

G Power 소프트웨어를 이용한 반복측정디자인의 표본수 산출

중앙대학교 의과대학 마취통증의학과

강 현

Sample size determination for repeated measures design using G Power software

Hyun Kang

Department of Anesthesiology and Pain Medicine, Chung-Ang University College of Medicine, Seoul, Korea

Repeated measures designs are widely used in the field of anesthesiology because they allow the detection of within-person change over time and provide a higher statistical power for detecting differences than a single measure design while reducing the costs and efforts to conduct a study. However, the complex process of calculating the sample size for repeated measures design requires profound statistical knowledge and also programming skills in some instances. In the present article, the author describes 1) the basic statistics for repeated measures design, 2) the explanation for G Power software, and 3) how to calculate the sample size using an example. (*Anesth Pain Med* 2015; 10: 6-15)

Key Words: G Power, Power calculation, Repeated measures design, Sample size determination, Sphericity.

서 론

모든 임상 연구자들은 임상적으로 의미 있는 연구를 올바르게 시행하여 연구 결과가 임상에 적절히 적용되어 환자와 의료진에게 조금이라도 도움이 되고자 노력한다. 올바른 임상연구를 시행하기 위하여, 우리는 연구의 가설을 설정하고, 계획을 세우며, 대상을 설정하여 연구를 진행한

Received: October 8, 2014.
Accepted: December 19, 2014.

Corresponding author: Hyun Kang, M.D., Ph.D., Department of Anesthesiology and Pain Medicine, Chung-Ang University College of Medicine, 102, Heukseok-ro, Dongjak-gu, Seoul 156-755, Korea. Tel: 82-2-6299-2586, Fax: 82-2-6299-2575, E-mail: roman00@naver.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

다. 연구대상 전체(모집단)를 대상으로 연구하면, 가설에 대한 진실을 얻을 수 있지만, 연구대상 전체를 연구하기는 어려우며, 실제로 모집단 전체를 조사하는 것은 대부분의 경우 불가능하며 또한 비효율적이다. 심지어는 모집단 전체를 조사하는 경우가 표본을 조사하는 경우보다 부정확할 경우도 있다. 그렇기 때문에 연구 대상 전체를 대표할 수 있는 표본을 뽑아, 표본의 자료를 분석하여 모집단의 정보(모수)를 추측하는 방법을 사용하게 된다.

이 때 연구 가설 및 질문에 답할 수 있는 적절한 표본의 수를 결정하는 것은 대단히 중요한 문제가 된다. 많은 경우에 연구자는 자신의 임의로, 아니면 자원이 허락하는 범위 내에서만 표본수를 결정하여 연구를 진행하곤 한다. 이는 자체로 논리적이지 못하기도 하지만, 윤리적이지 못하기도 하다.

필요한 수 이상으로 표본수를 증가시키는 것은 시간, 돈, 그리고 제한된 자원의 낭비를 초래하며, 연구대상자를 필요 이상의 위험에 노출 시킬 수 있다. 또한 연구자가 연구에서 보고자 하는 평가 변수뿐만 아니라 다른 변수까지 통계적으로 유의하게 만들어, 우리가 연구에서 진실로 유효성을 평가 하고자 하는 변수가 무엇인지 밝혀내기 어려운 일이 발생할 수 있다. 반대로 표본수가 충분하지 않으면 연구 결과의 검정력이 충분하게 커지지 않아 의미 있는 차이를 발견하기 어려우며, 적절한 표본수를 가진 연구에서 시행되었다면 유의성이 확보되어 의료현장에 적용될 수 있었던 결과가 유의성이 확보되지 않아 의료현장에 적용되지 않는 오류를 범할 수 있다. 연구자들이 연구에서 얻어내고자 하였던 질문에 대한 결과 또한 신뢰성이 떨어지게 된다. 따라서 연구를 진행하는데 있어 표본 수를 적절하게 산정하는 것은 연구가 비용 효과적으로 그리고 윤리적으로 시행되는데 있어 필수적인 요소이며, 연구 시행에서 보고자 했던 효과를 관찰할 수 있는 기회를 늘릴 수 있다[1,2].

