

A minimum combination t -test method for testing differences in population means based on a group of samples of size one

Miyoung Heo^a · Changwon Lim^{a,1}

^aDepartment of Applied statistics, Chung-Ang University

(Received February 24, 2017; Revised February 26, 2017; Accepted February 26, 2017)

Abstract

It is often possible to test for differences in population means when two or more samples are extracted from each N population. However, it is not possible to test for the mean difference if one sample is extracted from each population since a sample mean does not exist. But, by dividing a group of samples extracted one by one into two groups and generating a sample mean, we can identify a heterogeneity that may exist within the group by comparing the differences of the groups' mean. Therefore, we propose a minimum combination t -test method that can test the mean difference by the number of combinations that can be divided into two groups. In this paper, we proposed a method to test differences between means to check heterogeneity in a group of extracted samples. We verified the performance of the method by simulation study and obtained the results through real data analysis.

Keywords: heterogeneity, minimum combination t -test method, comparison of means, hypothesis testing

1. 서론

1908년에 William Sealy Gosset이 Student (1908)라는 필명으로 t -통계량을 개발하여 발표한 이래로 t -검정은 과학의 거의 모든 분야에서 한 개 또는 두 개의 모집단에서 모평균을 비교하는 데 쓰여왔다. 모집단이 하나일 때에는 모평균이 어떤 주어진 특정한 값과 같은지를 검정하기 위해, 그리고 모집단이 두 개일 때에는 각 모집단의 모평균이 같은지를 검정하기 위해 쓰인다. 그러나 모집단의 개수가 3개 이상인 경우에는 일반적으로 t -검정은 사용할 수 없고 분산분석을 통하여 모평균을 비교한다. 분산분석은 1918년에 Ronald A. Fisher에 의해 처음으로 제안되었고 (Fisher, 1918), 역시 그 이후로 과학의 많은 분야에서 사용되어 왔다.

모집단의 평균을 비교하기 위한 t -검정과 분산분석 모두 중요한 전제조건이 하나 만족되어야 하는데 그것은 각 모집단에서 추출한 표본의 크기가 2 이상이어야 한다는 것이다. 추출된 표본으로부터 모평균과 모분산을 추정하고 그 추정값을 사용하여 가설검정을 위한 검정통계량을 계산하는데, 이를 위해서는 표본의 크기가 최소한 2 이상이 되어야 하는 것이 필수적이다. 일반적으로는 모집단의 분포가 정규분포라

This research was supported by the Chung-Ang University Research Scholarship Grants in 2015.

¹Corresponding author: Department of Applied Statistics, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: clim@cau.ac.kr

는 가정을 만족하기 위해서 표본의 크기가 10 이상은 되어야 한다고 알려져 있다.

일반적으로 N 개의 모집단으로부터 추출된 크기가 2 이상인 표본에 대하여 모평균이 동일하지 검정할 때에는 일원배치 분산분석을 실시하게 된다. 즉, $H_0: \mu_1 = \dots = \mu_N$ 의 가설을 검정하게 되고 N 개의 처리마다 반복의 수도 달라질 수 있는데, 만약 반복의 수가 1로 N 개의 모집단으로부터 표본이 1개씩 추출되었다면 모평균에 대하여 $H_0: \mu_1 = \dots = \mu_N$ 의 가설은 일원배치 분산분석을 통하여 검정하는 것이 불가능하다. 평균이 동일한지에 대한 검정은 각 표본으로부터 표본평균인 \bar{X} 를 계산함으로써 그것으로 모평균을 추정하고 가설검정에 사용하게 되는데, 한 개의 표본에 대해서는 표본평균을 계산할 수 없고 따라서 가설검정이 불가능하게 된다.

