

# Visual inspection of overlapping confidence intervals for comparison of normal population means

Sookhee Choi<sup>a</sup> · Kyungsoo Han<sup>b,1</sup>

<sup>a</sup>Department of Psychology, Woosuk University;

<sup>b</sup>Department of Statistics, Chonbuk National University

(Received July 17, 2017; Revised August 19, 2017; Accepted September 23, 2017)

---

## Abstract

Data analysts sometimes test the equality of two normal population means by the inspection of the overlapping of two confidence intervals. This method seems simple to use; however, it is a common statistical misconception to suppose that two normal means are not significantly different because of no overlapping. This article will present transforming the confidence interval of the mean difference to individual confidence intervals that are visualized to inspect overlapping. It will also be shown that this technique can be extended when comparing the  $k$  normal population means with equal variances.

Keywords: confidence interval, overlapping, statistical significance

---

## 1. 서론

독립인 두 개의 모집단의 평균을 비교하는 통계적 추론은 다양한 연구 분야에서 가장 많이 사용되는 통계 분석 기법 중의 하나이다. 대학의 모든 교양 통계학 수업에서는 이 문제를 신뢰구간 또는 가설검증을 이용하여 상세히 다루고 있다. 가설검증보다 신뢰구간으로 모평균을 추론할 때의 장점은 점 추정치와 모집단의 모평균으로 생각될 수 있는 값들의 구간을 같이 제시한다는 데 있다.

데이터를 점도표로 표시할 수 있을 정도로 많지 않은 경우에는 Figure 1.1처럼 점도표와 신뢰구간을 같이 제시하는 것이 바람직하다 (Gardner과 Altman, 1986). 그림에서 왼쪽 두 개의 신뢰구간은 두 집단 각각의 모평균에 관한 95% 개별 신뢰구간이다. 통계학 초심자나 비전공자는 두 개의 신뢰구간이 겹치는 부분이 있으므로 5% 유의수준에서 통계적 유의성이 없다고 판단하는 오류를 범할 수 있다. 그러나 그림의 맨 오른쪽 신뢰구간은 두 집단의 모평균 차에 대한 95% 신뢰구간으로 0을 포함하지 않는다. 따라서 유의수준 5%에서 통계적으로 유의하게 다르고  $p$ -값을 계산하면 0.028이다.

각 집단의 모평균에 대한 개별 신뢰구간의 그래프 없이 표본 평균 차의 신뢰구간 그래프만 보면 각 집단의 표본 평균과 표준오차에 대한 정보가 없고 평균 차의 축에서 영점이 달라지므로 각 평균이 어느 범위에 위치하는지 파악하기 곤란하다는 단점이 따른다. 이러한 단점을 의식했기 때문인지 모르지만 통계 비전공자는 Figure 1.1에서 맨 오른쪽의 평균 차에 대한 신뢰구간은 생략하고 왼쪽 두 개의 신뢰구간만

---

<sup>1</sup>Corresponding author: Department of Statistics (Institute of Applied Statistics), Chonbuk National University, 567 Baekje-daero, deokjin-gu, Jeonju-si, Jeollabuk-do 54896, Korea. E-mail: kshan@jbnu.ac.kr

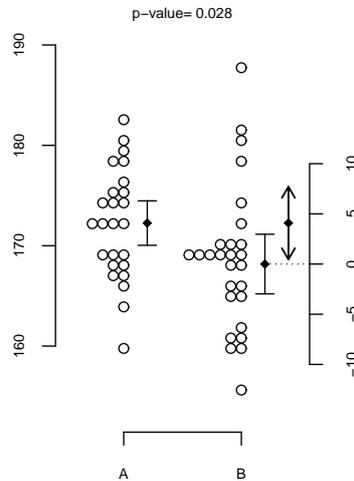


Figure 1.1. 95% confidence intervals.

가지고 서로 겹치는 부분의 유무에 따라 통계적 유의성을 판단하려고 한다. 이 방법은 매우 쉬워 편리해 보이지만 통계적 가설검증의 검정력을 심각하게 떨어뜨린다. Schenker와 Gentleman (2001)은 의학 분야 학술지에서 이러한 오류를 범한 논문을 60개 넘게 발견했다고 보고한다. 신뢰구간이 겹치는 비율과  $p$ -값 사이의 관계를 대략적으로 알 수 있는 몇 가지 규칙은 Cumming (2009)이 제안하였다.

