

Sample size of bootstrap method used in P300 concealed information test

Jin-Sup Eom^{1†}, Hajung Jeon²

¹Chungbuk National University, ²Chungnam National University

It is evaluated whether the P300 amplitude for the probe is greater than the P300 amplitude for the irrelevant in the P300 concealed information test. However, there is a problem that the P300 amplitude for the probe is overestimated because the number of trials of the irrelevant is much larger than that of the probe. Rosenfeld et al. (2008) attempted to solve this problem by reducing the bootstrap sample size of the irrelevant to the sample size of the probe. In general, the bootstrap sample size must be the same as the original sample size and the type 1 error rate becomes smaller than the significance level if the bootstrap sample size is smaller than the original sample size. The purpose of this study is to evaluate the type 1 error rate of the modified bootstrap method that reduces the bootstrap sample size of irrelevant through Monte Carlo studies and to check whether this error can be corrected. As a result of experiment 1, the type 1 error rate of the modified bootstrap method was about .073, which was lower than the significance level .10. The type 1 error rate of the adjusted bootstrap method with corrected the significance level using the standard error was about .140 which was higher than the significance level .10. Consequently, the error of the modified bootstrap method was not corrected. In order to investigate the reason why the error of the modified bootstrap method was not corrected, a Monte Carlo study using numbers was performed. In the results of experiment 2, the type 1 error rate of the modified bootstrap method was about .054, which was less than the significance level .10, and that of the adjusted bootstrap method was about .10, which was the same as the significance level. It was found that the reason why the error of the modified bootstrap method is not corrected was due to the specificity of the EEG data. The reasons why these errors are not corrected and how to solve these errors were discussed.

Keywords: P300, concealed information test, bootstrap, sample size

1차원고접수 21.05.03; 수정본접수: 21.06.09; 최종게재결정 21.06.10

조사대상자가 특정 정보를 기억하고 있는지 판단할 때 P300 숨긴정보검사(P300-based concealed information test: P300 CIT)를 사용할 수 있다(Rosenfeld, 2011). P300 CIT에서는 조사대상자에게 탐지하려고 하는 정보(관련자극 또는 탐침자극; probe)와 탐지하려는 정보와 직접 관련이 없는 정보들(무관련자극; irrelevant)을 섞어 제시하면서 뇌파를 측정한다. 보통 관련자극과 무관련자극의 비율을 1:4 이상으로 하며, 관련자극의 제시 횟수가 최소 30회 이상(따라서 무관련

자극의 제시횟수는 120회 이상) 되도록 전체 자극들을 반복 제시한다. 검사가 끝난 후에 관련자극과 무관련자극에 대한 사건관련전위(event-related potential: ERP)를 산출하고 P300 진폭을 측정한다. 관련자극에 대한 P300 진폭이 무관련자극에 대한 P300 진폭보다 더 크게 나타나면 조사대상자가 특정 정보를 기억하고 있다고 판단한다.

특정 조사대상자에서 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 큰지를 판단할

† 교신저자: 엄진섭, 충북대학교 심리학과, (28644) 충북 청주시 서원구 충대로 1
E-mail: jseom2003@hanmail.net

때 부트스트랩(bootstrap) 방법이 가장 많이 사용되고 있다 (Cutmore, Djakovic, Kebbell, & Shum, 2009; Rosenfeld, Soskins, Bosh, & Ryan, 2004). P300 CIT에서 부트스트랩 방법은 백분위를 이용하는 방법(Cutmore et al., 2009)과 표준오차를 이용하는 방법(Rosenfeld et al., 2004) 두 가지가 사용된다. 백분위를 이용하는 방법은 다음과 같다. ① 관련 자극에 대한 30개의 단일 시행 뇌파(single sweep)에서 30개를 복원추출하고, 무관련자극에 대한 120개의 단일 시행 뇌파에서 120개를 복원추출한다. ② 복원추출된 뇌파에서 관련 자극과 무관련자극의 ERP를 각각 산출하고, P300 진폭을 측정한다. ③ 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 크면 부트스트랩 지수에 1을 더한다. ④ ①~③의 과정을 1,000번 반복한다. ⑤ 부트스트랩 지수가 900 이상이면 유의수준(α) .10에서 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 크다고 판단한다. 표준오차를 이용하는 방법은 다음과 같다. ①번과 ②번은 백분위를 이용하는 방법과 동일하다. ③ 관련자극의 P300 진폭에서 무관련자극의 P300 진폭을 뺀다. ④ ①~③의 과정을 1,000번 반복하여 부트스트랩 표집분포를 형성한다. ⑤ 부트스트랩 표집분포의 평균을 M_B 라고 하고 표준편차를 SD_B 라고 할 때, $M_B - 1.29 \times SD_B$ 가 0보다 크면 유의수준(α) .10에서 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 크다고 판단한다.¹⁾

이러한 부트스트랩 방법에는 한 가지 문제점이 있다. 무관련자극의 부트스트랩 표본크기가 관련자극의 부트스트랩 표본크기보다 훨씬 더 크다는 점이다. 표본크기가 큰 자극에 비하여 표본크기가 작은 자극의 ERP에 고주파 노이즈가 더 많이 포함되므로, 관련자극의 P300 진폭이 과대추정되어 1종 오류가 증가하게 된다(Luck, 2014). P300 CIT에서는 이러한 문제에 대처하기 위하여 4~6 Hz의 저역통과필터를 이용하여 고주파 성분을 제거하며, P300 정점을 측정하는 대신에 P300 정점이 포함된 100 ms 구간의 평균 진폭을 측정한다(Rosenfeld, Shue, & Singer, 2007). 그러나 이러한 과정을 거친 후에도 고주파 노이즈에 의한 영향은 여전히 남아 있다(Jeon, Sohn, & Eom, 2020).