본 논문에서 제안하는 새로운 가설검정 방법은 폐암 전이성 종양에 대한 연구에서 측정된 자료로부터 잠재적으로 존재할 수 있는 종양의 이질성(tumor heterogeneity)을 검증하기 위해서 개발되었다. 종양의 이질성은 한 환자가 체내에 퍼져있는 종양들이 다양한 요인으로 인해 성질이 다른 종양세포가 형성되는 것으로 이질성을 파악하는 것은 중요한 문제이다 (Marusyk와 Polyak, 2010). 하지만 만약 몇 개의 종양만을 이용하여 이를 확인한다면 전체적인 이질성을 나타내는 데에 제한적일 수 있기 때문에 우리는 한 환자가 가진 모든 종양을 이용하여 이질성이 존재하는지 확인하기 위한 방법을 제안하였다. 종양의 이질성은 종양의 크기를 나타내는 여러 가지 변수들, 예를 들면 종양의 최대 지름 또는 종양의 부피를 가지고 판단할 수 있다. 한 환자에게 나타나 있는 종양들이 서로 이질적이라면 종양의 크기가 변하는 정도가 서로 다를 것이라고 가정할 수 있다. 한 환자가 체내에 N 개의 종양을 가지고 있다고 할 때, 그 환자의 i 번째 병변에 나타나 있는 종양의 크기를 서로 다른 시점에 두 번 측정하여 그 변화율을 X_i 라고 하고 각 X_i 가 $N(\mu_i, \sigma^2)$ 로부터 추출되었다고 가정한다면(여기서, i 는 $1, \dots, N$), 우리는 각 모평균이 동일한지를 검정함으로써 종양의 이질성 여부를 파악할 수 있을 것이다. 그러나 앞에서 언급한 바와 같이 N 개의 서로 다른 모집단으로부터 추출된 표본의 크기가 모두 1인 경우에는 $H_0: \mu_1 = \dots = \mu_N$ 와 같은 가설 검정은 불가능하다. 하지만 만약 N 개의 모집단에서 크기가 1인 표본이 각각 추출된 자료에 대하여 모집단의 평균이 같은지를 검정할 수 있는 통계적 방법이 제안된다면, 그 방법을 활용하여 종양의 이질성이 존재하는지를 확인할 수 있을 것이다.

본 논문은 총 5장으로 구성되어 있다. 2장에서는 우리가 제안하는 방법을 설명하였다. 또한, 3장에서는 모의실험 연구를 통하여 제안한 방법의 성능을 확인할 것이다. 4장에서는 실제자료를 제안한 방법으로 적용시킨 결과를 정리하였다. 마지막으로 5장에서는 본 논문에 대하여 전체적으로 정리하였다.

2. 방법론

우리는 N 개의 모집단으로부터 표본이 한 개씩 추출될 때, $H_0: \mu_1 = \dots = \mu_N$ 의 귀무가설을 검정함으로써 모평균의 차이가 존재하여 구성된 N 개의 표본들의 이질성을 확인할 수 있는 최소 조합 t -검정(minimum combination t -test) 방법에 대해 제안한다.

여기서, 각 X_i 가 $N(\mu_i, \sigma^2)$ 로부터 추출되었다고 가정했을 때 $H_0: \mu_1 = \dots = \mu_N$ 의 가설이 참이라면 μ_1, \dots, μ_N 가 모두 같게 되는데, 이를 μ 라고 하면 N 개의 표본들은 $N(\mu, \sigma^2)$ 의 동일한 분포로부터 추출되었다고 할 수 있다. 따라서 N 개의 표본을 크기가 비슷한 임의의 두 집단인 X_{11}, \dots, X_{1n_1} 과 X_{21}, \dots, X_{2n_2} 로 나누어도 귀무가설이 참이기 때문에 두 집단의 분포는 여전히 $N(\mu, \sigma^2)$ 의 분포로부터 추출됨을 알 수 있다. 여기에서, X_{11}, \dots, X_{1n_1} 과 X_{21}, \dots, X_{2n_2} 은 X_1, \dots, X_N 의 한 분할(partition)이다. 또한, N 이 짝수인 경우, $n_1 = n_2 = N/2$ 이고, N 이 홀수인 경우, $n_1 = (N-1)/2$, $n_2 = (N+1)/2$ 이다. 이를 이용하여 우리는 N 개의 표본인 X_1, \dots, X_N 을 임의의 두 집단으로 나누고 가설검정을 실시한다. 여기서, 모인 표본에 대하여 크기가 비슷한 두 집단으로 나누게 되는데, 모든 경

우의 수로 두 집단을 나눌 수 있는 조합을 고려하게 된다. 조합의 수(m)는 집단의 크기에 따라 다음과 같이 정의된다:

$$m = \begin{cases} {}^N C_{\frac{N}{2}}, & \text{if } N = \text{짝수}, \\ {}^N C_{\frac{(N-1)}{2}}, & \text{if } N = \text{홀수}, \end{cases}$$