비교할 집단이  $k$ 개일 때 모든 쌍의 평균 차에 대한 신뢰구간은  $k(k-1)/2$ 개이다.  $k$ 가 조금만 커져도 이 신뢰구간들을 모두 그래프로 표시하기엔 너무 많다. 각 집단의 모평균에 대한 신뢰구간을 그래프로 각각 그리고 이들이 서로 겹치는가 조사함으로써 통계적 유의성을 시각적으로 판단하게 한다면 통계학 초심자나 비전공자도 쉽게 이해할 수 있다. Browne (1979)는 두 집단의 표본 크기나 표준오차의 비에 따라서 신뢰구간의 축소 비율을 표로 만들었으나 일반적인 규칙을 발견하지는 못했다. Browne (1979) 이후 30여년 동안 정규 모집단의 평균을 비교하기 위한 신뢰구간 축소 기법을 다양한 학문 분야의 여러 연구자들이 제안하였다 (Goldstein과 Healy, 1995; Schenker와 Gentleman, 2001; Tryon, 2001; Payton 등, 2003; Cumming, 2009; Maghsoodloo와 Huang, 2010). 이들의 주요 연구 결과를 간단히 요약하면, 정규 모집단의 두 모평균을 비교할 경우에 95% 신뢰수준 대신에 83.4% 신뢰수준을 사용한 두 개의 신뢰구간이 겹치는 부분이 없을 때 5% 유의수준에서 두 모평균은 통계적으로 유의하게 다르다는 것이다.

기존 연구자들이 제안한 통계적 유의성 판단을 위한 신뢰구간 축소 기법은 대학의 교양 통계학 수준에서 다루기가 쉽지 않다. 또한 3개 이상의 모평균 비교 문제로 확장하기가 어렵다는 단점도 있다. 2절에서 두 모평균 차의 신뢰구간은 간단한 부등식을 사용하여 평균 비교를 위한 각각의 모평균에 대한 두 개의 신뢰구간으로 표현할 수 있음을 밝힌다. 같은 방법으로  $k$ 개의 모평균을 비교하는 튜키-크레머(Tukey-Kramer)의 다중비교 결과인 평균 차에 대한  $k(k-1)/2$  쌍의 동시 신뢰구간도  $k$ 개의 평균에 대한 개별 신뢰구간으로 나타낼 수 있어서 신뢰구간이 겹치는 것으로 통계적 유의성을 판단할 수 있음을 보인다.

## 2. 평균 차의 신뢰구간을 각 평균의 신뢰구간으로 분해하기

### 2.1. 분산이 다른 두 개의 정규 모집단

독립인 두 개의 정규 모집단  $N(\mu_i, \sigma_i^2)$ ,  $i = 1, 2$ 에서 표본  $n_i$ 개를 각각 뽑았을 때 표본 평균 차  $\bar{X}_1 -$

$\bar{X}_2$ 의 표집분포는  $N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$ 이다.  $\sigma_i^2$ 에 대한 추정치가  $S_i^2$ 일 때  $\mu_1, \mu_2, \mu_1 - \mu_2$ 에 대한  $100(1 - \alpha)\%$  신뢰구간은 각각 다음과 같다.

$$\bar{X}_1 \pm t_{\frac{\alpha}{2}, n_1-1} \frac{S_1}{\sqrt{n_1}}, \tag{2.1}$$

$$\bar{X}_2 \pm t_{\frac{\alpha}{2}, n_2-1} \frac{S_2}{\sqrt{n_2}}, \tag{2.2}$$

$$\bar{X}_1 - \bar{X}_2 \pm t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \tag{2.3}$$

여기서 자유도 df는

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 / (n_1 - 1) + \left(\frac{s_2^2}{n_2}\right)^2 / (n_2 - 1)}$$

로 정의된다.

