 t-검정은 모집단이 정규 분포를 따른다는 가정 하에 도출됩니다. 데이터가 정규 분포를 따르지 않더라도 Welch의 t-검정은 표본이 충분히 클 경우 검정력이 우수합니다.

목적

당사는 Welch 방법과 기존 2-표본 t-검정의 모의 유의 수준이 목표 유의 수준(제1종 오류율) 0.05와 얼마나 가깝게 일치하는지 확인하고자 했습니다.

방법

정규 분포를 따르는 모집단, 치우친 모집단 및 오염된 정규 분포를 따르는 (등분산 및 이분산) 모집단으로부터 생성된 10,000쌍의 독립 표본에 대해 Welch의 t-검정과 기존 2-표본 t-검정의 시뮬레이션을 수행했습니다. 표본의 크기는 다양했습니다. 정규 분포 모집단은 비교를 위한 대조 모집단으로 사용됩니다. 각 조건에 대해 모의 유의 수준을 계산하고 이를 목표 또는 명목 유의 수준 0.05와 비교했습니다. 검정이 효과적이면 모의 유의 수준이 0.05와 가까울 것입니다.

결과

표본 크기가 중간 이상일 때 Welch의 t-검정은 정규 및 비정규 데이터에 대해 모두 일정한 제1종 오류율을 유지합니다. 두 표본의 크기가 모두 15 이상이면 모의 유의 수준은 목표 유의 수준에 가깝습니다. 자세한 내용은 부록 A를 참조하십시오.

이 검정은 표본이 상대적으로 작을 때도 효과적이기 때문에 보조 도구는 데이터의 정규성을 검사하지 않습니다. 대신 표본 크기를 확인하고 보고서 카드에 다음과 같은 상태를 표시합니다.

상태	조건
	두 표본 크기가 모두 15 이상이므로 정규성은 문제가 되지 않음.
	하나 이상의 표본 크기가 15 미만이므로 정규성이 문제가 될 수 있음.

비정상 데이터

비정상 데이터는 특이치라고도 하는 극도로 크거나 작은 데이터 값입니다. 비정상 데이터는 분석 결과에 심각한 영향을 미칠 수 있습니다. 표본이 작을 때 비정상 데이터는 통계적으로 유의한 결과를 찾을 확률에 영향을 미칠 수 있습니다. 비정상 데이터는 데이터 수집의 문제나 프로세스의 비정상적인 동작을 나타낼 수 있습니다. 따라서 이런 데이터 점은 조사할만한 가치가 많으며, 가능하면 수정해야 합니다.

목적

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

방법

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

결과

보조 도구는 분포의 하위 또는 상위 사분위수로부터 사분위간 범위의 1.5배 이상 벗어난 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.

특이치는 표본 크기가 매우 작을 때만 검정력 함수에 영향을 미치는 경향이 있습니다. 일반적으로, 특이치가 있을 때 관찰되는 검정력 값은 목표 이론적 검정력 값보다 약간 높은 경향이 있습니다. 이 패턴은 부록 C의 그림 10에서 확인할 수 있습니다. 여기서 모의 및 이론적 검정력 곡선은 최소 표본 크기가 15에 도달할 때까지 합당한 수준으로 가까워지지 않습니다.

비정상 데이터가 있는지 검사할 때 2-표본 t-검정에 대해 보조 도구 보고서 카드에는 다음과 같은 상태가 표시됩니다.

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

표본 크기

일반적으로 가설 검정을 수행하여 "차이가 없음" 귀무 가설을 기각하기 위한 증거를 수집합니다. 표본이 너무 작으면 검정의 검정력이 실제로 존재하는 평균의 차이를 탐지하는데 충분하지 않아서 제2종 오류가 발생할 수 있습니다. 따라서 실제로 중요한 차이를 높은 확률로 탐지할 만큼 큰 표본 크기를 사용하는 것이 중요합니다.

목적

현재 데이터에서 귀무 가설을 반박하는 증거를 충분히 얻을 수 없는 경우에는 표본 크기가 검정에서 탐지하고자 하는 실제 차이를 높은 확률로 탐지하기에 충분히 크지 확인하고자 했습니다. 표본 크기 계획의 목적은 중요한 차이를 높은 확률로 탐지할 수 있는 충분한 크기의 표본을 사용하기 위한 것이지만, 표본이 너무 커서 무의미한 차이를 통계적으로 유의하다고 판단내릴 확률이 높아져서도 안 됩니다.

방법

검정력 및 표본 크기 분석은 통계적 분석을 수행하는 데 사용되는 특정 검정의 이론적 검정력 함수에 기초합니다. Welch t-검정의 경우 이 검정력 함수는 표본 크기와 두 모집단의 평균 차이, 그리고 두 모집단의 실제 분산에 의해 결정됩니다. 자세한 내용은 부록 B를 참조하십시오.

결과

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 탐지하려는 실제 차이를 사용자가 제공할 경우 차이를 80% 및 90% 확률로 탐지하는 데 필요한 표본 크기를 계산합니다.

결과는 구체적인 표본에 따라 달라지므로 보고할 일반적인 결과는 없습니다. 그러나 Welch 검정의 검정력 함수에 대한 자세한 내용은 부록 B와 C를 참조하십시오.

검정력과 표본 크기를 확인할 때 2-표본 t-검정에 대해 보조 도구 보고서 카드에는 다음과 같은 상태가 표시됩니다.

상태	조건
	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.
	검정에서 평균의 차이를 찾지 못함. 탐지할 평균의 실제 차이를 지정하지 않았으므로 보고서에는 표본 크기, 표준 편차 및 알파를 기준으로 80% 및 90% 확률로 탐지할 수 있는 차이가 표시됨.

참고 문헌

- Arnold, S. F. (1990). *Mathematical Statistics*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Aspin, A. A.A. (1949). Tables for Use in Comparisons whose Accuracy Involves Two Variances, Separately Estimated, *Biometrika*, 36, 290-296.
- Bartlett, M. S. (1935). The effect of non-normality on the t-distribution. *Proceedings of the Cambridge Philosophical Society*, 31, 223-231.
- Box, G. E. P. (1953). Non-normality and Tests on Variances, *Biometrika*, 40, 318-335.
- Geary, R. C. (1947). Testing for Normality, *Biometrika*, 34, 209-242.
- 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.
- Horsnell, G. (1953). The effect of unequal group variances on the F test for homogeneity of group means. *Biometrika*, 40, 128-136.
- James, G. S. (1951). The comparison of several groups of observations when the ratios of the populations variances are unknown, *Biometrika*, 38, 324-329.
- 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.
- Lehmann, E. L. (1959). *Testing statistical hypotheses*. New York, NY: Wiley.
- Neyman, J., Iwazskiewicz, K. & Kolodziejczyk, S. (1935). Statistical problems in agricultural experimentation, *Journal of the Royal Statistical Society, Series B*, 2, 107-180.
- Pearson, E. S. (1931). The Analysis of variance in case of non-normal variation, *Biometrika*, 23, 114-133.
- Pearson, E.S. & Hartley, H.O. (Eds.). (1954). *Biometrika Tables for Statisticians*, Vol. I. London: Cambridge University Press.
- Srivastava, A. B. L. (1958). Effect of non-normality on the power function of t-test, *Biometrika*, 45, 421-429.
- Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38, 330-336.
- Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34, 28-35.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal, *Biometrika*, 29, 350–362.

Wolfram, S. (1999). *The Mathematica Book* (4th ed.). Champaign, IL: Wolfram Media/Cambridge University Press.

부록 A: 비정규성과 이종성이 기존 2-표본 t-검정과 Welch t-검정에 미치는 영향

서로 다른 모형 가정 하에서 기존 2-표본 t-검정과 Welch의 t-검정을 비교하기 위해 고안된 여러 시뮬레이션 연구를 실시했습니다.

시뮬레이션 연구 A

연구는 다음의 세 부분으로 나뉘어 실시되었습니다.

- 연구의 첫 번째 부분에서는 정규성 가정이 참일 때 기존 2-표본 t-검정과 Welch의 t-검정의 등분산 가정에 대한 민감도를 조사했습니다. 2개의 독립적인 정규 분포로부터 표본 2개를 생성했습니다. 기존 표본인 첫 번째 표본은 평균이 0이고 표준 편차가 $\sigma_1 = 2$, $N(0,2)$ 인 정규 모집단에서 추출했습니다. 두 번째 표본도 평균이 0인 정규 분포에서 추출되었지만, 비율이 0.5, 1.0, 1.5 및 2가 되도록 표준 편차를 선택했습니다. 다시 말해, 두 번째 표본은 각각 $N(0, 1)$, $N(0, 2)$, $N(0, 3)$, 및 $N(0, 4)$ 모집단에서 추출되었습니다. 또한 각각의 경우에 기존 표본 크기는 20로 고정되었고 각각의 주어진 μ_2 에 대해 표본 크기의 비율이 약 0.5, 1, 1.5 및 2.0과 같도록 두 번째 표본 크기를 선택했습니다.

이들 2-표본 설계에 대해서는 각 모집단으로부터 10,000쌍의 독립적인 표본을 각각 생성했습니다. 그런 다음, 10,000쌍의 표본에 대해 각각 2-표본 t-검정과 Welch의 t-검정을 수행하여 평균의 차이가 없다는 귀무 가설을 검정했습니다. 평균의 실제 차이가 없으므로, 10,000회의 반복실험 중에서 귀무 가설이 기각된 비율이 검정의 모의 유의 수준을 나타냅니다. 각 검정의 목표 유의 수준은 $\alpha = 0.05$ 이므로, 각 검정 및 각 실험과 연관된 시뮬레이션 오류는 약 0.2%입니다.

- 두 번째 부분에서는 비정규성, 특히 치우침이 두 검정의 모의 유의 수준에 미치는 영향을 조사했습니다. 이 시뮬레이션은 기존 표본을 자유도가 2인 카이-제곱 분포 $\text{Chi}(2)$ 에서 추출하고 두 번째 표본을 다른 카이-제곱 분포에서 추출하여 $\rho = \sigma_2/\sigma_1$ 의 값이 0.5, 1.0, 1.5 및 2가 되도록 했다는 점을 제외하고 이전 시뮬레이션과 같은 방법으로 설정되었습니다. 가설 평균 차이는 모집단 평균의 실제 차이가 되도록 설정되었습니다.
- 세 번째 부분에서는 특이치가 두 t-검정의 성능에 미치는 영향을 조사했습니다. 이런 이유로, 표본 2개는 오염된 정규 분포에서 추출되었습니다. 오염된 정규 분포 $CN(p, \sigma)$ 는 2개의 정규 모집단, 즉 $N(0,1)$ 모집단과 정규 $N(0, \sigma)$ 모집단을 혼합한 것입니다. 오염된 정규 분포는 다음과 같이 정의됩니다.