마취통증의학과 의사들이 시행하는 연구에는 자료들이 서로 연관되어 있으며, 반복적으로 측정된 자료(repeated measures data)를 대상으로 시행한 경우를 흔히 볼 수 있다. 이러한 자료에는 1) 동일한 개체(subject)를 대상으로 시간의 진행에 따라 여러 번 측정을 하는 연구, 2) 동일한 개체에

서 각각 다른 치료를 받게 한 후 측정하는 연구, 3) 동일한 개체에서 몸의 다른 부분에서 측정한 연구, 4) 각각의 사례군(case)와 대조군(control)에서 개체를 개별적으로 짝짓기한 연구(individual matched study) 등이 포함된다. 위의 예들은 연구계획들에 있어 서로 본질적으로 다르지만, 모두 가지고 있는 공통점은 자료들간에 서로 연관(correlated)되어 블록(block)을 이루고 있다는 점이다. 블록은 시간의 진행에 따라 여러 번 측정된 경우, 여러 번 다른 치료를 시행한 경우, 몸의 다른 부분에서 관찰한 경우에는 개체(subject)내에서, 개별적으로 짝지어진 연구에서는 사례군과 대조군에서의 개체간에 블록을 이루고 있다.

측정값이 블록을 이루고 있을 경우 자료들은 블록 간보다는 블록 내에서 좀 더 비슷한 양상을 보이게 된다. 일반적으로 결과가 한 번 발생한 블록 내에서는 같은 결과가 다시 발생할 가능성이 높으며, 블록내의 측정된 자료들은 일정한 양상을 가지고 있을 가능성이 크다. 예를 들어 수술 후 구토가 발생한 환자에서 시간이 진행되더라도 다시 구토가 발생할 가능성이 높으며, 고혈압환자에서 측정된 혈압은 지속적으로 높은 혈압을 가질 가능성이 크다. 여기에서 동일한 개체는 대상자간에 비해 비슷한 반응을 나타낼 가능성이 높으므로, 반응은 연관(correlated) 되어있다고 할 수 있다.

이와 같은 연관되어 블록을 이루고 있는 자료를 수집하는 데에는 다음과 같은 중요한 장점들이 있다. 첫째는 동일한 개체에서 반복되어 측정된 자료를 모으면 대상자의 수를 늘리지 않고, 즉 표본의 수를 늘리지 않고, 연구의 검정력을 높일 수 있다. 이는 연구에 사용되는 비용, 자원 및 노력의 소모를 줄일 수 있게 한다.

둘째로 주요 변수를 시간의 흐름에 따라 반복해서 수집하는 것은 개체의 시간에 따른 변화를 좀더 명확하게 평가하게 할 수 있다. 셋째로 자료들이 연관되어 있으므로 그 자체로 혼란변수(confounding factor)의 영향을 줄일 수 있다. 이렇듯 반복되어 측정된 자료를 이용한 연구는 결과를 한 번만 측정하는 연구보다 많은 이점을 가지고 있지만, 통계나 프로그래밍에 대한 상당한 지식을 요구하며, 때로는 분석에 필요한 통계방법에 따른 표본 수 산정 방법을 찾을 수 없어, 통계를 전공하지 않은 연구자들에게 있어 표본 수의 산출은 아주 어려운 일이 될 수 있다. 또한 독립적인 측정을 시행한 연구들과는 달리 동일한 개체에서 얻어진 반복 측정된 자료들 간에는 상관관계를 가지며, 이러한 상관관계는 표본수를 산정하는데 있어 반드시 고려되어야만 한다[3,4].

본 문헌에서는 첫째, 반복 측정 자료 분석에 있어 표본수 산출을 위해 필요한 기본적인 통계에 대한 설명 둘째, G Power 프로그램에 대한 설명, 마지막으로 G Power를 이용한 반복 측정 자료를 위한 표본수 산정을 예를 들어 설명

하기로 한다.

반복측정자료(repeated measures data)의 분석을 위한 통계

실험의 대상이 되는 단위를 개체(subject)라 부르고, 동일한 개체에 대하여 2회 이상 반복해서 측정이 이루어지는 자료를 반복측정자료(repeated measures data)라 한다. 반복측정을 시행한 연구에서 측정된 자료에 영향을 미치는 원인을 반복요인이라 부른다. 반복요인으로는 시간, 용량, 농도, 온도 등이 있을 수 있다.