Rosenfeld 등(2008)의 연구에서는 이 문제에 대처하기 위하여 무관련자극의 부트스트랩 표본크기를 관련자극의 부트스트랩 표본크기와 비슷하게(또는 동일하게) 설정하는 방법을 사용하였다. 예를 들면, 부트스트랩 절차에서 무관련자극에 대한 120개의 단일 시행 뇌파에서 30개만 복원추출하

는 것이다. Rosenfeld와 동료들은 대부분의 연구에서 이렇게 수정된 부트스트랩 절차를 사용하고 있으며(Davydova, Rosenfeld, & Labkovsky, 2020; Lu et al., 2018; Olson, Rosenfeld, & Perrault, 2019), 최근 다른 연구자들도 이 방법을 사용하고 있다(Ham, Kim, Jeong, & Kim, 2020; Klein Selle, Gueta, Harpaz, Deouell, & Ben-Shakhar, 2021; Lukács et al., 2016; Song, Kim, Lee, Chang, & Kim, 2018). 그러나 Rosenfeld 등(2008)은 무관련자극의 부트스트랩 표본크기를 축소하는 수정된 부트스트랩 방법의 타당성에 대해서 언급하지 않았다.

부트스트랩 방법은 Efron(1979)에 의해 개발된 것으로, 부트스트랩 표본크기는 원래의 표본크기와 동일해야 한다(Arcones, 2003; Hernández et al., 2017). 원래의 표본크기보다 작은 부트스트랩 표본크기를 사용하면 검증력이 낮아진다(Bickel, Götze, & van Zwet, 2012). 예외적으로 사분위수나 극단값에 대한 부트스트랩, 변량이 무한히 큰 경우의 부트스트랩에서는 원래의 표본크기보다 작은 부트스트랩 표본크기를 사용하는 것이 더 좋은 것으로 보고되기도 한다(Arcones, 2003; Athreya, 1987; Bickel et al., 2012; Shao, 1994).

P300 진폭은 단일 시행 뇌파들을 평균한 ERP로부터 계산하므로 P300 CIT에서 사용하는 부트스트랩 방법은 두 평균의 차이 검증과 같은 분석이다. 따라서 앞서 언급한 예외적 상황에 포함되지 않으므로 무관련자극의 부트스트랩 표본크기를 축소하는 수정된 부트스트랩 방법은 효율성이 낮을 것으로 예상된다. 정규성 가정과 변량의 동질성 가정이 충족된다면 부트스트랩 방법과 t 검증은 매우 유사한 결과를 보이므로(Ahad, Abdullah, & Lai, 2012), t 검증의 표준오차 수식을 활용하여 부트스트랩 표집분포의 표준편차를 추정해볼 수 있다. 무관련자극의 시행수가 관련자극의 4배이고 시행수와 동일한 부트스트랩 표본크기를 사용한다면, 부트스트랩 표집분포의 표준편차를 수식 (1)과 같이 정리할 수 있다. 무관련자극의 부트스트랩 표본크기를 관련자극의 시행수와 동일하게 축소한 경우, 부트스트랩 표집분포의 표준편차는 수식 (2)와 같다. 두 수식을 비교해보면, 원래의 표본크기를 사용한 부트스트랩 방법의 표준오차에 비교하여 부트스트랩 표본크기를 축소한 수정된 부트스트랩 방법의 표준오차가 약 1.265배 더 크다는 것을 알 수 있다. 따라서 수정된 부트스트랩 방법에서 임계치로 1.29를 사용한다면, 실제로는 1.632(1.29×1.265)의 임계치를 사용하는 것과 같은 효과가 나타날 것이다. 자유도가 100인 t 분포에서 임계치 1.632에 해당하는 일방검증의 확률은 .053이므로, 임계치 1.29를 사

1) 자유도가 100인 t 분포에서 1.29보다 큰 면적이 .10 이므로, Rosenfeld와 동료들은 1.29를 임계치로 사용한다.