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

여기서 p 는 혼합 모수이고 $1-p$ 는 오염의 비율, 즉 특이치의 비율입니다. X 가 $CN(p, \sigma)$ 로 분포된 경우 평균은 $\mu_X = 0$ 이고 표준 편차는 $\sigma_X = \sqrt{p + (1-p)\sigma^2}$ 임을 증명하는 것은 쉽습니다.

기준 표본은 $CN(.8, 4)$ 에서 도출되었고 두 번째 표본은 오염된 정규 분포 $CN(.8, \sigma)$ 에서 추출되었습니다. I부 및 II부와 마찬가지로, (오염된) 두 모집단의 표준 편차 비율 $\rho = \sigma_2/\sigma_1$ 이 0.5, 1.0, 1.5 및 2와 같도록 모수 σ 를 선택했습니다. $\sigma_1 = \sqrt{.8 + (1 - .8) * 16} = 2.0$ 이기 때문에 이는 각각 $\sigma = 1, 4, 6.40, 8.72$ 를 선택하는 결과로 이어집니다. 다시 말해 두 번째 표본은 $CN(.8, 1)$, $CN(.8, 4)$, $CN(.8, 6.4)$ 및 $CN(.8, 8.72)$ 에서 추출되었습니다. 그 다음으로, I부에서 설명한 것과 같은 방법으로 시뮬레이션을 실시했습니다.

연구 결과는 표 1에 정리되고 그림 1, 2, 3에 표시되었습니다.

결과 및 요약

시뮬레이션 결과는 대체로 기준 2-표본 t-검정이 정규성과 등분산 가정 하에서 표본 크기가 작을 때도 목표 수준에 가까운 유의 수준을 제공한다는 이론적인 결과를 뒷받침합니다. 그림 1의 두 번째 열 그림은 두 정규 모집단의 분산이 동일한 설계의 모의 유의 수준을 표시합니다. 기준 2-표본 t-검정에 따른 모의 유의 수준은 목표 수준 선과 구별되지 않습니다.

아래 표에는 정규 모집단, 치우친 모집단(카이-제곱) 및 오염된 정규 모집단으로부터 각각 생성된 표본의 쌍을 기준으로 $\alpha = 0.05$ 인 기준 2-표본 t-검정과 Welch의 t-검정 모두에 대한 양측 검정의 모의 유의 수준이 나와 있습니다. 표본의 쌍은 같은 분포 유형에서 추출되었지만, 각 모집단의 분산은 같지 않을 수 있습니다.

표 1 $n = 5$ 일 때 양측 검정(각각 $\alpha = 0.05$ 인 기준 2-표본 t-검정과 Welch의 t-검정)의 모의 유의 수준

			기준 모집단: $N(0, 2)$ 2차 모집단: $N(0, \sigma_2)$				기준 모집단: $\text{Chi}(2)$ 2차 모집단: 카이-제곱				기준 모집단: $CN(.8, 4)$ 2차 모집단: $CN(8, \sigma)$			
			.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5	2.0
n_2	$\frac{n_2}{n_1}$	방법	$n_1 = 5$				$n_1 = 5$				$n_1 = 5$			
3	.6	2T	.035	.050	.079	.105	.058	.042	.078	.113	.031	.036	.035	.034
		Welch	.035	.039	.049	.055	.048	.029	.055	.063	.029	.024	.021	.020
5	1.0	2T	.061	.052	.054	.058	.086	.036	.054	.064	.035	.031	.025	.023
		Welch	.048	.042	.044	.047	.066	.021	.040	.050	.027	.023	.018	.016
8	1.6	2T	.096	.048	.033	.027	.133	.041	.033	.032	.059	.037	.029	.024
		Welch	.050	.045	.043	.042	.094	.034	.032	.041	.034	.029	.026	.022
10	2.0	2T	.118	.055	.034	.025	.139	.041	.028	.024	.073	.041	.028	.023
		Welch	.052	.051	.050	.051	.097	.041	.033	.042	.035	.032	.028	.025

표 2 n = 10일 때 양측 검정(각각 $\alpha = 0.05$ 인 기준 2-표본 t-검정과 Welch의 t-검정)의 모의 유의 수준

			기준 모집단: N(0,2) 2차 모집단: N(0, σ_2)				기준 모집단: Chi(2) 2차 모집단: 카이-제곱				기준 모집단: CN(.8,4) 2차 모집단: CN(8, σ)			
			.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5	2.0
n_2	$\frac{n_2}{n_1}$	방법	$n_1 = 10$				$n_1 = 10$				$n_1 = 10$			
5	.5	2T	.020	.050	.081	.112	.039	.044	.091	.123	.021	.035	.045	.047
		Welch	.046	.048	.050	.050	.043	.047	.067	.063	.034	.028	.022	.019
10	1.0	2T	.057	.051	.053	.055	.068	.044	.053	.054	.043	.042	.037	.032
		Welch	.051	.049	.049	.049	.062	.037	.046	.049	.039	.038	.032	.027
15	1.5	2T	.088	.048	.034	.029	.100	.043	.032	.032	.064	.040	.028	.021
		Welch	.050	.048	.047	.048	.074	.044	.041	.046	.035	.037	.035	.031
20	2	2T	.110	.048	.026	.019	.133	.042	.026	.022	.093	.046	.029	.019
		Welch	.048	.047	.045	.046	.083	.050	.044	.049	.036	.039	.040	.038

표 3 n = 15일 때 양측 검정(각각 $\alpha = 0.05$ 인 기준 2-표본 t-검정과 Welch의 t-검정)의 모의 유의 수준

			기준 모집단: N(0,2) 2차 모집단: N(0, σ_2)				기준 모집단: Chi(2) 2차 모집단: 카이-제곱				기준 모집단: CN(.8,4) 2차 모집단: CN(8, σ)			
			.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5	2.0
n_2	$\frac{n_2}{n_1}$	방법	$n_1 = 15$				$n_1 = 15$				$n_1 = 15$			
8	.53	2T	.021	.050	.083	.110	.036	.041	.089	.114	.022	.044	.056	.062
		Welch	.050	.051	.051	.050	.047	.049	.067	.062	.044	.036	.027	.022
15	1.0	2T	.049	.047	.050	.053	.064	.046	.051	.061	.045	.045	.041	.037
		Welch	.045	.046	.049	.048	.060	.042	.048	.057	.042	.043	.039	.033
23	1.53	2T	.081	.049	.033	.028	.103	.042	.036	.030	.075	.048	.033	.024
		Welch	.048	.049	.048	.050	.071	.042	.048	.050	.042	.045	.044	.041

			기준 모집단: N(0,2) 2차 모집단: N(0, σ_2)				기준 모집단: Chi(2) 2차 모집단: 카이-제곱				기준 모집단: CN(.8,4) 2차 모집단: CN(8, σ)			
		$\frac{\sigma_2}{\sigma_1}$.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5	2.0
n_2	$\frac{n_2}{n_1}$	방법	$n_1 = 15$				$n_1 = 15$				$n_1 = 15$			
30	2.0	2T	.111	.050	.028	.018	.123	.049	.027	.020	.100	.046	.025	.016
		Welch	.049	.051	.051	.053	.074	.056	.045	.047	.039	.044	.042	.040

표 4 $n = 20$ 일 때 양측 검정(각각 $\alpha = 0.05$ 인 기준 2-표본 t-검정과 Welch의 t-검정)의 모의 유의 수준

			기준 모집단: N(0,2) 2차 모집단: N(0, σ_2)				기준 모집단: Chi(2) 2차 모집단: 카이-제곱				기준 모집단: CN(.8,4) 2차 모집단: CN(8, σ)			
		$\frac{\sigma_2}{\sigma_1}$.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5	2.0
n_2	$\frac{n_2}{n_1}$	방법	$n_1 = 20$				$n_1 = 20$				$n_1 = 20$			
10	.5	2T	.019	.052	.087	.115	.028	.048	.087	.119	.021	.048	.067	.079
		Welch	.050	.054	.053	.053	.044	.054	.061	.061	.048	.042	.035	.028
20	1.0	2T	.048	.049	.052	.053	.057	.046	.052	.056	.049	.044	.042	.040
		Welch	.045	.049	.051	.050	.055	.044	.050	.052	.047	.042	.040	.037
30	1.5	2T	.086	.054	.039	.032	.098	.047	.035	.033	.075	.047	.033	.022
		Welch	.054	.054	.053	.052	.068	.047	.051	.053	.041	.043	.044	.042
40	2.0	2T	.107	.049	.026	.016	.123	.046	.027	.019	.107	.047	.026	.016
		Welch	.048	.049	.046	.047	.070	.054	.046	.045	.044	.043	.043	.042

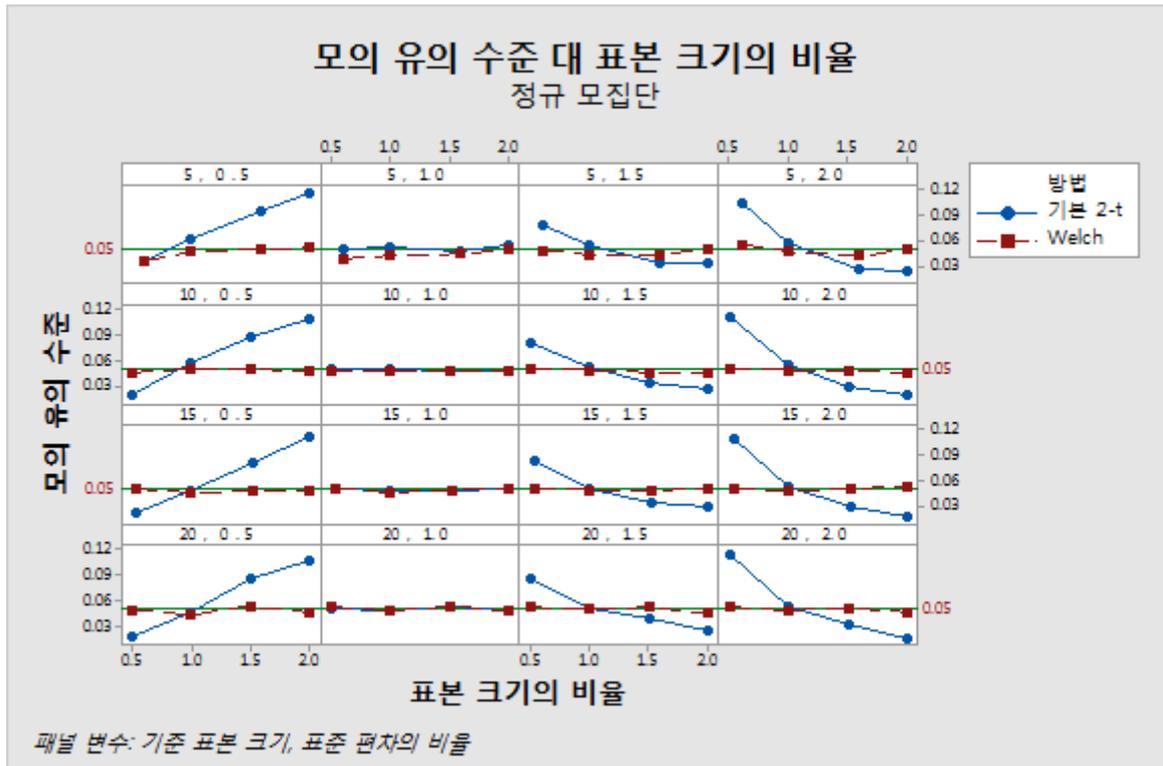


그림 1 표본 크기의 비율과 분산이 같거나 다른 2개의 정규 분포로부터 생성된 표본 쌍에 기초한 두 양측 검정(각각 $\alpha = 0.05$ 인 기본 2-표본 t-검정과 Welch의 t-검정)의 모의 유의 수준 그래프.