여기서 조합의 수는 최대 4,000개로 제한한다. 이때, 다중 검정의 개념에 착안하여 새롭게 나누어진 두 그룹에 대하여 조합의 수만큼 t -검정을 실시하여 나누어진 두 집단 간 평균을 비교하게 된다. 제안한 방법에서의 t -검정을 위한 i 번째 조합에서의 귀무가설과 대립가설은 다음과 같이 설정된다:

$$H_{0i} : \mu_{Ai} = \mu_{Bi} \quad \text{vs.} \quad H_{1i} : \mu_{Ai} \neq \mu_{Bi}, \quad i = 1, \dots, m.$$

여기에서, μ_{Ai} 와 μ_{Bi} 는 i 번째 조합에서 나누어진 두 집단의 모평균들이다. 모든 조합에서 귀무가설 하에서는 $\mu_{Ai} = \mu_{Bi} = \mu$ 이다. 만약 여기서 귀무가설을 기각하지 못한다면 임의로 나누어진 그룹들의 모평균이 모두 같기 때문에 동일한 분포로부터 추출되었다는 의미를 가지므로 원래 설정된 $H_0: \mu_1 = \dots = \mu_N$ 의 가설을 기각하지 못한다는 의미와 동일하게 된다. 또한, 귀무가설을 기각한다면 최소한 한 쌍의 두 그룹이 서로 상이한 분포로부터 추출되었으므로 마찬가지로 이는 $H_0: \mu_1 = \dots = \mu_N$ 의 가설을 기각한다고 할 수 있다.

이 방법은 모든 가능한 조합의 개수만큼 나누어진 두 집단에 대하여 t -검정을 실시하기 때문에 유의확률인 p -값 또한 그 개수만큼 계산되어 도출되는데, 그 중 적어도 한 개의 p -값이 유의수준에 대해 유의하다면 귀무가설을 기각하여 나누어진 두 집단이 서로 다른 평균을 가진다고 할 수 있으므로 p -값 중 최솟값을 유의수준과 비교한다. 여기서 우리는 유의수준에 대하여 본페르니 교정을 이용하여 다중 검정에서 생길 수 있는 오류를 보정할 것이다 (Dunn, 1961). 따라서 유의수준은 $\alpha = 0.05/m$ 로 설정된다. 제안된 새로운 가설검정 방법은 위에서 설명한 바와 같이 가설 검정 시 모든 가능한 조합의 개수만큼 t -검정을 하고 각 t -검정으로부터 p -값이 얻어지는 데, 그 중에서 최솟값을 구하고 그 값을 이용하여 귀무가설의 기각 여부를 판단하기 때문에 최소 조합 t -검정이라고 부르기로 하였다.

3. 모의실험 연구

3.1. 실험 계획

우리는 모의실험을 진행하기 위해 두 종류의 자료를 생성하였다. 표본의 개수인 N 은 10, 20, 30, 40, 50, 100로 설정하였다. 먼저, $H_0: \mu_1 = \dots = \mu_N$ 의 귀무가설이 참일 경우를 고려하기 위해 정규분포를 따르는 동일한 분포로부터 자료를 생성하였다. 정규분포 $N(10, \sigma^2)$ 에서 σ 가 0.1, 1, 2인 분포를 따르고 평균이 동일한 자료를 랜덤하게 생성하여 세 가지 경우를 고려하였다. 둘째로, H_1 : 어떤 $i \neq j$ 에 대하여 $\mu_i \neq \mu_j$ 의 대립가설이 참일 경우를 고려하기 위한 자료를 생성하였다. 이는 서로 다른 분포로부터 추출된 표본들을 의미하므로 표본의 개수인 N 에 대하여 각각 자료의 절반을 평균이 상이한 세 가지 종류의 자료를 생성하였다. 먼저, 절반은 $N(5, 0.1^2)$ 분포로부터, 나머지는 $N(0, 0.1^2)$ 분포로부터 생성되어 혼합된 자료와 절반은 $N(10, 0.1^2)$ 분포로부터, 나머지는 $N(0, 0.1^2)$ 분포로부터 생성된 자료를 고려하였다. 마지막으로 절반은 $N(15, 0.1^2)$ 분포로부터, 나머지는 $N(0, 0.1^2)$ 분포로부터 생성된 평균이 상이한 자료를 표현하여 모의실험을 진행하였다.