용한 수정된 부트스트랩 방법의 1종 오류율은 .053이 산출될 것으로 예상된다.

$$\begin{aligned}\sqrt{\frac{s^2}{N_1} + \frac{s^2}{N_2}} &= \sqrt{\frac{s^2}{N_1} + \frac{s^2}{4N_1}} & (1) \\ &= \sqrt{\frac{s^2}{N_1} \left(1 + \frac{1}{4}\right)} \\ &= \sqrt{1.25 \frac{s^2}{N_1}}\end{aligned}$$

$$\sqrt{\frac{s^2}{N_1} + \frac{s^2}{N_1}} = \sqrt{2 \frac{s^2}{N_1}} \quad (2)$$

본 연구는 두 가지 목적을 가진다. 첫 번째 목적은 P300 CIT에서 무관련자극의 부트스트랩 표본크기를 축소하는 수정된 부트스트랩 방법의 1종 오류율이 유의수준보다 낮게 나타나는지 확인하는 것이다. 본 연구의 두 번째 목적은 수정된 부트스트랩 방법의 1종 오류율을 교정하는 방법의 정확도를 확인하는 것이다. 표준오차를 이용하는 방법에서는 수정된 부트스트랩 방법의 표준오차가 부트스트랩 방법의 표준오차보다 1.265배 더 크게 나타날 것으로 추정되므로, 수정된 부트스트랩 방법의 표준오차를 $1/1.265=0.791$ 배 하여 사용하면 유의수준과 동일한 1종 오류율이 나타날 것이다. 백분위를 이용한 방법에서는 다음과 같이 교정할 수 있다. t 값이 1.29일 때 일방검증의 확률이 .10이다. 수정된 부트스트랩 방법에서 과대추정된 표준오차를 수정하기 위하여 임계치 1.29에 0.791을 곱하면, 임계치는 1.020이 된다. 이때 일방검증의 확률은 .154이므로, 부트스트랩 지수 846을 기준으로 판단하면 유의수준과 동일한 1종 오류율이 나타날 것이다.

본 연구에서는 몬테 카를로(Monte Carlo) 연구를 통하여 수정된 부트스트랩 방법의 1종 오류율을 확인하였다. 실험 1에서는 단일 시행 뇌파 모집단을 이용한 몬테 카를로 연구를 수행하였으며, 실험 2에서는 실험 1의 결과 해석을 위하여 숫자를 이용한 몬테 카를로 연구를 수행하였다.

실험 1

실험 1에서는 사인곡선을 합성하여 만들어낸 단일 시행 뇌파를 모집단으로 하는 몬테 카를로 연구를 실시하였다. 이 모집단에 대해서 표본크기가 N_1 과 $4N_1$ 인 두 개의 표본을 추출한 후 일반적인 부트스트랩 방법과 부트스트랩 표본크기를 축소한 수정된 부트스트랩 방법, 이론적인 표준오차를 이

용하여 유의수준을 교정한 부트스트랩 방법을 적용하여 1종 오류율을 확인하였다. 부트스트랩 방법은 백분위를 이용하는 방법과 표준오차를 이용하는 방법 두 가지를 사용하였다. 그리고 원래의 표본크기가 부트스트랩 방법의 1종 오류율에 미치는 영향을 확인하기 위하여 관련자극의 최소 시행수로 생각되는 30과 이것의 두 배인 60을 사용하여, 관련자극과 무관련자극의 표본크기가 30과 120인 조건과 60과 240인 조건에서 몬테 카를로 연구를 수행하였다.

방 법

실험조건

본 연구는 1종 오류율을 확인하는 것이 목적이므로, 관련자극에 대한 모집단과 무관련자극에 대한 모집단을 동일하게 정의하였다. 분석방법으로 전형적인 부트스트랩 방법과 부트스트랩 표본크기를 축소한 수정된 부트스트랩 방법, 수정된 부트스트랩 방법의 편향을 교정한 방법을 사용하였으며, 표본크기로 30과 120인 조건과 60과 240인 조건을 설정하였다.

모집단 정의

단일 시행 뇌파의 모집단은 선행연구의 절차에 따라서 사인곡선을 이용하여 정의하였다(Yeung, Bogacz, Holroyd, & Cohen, 2004; Zoumpoulaki, Alsufyani, & Bowman, 2015). 뇌파는 다양한 주파수를 가지는 사인곡선의 합으로 표현될 수 있으므로(Luck, 2014), 다양한 주파수의 사인곡선을 합하여 단일 시행 뇌파의 모집단을 구성하였다. 단일 시행 뇌파는 배경뇌파에 순수한 P300 파형이 더해진 것이므로, 먼저 순수한 P300 파형을 정의하였다. Figure 1의 왼쪽 그림과 같이 사인곡선을 이용하여 순수한 P300 파형을 정의하였다. 선행연구의 무죄조건에서 관련자극과 무관련자극의 P300 진폭은 $5 \mu V$ 내외였으므로(Rosenfeld et al., 2008) 순수한 P300 정점-정점의 진폭을 $5 \mu V$ 로 하였다. 또한 P300 정적 파형의 폭과 P300 이후 부적 파형의 폭은 각각 약 200~300 ms였으므로(Eom, Jeon, & Sohn, 2018), 2 Hz 사인곡선의 정적 파형 250 ms와 2 Hz 사인곡선의 부적 파형 250 ms를 더하여 총 500 ms 길이의 순수한 P300 파형을 만들었다. 선행연구에서 뇌파를 측정할 때 0.3 Hz 고역통과필터를 사용하고 뇌파를 분석할 때 4.23 Hz 저역통과필터를 사용하여 뇌파에서 고주파 성분을 제거하였다(Rosenfeld et al., 2008). 따라서 본 연구에서도 주파수가 0.3 Hz, 1.3 Hz, 2.3 Hz, 3.3 Hz, 4.3 Hz이고 무선적 위상(phase)을 가지는 사인곡선