시뮬레이션에서는 표본이 비교적 작을 경우 기본 2-표본 t-검정이 비정규성에 대해서는 강하지만 2-표본 설계가 불균형인 경우 등분산 가정에는 민감하다는 결과가 나왔습니다. 이 내용은 그림 1, 2, 3에 그래프로 나와 있습니다. 기본 2-표본 t-검정에 기초한 모의 유의 수준 곡선은 분산이 매우 다를 때도 표본 크기의 비율이 1.0인 지점에서 목표 수준 곡선과 교차합니다. 세 가지 분포 유형(정규, 카이-제곱 및 오염된 정규 분포)의 경우에 모두 표본 크기가 다르면 기본 2-표본 t-검정의 모의 유의 수준은 분산이 같을 경우에만 목표 수준에 근접합니다. 이 내용은 그림 1, 2, 3의 두 번째 열 그림에 각각 묘사되어 있습니다.

기본 t-검정은 설계가 불균형이고 분산이 같지 않을 때 효과적이지 않습니다. 분산에 작은 차이만 있어도 문제가 될 수 있습니다. 이처럼 분산이 다른 불균형 설계에서는 데이터의 정규성에 의해 모의 유의 수준이 개선되지 않습니다. 실제로 모의 유의 수준은 표본 크기가 증가할 때 모집단에 관계 없이 목표 수준에서 멀어집니다. 분산이 더 큰 모집단에서 더 큰 표본을 추출하면 모의 유의 수준이 목표 수준보다 더 작습니다. 분산이 더 작은 모집단에서 더 큰 표본을 추출하면 모의 유의 수준이 목표 수준보다 더 큼니다. Arnold(1990, 372페이지)는 이분산 가정 하에서 기본 2-표본 t-검정 통계량의 점근 분포를 조사할 때 유사한 의견을 제시했습니다.

반면에 Welch 2-표본 t-검정은 그림 1, 2, 3에 나와 있는 것처럼 등분산 가정 이탈의 영향도 받지 않습니다. Welch t-검정은 등분산 가정 하에서 도출되지 않기 때문입니다.

Welch의 t-검정을 도출하는 데 사용된 정규 가정은 두 표본 크기 중 최소값이 매우 작을 때만 중요한 것으로 보입니다. 그러나 표본이 큰 경우 검정은 정규 가정 이탈의 영향을 받지 않습니다. 이 내용은 두 표본의 최소 크기가 15일 때 모의 유의 수준이 목표 수준과 계속 가깝게 유지되는 그림 2 및 3에 설명되어 있습니다. 두 표본이 모두 자유도가 2인 카이-

제곱 분포에서 생성되고 두 표본의 크기가 모두 15일 때 모의 유의 수준은 0.042입니다(표 3 참조).

특이치 역시 두 표본의 최소 크기가 충분히 크면 Welch의 t-검정의 성능에 영향을 미치지 않습니다. 표 3과 그림 3에서는 두 표본의 최소 크기가 15 이상이면 모의 유의 수준이 목표 수준에 가깝다는 것을 볼 수 있습니다.(표준 편차 비율이 0.5, 1.0, 1.5 및 2.0일 때 모의 유의 수준은 각각 0.045, 0.045, 0.041, 0.037입니다.)

이런 결과는 실제로 대부분 Welch 2-표본 t-검정이 모의 유의 수준 또는 제1종 오류율을 기준으로 기존 2-표본 t-검정보다 더 우수함을 보여줍니다.

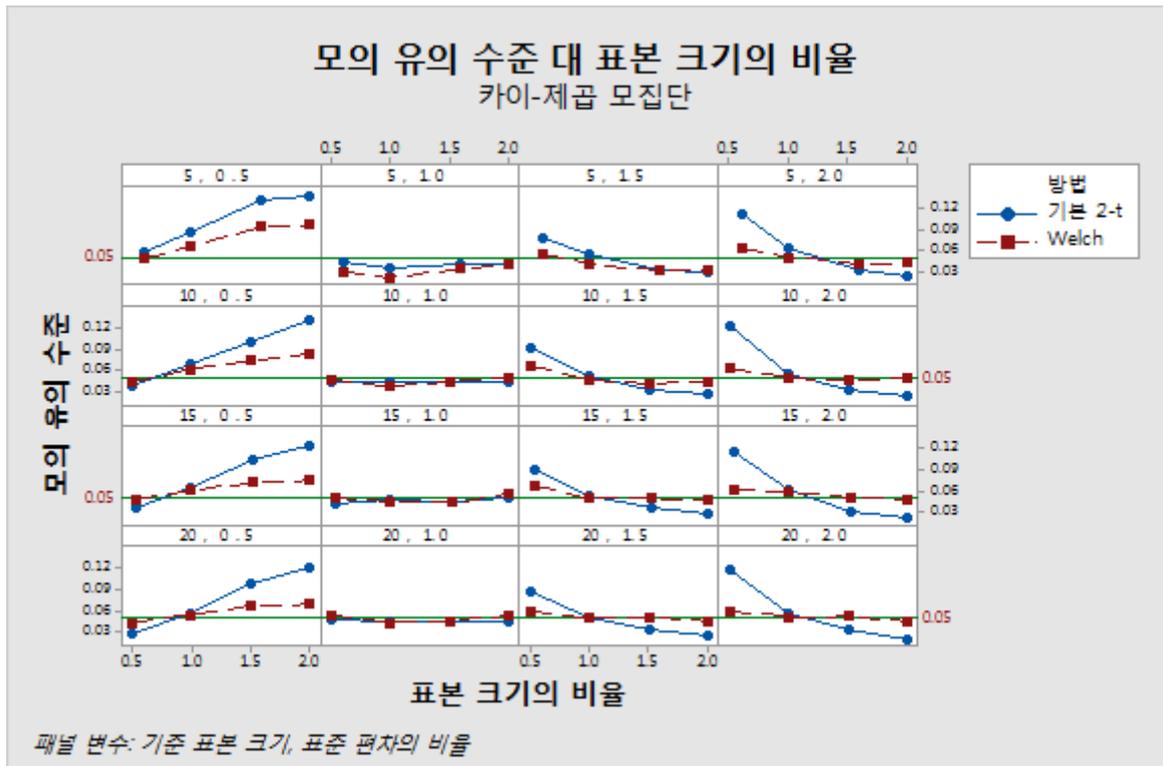


그림 2 표본 크기 비율과 분산이 같거나 다른 2개의 정규 분포로부터 생성된 표본 쌍에 기초한 두 양측 검정(기본 2-표본 t-검정과 Welch의 t-검정)의 모의 유의 수준 그래프.

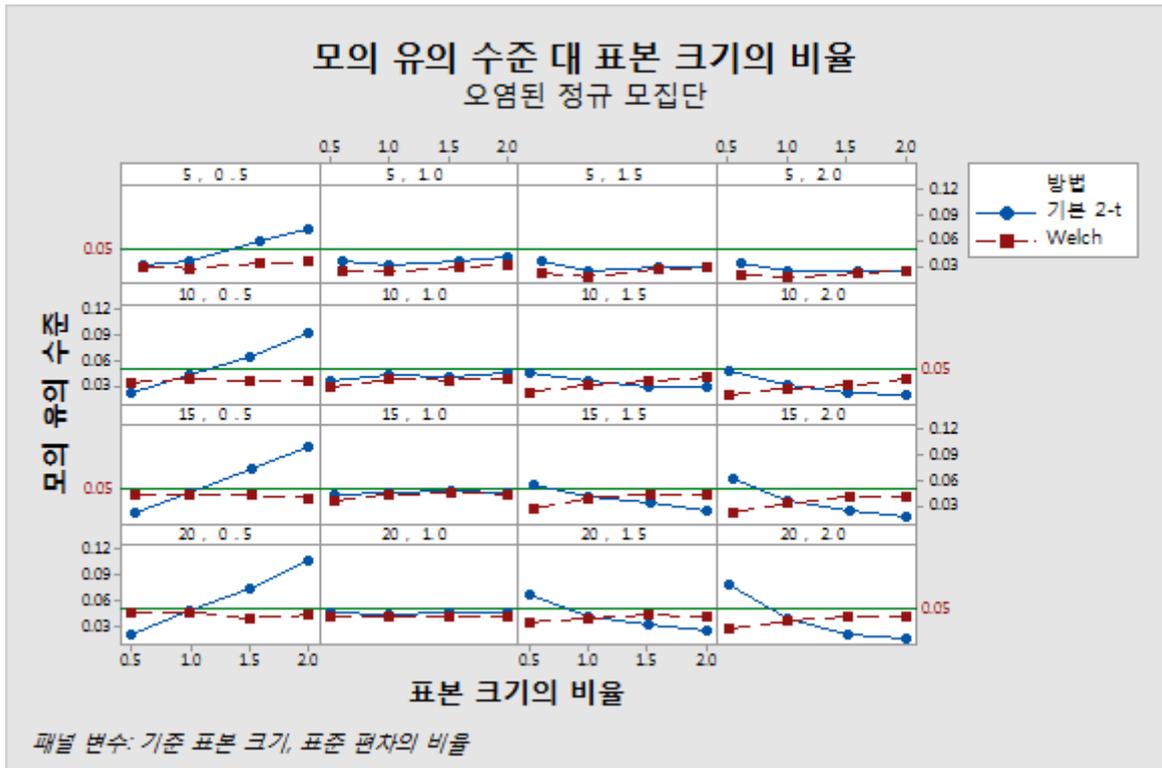


그림 3 표본 크기 비율과 분산이 같거나 다른 2개의 정규 분포로부터 생성된 표본 쌍에 기초한 두 양측 검정(기본 2-표본 t-검정과 Welch의 t-검정)의 모의 유의 수준 그래프.