두 가지 실험군의 효과를 비교하는 연구가 있다고 가정하자. 임의로 집단을 나누어 한 집단은 대조군(control group)으로 약을 복용하지 않고, 다른 집단은 실험군(experimental group)으로 약을 복용한 후 통증을 측정하였다고 하자. 이 경우 두 집단을 구성하는 개체들은 물론 다르다. 이 자료를 분석할 경우, 이 표본 t검정(unpaired t-test)가 적당하다고 할 수 있다.

두 가지 실험군의 효과를 비교하는 다른 연구를 생각해 보도록 하자. 개체들 간의 통증에는 서로 차이가 있을 수 있으므로 개체 내에서 실험군의 복용 전, 후의 통증의 차이를 보는 것이 더 효율적이라 생각할 수 있다. 즉 개체간에는 통증의 차이가 존재하므로, 각 개체를 고정하여 연구를 진행할 경우 개체 자체가 블록의 효과를 갖기 때문에 자료들은 연관되어 있다고 할 수 있다. 이 경우 약을 복용하지 않은 경우와 약을 복용한 경우는 동일한 개체에서 측정되었으므로 독립일 수 없어 대응 t검정(paired t-test)로 분석하는 것이 적절하다. 반복이 들인 경우는 대응 t검정으로 분석할 수 있으나, 반복수준의 수가 둘을 넘어서는 경우는 반복측정분산분석을 시행하게 된다. 이것은 짝 지워 지지 않은 자료에서 군이 2개일 경우 이 표본 t검정을 사용하고 군이 둘을 넘어서면 분산분석(analysis of variance)을 시행하는 것과 같은 원리이다.

반복측정자료분산분석에서는 처리들간의 비교에 있어 개체내 변동을 분리하여 분석하게 되어 각각의 개체가 대조 그룹(블록)의 역할을 함으로 비교성이 높아지는 장점을 지닌다. 즉 개체간의 변이를 줄이기 위해 스스로가 블록이 되는 것이다. 따라서 반복측정에 대한 연구를 계획할 때 분산 및 반복측정된 자료들 간의 상관관계양상을 파악하여 지정하는 것이 대단히 중요하다. 이를 제대로 파악하여 지정하지 못하면 표본수 및 검정력 계산뿐 아니라 결과의 분석까지 정확하지 않게 된다.

상관관계의 양상은 복잡도(complexity)의 정도에 따라 1) 상관관계가 없음(독립적 관찰), 2) 상관관계가 동일함, 3) 상관관계 내에 일정한 법칙이 있음 4) 일정한 양상이 없음 등의 4가지 형태로 분류해 볼 수 있다(Table 1). 독립적 관찰

Table 1. Patterns for Correlation

1) None	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
2) Constant	$\begin{pmatrix} 1 & 0.6 & 0.6 & 0.6 \\ 0.6 & 1 & 0.6 & 0.6 \\ 0.6 & 0.6 & 1 & 0.6 \\ 0.6 & 0.6 & 0.6 & 1 \end{pmatrix}$
3) 1 st Order Correlation	$\begin{pmatrix} 1 & 0.6 & 0.36 & 0.216 \\ 0.6 & 1 & 0.6 & 0.36 \\ 0.36 & 0.6 & 1 & 0.6 \\ 0.216 & 0.36 & 0.6 & 1 \end{pmatrix}$
4) No pattern	$\begin{pmatrix} 1 & 0.6 & 0 & 0.4 \\ 0.6 & 1 & 0.5 & 0.1 \\ 0 & 0.5 & 1 & 0.3 \\ 0.4 & 0.1 & 0.3 & 1 \end{pmatrix}$