우리는 다음과 같은 기준에 따라 성능을 확인하였다: (1) 귀무가설이 참일 경우의 제 1종의 오류를 범할 확률(type I error rate); (2) 대립가설이 참일 경우의 검정력(power). 각 모의실험의 시나리오별로 자

료를 100번 생성시켜서 그 중에서 귀무가설을 몇 번 기각시키는 지를 센 후 제 1종의 오류를 범할 확률 또는 검정력을 구하고, 그 과정을 다시 100번 반복하여서 그 평균값을 구하였다. 이를 식으로 표현하면 다음과 같다:

$$\text{제 1종의 오류를 범할 확률(또는 검정력)} = \frac{1}{100} \sum_{i=1}^{100} \left(\frac{\text{귀무가설을 기각한 횟수}}{100} \right)_i,$$

여기서 우리는 제 1종의 오류를 범할 확률과 검정력을 이용하여 성능을 확인할 때 높은 값의 검정력보다 제 1종의 오류를 범할 확률이 0.05에 근사하는 경우로 초점을 두었다. 또한, 귀무가설이 참일 경우에서 유의수준이 $\alpha = 0.1/m$ 로 설정한 모의실험도 함께 진행하였다.

3.2. 결과

Figure 3.1을 통해 모의실험 결과를 확인해보자. 우선 Figure 3.1(a), (c), (e)를 볼 때, 귀무가설이 참이라는 가정 하에 유의수준이 $0.05/m$ 일 때 σ 가 0.1, 1, 2로 설정된 경우 모두 0.05보다 작거나 근사한 제 1종 오류 확률을 보인다. 또한, 유의수준이 $0.1/m$ 으로 설정된 경우에도 제 1종의 오류를 범할 확률이 0.1에 가깝게 나타나므로 귀무가설이 참이라는 가정 하에서 두 가지 다른 값의 유의수준을 지정했을 때 두 경우 모두 제안된 방법이 좋은 성능을 보임을 확인할 수 있다.

Figure 3.1(b), (d), (f)를 볼 때, 대립가설이 참이라는 가정 하에서는 세 가지 경우 모두 유의수준이 $0.05/m$ 일 때 검정력을 확인하면 N 이 10인 경우에만 1이고 다른 경우에는 모두 0.5 근처로 나온다. 이 값은 일반적으로 검정력이 높다고 할 때 고려되는 0.8보다는 작지만 제 1종 오류를 범할 확률을 고려한다면 어느 정도 좋은 성능임을 확인할 수 있다. 모의실험을 통해 도출된 결과에서 제 1종 오류를 범할 확률의 자세한 값은 Table 3.1을, 검정력에 대한 값은 Table 3.2를 보라.

4. 실제자료 예시

본 장에서는 전이성 폐암의 종양 연구로부터 얻어진 실제자료에 제안한 방법을 적용하였다. 이 자료는 전이성 폐암인 환자를 대상으로 각각 흉부 CT를 통해 최대 지름(maximum diameter)과 부피(volume)를 측정하고, 항암치료 후에도 후속 촬영을 실시하여 전이성 종양의 성장패턴 차이와 변화율에 따른 종양 이질성(tumor heterogeneity)을 파악하기 위한 연구(Yoo 등, 2017)로부터 생성되었다. 종양 이질성은 환경적, 유전적 요인에 의하여 표현형, 행동성향 등 특징이 다양한 종류의 종양세포를 형성하는 것이다(Marusyk와 Polyak, 2010). 따라서 투여되는 약물의 민감성이 다른 종양세포가 모여 조직을 형성하기 때문에 약물에 대한 반응정도가 달라 치료가 어려우므로 이를 연구하는 것이 중요하다. 우리가 제안한 방법을 적용시켜 종양 이질성을 지닌 환자를 파악할 수 있다.

이 자료는 2010년 1월부터 2014년 12월까지 폐암 전이성을 지닌 10명의 환자들을 대상으로 두 번의 CT 촬영을 실시하여 종양의 최대 지름(cm)과 부피(cm^3)를 측정하였다. 10명의 환자들은 1명의 여성 환자와 9명의 남성 환자로 구성되어 있고, 각 환자들이 지닌 종양(nodule)의 수는 4개부터 52개까지 존재한다.