부록 B: 두 검정의 검정력 함수 비교

당사는 Welch의 t-검정의 검정력 함수가 기존 2-표본 t-검정의 검정력 함수와 같거나 거의 같을 수 있는 조건을 확인하고자 했습니다.

일반적으로 t-검정(1-표본 또는 2-표본)의 검정력 함수는 잘 알려져 있고 여러 출판물에서 논의되었습니다(Pearson & Hartley, 1952; Neyman 외, 1935; Srivastava, 1958). 다음 항목에서는 2-표본 설계의 세 가지 대립 가설에 대한 검정력 함수에 대해 각각 서술합니다.

정리 B1

정규성 및 등분산 가정 하에서 명목 크기가 α 인 양측 2-표본 t-검정의 검정력 함수는 표본 크기의 함수로 표현할 수 있으며, 차이 $\delta = \mu_1 - \mu_2$ 는 다음 식으로 나타낼 수 있습니다.

$$\pi(n_1, n_2, \delta) = 1 - F_{d_C, \lambda}(t_{d_C}^{\alpha/2}) + F_{d_C, \lambda}(-t_{d_C}^{\alpha/2})$$

여기서 $F_{d_C, \lambda}(\cdot)$ 는 자유도가 $d_C = n_1 + n_2 - 2$ 고 비중심 모수가 다음과 같은 비중심 t 분포의 C.D.F입니다.

$$\lambda = \frac{\delta}{\sigma \sqrt{1/n_1 + 1/n_2}}$$

또한 대립 가설 $\mu_1 > \mu_2$ 와 연관된 검정력 함수는 다음 식으로 표현됩니다.

$$\pi(n_1, n_2, \delta) = 1 - F_{d_C, \lambda}(t_{d_C}^{\alpha})$$

반면에 $\mu_1 < \mu_2$ 대립 가설에 대해 검정할 때의 검정력은 다음 식으로 표현됩니다

$$\pi(n_1, n_2, \delta) = F_{d_C, \lambda}(-t_{d_C}^{\alpha})$$

위 정리의 결과는 잘 알려져 있지만, Welch의 수정된 t-검정에 기초한 검정의 검정력 함수는 문헌에서 구체적으로 논의된 적이 없습니다. 일원 분산 분석 모형에 대해 주어진 근사 검정력 함수로부터 근사를 도출할 수 있습니다(Kulinskaya 외, 2003 참조). 그러나 이 검정력 함수는 양측 대립 가설에만 적용할 수 있습니다. 그러나 2-표본 설계는 매우 특별하므로 Welch t-검정의 (정확한) 검정력 함수를 세 개의 대립 가설에 대해 각각 구할 수 있는 다른 방법을 적용할 수 있습니다. 이러한 함수는 다음 정리에 제시되어 있습니다.

정리 B2

모집단이 (분산이 같지 않을 수 있지만) 정규 분포를 따른다는 가정 하에 명목 크기가 α 인 양측 Welch t-검정의 검정력 함수는 표본 크기 및 차이 $\delta = \mu_1 - \mu_2$ 의 함수로 표현할 수 있습니다.

$$\pi_W(n_1, n_2, \delta) = 1 - G_{d_W, \lambda_W}(t_{d_W}^{\alpha/2}) + G_{d_W, \lambda_W}(-t_{d_W}^{\alpha/2})$$

여기서 $G_{d, \lambda}(\cdot)$ 는 다음 식으로 표현되는 자유도가 d_W 고

$$d_W = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1 - 1)} + \frac{\sigma_2^4}{n_2^2(n_2 - 1)}}$$

비중심 모수가 다음과 같은 비중심 t 분포의 C.D.F입니다.

$$\lambda_W = \frac{\delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

단측 대립 가설에 대한 검정력 함수의 식은 다음과 같습니다.

$$\pi_W(n_1, n_2, \delta) = 1 - G_{d_W, \lambda_W}(t_{d_W}^\alpha)$$

및

$$\pi_W(n_1, n_2, \delta) = G_{d_W, \lambda_W}(-t_{d_W}^\alpha)$$

위의 공식은 귀무 가설을 대립 가설 $\mu_1 > \mu_2$ 와 $\mu_1 < \mu_2$ 에 대해 검정하는 경우에 각각 해당됩니다.

결과 증명은 부록 D에 나와 있습니다.

이 검정력 함수 2개를 비교하기 전에 기존 2-표본 t-검정은 모집단의 분산이 같다는 추가적인 가정 하에 도출되기 때문에 두 검정의 이론적 검정력 함수는 두 번째 가정이 Welch의 t-검정에서 성립할 때 비교해야 한다는 점에 유의하십시오.

이론적으로, 정규성 및 등분산 가정 하에서

모든 n_1, n_2, δ 에 대해 $\pi(n_1, n_2, \delta) \geq \pi_W(n_1, n_2, \delta)$ 임이 알려져 있습니다.

다음 결과는 두 함수가 (대략적으로) 같은 조건을 서술합니다.

정리 B3

정규성 및 등분산 가정 하에서 다음이 성립됩니다.

1. $n_1 \sim n_2$ 이면 각 차이 δ 에 대해 $\pi(n_1, n_2, \delta) \sim \pi_W(n_1, n_2, \delta)$ 입니다. 특히 $n_1 = n_2$ 면 각 차이 δ 에 대해 $\pi(n_1, n_2, \delta) = \pi_W(n_1, n_2, \delta)$ 이므로 Welch t-검정은 기존 2-표본 t-검정만큼 강력합니다.
2. n_1 과 n_2 가 작고 $n_1 \neq n_2$ 면 Welch t-검정의 검정력은 기존 2-표본 t-검정보다 낮습니다. 그러나 n_1 과 n_2 가 크면 (표본 크기의 차이에 관계 없이) $\pi(n_1, n_2, \delta) \sim \pi_W(n_1, n_2, \delta)$ 입니다.

결과 증명은 부록 E에 제시되어 있습니다.

등분산 가정 하에서 두 검정의 검정력 함수와 연관된 비중심 모수는 동일합니다. 검정력 함수가 차이가 나는 유일한 이유는 자유도가 각각 차이가 나기 때문입니다. 이론적으로, 언급된 가정 하에서 기존 t-검정이 UMP(uniformly most powerful)이므로 자유도가 더 높음을 알 수 있습니다. 그러나 위 결과의 핵심은 설계가 균형이거나 균형에 가까울 경우 검정력 함수가 동일하거나 거의 동일하다는 것입니다. 기존 t-검정은 설계가 현저히 불균형하고 표본이 작을 때만 Welch t-검정보다 검정력이 훨씬 더 우수합니다. 부록 A에 나와 있는 것처럼 기존 2-표본 t-검정이 등분산 가정에 특별히 민감하게 반응하는 경우에도 마찬가지입니다. 따라서 실제로는 Welch t-검정의 검정력 함수가 더 신뢰할 수 있는 함수입니다.

여기서는 정리 B3의 결과를 정규 모집단 2개의 표준 편차가 똑같이 3인 다음 예를 사용하여 설명합니다. 정리 B1과 정리 B2의 (양측) 검정력 함수에 기초한 검정력 값은 다음 네 가지 시나리오에서 계산됩니다.

1. 두 표본이 모두 작지만 크기가 같은 경우($n_1 = n_2 = 10$).
2. 두 표본이 모두 작지만 한 표본이 다른 표본보다 2배 더 큰 경우 ($n_1 = 10, n_2 = 20$).

3. 한 표본은 작고 다른 표본은 중간 크기지만 중간 크기 표본이 작은 표본보다 4배 더 큰 경우($n_1 = 10, n_2 = 40$).
4. 한 표본이 중간 크기고 다른 표본은 크지만 더 큰 표본이 중간 크기 표본보다 4배 더 큰 경우($n_1 = 50, n_2 = 200$).

두 검정에 대해 모두 $\alpha = 0.05$ 라고 가정하여 각 시나리오에 대해 검정력 함수를 차이 $\delta = 0.0, 0.5, 1.0, 1.5, 2.0, \dots, 5.0$ 에서 평가합니다. 결과는 표 5에 제시되어 있으며, 함수는 그림 4에 표시되어 있습니다.

표 5 양측 기준 2-표본 t-검정과 양측 Welch t-검정의 이론적 검정력 함수 비교, $\alpha = 0.05$. 표본 크기 n_1 과 n_2 는 고정이며, 검정력 함수를 차이 δ (범위: 0.0 - 5.0)에서 평가했습니다.

δ	0.0	0.5	1.0	1.5	2.0	2.5	3	3.5	4	4.5	5.0
$n_1 = n_2 = 10$											
$\pi(n_1, n_2, \delta)$.05	.064	.109	.185	.292	.422	.562	.694	.805	.887	.941
$\pi_W(n_1, n_2, \delta)$.05	.064	.109	.185	.292	.422	.562	.694	.805	.887	.941
$n_1 = 10, n_2 = 20$											
$\pi(n_1, n_2, \delta)$.05	.070	.132	.239	.383	.547	.703	.828	.913	.962	.986
$\pi_W(n_1, n_2, \delta)$.05	.070	.129	.231	.371	.531	.686	.813	.902	.955	.982
$n_1 = 10, n_2 = 40$											
$\pi(n_1, n_2, \delta)$.05	.075	.152	.283	.455	.637	.791	.899	.959	.986	.996
$\pi_W(n_1, n_2, \delta)$.05	.072	.142	.261	.419	.592	.748	.865	.938	.976	.992
$n_1 = 50, n_2 = 200$											
$\pi(n_1, n_2, \delta)$.05	.182	.556	.883	.987	.999	1.	1.	1.	1.	1.
$\pi_W(n_1, n_2, \delta)$.05	.180	.548	.877	.986	.999	1.	1.	1.	1.	1.