은 상관관계가 전혀 없는 것으로, 상관관계를 지니고 있을 것으로 생각되는 반복측정자료에서는 의미가 없는 경우가 많다. 하지만 블록의 수가 반복관찰의 횟수보다 상대적으로 많을 때와 같이 상관관계의 영향이 무시할 수 있을 정도로 작을 경우 생각해 볼 수 있다. 상관관계가 동일한 경우는 모든 측정된 자료내에서 일정한 상관관계를 가정하는 것으로 이때의 상관계수를 급내 상관계수(intra-class correlation) 이라고 한다. 상관관계 내에서 발생하는 일정한 법칙의 경우 이에 대한 많은 모델이 발전되어 왔다. 이중 중 가장 흔한 예는 일차자기 상관(first-order autoregressive, AR1)으로 선형 자기상관관계의 특별한 경우라고 할 수 있다. 일차 자기상관은 시간이 지나감에 따라, 혹은 시간 사이의 거리의 증가에 따라 반복측정자료간의 상관관계가 지수함수적으로 감소한다는 가정을 바탕으로 한다. 일정한 양상이 없는 경우는 자료가 어떻게 상관되어 있는지에 대한 가정이 없으며, 따라서 각각의 측정 사이의 상관관계는 측정마다 다르게 나타날 수 있다. 일정한 패턴이 없을 경우 우리는 $n \times (n - 1)/2$ 개 (n 은 측정 횟수) 만큼의 상관관계를 측정할 필요가 있으며, 다양한 횟수의 관찰이 있게 될수록 상관관계에 대한 정확한 추정 은 어려워진다.

어떤 변수를 반복해서 측정할 경우 분산(variance)은 다양한 양상을 지닐 수 있다. 예를 들어 실험 환경에 따라서, 혈액내의 대사물의 농도의 분산은 시간에 따라 증가하기도, 감소하기도, 혹은 변하지 않기도 할 것이다. 따라서 분산을 정확하게 파악하거나 예측하는 것이 중요하며, 이전의 데이터가 없어서 분산의 양상을 알 수 없을 경우에는 분산의 양상에 대한 가정이 필요할 수 있다. 이들 중 분산 양상에

대한 가장 간단한 가정은 반복측정간의 동일한 분산을 가정하는 것이다.

반복측정자료의 분석방법에는 크게 두 가지 접근방법이 있다. 첫째 방법은 정해져 있는 선형모형에서 분석이 출발한 일변량(univariate) 분석이고, 둘째 방법은 다변량(multivariate) 분석이 있다.

두 방법의 선택은 분석에서 종속변수의 앞서 설명한 분산과 상관관계에 의해 결정되는 분산-공분산 행렬(혹은 표준편차-상관계수 행렬)의 형태에 대한 구형성(sphericity) 가정에 따라 달라진다. 구형성 가정을 만족하면(mauchly 등의 구형성 검정 등을 통해 검정 가능하다.) 분산-공분산 행렬의 각 요소가 동일한 값으로 표현되며, 통계량 F (검정통계량 분산비)가 해당하는 자유도를 정확히 가지고 F분포하게 되어 이 경우는 일변량 분석을 시행하는 것이 더욱 적합하다. 반면에 구형성 가정이 만족 안될 경우 일변량 분석을 행하면 통계량 F는 정확하게 F분포를 따르지 않게 되어 검정력(처리 효과에 유의한 차이가 있을 경우 이를 감지해내는 능력)이 떨어지게 된다. 따라서 이 경우 다변량 분석을 행하게 된다. 반복측정자료의 분석을 위한 방법에서 다변량 분석 방법으로는 Pillai의 trace, Wilks의 lamda, Hoelling의 trace, Roy의 최대근(maximum root) 등이 있다. 다변량 분석은 분산-공분산행렬에 대한 어떤 가정이 없이도 가능하지만, 개체의 수가 적은 경우는 F의 분포의 자유도가 떨어지게 되어 검정력이 매우 떨어지게 된다. 따라서 구형성가정이 만족이 되지 않더라도, 아주 크게 벗어나지 않는다면 어느 정도 수정을 거쳐 일변량 분석을 시행할 수 있는데, 이에 대한 대응책으로 구형성가정이 크게 벗어나지 않는 범위 안에서 자유도의 손실을 줄이고 일변량 분석을 행하는 Greenhouse-Geisser, Hyunh-Feldt방법 등이 있다[5].

G Power를 이용한 표본수 산정

반복측정자료를 수집하는 연구를 계획하고 수행하는데 있어, 표본 수를 산정하기 위해 우리는 반복측정자료에 대한 분석 방법에 따라 다양한 표본수 산출 공식을 적용할 수 있다.

이 중에는 간단한 공식을 적용하여 계산할 수 있는 경우도 있지만, 때로는 계획된 분석에 대한 표본수 산정 및 검정력 계산에 대한 입증된 공식이 없는 경우도 있다. 이러한 경우에는 컴퓨터시뮬레이션을 시행하여 표본수 및 검정력 계산을 시행하는 수 밖에 없다.