식 (2.3)에서 주어진 평균 차에 대한 신뢰구간의 신뢰수준은  $1 - \alpha$ 로 유지한 채 각각  $\mu_1, \mu_2$ 에 대한 두 개의 신뢰구간으로 분리하여 표현할 수 있다.  $r_1 + r_2 = 1$  ( $0 \leq r_1 \leq 1$ )일 때 식 (2.3)의 오차한계를 각각  $r_1$ 과  $r_2$ 의 비율만큼 축소한  $\mu_1$ 과  $\mu_2$ 의 신뢰구간

$$\bar{X}_1 \pm r_1 \times t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \tag{2.4}$$

$$\bar{X}_2 \pm r_2 \times t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \tag{2.5}$$

은 귀무가설( $H_0 : \mu_1 = \mu_2$ ) 하에서 겹치지 않을 확률이  $\alpha$ 가 된다. 왜냐하면 식 (2.4)와 (2.5)에서

$$m = t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

이라 표기하면 귀무가설 하에서 두 개의 신뢰구간 (2.4)와 (2.5)가 겹치는 부분이 없을 확률은

$$\begin{aligned} P(\bar{X}_1 + r_1 m < \bar{X}_2 - r_2 m, \bar{X}_2 + r_2 m < \bar{X}_1 - r_1 m) &= P(|\bar{X}_1 - \bar{X}_2| > (r_1 + r_2)m) \\ &= P(|\bar{X}_1 - \bar{X}_2| > m) \\ &= \alpha \end{aligned}$$

가 항상 성립하기 때문이다.

식 (2.1)과 (2.2)에서 신뢰구간의 길이는 각각 표준오차  $S_i/\sqrt{n_i}$  ( $i = 1, 2$ )에 비례한다. 따라서 식 (2.4)와 (2.5)에서  $r_1, r_2$ 는 표준오차에 비례하게

$$r_1 = \frac{S_1/\sqrt{n_1}}{S_1/\sqrt{n_1} + S_2/\sqrt{n_2}}, \tag{2.6}$$

$$r_2 = 1 - r_1 = \frac{S_2/\sqrt{n_2}}{S_1/\sqrt{n_1} + S_2/\sqrt{n_2}} \tag{2.7}$$

로 결정하는 것이 합리적이다.

식 (2.6)의  $r_1$ 을 식 (2.4)에 대입하여 정리하면

$$\begin{aligned}\bar{X}_1 \pm r_1 \times t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} &= \bar{X}_1 \pm \frac{S_1/\sqrt{n_1}}{S_1/\sqrt{n_1} + S_2/\sqrt{n_2}} \times t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \\ &= \bar{X}_1 \pm t_{\frac{\alpha}{2}, df} \frac{S_1}{\sqrt{n_1}} \frac{\sqrt{S_1^2/n_1 + S_2^2/n_2}}{S_1/\sqrt{n_1} + S_2/\sqrt{n_2}} \\ &= \bar{X}_1 \pm t_{\frac{\alpha}{2}, df} \frac{S_1}{\sqrt{n_1}} \frac{\sqrt{1 + \frac{S_2^2/n_2}{S_1^2/n_1}}}{1 + \frac{S_2/\sqrt{n_2}}{S_1/\sqrt{n_1}}}\end{aligned}\quad (2.8)$$

이 된다. 마찬가지로 식 (2.7)의  $r_2$ 를 식 (2.5)에 대입하여 정리하면

$$\bar{X}_2 \pm t_{\frac{\alpha}{2}, df} \frac{S_2}{\sqrt{n_2}} \frac{\sqrt{1 + \frac{S_2^2/n_2}{S_1^2/n_1}}}{1 + \frac{S_2/\sqrt{n_2}}{S_1/\sqrt{n_1}}}\quad (2.9)$$

이 된다. 식 (2.8)과 (2.9)에서

$$\gamma = \frac{S_2/\sqrt{n_2}}{S_1/\sqrt{n_1}}$$

라고 표기하면 식 (2.4)와 (2.5)의 신뢰구간은

$$\begin{aligned}\bar{X}_1 \pm t_{\frac{\alpha}{2}, df} \frac{\sqrt{1 + \gamma^2}}{1 + \gamma} \frac{S_1}{\sqrt{n_1}}, \\ \bar{X}_2 \pm t_{\frac{\alpha}{2}, df} \frac{\sqrt{1 + \gamma^2}}{1 + \gamma} \frac{S_2}{\sqrt{n_2}}\end{aligned}\quad (2.10)$$

이 된다. 식 (2.10)에서 두 개의 신뢰구간이 겹치지 않을 때 두 모평균은 유의수준  $\alpha$ 에서 서로 다르다.