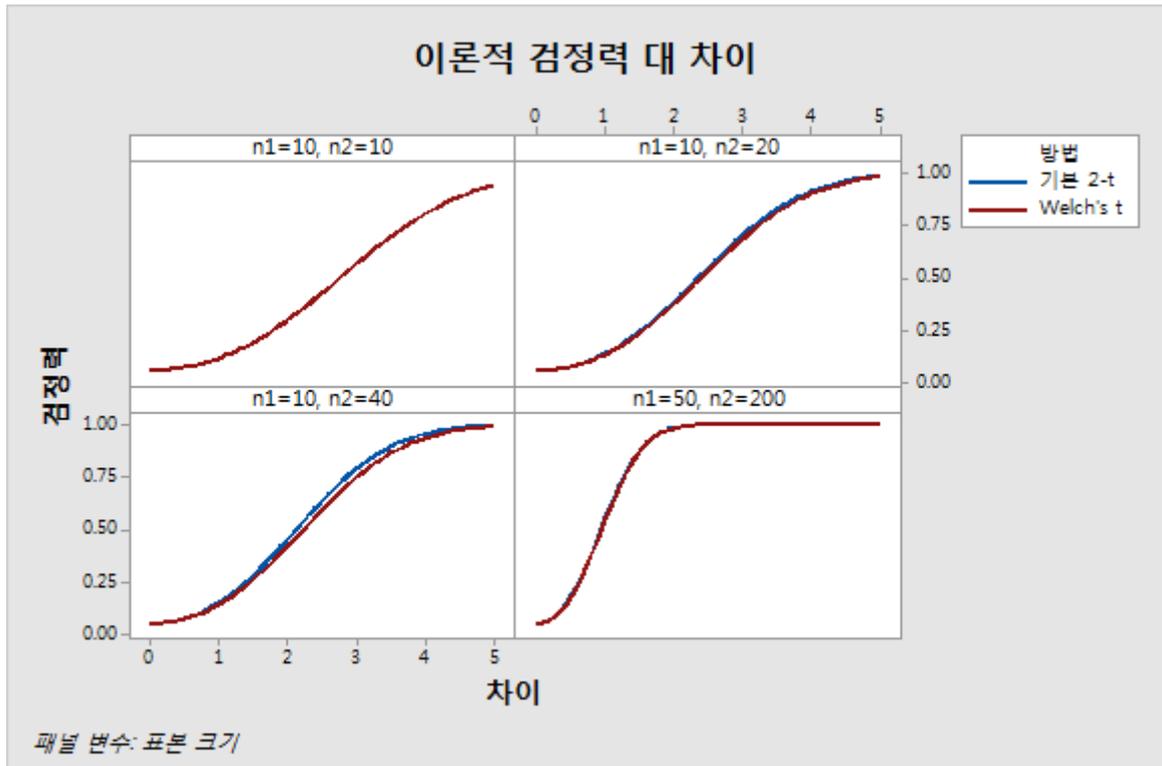


그림 4 양측 기준 2-표본 t-검정과 양측 Welch t-검정의 이론적 검정력 함수를 탐지할 평균의 차이 δ 와 비교하여 표시한 그래프. 두 검정에서 모두 $\alpha = 0.05$ 를 사용합니다. 가정된 모집단은 표준 편차가 똑같이 3인 정규 분포를 따릅니다.

시뮬레이션 연구 B

이 시뮬레이션 연구의 목적은 분산이 같지 않다고 가정되는 균형 설계에서 기준 2-표본 t-검정과 연관된 검정력 수준을 Welch의 2-표본 t-검정과 연관된 검정력 수준과 비교하는 것입니다. 해당 연구의 실험은 부록 A에 설명되어 있는 것과 유사합니다.

첫 번째 실험 그룹에서는 이분산 정규 모집단으로부터 크기가 같은 표본 쌍을 생성했습니다. 기준 모집단은 $N(0,2)$ 가 되도록 고정되었고 두 번째 정규 모집단은 표준 편차 비율 $\rho = \sigma_2/\sigma_1$ 이 0.5, 1.5 및 2가 되도록 선택되었습니다. 마찬가지로, 두 번째 집단에서 두 표본은 이분산 카이-제곱 분포에서 추출되었습니다(기준 모집단은 Chi(2)임). 마지막 실험 집합에서는 앞의 부록 A에서 정의한 것과 같은 오염된 정규 분포(기준 모집단 $CN(8,4)$)로부터 표본 쌍을 생성했습니다.

각 실험 집합에 대해 표본 크기가 $n = n_1 = n_2 = 5, 10, 15, 20, 25, 30$ 인 각 검정과 연관된 모의 검정력 수준을 (주어진 탐지 가능한 차이 δ 에서) 계산했습니다. 각 실험에서는 모의 검정력 수준을 귀무 가설이 거짓이었을 때 기각된 횟수의 비율로 계산했습니다. 모든 실험에 대해 평균의 차이를 기준 모집단(두 표본의 첫 번째)에 표준 단위로 지정했습니다. 더 구체적으로 말하면 본 연구의 세 가지 분포 유형에 대해 모두 상대적으로 작았기 때문에 $\delta = 1.0 \times \sigma_1 = 2.0$ 을 고정했습니다. 시뮬레이션 결과는 표 2.2에 보고되고 그림 2.2a, 그림 2.2b 및 그림 2.2c에 표시되었습니다.

결과 및 요약

표 6과 그림 4의 결과는 등분산 가정 하에서 이론적 검정력 함수가 정리 2.3에 나와 있는 것처럼 균형 설계에서 동일함을 보여줍니다. 또한 표본 크기가 비교적 작지만 거의 같으면 두 함수는 대체로 같은 검정력 값을 나타냅니다. 표본이 비교적 작고 표본 하나가 다른 표본보다 약 4배 정도 클 때(예를 들어 $n_1 = 10, n_2 = 40$ 일 때)만 검정력 함수 간에 어느 정도 눈에 띄는 차이가 나타나기 시작합니다. 이 경우에도 기존 2-표본 t-검정에 기초한 이론적 검정력 값은 Welch의 t-검정에 기초한 검정력 값보다 약간 더 높을 뿐입니다. 마지막으로, 설계가 현저히 불균형이지만 표본이 (비교적) 크면 두 검정력 함수는 정리 B3에 서술된 것처럼 사실상 동일합니다.

또한 이분산 균형 설계에서 두 검정은 사실상 동일한 검정력 값을 나타냅니다. 그러나 매우 작은 표본($n < 10$)에서는 기존 2-표본 t-검정이 약간 더 효과적입니다.

표 6 이분산 균형 설계에서 기존 2-표본 t-검정과 Welch의 t-검정의 모의 검정력 수준 비교

n	$\frac{\sigma_2}{\sigma_1}$	기준 모집단: N(0,2)			기준 모집단: Chi(2)			기준 모집단: CN(.8,4)		
		.5	1.5	2.0	.5	1.5	2.0	.5	1.5	2.0
5	2T	0.431	0.196	0.152	0.555	0.281	0.215	0.579	0.373	0.335
	Welch	0.366	0.166	0.119	0.424	0.25	0.184	0.521	0.32	0.283
10	2T	0.77	0.385	0.27	0.846	0.438	0.324	0.79	0.51	0.435
	Welch	0.747	0.372	0.253	0.832	0.427	0.308	0.776	0.493	0.417
15	2T	0.916	0.539	0.387	0.948	0.565	0.424	0.898	0.615	0.508
	Welch	0.908	0.532	0.375	0.945	0.557	0.413	0.891	0.605	0.497
20	2T	0.971	0.682	0.497	0.982	0.68	0.521	0.952	0.702	0.573
	Welch	0.969	0.677	0.487	0.981	0.676	0.511	0.947	0.697	0.563
25	2T	0.99	0.779	0.591	0.994	0.765	0.605	0.98	0.783	0.641
	Welch	0.99	0.777	0.582	0.994	0.762	0.597	0.979	0.778	0.636
30	2T	0.998	0.851	0.675	0.998	0.826	0.676	0.994	0.839	0.699
	Welch	0.998	0.849	0.67	0.998	0.824	0.668	0.994	0.836	0.694

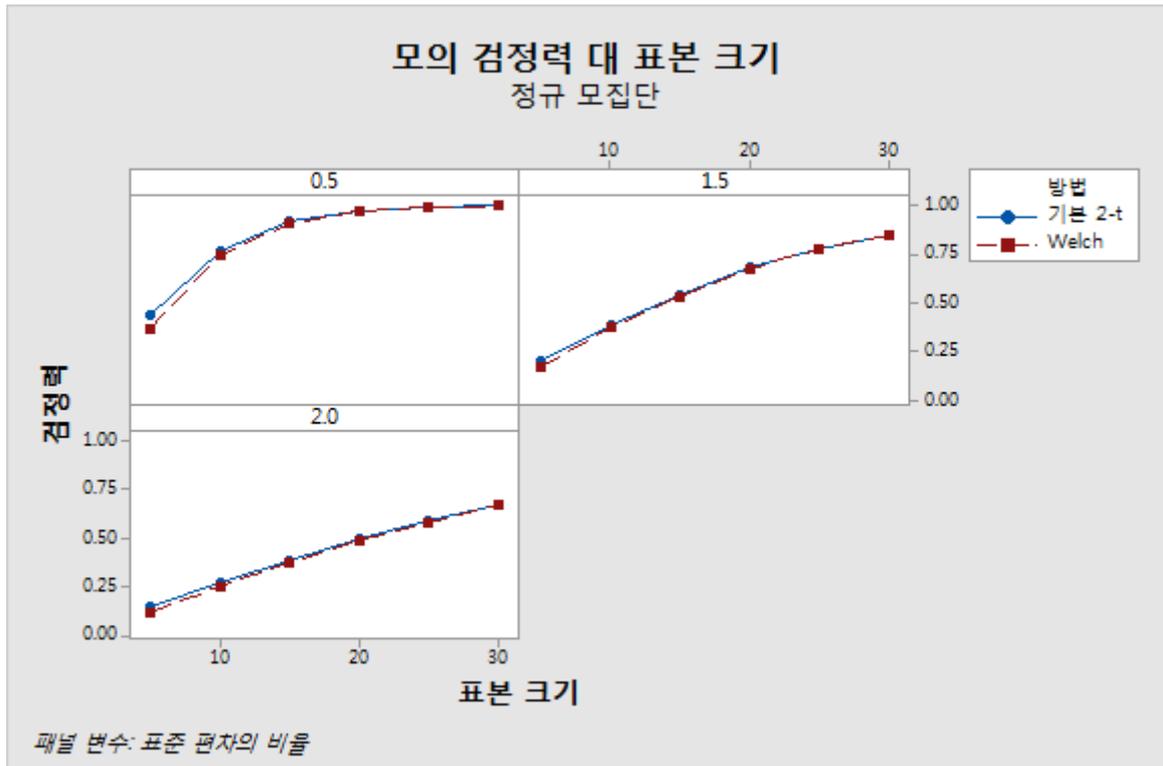


그림 5 이분산 균형 설계에서 기본 2-표본 t-검정과 Welch의 2-표본 t-검정의 모의 검정력 수준 비교 표본은 표준 편차 비율이 0.5, 1.5 및 2.0이 되도록 이분산 정규 분포를 따르는 모집단에서 추출되었습니다.