또한 대부분의 경우에서 계산 과정이 복잡하여 일반적인 프로그램으로는 계산이 불가능하며, 계산이 가능한 프로그램 중에는 폭넓은 통계에 대한 이해와 소프트웨어의 프로그래밍에 대한 지식을 필요한 경우들이 많다. 아니면 상업적 프로그램들로 값이 비싸 실제 사용이 어려운 경우가 많

다.

저자는 폭넓은 통계와 프로그래밍에 대한 지식이 필요하지 않도록 GUI (graphic user interface) 기법이 적용된 프로그램 중에서 값이 비싸 실제 사용이 어렵지 않도록 무료로 다운받을 수 있는 G Power (version 3.0)를 이용하여 표본수 산출 및 검정력 계산에 대한 과정을 보여주고자 한다.

G Power는 다양한 통계분석법에 대하여 사용이 간편한 표본수 산정 및 검정력 계산을 쉽게 할 수 있는 프로그램이다. 또한 www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3에서 무료로 다운받을 수 있다. G Power에서는 다양한 통계학적 분석(F test, t test, χ^2 test, Z test와 exact test)에 대한 표본수 및 검정력의 계산이 가능하다.

일반적으로 G Power를 이용한 표본수 산출 및 검정력의 계산은 다음과 같은 순서로 진행한다.

1. 적절한 통계학적 검정 방법을 선택한다.

표본 수를 산정하는데 있어 첫 번째 과정 중의 하나는 연구 종료 후 데이터를 분석할 통계적 방법을 선택하는 것이다. 그리고 분석할 통계적 방법에 맞는 표본수 산출 및 검정력 계산 방법을 선택하는 것이다. 예를 들어 연구자들이 진통제에 대한 반응을 보는 연구를 시행하는데 연령에 따른 보정을 한 분석을 하기로 하였다. 연구자들은 연령을 공변량(covariate)로 설정한 공분산분석(analysis of covariance, ANCOVA)을 이용하여 분석하기로 계획하였다. 이 경우 계획된 분석 방법이 t-test가 아니기 때문에, t-test에 근거한 표본수 산출을 하는 것은 부적절하다. 표본수의 산정에 이용되는 통계 방법과 데이터 분석에 사용된 통계방법이 다르다면 표본수가 지나치게 크거나, 작아지게 되며, 이는 연구 결과를 적절하게 해석하는 일을 힘들게 만들어, 시험 자체를 다시 시행 하는 편이 낫거나, 종료된 연구 결과를 출판하지 못하게 하는 일이 발생할 수 있다.

G Power에서는 크게 두 가지 방법으로 통계검정 방법의 선택이 가능하다.

1) 분포에 근거한 통계검정방법의 선택

Test family 메뉴를 이용하여 Exact, F tests, t tests, χ^2 tests, z tests를 선택한다.

2) 디자인에 근거한 통계검정방법의 선택

Statistical test 메뉴를 이용하여 선택한다. 이 메뉴는 화면의 맨위에 위치한 toolbar의 test 메뉴에서, 통계학적 검정을 시행하고자 하는 변수(correlation and regression, means, proportions, variance, generic)와 연구의 디자인을 선택 함으로서 시행할 수 있다.

2. 가능한 5가지 검정력 계산 방법 중 한가지를 선택한다.

이 과정은 구하고자 하는 변수와 주어진 변수를 고려하여 선택할 수 있다. 이 과정은 Type of power analysis 탭을 다음 5가지 중 한 가지로 선택함으로써 선정할 수 있다.

Table 2. Input Parameters for Sample Size Calculation for Repeated Measures Design

Input parameter	Explanation
α err prob	The probability for type I error, The probability of reject the null hypothesis when the null hypothesis is true.; usually set at 0.01 or 0.05.
Power (1- β err prob)	The probability of reject the null hypothesis when the null hypothesis is false, 1- β (type II error); usually set at 0.8 or 0.9.
Effect size f	Minimum effect size difference to be detected by the investigator
Number of groups	Number of groups
Number of measurement	Number of repeated measurement
Corr among rep measures	Correlations among pairs of the repeated measurements
Nonsphericity correction (ϵ)	Correlation factor ($0 \leq \epsilon \leq 1$), meet sphericity assumption as ϵ approach 1