우리가 제안한 방법을 적용하여 종양 이질성을 파악하기 위해, 먼저 환자들의 두 번의 방문을 통해 얻어진 관측값들을 이용하여 변화율을 계산한다. 변화율은 다음과 같이 계산된다:

$$\text{변화율} = \frac{V_2 - V_1}{V_1} \times 100(\%),$$

여기서 V_1 , V_2 는 각각 첫 번째 방문 시 관측값, 두 번째 방문 시 관측값에 해당된다. 이는 각 환자가 지닌 종양의 개수만큼 계산된다. 우리가 제안한 방법을 적용하여 얻은 결과는 Table 4.1을 보라.

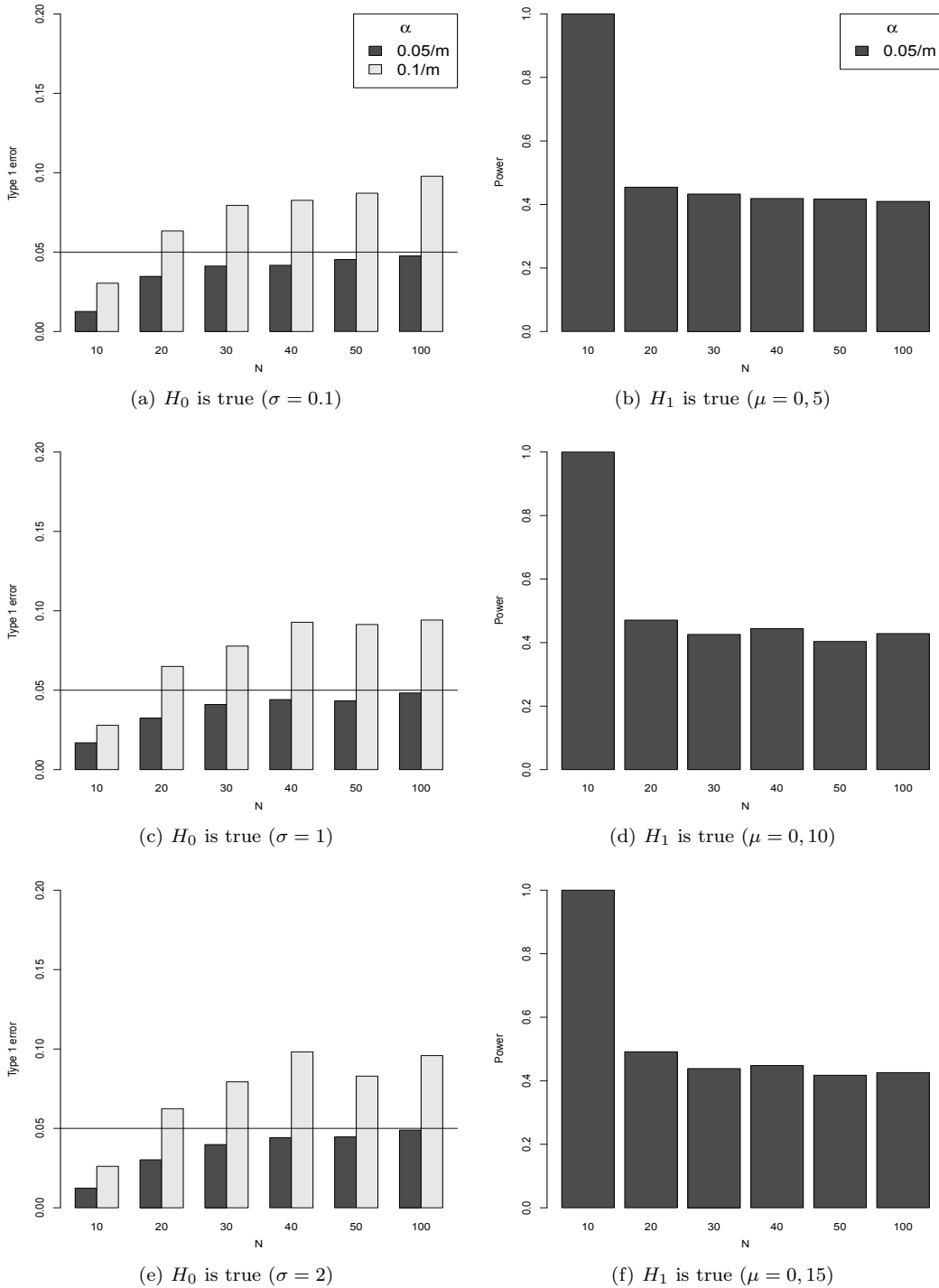


Figure 3.1. Estimated type I error rate when H_0 is true, and power when H_1 is true for minimum combination t -test method.