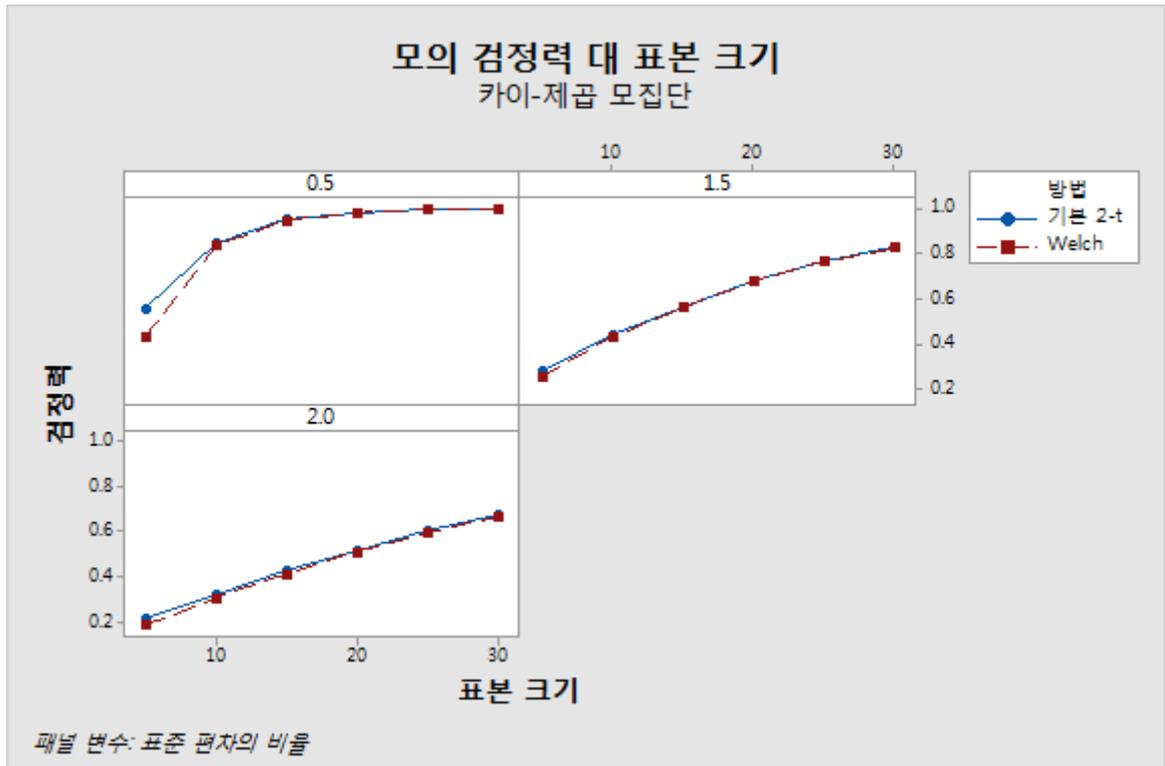


그림 6 이분산 균형 설계에서 기본 2-표본 t-검정과 Welch의 2-표본 t-검정의 모의 검정력 수준 비교 표본은 표준 편차 비율이 0.5, 1.5 및 2.0이 되도록 이분산 카이-제곱 분포를 따르는 모집단에서 추출되었습니다.

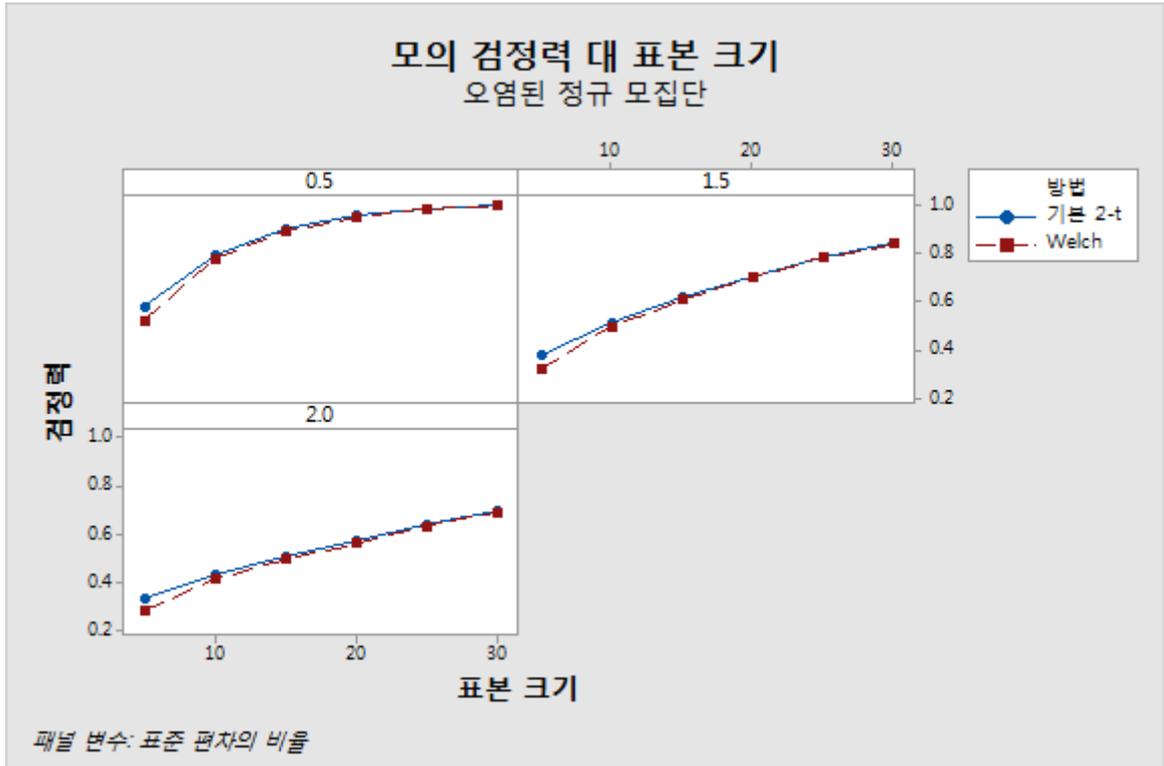


그림 7 이분산 균형 설계에서 기본 2-표본 t-검정과 Welch의 2-표본 t-검정의 모의 검정력 수준 비교 표본은 표준 편차 비율이 0.5, 1.5 및 2.0이 되도록 오염된 이분산 정규 분포를 따르는 모집단에서 추출되었습니다.

부록 C: 검정력 및 표본 크기와 정규성에 대한 민감도

보조 도구에서 두 모집단의 평균을 비교하기 위한 검정력 분석은 Welch t-검정의 검정력 함수를 토대로 수행됩니다. 이 함수가 함수 도출에 적용된 정규 가정에 민감하게 반응할 경우, 검정력 분석에서 잘못된 결론이 나올 수 있습니다. 이런 이유로 이 함수의 정규 가정에 대한 민감도를 조사하기 위한 시뮬레이션 연구를 실시했습니다. 민감도는 표본을 비정규 분포로부터 생성할 때 모의 검정력 수준과 이론적 검정력 함수로부터 계산된 검정력 수준이 일치하는 정도를 기준으로 평가되었습니다. 정리 B2에 따르면 표본이 정규 분포로부터 생성될 때 모의 검정력 수준과 이론적 검정력 수준이 가장 가깝기 때문에 정규 분포는 대조 모집단의 역할을 합니다.

시뮬레이션 연구 C

연구는 정규, 카이-제곱 및 오염된 정규 분포의 세 가지 분포를 사용하여 세 부분으로 나누어 실시했습니다. 자세한 내용은 부록 A를 참조하십시오. 이 연구의 각 부분에서는 모의 검정력을 (주어진 표본 크기 n_1 과 n_2 에 대해 주어진 탐지 가능한 차이 δ 에서) 귀무 가설이 거짓일 때 기각된 횟수의 비율로 계산했습니다. 모든 경우 탐지할 차이는 기준 모집단의 표준 단위로 지정되었습니다. 즉, 본 연구에 사용되는 세 가지 분포 유형에 대해 모두 $\delta = 1.0 \times \sigma_1 = 2.0$ 입니다. Welch의 t-검정에 기초한 이론적 검정력 값도 비교를 위해 계산했습니다.

시뮬레이션 결과 및 요약

시뮬레이션에서는 표본 크기가 비교적 작을 경우 Welch t-검정의 검정력 함수가 정규성 가정에 대해 강하다는 결과가 나왔습니다. 일반적으로 두 표본의 크기가 최소 15개에 불과하면 모의 검정력 값이 그에 해당되는 목표 이론적 검정력 수준과 가깝습니다(표 7-10과 그림 8-10 참조).

표 7-10에는 정규 모집단, 치우친 모집단(카이-제곱) 및 오염된 정규 모집단으로부터 생성된 표본의 쌍에 기초한 $\alpha = 0.05$ 인 양측 Welch t-검정의 모의 검정력 수준이 나와 있습니다. 표본의 쌍은 같은 분포 유형에서 추출되었지만, 모집단의 분산은 같지 않을 수 있습니다. 비교를 위해 이론적 검정력 값을 계산했습니다.