- 1) A priori (검정력 1- β , 유의수준 α , 선별하고자 하는 모집단의 효과 크기를 이용하여 표본수 N을 산출함)
- 2) Compromise (효과크기, 표본수 N, 오차확률비 $q = \beta/\alpha$ (두 오차의 상대적 심각도를 나타냄)를 이용하여 유의수준 α , 검정력 1- β 를 계산함)
- 3) Criterion (검정력 1- β , 효과크기, 표본수를 이용하여 유의수준 α , 그리고 이와 연관된 결정 criterion을 제공함)
- 4) Post-hoc (유의수준 α , 효과크기, 표본수 N을 이용하여 검정력 1- β 를 제공함)
- 5) Sensitivity (유의수준 α , 검정력 1- β , 표본수 N을 이용하여 모집단의 효과크기를 계산함)

3. 분석을 위해 필요한 변수를 입력하고 calculate를 선택한다.

변수는 분석의 종류에 따라 달라지나, 반복측정분산분석에서 필요한 입력변수는 Table 2과 같다. 화면의 왼편에 있는 Determine버튼을 누르면, 선택된 통계 기법에 해당되는 효과크기를 계산할 수 있는 Effect size drawer 창이 열리게 되며, 이를 통해 효과크기를 계산할 수 있다.

G Power를 이용한 반복측정자료 표본수 산정의 예

저자는 반복측정자료의 분석에 대한 가장 간단한 단변량 분석에서부터, 복잡한 다변량 분석에 이르기까지, 표본수 및 검정력 계산 과정을 통증추정을 한 가상의 연구를 예로 들어 설명을 하고자 한다.

앞서 언급하였듯이 단변량 분석의 시행을 위해서는 구형성 가정을 만족한다는 가정에 만족하여야 한다. 구형성 가

정은 1) 반복 측정된 모든 자료의 분산이 동일하고 2) 반복 측정된 자료들의 쌍들 사이의 상관관계가 동일하다는 요소를 만족한다면, 성립된다. 구형성 가정이 만족될 경우, 다변량 분석을 시행하는 것 보다 단변량 분석을 사용하는 편이 검정력을 높일 수 있다[6].

하지만, 구형성 가정은 매우 강력한 가정으로, 가정이 위반되는 경우가 흔하며, 가정에 대한 두 가지 요소 중 특히 반복 측정된 자료들의 쌍들 간의 상관관계가 동일하다는 가정이 위반되는 경우가 흔히 발생한다. 예를 들어 시간이 진행됨에 따라 다섯 번 반복 측정된 자료가 있다고 가정했을 때, 첫 번째와 두 번째 측정값들 사이의 상관관계가 첫 번째와 다섯 번째 측정값들의 상관관계보다 크다는 것은 거의 확실하다고 할 수 있다.

만일 구형성 가정이 위반된다면, F-value 값이 커지게 되어 군내, 군간, 군내-군간의 교호작용에 있어 1종 오류를 범

하게 될 가능성이 높아지게 된다. 이로 인해 유의하지 않은 자료를 유의하게 잘못 분석, 해석하는 오류를 범하게 될 수 있다. 군내 혹은 군내-군간의 교호작용의 분석 및 해석에서 이러한 잘못을 보정하기 위하여 noncentrality parameter와 F-분포의 자유도에 Hyunh 등이 제시한 상관인자(correlation factor) ϵ 를 곱해줄 수 있다[5]. 구형성 가정을 만족하게 된다면 $\epsilon = 1$ 이 되며, 구형성가정의 위배가 심해질수록 $\frac{1}{n-1}$ 에 접근하게 된다(여기에서 n은 반복측정 횟수를 의미한다). 따라서 반복측정 분석에 대한 표본수 산출 및 검정력 계산에 있어 구형성 가정에 대한 고려는 필수적이며, 이에 대한 고려는 계산 결과에 큰 영향을 미친다.

단변량 분석

G Power는 군간, 군내, 군내-군간의 교호작용에 대한 단변량 분석에서 표본수 산정 및 검정력 계산을 할 수 있는 매우 유사한, 각각의 분석에 대한 별도의 방법 및 과정들을 제공한다. 저자는 각각의 방법에 대하여 가상의 예를 들어 G Power를 이용한 계산 과정에 대하여 설명 하고자 한다.