Figure 2.1의 왼쪽 신뢰구간은 각 집단의 모평균에 대한 95% 개별 신뢰구간이다.  $p$ -값은 유의수준 5%보다 작지만 신뢰구간은 서로 겹치므로 시각적으로 통계적 유의성을 판단하기 어렵다. 그러나 식 (2.10)에 기초한 오른쪽 신뢰구간은 서로 겹치지 않은 경우에는  $p$ -값도 5%보다 작게 되므로 통계적 유의성을 시각적으로 판단할 수 있다.

두 정규 모집단의 공통분산이  $\sigma^2$ 이고, 표본 크기도 같은  $n$ 인 경우에  $\sigma^2$ 의 공통분산추정치를  $S_p^2$ 으로 표기한다면 식 (2.10)의 신뢰구간은 다음과 같다.

$$\begin{aligned}\bar{X}_1 \pm t_{\frac{\alpha}{2}, 2n-2} \frac{1}{\sqrt{2}} \frac{S_p}{\sqrt{n}}, \\ \bar{X}_2 \pm t_{\frac{\alpha}{2}, 2n-2} \frac{1}{\sqrt{2}} \frac{S_p}{\sqrt{n}}\end{aligned}\quad (2.11)$$

Figure 2.2는 95% 신뢰수준에서 식 (2.1)과 (2.11)을 비교했을 때 개별 신뢰구간의 축소비율과 축소된 신뢰구간의 신뢰수준을 표본 크기에 따라 그래프로 그린 것이다.  $n = 5$ 일 때 축소비율은 58.7%이고 신뢰수준은 82.2%이고  $n = 100$ 일 때의 축소비율은 70.3%이고 신뢰수준은 83.4%이다 (Payton 등, 2003).

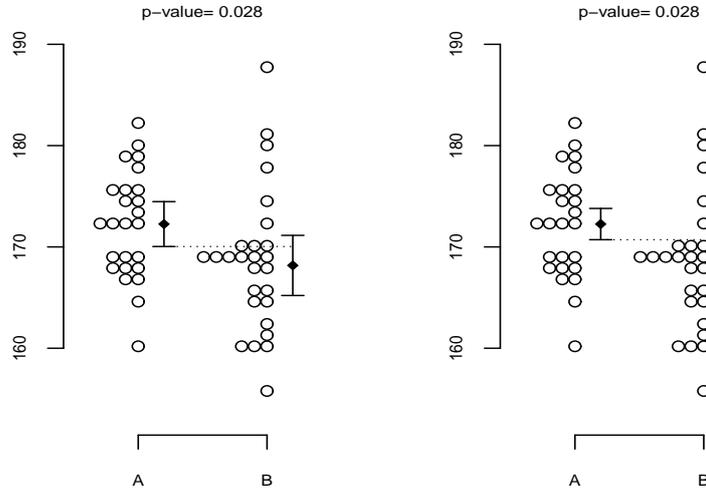


Figure 2.1. Visual inspection of overlapping confidence intervals.

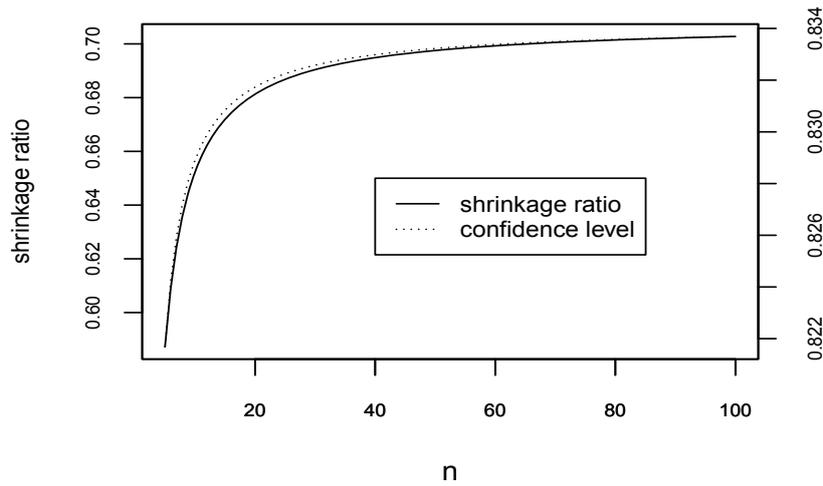


Figure 2.2. Shrinkage ratio and confidence level.