Table 3.1. Simulation results (estimated type I error rate with standard errors within parentheses) for data when H_0 is true

N	$\sigma = 0.1$		$\sigma = 1$		$\sigma = 2$	
	$0.05/m$	$0.1/m$	$0.05/m$	$0.1/m$	$0.05/m$	$0.1/m$
10	0.0126 (0.0011)	0.0304 (0.0014)	0.0169 (0.0011)	0.0279 (0.0013)	0.0124 (0.0013)	0.0261 (0.0013)
20	0.0347 (0.0015)	0.0633 (0.0025)	0.0324 (0.0014)	0.0650 (0.0027)	0.0302 (0.0015)	0.0624 (0.0025)
30	0.0412 (0.0021)	0.0794 (0.0029)	0.0410 (0.0019)	0.0778 (0.0028)	0.0398 (0.0020)	0.0793 (0.0028)
40	0.0417 (0.0020)	0.0826 (0.0029)	0.0441 (0.0018)	0.0928 (0.0030)	0.0442 (0.0019)	0.0982 (0.0032)
50	0.0453 (0.0022)	0.0871 (0.0031)	0.0433 (0.0021)	0.0914 (0.0031)	0.0446 (0.0021)	0.0829 (0.0030)
100	0.0476 (0.0023)	0.0978 (0.0032)	0.0483 (0.0023)	0.0942 (0.0031)	0.0489 (0.0022)	0.0958 (0.0031)

Table 3.2. Simulation results (estimated power with standard errors within parentheses) for data when H_1 is true

N	$\mu = 0, 5$	$\mu = 0, 10$	$\mu = 0, 15$
	$0.05/m$	$0.05/m$	$0.05/m$
10	1.0000 (0.0000)	1.0000 (0.0000)	1.0000 (0.0000)
20	0.4543 (0.0051)	0.4708 (0.0049)	0.4912 (0.0049)
30	0.4328 (0.0052)	0.4257 (0.0046)	0.4382 (0.0050)
40	0.4188 (0.0049)	0.4439 (0.0046)	0.4476 (0.0045)
50	0.4169 (0.0048)	0.4035 (0.0043)	0.4171 (0.0048)
100	0.4096 (0.0046)	0.4283 (0.0045)	0.4256 (0.0045)

Table 4.1. The results of applying the proposed method for each patient's nodules in 10 patients

Patient No.	No. of nodules	Min(p)		Min(Min(p))	Significant level ($0.05/m$)	Overall heterogeneity
		Maximum diameter	Volume			
1	27	0.0006	0.0005	0.0005	0.0000125	No
2	8	0.1463	0.3292	0.1463	0.0007142	No
3	52	0.0006	0.0028	0.0006	0.0000125	No
4	7	0.0878	0.1879	0.0878	0.0014285	No
5	25	0.0333	0.1324	0.0333	0.0000125	No
6	8	0.0324	0.1955	0.0324	0.0007142	No
7	5	0.0966	0.4226	0.0966	0.0050000	No
8	10	0.0305	0.0283	0.0283	0.0001984	No
9	9	0.0399	0.2978	0.0399	0.0003968	No
10	4	0.2815	0.3005	0.2815	0.0083333	No

우선, 각각 측정된 최대 지름과 부피에 대한 변화율을 대상으로 p -값의 최솟값(Min(p))을 도출하여 α 가 $0.05/m$ 인 유의수준에서 검정할 수 있다. 여기서 유의수준보다 작은 값을 가진다면 $H_0: \mu_1 = \dots = \mu_N$ 의 귀무가설을 기각하므로 이질성이 존재한다고 판단한다. Table 4.1의 결과를 보면 환자 10명에 대한 최대 지름과 부피의 변화율에 대해서 p -값의 최솟값이 각 환자들의 유의수준보다 크므로 모두 이질성을 보이지 않는다고 할 수 있다. 또한, 전체적인 이질성을 확인하기 위해 최대 지름과 부피에 대하여 도출된 두 최솟값을 이용할 수 있는데, 그 중 더 작은 값이 유의수준보다 작다면 환자가 지닌 종양에 대하여 전체적인 이질성이 존재한다고 할 수 있다. 결과, 10명의 환자 모두 전체적인 이질성을 보인다고 할 수 없다.