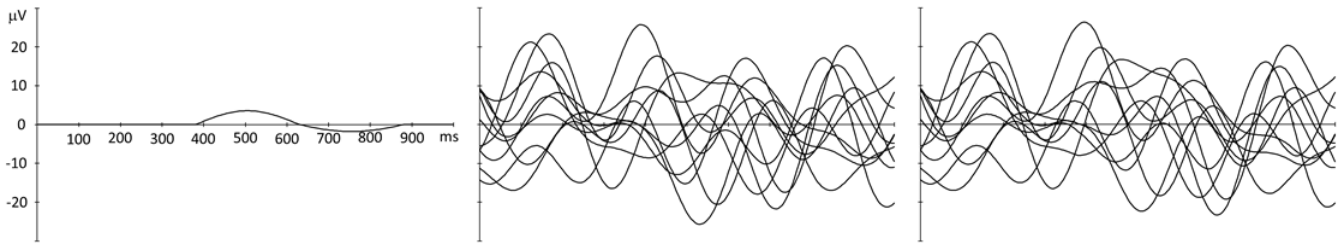


Figure 1. The P300 wave (left), background EEGs with an length of 1000 ms (middle), and single sweeps (right). The peak to peak P300 amplitude is 5.0 μV

5개를 평균한 값에 30을 곱하여 진폭 범위가 $-30 \mu V$ 에서 $+30 \mu V$ 이고 길이가 1초인 배경뇌파를 구성하였다. Figure 1의 가운데 그림에 10개의 배경뇌파가 제시되어 있다. 마지막으로 P300 CIT에서 P300 정점의 잠재시간이 500 ms 전후이므로(Jeon et al., 2020), 배경뇌파의 380 ms ~ 880 ms 지점에 순수한 P300 파형을 더하여 Figure 1의 오른쪽 그림과 같이 단일 시행 뇌파를 구성하였다. 10,000개의 단일시행 뇌파를 만들어 P300 진폭을 측정하고, 모집단의 P300 진폭인 $5.00 \mu V$ 와 유사한 $5.05 \mu V$ 가 산출되었다. P300 진폭이 나타난 위치에서 단일시행 뇌파의 P300 진폭을 측정하고, 표준편차 11.18, 왜도 -0.01 , 첨도 -0.55 였다. 왜도와 첨도가 0에 가까워 단일 시행 P300 진폭의 분포는 정규분포와 비슷한 것으로 나타났다.

여러 개의 사인곡선의 합으로 만들어진 단일 시행 뇌파가 실제 단일 시행 뇌파와 유사한지 평가하였다. 뇌파 관련 수업을 이수하였거나 뇌파연구를 수행한 경험이 있는 5명의 석/박사들이 평가에 참가하였다(여자 4명, 평균연령 30.2세). 평가자들에게 사인곡선으로 합성한 단일 시행 뇌파 100개와 실제 실험에서 측정된 단일 시행 뇌파 100개를 섞어서 제시한 후 실제 뇌파를 선택하도록 하였다. 평가 결과, 평균 정답율은 $48\%(SD=18\%)$ 로 우연수준과 매우 유사하였다. 즉, 사인곡선으로 합성한 단일 시행 뇌파와 실제 단일 시행 뇌파는 형태상 매우 유사하였다.

무선표본추출

단일 시행 뇌파 모집단에서 표본크기 30(또는 60)의 표본을 무선 추출하여 관련자극의 단일 시행 뇌파들을 구성하였으며, 동일한 모집단에서 표본크기 120(또는 240)의 표본을 무선 추출하여 무관련자극의 단일 시행 뇌파들을 구성하였다.

분석방법

표본 추출한 단일 시행 뇌파 자료에 대해서 일반적 부트스트랩 방법과 수정된 부트스트랩 방법, 교정된 부트스트랩 방법

으로 P300 진폭 차이 검증을 실시하였다. 각 부트스트랩 방법은 다시 백분위를 이용하는 방법과 표준오차를 이용하는 방법 두 가지로 나누어 적용하였다. 단일 시행 뇌파 자료에 대하여 백분위를 이용한 부트스트랩 방법은 다음과 같다. 관련자극에 대한 30개(또는 60개)의 단일 시행 뇌파에서 30개(또는 60개)를 복원추출한다. 무관련자극에 대한 120개(또는 240개)의 단일 시행 뇌파에서 120개(240개)를 복원추출한다. 각 자극별로 단일 시행 뇌파들을 평균하여 ERP를 산출하고, ERP에서 P300 진폭을 계산한다. P300 진폭은 선행연구와 동일하게 정점-정점 방법으로 측정하였다. 구체적으로 P300 진폭은 300 ms에서 700 ms까지 100 ms 구간의 평균전위가 가장 큰 값에서 정점 이후부터 1,000 ms까지 100 ms 구간의 평균전위가 가장 낮은 값을 뺀 값으로 정의하였다. 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 더 크면 부트스트랩 지수에 1을 추가한다. 이 과정을 1,000번 반복한 후 부트스트랩 지수가 900 이상이면, 유의수준 .10에서 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 크다고 판단한다. 표준오차를 이용한 방법은 부트스트랩 표본을 추출하고 P300 진폭을 계산한 후 다음과 같이 진행한다. 관련자극의 P300 진폭에서 무관련자극의 P300 진폭을 뺀 차이값을 계산한다. 이 과정을 1,000번 반복하여 P300 진폭 차이값의 평균과 표준편차를 계산한다. 평균- $1.29 \times$ 표준오차가 0보다 크면, 유의수준 .10에서 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 크다고 판단한다. 수정된 부트스트랩 방법은 120개(또는 240개)의 무관련자극에 대한 단일 시행 뇌파에서 30개(또는 60개)만 복원추출한다는 점을 제외하고는 부트스트랩 방법과 동일하다. 교정된 부트스트랩 방법은 표준오차에 0.79를 곱하여 교정한 표준오차를 사용하는 것과 판단 기준으로 부트스트랩 지수 900을 사용하는 대신에 부트스트랩 지수 846을 사용하는 것이었다. 각 모집단에서 표본을 추출하고 부트스트랩 방법을 적용하는 과정을 100,000번 반복한 후 각 부트스트랩 방법의 1종 오류율을 계산하였다.