2.2. 분산이 같은 k개의 정규 모집단

분산이 모두 같은  $\sigma^2$ 인 k개의 정규 모집단에서 각각  $n_i$  ( $1 \leq i \leq k$ )개의 표본을 뽑았을 때 쌍별 모평균 차이에 대한 동시 신뢰구간을 튜키-크래머 방법으로 구하면 다음과 같다.

$$\bar{X}_i - \bar{X}_j \pm \frac{q_{\alpha;k;N-k}}{\sqrt{2}} S_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}, \quad i, j = 1, \dots, k; i \neq j, \tag{2.12}$$

여기서  $N = \sum n_i$ 이고  $q_{\alpha;k;N-k}$ 는 스튜던트화 범위(studentized range) 분포의 분위수이다.

식 (2.10)을 유도한 방법을 식 (2.12)에 적용하면 평균 차에 대한 신뢰구간을 식 (2.13)과 같이 각 평균

에 대한 신뢰구간으로 바꿀 수 있다.

$$\begin{aligned}\bar{X}_i &\pm \frac{q_{\alpha;k;N-k}}{\sqrt{2}} \frac{\sqrt{n_i+n_j}}{\sqrt{n_i}+\sqrt{n_j}} \frac{S_p}{\sqrt{n_i}}, \\ \bar{X}_j &\pm \frac{q_{\alpha;k;N-k}}{\sqrt{2}} \frac{\sqrt{n_i+n_j}}{\sqrt{n_i}+\sqrt{n_j}} \frac{S_p}{\sqrt{n_j}}.\end{aligned}\quad (2.13)$$

식 (2.13)에서 표본 크기  $n_i$ 가 모두 같지 않을 때는  $i$ 번째 모평균에 대한 신뢰구간이 최대  $(k-1)$ 개 존재할 수 있으므로 유일하게 결정되지 않는다. 그러나 표본 크기  $n_i$ 가 모두 같은  $n$ 일 때는  $\mu_i$ 에 대한 신뢰구간은

$$\bar{X}_i \pm \frac{q_{\alpha;k;N-k}}{2} \frac{S_p}{\sqrt{n}}$$

이 되므로 각 신뢰구간은 유일하게 결정된다.

식 (2.13)에서 표본 크기  $n_i$ 가 서로 다른 경우  $\mu_i$ 에 대한 신뢰구간이  $j$ 번째 집단의 표본 크기  $n_j$  ( $j = 1, \dots, k, j \neq i$ )에 의존하므로  $\sqrt{n_i+n_j}/(\sqrt{n_i}+\sqrt{n_j})$  ( $j = 1, \dots, k, j \neq i$ )의 최댓값 또는 평균을 사용하여 다른 집단의 표본 크기에 대한 의존성을 없앨 수 있다. 식 (2.14)에서는 최댓값을 사용하였다.

$$\bar{X}_i \pm \frac{q_{\alpha;k;N-k}}{\sqrt{2}} r_i \frac{S_p}{\sqrt{n_i}}, \quad (2.14)$$

여기서  $r_i$ 는 다음과 같이 정의된다.

$$r_i = \max_{j=1, \dots, k, j \neq i} \frac{\sqrt{n_i+n_j}}{\sqrt{n_i}+\sqrt{n_j}}. \quad (2.15)$$