5. 결론

본 연구에서는 N 개의 모집단에서 표본이 한 개씩 추출되었을 때 평균 비교를 하기 위해 크기가 유사한 두 집단으로 나눌 수 있는 조합의 수를 고려해서 두 집단으로 나누어 각 집단의 \bar{X} 를 계산하고 평균 차이 검정을 통해 집단에 속한 표본들의 이질성을 확인할 수 있는 최소 조합 t -검정 방법을 제안하였다. 일반적으로 각 모집단으로부터 추출된 2개 이상의 표본에 대하여 모평균이 동일한지 검정할 때에는 일원배치 분산분석을 실시하게 되는데, 이 방법을 통해 반복의 수가 1인 경우 $H_0: \mu_1 = \dots = \mu_N$ 의 가설 검정이 가능하다.

우리는 모의실험을 통해 우리가 제안한 방법이 제 1종의 오류를 범할 확률을 작게 유지하는 결과를 보여 좋은 성능을 보임을 확인하였다. 그러나 검정력이 높지 않은 한계가 있음도 확인할 수 있었다. 또한, 표본들의 이질성을 파악하기 위하여 우리가 제안한 방법을 실제 자료에 적용시켰다. 대립가설이 참인 모의실험의 시나리오에서 각 모집단의 평균의 값을 절반은 0으로 하고 나머지 절반은 5, 10, 15 등 0보다 굉장히 큰 값으로 하였다. 이런 극단적인 설정에도 불구하고 검정력이 거의 모든 경우에 0.5가 나오는 것은 N 의 크기가 커짐에 따라 크기가 비슷한 두 집단으로 나누는 조합의 수가 급격히 커지고 이에 따라 수행해야 하는 t -검정의 횟수가 급격히 증가하기 때문에 사용자의 편의를 위해 t -검정의 최대 수행 횟수, 즉 m 의 최댓값을 4,000으로 제한을 두었기 때문이다. 모의실험에서 N 이 10인 경우에는 가능한 조합의 수가 252이기 때문에 m 의 값 또한 252가 되고 이 경우는 모의실험에서 검정력이 1이 나오게 된다. 그러나 다른 경우는 모두 조합의 수가 4,000을 훨씬 넘고 이런 경우에는 m 의 값이 4,000이 되고 검정력이 0.5에 가깝게 나오는 것이다. 각 모집단으로부터 추출된 표본의 크기가 1인 경우에 평균을 비교하는 가설검정 방법에 대한 연구이기 때문에 연구의 초점이 제 1종의 오류를 범할 확률을 조절하는 것에 더 맞춰져 있다. 게다가 실제 자료에서 N 에 해당하는 각 환자의 종양의 수는 10명의 환자 중 7명이 10 이하이기 때문에 우리의 방법을 적용했을 때의 결과가 모의실험 결과에 비추어 봤을 때 신뢰할 수 있다고 할 수 있겠다.