결과 및 논의

각 실험조건 별 1종 오류율이 Table 1에 제시되어 있다. 전반적으로 백분위를 이용하는 방법과 표준오차를 이용하는 방법 간에 1종 오류율에서 큰 차이는 나타나지 않았다. 관련자극과 무관련자극의 시행수가 각각 30과 120일 때 부트스트랩 방법의 1종 오류율은 약 .31로 유의수준보다 높았다. 이러한 결과는 선행연구와 매우 유사한 수치로 시행수 차이에 따른 노이즈의 차이에 의한 것이다(Jeon et al., 2020). 시행수가 60과 240일 때 부트스트랩 방법의 1종 오류율은 약 .27로 시행수가 작은 경우와 비교하여 다소 감소하였지만 여전히 매우 높은 1종 오류율을 보였다. 뇌파에 5 Hz 이상의 고주파 성분이 포함되어 있지 않고 100 ms 평균전위를 사용하였음에도 불구하고, 관련자극의 P300 진폭이 과대추정되었다는 것은 5 Hz 이하의 저주파 성분도 시행수에 따라 P300 진폭에 다르게 영향을 미친다는 것을 의미한다. 따라서 시행수가 서로 다른 자극의 P300 진폭을 비교하기 위해서는 새로운 접근법이 필요할 것이다. 한편, 부트스트랩 방법의 1종 오류율이 선행연구와 유사하게 나타났다는 것은 사인곡선을 이용하여 뇌파를 생성하는 방법의 타당성을 보여주는 것이기도 하다.

무관련자극의 부트스트랩 표본크기를 관련자극의 시행수와 동일하게 축소한 수정된 부트스트랩 방법의 1종 오류율은 실험조건에 따라 .072~.074를 보였다. 이러한 수치는 서론에서 이론적으로 예측한 .053보다 약 .02 더 높은 값이며, 유의수준 .10보다는 약 .03 낮은 값이다. 즉, 서론에서 예측한 바와 같이 부트스트랩 표본크기를 축소한 경우에 유의수준보다 낮은 1종 오류율을 보였다. 1종 오류율이 낮으면 검증력도 낮아지므로, 수정된 부트스트랩 방법의 검증력은 이상적인 분석방법보다 낮을 것이다. 수정된 부트스트랩 방법의 오류를 교정한 부트스트랩 방법의 1종 오류율은 실험조건에 따라 .138~.142로 나타나, 유의수준 .10보다 더 높았다. 즉, 수정된 부트스트랩 방법의 오류가 적절하게 교정되지 않았다. 수정된 부트스트랩 방법의 1종 오류율이 예상보다 높게 나온 이유와 교정된 부트스트랩 방법의 1종 오류율

이 유의수준보다 높은 이유는 동일할 것이다.

수정된 부트스트랩 방법의 1종 오류율이 예상보다 더 높게 나타난 이유를 추정해보면 두 가지를 꼽을 수 있다. 한 가지 가능성은 예측이 잘못된 경우이다. 본 연구에서는 두 평균 차이의 표준오차 공식을 이용하여 수정된 부트스트랩 방법의 1종 오류율을 예측하였다. 만약, 이론적인 표준오차와 부트스트랩 분포의 표준오차가 다르다면 수정된 부트스트랩 방법의 1종 오류율이 예상과 다를 수 있다. 다른 한 가지는 뇌파 자료의 독특한 특성 때문일 가능성이 있다. 수정된 부트스트랩 방법의 1종 오류율이 예상보다 높게 나타난 원인이 부트스트랩 방법 자체의 문제인지 또는 뇌파자료의 독특한 특성 때문인지 평가하기 위하여 실험 2를 실시하였다.