표 7 n = 5일 때 $\alpha = 0.05$ 인 양측 Welch t-검정의 모의 검정력 수준

			기준 모집단: N(0,2)				기준 모집단: Chi(2)				기준 모집단: CN(.8,4)			
			.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5	2.0
n_2	$\frac{n_2}{n_1}$	$\frac{\sigma_2}{\sigma_1}$	$n_1 = 5$				$n_1 = 5$				$n_1 = 5$			
3	.6	관측치	.288	.158	.113	.091	.432	.305	.211	.149	.361	.257	.234	.220
		목표값	.353	.192	.116	.092	.353	.192	.116	.092	.353	.192	.116	.092

		기준 모집단: N(0,2)				기준 모집단: Chi(2)				기준 모집단: CN(.8,4)				
		$\frac{\sigma_2}{\sigma_1}$.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5	2.0
n_2	$\frac{n_2}{n_1}$		$n_1 = 5$				$n_1 = 5$				$n_1 = 5$			
5	1.0	관측치	.370	.252	.169	.121	.427	.334	.248	.189	.522	.380	.319	.284
		목표값	.389	.286	.190	.137	.389	.286	.190	.137	.389	.286	.190	.137
8	1.6	관측치	.387	.326	.242	.179	.427	.364	.286	.225	.573	.453	.374	.319
		목표값	.400	.345	.260	.193	.400	.345	.260	.193	.400	.345	.260	.193
10	2.0	관측치	.390	.351	.272	.208	.421	.373	.296	.235	.590	.483	.394	.336
		목표값	.402	.364	.291	.223	.402	.364	.291	.223	.402	.364	.291	.223

표 8 $n = 10$ 일 때 $\alpha = 0.05$ 인 양측 Welch t-검정의 모의 검정력 수준

		기준 모집단: N(0,2)				기준 모집단: Chi(2)				기준 모집단: CN(.8,4)				
		$\frac{\sigma_2}{\sigma_1}$.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5	2.0
n_2	$\frac{n_2}{n_1}$		$n_1 = 10$				$n_1 = 10$				$n_1 = 10$			
5	.5	관측치	.651	.346	.197	.131	.768	.493	.320	.221	.689	.484	.404	.358
		목표값	.666	.364	.206	.139	.666	.364	.206	.139	.666	.364	.206	.139
10	1.0	관측치	.742	.556	.369	.254	.831	.612	.430	.308	.776	.619	.496	.419
		목표값	.745	.562	.337	.259	.745	.562	.337	.259	.745	.562	.337	.259
15	1.5	관측치	.765	.641	.483	.358	.865	.679	.511	.377	.792	.679	.547	.456
		목표값	.767	.643	.483	.352	.767	.643	.483	.352	.767	.643	.483	.352
20	2	관측치	.774	.683	.549	.417	.898	.737	.565	.448	.797	.716	.596	.490
		목표값	.777	.686	.551	.422	.777	.686	.551	.422	.777	.686	.551	.422

표 9 n = 15일 때 $\alpha = 0.05$ 인 양측 Welch t-검정의 모의 검정력 수준

			기준 모집단: N(0,2)				기준 모집단: Chi(2)				기준 모집단: CN(.8,4)			
			$\frac{\sigma_2}{\sigma_1}$.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5
n_2	$\frac{n_2}{n_1}$		$n_1 = 15$				$n_1 = 15$				$n_1 = 15$			
8	.53	관측치	.857	.569	.342	.229	.871	.651	.421	.293	.853	.632	.505	.428
		목표값	.861	.568	.338	.221	.861	.568	.338	.221	.861	.568	.338	.221
15	1.0	관측치	.906	.745	.535	.368	.942	.763	.563	.415	.891	.760	.611	.500
		목표값	.910	.753	.541	.379	.910	.753	.541	.379	.910	.753	.541	.379
23	1.53	관측치	.928	.831	.667	.502	.975	.858	.676	.517	.898	.825	.698	.572
		목표값	.925	.830	.670	.509	.925	.830	.670	.509	.925	.830	.670	.509
30	2.0	관측치	.933	.861	.737	.589	.984	.903	.750	.598	.902	.847	.742	.619
		목표값	.931	.863	.736	.589	.931	.863	.736	.589	.931	.863	.736	.589

표 10 n = 20일 때 $\alpha = 0.05$ 인 양측 Welch t-검정의 모의 검정력 수준

			기준 모집단: N(0,2)				기준 모집단: Chi(2)				기준 모집단: CN(.8,4)			
			$\frac{\sigma_2}{\sigma_1}$.5	1.0	1.5	2.0	.5	1.0	1.5	2.0	.5	1.0	1.5
n_2	$\frac{n_2}{n_1}$		$n_1 = 20$				$n_1 = 20$				$n_1 = 20$			
10	.5	관측치	.938	.687	.426	.275	.920	.698	.486	.333	.923	.716	.568	.476
		목표값	.941	.686	.424	.277	.941	.686	.424	.277	.941	.686	.424	.277
20	1.0	관측치	.971	.866	.672	.485	.981	.858	.670	.506	.952	.856	.696	.567
		목표값	.971	.869	.673	.489	.971	.869	.673	.489	.971	.869	.673	.489
30	1.5	관측치	.977	.923	.791	.629	.995	.932	.785	.631	.960	.908	.798	.662
		목표값	.978	.922	.791	.628	.978	.922	.791	.628	.978	.922	.791	.628
40	2.0	관측치	.983	.950	.858	.724	.998	.966	.864	.726	.958	.929	.845	.725
		목표값	.981	.945	.854	.719	.981	.945	.854	.719	.981	.945	.854	.719

두 표본이 정규 모집단으로부터 생성되었을 때 모의 검정력 값은 표본이 매우 작을 경우에도 이론적 검정력 값과 일치합니다. 그림 7에 나와 있는 것처럼 이론 및 모의 검정력 곡선은 사실상 구분되지 않습니다. 이런 결과는 정리 B2와 일치합니다.

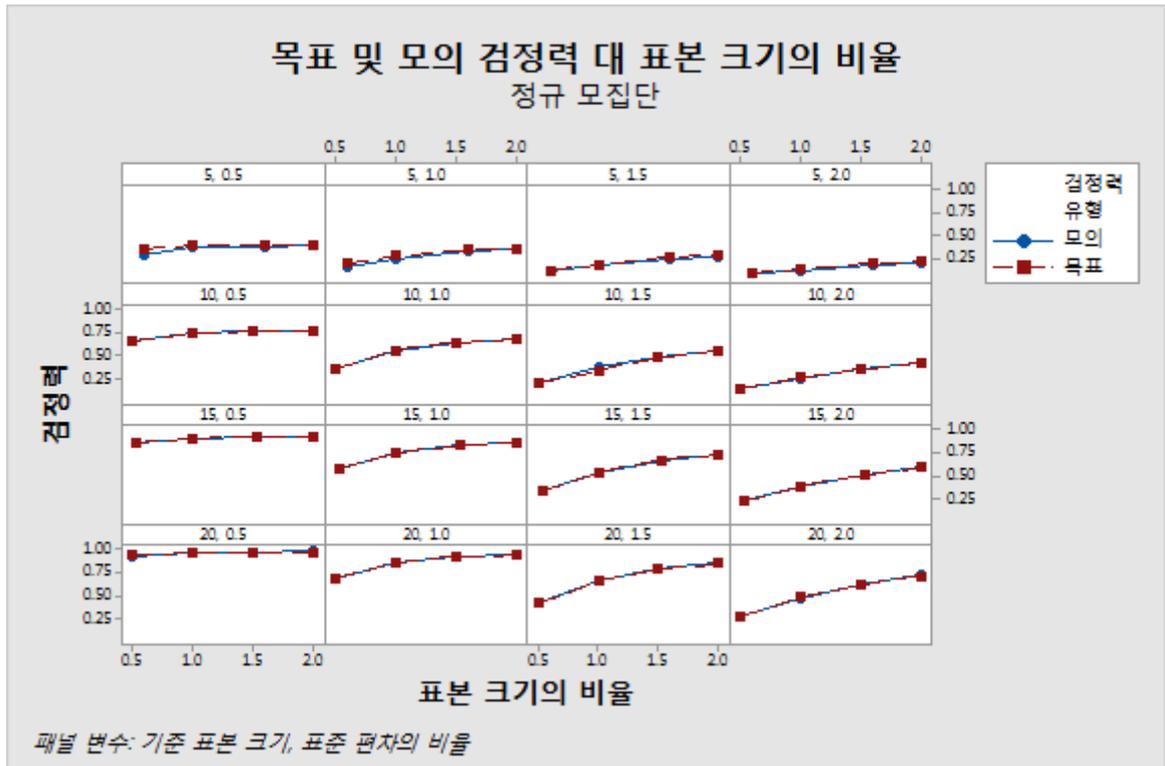


그림 8 표본 크기 비율과 2개의 등분산 또는 이분산 정규 모집단으로부터 생성된 표본 쌍에 기초한 $\alpha = 0.05$ 인 양측 Welch t-검정의 모의 및 목표 이론적 검정력 수준을 나타낸 그래프

표본이 치우친 카이-제곱 분포에서 생성되었을 때 표본이 매우 작을 경우 모의 검정력 값은 이론적 검정력 값보다 높습니다. 그러나 표본 크기가 커질수록 검정력 값은 점점 유사해집니다. 그림 9는 두 표본의 최소 크기가 10개 이상일 때 목표 이론적 검정력 곡선과 모의 검정력 곡선이 계속 가깝게 유지됨을 보여줍니다. 이는 표본이 비교적 작을 때도 치우친 데이터가 Welch t-검정의 검정력 함수에 큰 영향을 미치지 않음을 보여줍니다.

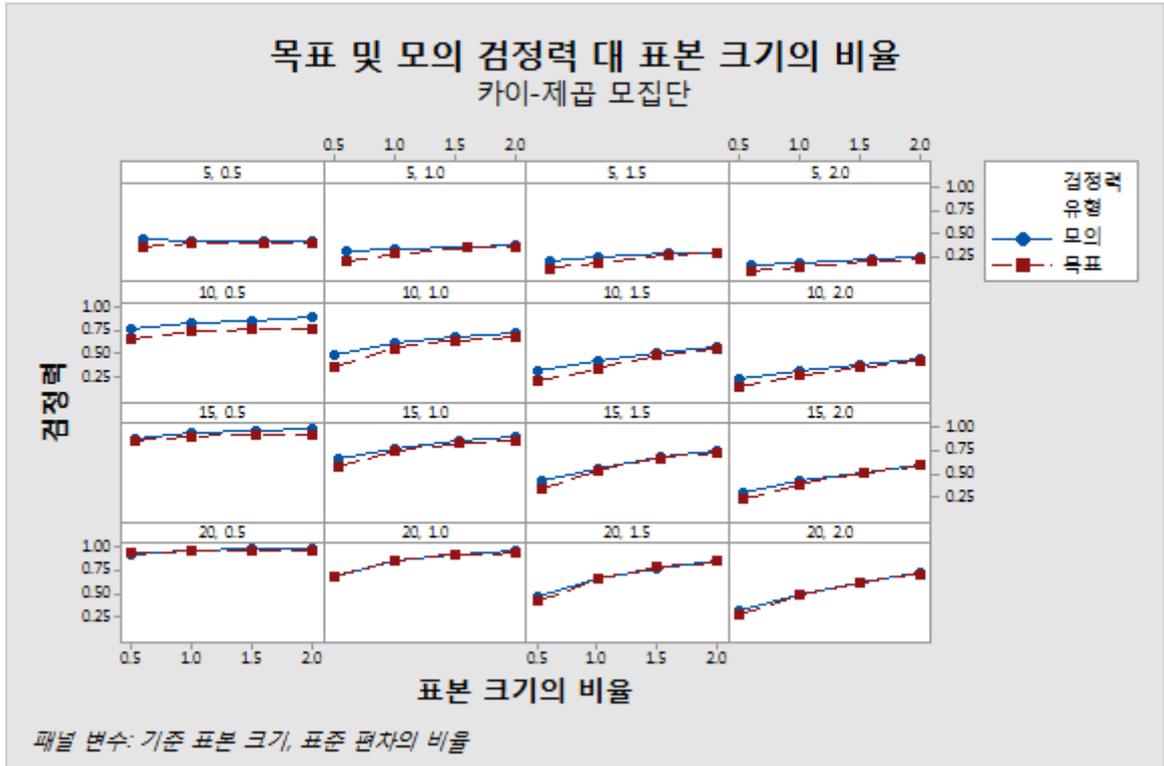


그림 9 표본 크기 비율과 2개의 등분산 또는 이분산 정규 모집단으로부터 생성된 표본 쌍에 기초한 $\alpha = 0.05$ 인 양측 Welch t-검정의 모의 및 목표 이론적 검정력 수준을 나타낸 그래프

또한 특이치는 표본 크기가 매우 작을 때만 검정력 함수에 영향을 미치는 경향이 있습니다. 일반적으로, 특이치가 있을 때 모의 검정력 값은 목표 이론적 검정력 값보다 약간 높은 경향이 있습니다. 이 내용은 모의 및 이론적 검정력 곡선이 최소 표본 크기가 15에 도달할 때까지 합당한 수준으로 가까워지지 않는 것으로 그림 10에 나와 있습니다.