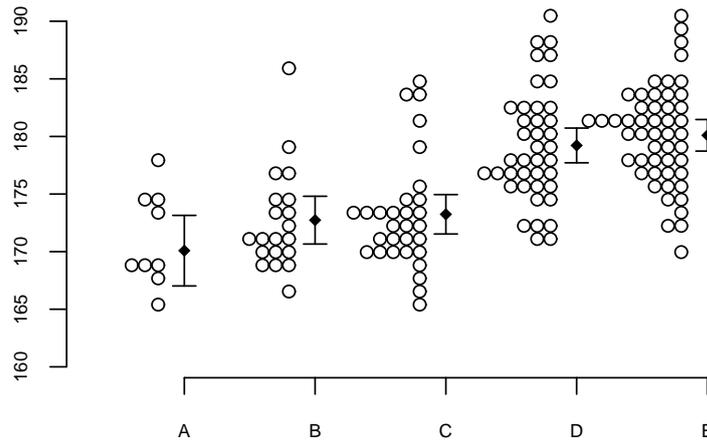
식 (2.14)를 이용하여 평균을 비교하는 방법을 R에서 생성한 모의실험 데이터로 설명한다. 분산은 모두 같은 25이며 5개의 정규모집단의 모평균은 172부터 180까지 2만큼 증가시켰다. 난수 초기치는 1234이며 `rnorm` 함수를 이용하여 생성한 데이터를 요약한 것이 Table 2.1이다. R의 `aov` 함수로 분산분석을 실행한 후, R이 제공하는 `TukeyHSD` 함수를 이용하여 평균 차에 대한 95% 동시 신뢰구간을 구한 결과는 다음과 같다.

	diff	lwr	upr	p adj
B-A	2.6505	-2.3054	7.6063	0.5788
C-A	3.1577	-1.5147	7.8301	0.3399
D-A	9.1364	4.6124	13.6605	0.0000
E-A	10.0157	5.5830	14.4483	0.0000
C-B	0.5072	-3.1867	4.2011	0.9956
D-B	6.4859	2.9816	9.9903	0.0000
E-B	7.3652	3.9797	10.7507	0.0000
D-C	5.9787	2.8882	9.0692	0.0000
E-C	6.8580	3.9029	9.8131	0.0000
E-D	0.8792	-1.8352	3.5937	0.8984

각 집단의 표본 크기가 서로 다른 경우에 튜키-크래머 방법은 설정된 유의수준보다 더 보수적으로 가설 검증을 한다는 사실은 잘 알려져 있다. 따라서 위에서 구한 동시 신뢰구간으로 평균을 비교하는 가설검

**Table 2.1.** Summary statistics of simulated data

Group	Sample size	Sample mean	Sample variance
A	10	170.1	24.8
B	20	172.7	19.2
C	30	173.2	23.0
D	40	179.2	24.3
E	50	180.1	18.6



**Figure 2.3.** Visualizing the overlap of confidence intervals.

증을 할 때 유의수준은 5%보다 작음에 유의해야 한다. 10쌍의 평균 차에 대한 동시 신뢰구간 결과를 해석하면 집단 A, B, C는 서로 간에 평균 차이가 없고, 집단 D, E도 평균 차이가 없다. 반면에 집단 A, B, C는 집단 D, E와 서로 평균이 다르다.

평균 차에 대한 95% 동시 신뢰구간을 식 (2.14)를 이용하여 각 평균에 신뢰구간으로 바꾸어 그래프로 그린 것이 Figure 2.3이며 각각의 평균에 대한 신뢰구간의 신뢰수준은 95%가 아님에 유의해야 한다. A, B, C의 그룹과, D, E의 그룹으로 나누어짐을 쉽게 알 수 있다.

### 3. 마무리

Fisher의 유의성검증과 Neyman-Pearson의 가설 검증 기법은 개발된 이후 다양한 학문 분야에서 잘못 사용된 수많은 사례들을 연구자들이 꾸준히 지적하였다. 그 결과 1980년대 이후 의학계의 학술지는 통계적 추론 기법으로 가설검증보다는 신뢰구간을 주로 이용하고 있다. 2010년 미국심리학회가 논문 작성 시에 점추정과 신뢰구간에 기초한 통계분석 결과 보고에 대한 지침을 공표한 이후 이 지침은 다른 학문 분야에도 급속히 퍼지고 있다. 가설검증 없이 신뢰구간만으로 통계적 추론을 다룬 기초통계학 교재인 Cumming과 Calin-Jageman (2017)도 출판되었다. 따라서 이제는 대학에서 교양통계학을 가르칠 때 가설검증 기법보다 신뢰구간을 더 중요하게 생각하고 더 많이 다루어야 할 때이다.