실험 2

실험 1에서 무관련자극의 부트스트랩 표본크기를 축소한 수정된 부트스트랩 방법의 1종 오류율이 예상보다 높게 나온 것이 표준오차를 이용한 예측의 문제인지 확인하기 위하여 숫자를 모집단으로 하는 몬테 카를로 연구를 수행하였다. 단일 시행 뇌파에서 P300 진폭의 분포는 정규분포와 유사하므로(Eom et al., 2018), 숫자 모집단도 정규분포를 이용하여 정의하였다. ERP 연구에서 특정 값을 넘어서는 시행(예, $\pm 50 \mu V$)은 분석에서 제외하므로, 숫자 모집단에서도 절단된 정규분포를 사용하였다.

방법

실험조건

실험 1과 동일하였다.

모집단 정의

모집단은 평균 0과 표준편차 1인 정규분포에서 -3부터 +3까지의 값들로 정의하였다.

Table 1. Type 1 error rates of experiment 1 ($\alpha=.10$)

N1, N2	Percentile			Standard error		
	Bootstrap method	Modified bootstrap method	Adjusted bootstrap method	Bootstrap method	Modified bootstrap method	Adjusted bootstrap method
30, 120	.316	.074	.142	.308	.074	.142
60, 240	.276	.073	.138	.270	.072	.139

무선표본추출

실험 1과 동일하였다.

분석방법

P300 진폭 대신에 표본의 평균을 계산한다는 점을 제외하고는 실험 1과 동일하였다.

결과 및 논의

각 실험조건 별 1종 오류율이 Table 2에 제시되어 있다. 모든 실험조건에서 백분위를 이용하는 방법과 표준오차를 이용하는 방법의 1종 오류율은 매우 비슷하였다. 부트스트랩 방법의 1종 오류율은 실험조건에 따라 .101~.104로 나타나 유의수준 .10과 매우 유사하였다. 이러한 결과는 다양한 분포 조건에서 부트스트랩 방법의 1종 오류율이 유의수준과 유사하다는 선행연구들과 일치하는 결과로(Dwivedi, Mallawaarachchi, & Alvarado, 2017; Konietschke, & Pauly, 2014), 일반적인 평균차이 검증에서 부트스트랩 방법이 1종 오류율을 적절하게 통제한다는 것을 의미한다. 집단 2의 부트스트랩 표본크기를 축소한 수정된 부트스트랩 방법의 1종 오류율은 실험조건에 따라 .053~.055로 나타나 서론에서 예측한 .053과 매우 유사하였다. 그리고 이러한 예측에 근거하여 교정된 부트스트랩 방법의 1종 오류율은 실험조건에 따라 .099~.101로 유의수준과 거의 일치하였다. 결과적으로 실험 1에서 단일 시행 뇌파자료에 대한 수정된 부트스트랩 방법의 1종 오류율이 예상보다 높게 나타난 이유는 뇌파자료의 독특한 특성 때문일 가능성이 높았다.

종합논의

실험 결과, P300 CIT에 대한 부트스트랩 방법에서 원래의 표본크기보다 작은 부트스트랩 표본크기를 사용하면 유의수준 .10보다 낮은 .073의 1종 오류율을 보였다. 1종 오류율이 유의수준보다 낮으면 검증력이 낮아진다는 문제가 생긴다. 즉 유죄인 조사대상자가 무죄 판정받을 가능성이 증가한다.

보통 모수검증에서 표본크기가 30 이상이면 모집단 분포의 모양에 관계없이 정규성 가정이 충족된 것으로 판단한다(Ross, 2010). 저자의 미발간 컴퓨터 모의실험에 의하면, 모집단의 왜도와 첨도가 각각 2.0과 6.0, 표본크기가 30, 유의수준 .05일 때 1종 오류율의 편파는 약 .01 이내인 것으로 나타났다. 반복측정 변량분석에서는 구형성 가정(sphericity assumption)을 중요한 가정으로 생각하며, 구형성 가정이 파기될 경우에 편파를 교정하는 절차를 거친다(Park, 2003). 구형성 가정이 파기될 경우 1종 오류율의 편파는 조건에 따라 .01~.03으로 나타난다(Haverkamp & Beauducel, 2017). 통계분석에서 현재의 관행을 고려하면 1종 오류율에서 .01 이내의 편파는 수용하는 것으로 생각하며 .01 이상의 편파는 교정해야할 정도의 큰 편파로 인식하는 것 같다. 무관련자극의 부트스트랩 표본크기를 축소한 수정된 부트스트랩의 1종 오류율은 유의수준과 .02~.03의 편파를 보이므로 교정해야할 정도의 큰 오류로 생각된다.

수정된 부트스트랩 방법의 1종 오류가 유의수준보다 작게 나타나는 편파를 교정하기 위하여 이론적인 표준오차를 이용하여 유의수준을 교정하였다. 교정된 부트스트랩 방법을 적용한 결과에서 유의수준보다 .04 높은 1종 오류가 산출되어 편파가 교정되지 않았다. 숫자 모집단을 대상으로 한 실험 2에서 교정된 부트스트랩 방법의 1종 오류가 유의수준과 동일하였다는 것을 고려하면, P300 CIT에서 수정된 부트스트랩 방법의 1종 오류가 교정되지 않은 이유는 뇌파자료의 특이성 때문인 것으로 생각된다.