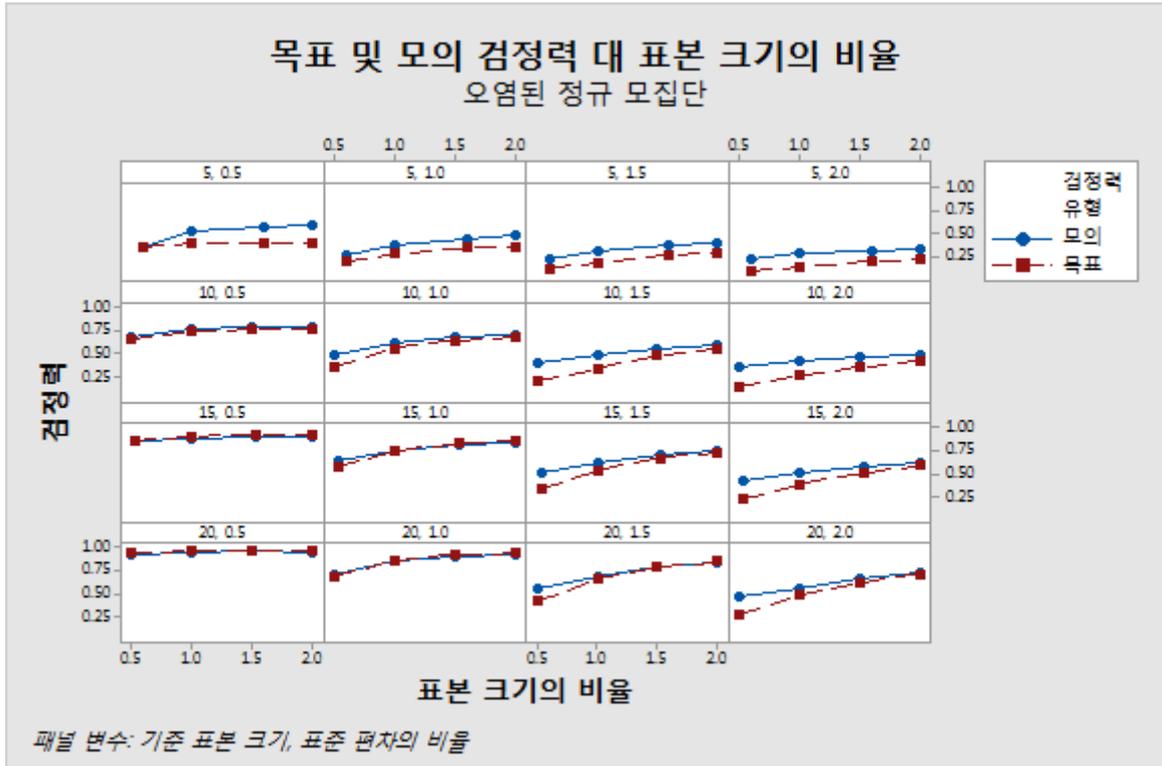


그림 10 표본 크기 비율과 2개의 등분산 또는 이분산 정규 모집단으로부터 생성된 표본 쌍에 기초한 $\alpha = 0.05$ 인 양측 Welch t-검정의 모의 및 목표 이론적 검정력 수준을 나타낸 그래프

부록 D: 정리 B2 증명

2-표본 모형의 경우 귀무 가설 하에서 다음 검정 통계량의 분포를 도출하는 Welch의 방식은

$$t_w(x, y) = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

다음 분포

$$V = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

를 카이-제곱 분포에 대한 비례로 근사하는 방법에 기초하고 있습니다. 더 구체적으로 말하면,

$$\frac{d_w V}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

는 자유도가 d_w 인 카이-제곱 분포로 근사 분포됩니다. 여기서:

$$d_w = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1-1)} + \frac{\sigma_2^4}{n_2^2(n_2-1)}}$$

(참고: 1-표본 설정에서 이는 $(n-1)s^2/\sigma^2 \sim \chi_{n-1}^2$ 이라는 잘 알려진 기존 결과로 감소합니다.)

대립 가설 $H_A: \mu_1 \neq \mu_2$ (또는 같은 말로 $\delta \neq 0$)에 대한 귀무 가설 $H_0: \mu_1 = \mu_2$ (또는 같은 말로 $\delta = 0$)의 검정을 고려하십시오.

귀무 가설 하에서 검정력 함수에 대해 다음 공식이 성립합니다.

$$\pi(n_1, n_2, \delta) = \pi(n_1, n_2, 0) = 1 - \Pr\left(-t_{d_w}^{\alpha/2} \leq \frac{\bar{x} - \bar{y}}{\sqrt{V}} \leq t_{d_w}^{\alpha/2}\right) \approx \alpha$$

여기서 t_d^α 는 자유도가 d 인 t-분포의 100α 상위 백분위 점을 나타냅니다.

대립 가설 하에서

$$\frac{\bar{x} - \bar{y}}{\sqrt{V}} = \frac{\frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} + \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\sqrt{\frac{d_w V}{d_w \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}$$

는 자유도가 d_w 고 비중심 모수가 다음과 같은 근사 비중심 t 분포를 따릅니다.

$$\lambda_w = \frac{\delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

그 이유는 앞에서 설명한 대로

$$\frac{d_W V}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

는 자유도가 d_W 인 카이-제곱 분포로 근사 분포되고

$$\frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

는 표준 정규 분포에 따라 분포되기 때문입니다.

따라서 대립 가설 하에서 다음이 성립됩니다.

$$\pi(n_1, n_2, \delta) = 1 - \Pr\left(-t_{d_W}^{\alpha/2} \leq \frac{\bar{x} - \bar{y}}{\sqrt{V}} \leq t_{d_W}^{\alpha/2}\right) \approx 1 - G_{d_W, \lambda_W}\left(t_{d_W}^{\alpha/2}\right) + G_{d_W, \lambda_W}\left(-t_{d_W}^{\alpha/2}\right)$$

여기서 $G_{d_W, \lambda}(\cdot)$ 는 자유도가 d_W 고 비중심 모수가 위와 같은 λ 인 비중심 t 분포의 C.D.F입니다.

부록 E: 정리 B3 증명

첫째, d_W 는 다음과 같이 다시 쓸 수 있습니다.

$$d_W = \frac{\left(\frac{1}{n_1} + \frac{\rho^2}{n_2}\right)^2}{\frac{1}{n_1^2(n_1-1)} + \frac{\rho^4}{n_2^2(n_2-1)}}$$

여기서 $\rho = \sigma_1/\sigma_2$ 입니다.

마찬가지로, Welch t-검정의 검정력 함수와 연관된 비중심 모수도 다음과 같이 다시 쓸 수 있습니다.

$$\lambda_W = \frac{\delta/\sigma_1}{\sqrt{1/n_1 + \rho^2/n_2}}$$

등분산 가정 하에서 기존 2-표본 t-검정과 Welch 검정의 검정력 함수와 연관된 비중심 모수는 동일합니다. 즉, 다음이 성립합니다.

$$\lambda = \lambda_W = \frac{\delta}{\sigma\sqrt{1/n_1 + 1/n_2}}$$

여기서 σ 는 두 모집단의 공분산입니다. 따라서 두 검정의 검정력 함수 간에 존재하는 유일한 차이점은 각각의 자유도뿐입니다. 그러나 등분산 가정 하에서 Welch t-검정의 검정력 함수와 연관된 자유도는 다음과 같습니다.

$$d_W = \frac{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^2}{\frac{1}{n_1^2(n_1-1)} + \frac{1}{n_2^2(n_2-1)}} = \frac{(n_1 + n_2)^2(n_1-1)(n_2-1)}{n_1^2(n_1-1) + n_2^2(n_2-1)}$$

정리 1에 따라 기존 2-표본 t-검정의 검정력 함수와 관련된 자유도는 $d_C = n_1 + n_2 - 2$ 입니다. 몇 가지 대수 조작을 거치면 다음 식이 얻어집니다.

$$d_C - d_W = \frac{(n_1 - n_2)^2(n_1 + n_2 - 1)^2}{n_1^2(n_1 - 1) + n_2^2(n_2 - 1)} \geq 0$$

$d - d_W \geq 0$ 이라는 사실은 놀랍지 않습니다. 그 이유는 등분산 가정 하에서 기존 2-표본 t-검정이 UMP(uniformly most powerful)이므로 해당 검정의 검정력 함수와 연관된 자유도가 더 높을 것이라고 기대할 수 있기 때문입니다.

이제 $n_1 \sim n_2$ 면 $d \sim d_W$ 이고, 그 결과 검정력 함수는 거의 같습니다. 특히 $n_1 = n_2$ 면 두 검정의 검정력 함수는 동일합니다. 이로써 정리 2.3의 첫 부분이 증명됩니다.

$n_1 \neq n_2$ 면 $d_C - d_W > 0$ 이므로 Welch t-검정의 검정력은 기존 2-표본 t-검정보다 낮습니다.

또한 표본이 크면, 즉 $n_1 \rightarrow \infty$ 고 $n_2 \rightarrow \infty$ 면 $d_C \rightarrow \infty$ 고 $d_W \rightarrow \infty$ 이므로 두 검정과 연관된 검정 통계의 점근 분포는 표준 정규 분포입니다. 따라서 두 검정은 점근적으로 동일하며 같은 점근 검정력 함수를 생성합니다.

© 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. All other marks referenced remain the property of their respective owners.