본 논문은 독립인 여러 개의 정규 모집단의 평균을 비교할 때 평균 차에 대한 동시 신뢰구간을 개별 평균에 대한 신뢰구간으로 변환하여 시각화하는 방법을 다루었다. 집단이 두 개인 경우는 간단한 부등식 연산만으로 두 개의 평균에 대한 신뢰구간을 만들 수 있으므로 현행 대학의 교양 통계학 수준에서 쉽게

다를 수 있다. 통계 비전공자가 교양 통계학을 1학기 배운 후에 몇 년이 지나면 평균 차에 대한 가설검증이나 신뢰구간 공식을 대부분 잊을 것이다. 그러나 각 평균에 대한 신뢰구간을 그래프로 표시한 후에 이 둘이 겹치지 않을 때 평균이 유의하게 다르고, 겹치면 평균이 유의하게 다르지 않다고 해석하는 방법은 쉽게 잊어버리지 않을 것이다. 이 방법은 매우 직관적이라 통계를 배우지 않은 사람이 처음 들어도 쉽게 이해할 수 있을 것이라 생각한다.

## References

- Browne, R. H. (1979). On visual assessment of the significance of a mean difference, *Biometrics*, **35**, 657–665.
- Cumming, G. (2009). Inference by eye: reading the overlap of independent confidence intervals, *Statistics in Medicine*, **28**, 205–220.
- Cumming, G. and Calin-Jageman, R. J. (2017). *Introduction to the New Statistics: Estimation, Open Science, and Beyond*, Routledge, New York.
- Gardner, M. J. and Altman, D. G. (1986). Confidence intervals rather than  $P$  values: estimation rather than hypothesis testing, *British Medical Journal*, **292**, 746–750.
- Goldstein, H. and Healy, M. J. R. (1995). The graphical presentation of a collection of means, *Journal of the Royal Statistical Society A*, **158**, 175–177.
- Maghsoodloo, S. and Huang, C. (2010). Comparing the overlapping of two independent confidence intervals with a single confidence interval for two normal population parameters, *Journal of Statistical Planning and Inference*, **140**, 3295–3305.
- Payton, M. E., Greenstone, M. H., and Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance?, *The Journal of Insect Science*, **3**, 34–39.
- Schenker, N. and Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals, *The American Statistician*, **55**, 182–186.
- Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: an integrated alternative method of conducting null hypothesis statistical tests, *Psychological Methods*, **6**, 371–386.

# 정규 모집단의 평균 비교를 위한 신뢰구간 겹치기 시각화

최속희<sup>a</sup> · 한경수<sup>b,1</sup>

<sup>a</sup>우석대학교 심리학과, <sup>b</sup>전북대학교 통계학과

(2017년 7월 17일 접수, 2017년 8월 19일 수정, 2017년 9월 23일 채택)

---

## 요약

두 개의 정규 모집단의 평균이 같은가를 검증할 때 두 개의 신뢰구간이 겹치는 지를 시각적으로 판단하여 결정하는 방법은 매우 직관적이면서도 사용하기 쉽다. 그러나 신뢰구간이 겹칠 때도 두 집단의 평균은 통계적으로 유의하게 다를 수 있으므로 가설검증 결과와 다를 수 있다. 평균 차에 대한 신뢰구간을 각 평균의 신뢰구간으로 변환한 후에 두 신뢰구간이 겹치는지 여부를 시각적으로 판단하여 가설검증을 할 수 있는 방법을 제안한다. 또한 이 방법은 분산이 같은  $k$ 개의 정규 모집단의 평균을 비교할 경우에도 사용할 수 있음을 보인다.

주요용어: 신뢰구간, 겹치기, 통계적 유의성

---

<sup>1</sup>교신저자: (54896) 전북 전주시 덕진구 백제대로 567, 전북대학교 통계학과 (응용통계연구소).  
E-mail: kshan@jbnu.ac.kr