숫자자료에서는 모집단에서 표본 추출한 숫자들을 분석한다. 그러나 뇌파자료에서는 모집단에서 단일 시행 뇌파들을 표본추출한 후 ERP에서 P300 진폭을 계산하여 분석한다. 이때 ERP에서 100 ms 구간의 평균전위가 가장 높은 값을 P300 진폭으로 사용하므로 표본마다 P300 진폭을 계산하는 구간이 달라지며, 부트스트랩 표본에서도 P300 진폭을 계산하는 구간이 달라진다. 즉, 표본마다 P300 진폭을 계산하는 구간이 달라진다는 점이 뇌파자료의 특이성이다. 현재로서는 뇌파자료의 이러한 특성이 어떠한 방식으로 부트스트랩의 결과에 영향을 미치는지는 파악하기 어렵다. 추후 연구를 통해

Table 2. Type 1 error rates of experiment 2 ($\alpha=.10$)

N1, N2	Percentile			Standard error		
	Bootstra method	Modified bootstrap method	Adjusted bootstrap method	Bootstrap method	Modified bootstrap method	Adjusted bootstrap method
30, 120	.104	.055	.101	.103	.054	.101
60, 240	.101	.053	.099	.101	.053	.099

여 P300 진폭을 측정하는 방법이 부트스트랩의 결과에 영향을 미치는 방식에 대한 이해가 필요할 것이다.

P300 CIT의 최종 판단을 위하여 부트스트랩 방법이 널리 사용된다. 그러나 일반적인 부트스트랩 방법에서 비교하려는 두 자극의 시행수 차이로 인하여 1종 오류가 상승하며, 무관련자극의 부트스트랩 표본크기를 축소한 수정된 부트스트랩 방법에서는 1종 오류가 유의수준보다 작아지는 편파가 발생한다. 그리고 이러한 편파는 교정되기 어려운 것으로 나타났다. 따라서 1종 오류가 편파되지 않는 분석방법이 필요하다. Jeon 등(2020)은 P300 CIT에서 순열검증을 사용하면 관련자극과 무관련자극 간의 시행수 차이로 인한 1종 오류의 편파가 발생하지 않는다는 것을 보였다. 부트스트랩 방법에서는 관련자극과 무관련자극 간의 P300 진폭차이에 대한 표집 분포를 형성하고 이 분포로부터 90% 신뢰구간을 산출하므로, 시행수 차이로 인한 편파가 신뢰구간에 영향을 미치게 된다. 그러나 순열검증에서는 영가설이 참일 때 관련자극과 무관련자극 간의 P300 진폭차이에 대한 표집분포를 형성하고, 관찰된 관련자극과 무관련자극 간 P300 진폭차이가 표집분포에서 차지하는 위치를 평가한다. 시행수 차이로 인한 편파가 반영된 표집분포에서 시행수 차이로 인한 편파가 존재하는 관찰된 P300 진폭차이의 위치를 평가하므로, 시행수 차이로 인한 편파가 없는 확률계산이 가능하다. 따라서 P300 CIT에서 여러 가지 부트스트랩 방법을 사용하는 대신 순열검증을 사용하여 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 더 크지를 평가한다면, 통계적으로 편파되지 않는 결과를 얻을 수 있을 것이다.

References

- Ahad, N. A., Abdullah, S., & Lai, C. H. (2012). Relative power performance of t-test and bootstrap procedure for two-sample. *Pertanika Journal of Science & Technology*, 20, 43-52.
- Arcones, M. A. (2003). On the asymptotic accuracy of the bootstrap under arbitrary resampling size. *Annals of the Institute of Statistical Mathematics*, 55, 563-583.
- Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *The Annals of Statistics*, 15, 724-731.
- Bickel, P. J., Götze, F., & van Zwet, W. R. (2012). Resampling fewer than n observations: gains, losses, and remedies for losses. In *Selected works of Willem van Zwet* (pp. 267-297). Springer, New York, NY.
- Cutmore, T. R., Djakovic, T., Kebbell, M. R., & Shum, D. H. (2009). An object cue is more effective than a word in ERP-based detection of deception. *International Journal of Psychophysiology*, 71, 185-192.
- Davydova, E., Rosenfeld, J. P., & Labkovsky, E. (2020). Necessity of the target discrimination in the P300 based complex trial protocol test for concealed information. *Psychophysiology*, 57, e13548.
- Dwivedi, A. K., Mallawaarachchi, I., & Alvarado, L. A. (2017). Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. *Statistics in medicine*, 36, 2187-2205.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1-26.
- Eom, J. S., Jeon, H., & Sohn, J. H. (2018). The false positive rate of P300-based concealed information test. *The Korean Journal of Cognitive and biological Psychology*, 30, 241-259.
- Ham, K. Kim, K., Jeong, H., & Kim, Y. (2020). Application of ERP-based memory identification using bootstrap technique for eyewitness: Comparison of recognition protocol and concealed information task protocol. *The Korean Journal of Cognitive and biological Psychology*, 32, 125-136.
- Haverkamp, N., & Beauducel, A. (2017). Violation of the sphericity assumption and its effect on type-I error rates in repeated measures ANOVA and multi-level linear models (MLM). *Frontiers in psychology*, 8, 1841.
- Hernández, W., Calderón-Córdova, C., González-Posada, V., Parra-Cerrada, Á., Jiménez, J. L., González-García, J. E., ... & Sergiyenko, O. Y. (2017). Bootstrap-based frequency estimation method. *Measurement*, 95, 193-200.
- Jeon, H., Sohn, J. H., & Eom, J. S. (2020). The accuracy of permutation test in P300-based concealed information test. *The Korean Journal of Cognitive and biological Psychology*, 32, 69-83.
- klein Selle, N., Gueta, C., Harpaz, Y., Deouell, L. Y., & Ben-Shakhar, G. (2021). Brain-based concealed memory detection is driven mainly by orientation to salient items. *Cortex*, 136, 41-55.
- Konietschke, F., & Pauly, M. (2014). Bootstrapping and permuting paired t-test type statistics. *Statistics and Computing*, 24, 283-296.
- Lu, Y., Rosenfeld, J. P., Deng, X., Zhang, E., Zheng, H., Yan, G., ... & Hayat, S. Z. (2018). Inferior detection