또한, 본 연구에서는 유의수준에 대하여 본페르니 교정을 이용하여 오류를 보정하였다. 본 연구는 각 모집단으로부터 추출된 표본의 크기가 모두 1인 경우에 다수의 모집단의 평균을 비교하기 위한 가설검정 방법을 제안하였다. 일반적인 통계학의 가설검정을 포함한 모든 추론에 대한 이론 또는 방법론에서 표본의 크기가 1인 경우에는 모집단의 평균에 대한 비교는 불가능하다. 그러나 본 연구에서는 이러한 특수한 경우에도 모집단의 평균을 비교할 수 있는 방법을 다중 검정의 맥락에서 제시하고 있다. 하지만 각 모집단으로부터 추출된 표본의 크기가 모두 1이기 때문에 제안된 가설검정 방법은 여전히 그 결과를 신뢰하기가 어려울 수 있다. 따라서 우리는 다중 검정에서 제 1종의 오류를 범할 확률을 조절하는 여러 가지 방법들 중 가장 보수적인 본페르니 교정을 통해 제안된 방법을 사용하여 도출된 결과에 대한 신뢰도를 높이하고자 하였다. 그러나 이 방법 외에도 p -값의 순서를 고려하여 false discovery rate (FDR)을 보정할 수 있는 Holm-Bonferroni 방법 (Holm, 1979)이나 Benjamini-Hochberg 방법 (Hochberg와 Benjamini, 1990) 등 다른 다중 검정의 방법을 고려할 수 있을 것이다. 또한 우리는 한 집단을 두 집단으로 나누었는데, 조합할 수 있는 만큼 세 집단 이상으로 나누어 평균 비교에 대한 검정을 실시할 수 있다. 여기서 t -검정이 아니라 등분산 가정이나 정규성 가정에 따른 일원배치 분산분석이나 Kruskal-Wallis 방법 (Kruskal과 Wallis, 1952)을 고려할 수 있을 것이다. 이처럼 모집단에서 표본이 한 개씩 추출되었을 때 평균 비교를 함으로써 이질성을 분석할 수 있는 다양한 연구가 진행되어야 할 것이다.

References

- Dunn, O. J. (1961). Multiple comparisons among means, *Journal of the American Statistical Association*,

56, 52–64.

- Fisher, R. (1918). The correlation between relatives on the supposition of Mendelian inheritance, *Transactions of the Royal Society of Edinburgh*, **52**, 399–433.
- Hochberg, Y. and Benjamini, Y. (1990). More powerful procedures for multiple significance testing, *Statistics in medicine*, **9**, 811–818.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics*, **6**, 65–70.
- Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis, *Journal of the American statistical Association*, **47**, 583–621.
- Marusyk, A. and Polyak, K. (2010). Tumor heterogeneity: causes and consequences, *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, **1805**, 105–117.
- Student. (1908). The probable error of a mean, *Biometrika*, **6**, 1–25.
- Yoo, J., Kim, Y., Lim, C., Heo, M., Hwang, I., and Chong, S. (2017). Assessment of Spatial Tumor Heterogeneity using CT Phenotypic Features Estimated by Semi-Automated 3D CT Volumetry of Multiple Pulmonary Metastatic Nodules: A Preliminary Study, *unpublished manuscript*.

크기가 1인 표본들로 구성된 집단에 기반한 모평균의 차이를 검정하기 위한 최소 조합 t -검정 방법

허미영^a · 임창원^{a,1}

^a중앙대학교 응용통계학과

(2017년 2월 24일 접수, 2017년 2월 26일 수정, 2017년 2월 26일 채택)

요약

일반적으로 각 N 개의 모집단에서 2개 이상의 표본이 추출되었을 때, $H_0 : \mu_1 = \dots = \mu_N$ 의 가설에 대하여 검정할 수 있지만 각 모집단으로부터 표본이 한 개씩 추출된다면 \bar{X} 가 존재하지 않으므로 모평균의 차이 검정은 불가능하다. 하지만 하나씩 추출된 표본으로 구성된 집단을 두 집단으로 나누어 임의의 평균을 생성함으로써 평균의 차이를 비교한다면 표본들 사이에 존재할 수 있는 이질성을 파악할 수 있다. 따라서 우리는 두 집단으로 나눌 수 있는 조합의 수 만큼 평균 차이를 검정할 수 있는 최소 조합 t -검정 방법을 제안하고자 한다. 최종적으로 본 논문에서는 한 개씩 추출된 표본들 사이의 이질성을 확인하기 위하여 평균 차이를 검정할 수 있는 방법을 제안하였고 모의실험 연구를 통해 성능을 확인하였고 실제 자료 분석을 통해 결과를 도출하였다.

주요용어: 이질성, 최소 조합 t -검정 방법, 평균 차이, 가설 검정

이 논문은 2015년도 중앙대학교 연구장학기금 지원에 의한 것임.

¹교신저자: (06974) 서울시 동작구 흑석로 84, 중앙대학교 경영경제대학 응용통계학과. E-mail: clim@cau.ac.kr