- of information from collaborative versus individual crimes based on a P300 Concealed Information Test. *Psychophysiology*, 55, e13021.
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Lukács, G., Weiss, B., Dalos, V. D., Kilencz, T., Tudja, S., & Csifcsák, G. (2016). The first independent study on the complex trial protocol version of the P300-based concealed information test: Corroboration of previous findings and highlights on vulnerabilities. *International Journal of Psychophysiology*, 110, 56-65.
- Olson, J., Rosenfeld, J. P., & Perrault, E. (2019). Deleterious effects of probe related versus irrelevant targets on the “CIT effect” in the P300 and RT based three stimulus protocol for detection of concealed information. *Psychophysiology*, 56, e13459.
- Park, K. (2003). *Analysis of variance and regression analysis*. Hakjisa.
- Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 63-89). Cambridge, UK: Cambridge University Press.
- Rosenfeld, J. P., Labkovsky, E., Winogard, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The complex trial protocol(CTP): A new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology*, 45, 906-919.
- Rosenfeld, J. P., Shue, E., & Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. *Biological psychology*, 74, 396-404.
- Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300 based tests of detection of concealed information. *Psychophysiology*, 41, 205-219.
- Ross, S. (2010). *Introductory Statistics*. Elsevier Science.
- Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society*, 122, 1251-1262.
- Song, I., Kim, H., Lee, K. E., Chang, E., & Kim, H. T. (2018). The comparison of Bootstrapping analyses in P300-CIT. *The Korean Journal of Forensic Psychology*, 9, 75-99.
- Yeung, N., Bogacz, R., Holroyd, C. B., & Cohen, J. D. (2004). Detection of synchronized oscillations in the electroencephalogram: an evaluation of methods. *Psychophysiology*, 41(6), 822-832.
- Zoumpoulaki, A., Alsufyani, A., & Bowman, H. (2015). Resampling the peak, some dos and don'ts. *Psychophysiology*, 52, 444-448.

P300 숨긴정보검사에 사용되는 부트스트랩 방법의 표본 크기

엄진섭¹, 전하정²

¹충북대학교, ²충남대학교

P300 숨긴정보검사에서 관련자극에 대한 P300 진폭이 무관련자극에 대한 P300 진폭보다 더 크지를 평가한다. 그런데 무관련자극의 시행수가 관련자극의 시행수보다 훨씬 더 크기 때문에 관련자극에 대한 P300 진폭이 과대추정된다는 문제점이 있다. Rosenfeld 등(2008)은 이 문제에 대처하기 위하여 무관련자극의 부트스트랩 표본크기를 관련자극의 표본크기로 축소하여 사용하였다. 일반적으로 부트스트랩 표본크기는 원래의 표본크기와 동일해야만 하며, 부트스트랩 표본크기가 원래의 표본크기보다 작으면 1종 오류율이 유의수준보다 작아지는 문제가 발생한다. 본 연구의 목적은 몬테카를로 연구를 통하여 무관련자극의 부트스트랩 표본크기를 축소하는 수정된 부트스트랩 방법의 1종 오류율을 평가하고, 이러한 오류가 교정될 수 있는지 확인하는 것이다. 실험 1의 결과, 수정된 부트스트랩 방법의 1종 오류율은 약 .073으로 유의수준 .10보다 낮았다. 표준오차를 이용하여 유의수준을 교정한 부트스트랩 방법의 1종 오류율은 약 .140으로 유의수준 .10보다 더 높게 나타나, 수정된 부트스트랩 방법의 오류가 교정되지 않았다. 실험 2에서 수정된 부트스트랩 방법의 오류가 교정되지 않는 이유를 평가하기 위하여 숫자를 이용한 몬테카를로 연구를 수행하였다. 연구결과, 수정된 부트스트랩 방법의 1종 오류율은 약 .054로 유의수준 .10보다 작았으며, 교정된 부트스트랩 방법의 1종 오류율은 약 .10으로 유의수준과 동일하였다. 따라서 수정된 부트스트랩 방법의 오류가 교정되지 않는 이유는 뇌파자료의 특수성 때문인 것으로 나타났다. 이러한 오류를 극복하는 방법에 대해서 논의하였다.

주제어: P300, 숨긴정보검사, 부트스트랩, 표본 크기