1. 효과크기가 정해져 있는 경우

두 가지 진통제가 수술 후 통증에 미치는 효과를 보는 연구가 있다고 가정한다. A라는 약은 수술 후 통증의 조절을 위하여 진통적으로 사용되어온 약이며, B라는 약은 새로 개발된 약이라고 한다. 두 가지 약의 효과를 판정하기 위하여, 수술 후 2시간, 4시간, 6시간에 100 mm 시각통증점수(visual analogue scale)를 사용하여 통증을 측정하기로 하였다. 이전의 연구에서 각 시간 사이에 시각통증점수들 사이의 상관관계는 0.5였으며, 연구자들은 $\alpha = .05$, 검정력 80%를 적용하여 표본수의 계산을 수행하고자 한다.

만일 찾아내고자 하는 효과크기가 알려져 있다면 이 가정은 매우 쉽게 적용할 수 있다. 본 예에서는 Cohen이 제시한 F-분포에서 중간크기의 효과크기(effect size = 0.25)를 선

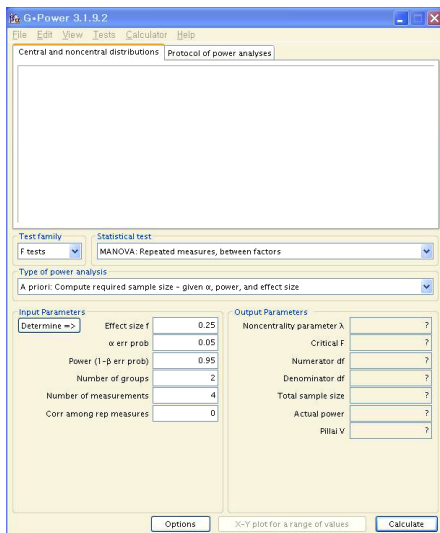


Fig. 1. G Power input window.

Table 3. Formulae for Sample Size Calculation of Repeated Measures Design (Univariate Case)

	Null hypothesis	Effect size	Degree of freedom	Noncentrality parameter
Between effects	$\mu_i - \mu = 0$ $i = 1, \dots, k$	$f = \frac{\sigma_\mu}{\sigma}$	$df_1 = k - 1$ $df_2 = N - k$	$\lambda = f^2 u N \epsilon$ $u = \frac{m}{1 + (m-1)\rho}$
Within effects	$\mu_i - \mu = 0$ $i = 1, \dots, m$	$f = \frac{\sigma_\mu}{\sigma}$	$df_1 = (m-1)e$ $df_2 = (N-k)(m-1)e$	$\lambda = f^2 u N$ $u = \frac{m}{1-\rho}$
Between-within interactions	$\mu_{ij} - \mu_i - \dots - \mu_j + \mu = 0$ $i = 1, \dots, k$ $j = 1, \dots, m$	$f = \frac{\sigma_\mu}{\sigma}$	$df_1 = (k-1)(m-1)e$ $df_2 = (N-K)(m-1)e$	$\lambda = f^2 u N \epsilon$ $u = \frac{m}{1-\rho}$

Levels of between factor: k, Levels of repeated measures factor: m, Population correlation among repeated measures: ρ , Nonsphericity correction; ϵ .

별하기 위한 표본수를 산정하고자 한다[7].

먼저 군간비교에서 표본수 산출을 해 보도록 하겠다. G Power를 연 후, Test > Means > Repeated measures, Between factors, ANOVA-approach로 들어가면 다음과 같은 화면이 나온다(Fig. 1). 여기에서 Type of power analysis를 A priori: Compute required sample size-given α , power, and effect size로 설정하고, 효과크기 0.25를 찾아내기로 하였으므로, Effect size f에 0.25를, $\alpha = .05$ 로 하였으므로, α err prob에 0.05를, 검정력이 80%이므로, Power 0.8를, A, B 두 군을 비교하고 있으므로, Number of groups에 2를, 2시간, 4시간, 6시간에 3번 측정하므로, Number of measurement에 3을, 각 시간 사이에 시각통증점수들 사이의 상관관계는 0.5였으므로, Corr among rep measures에 0.5를 넣고 calculate 버튼을 누르면, Total sample size 86이 계산된다. 화면의 우측에 있는 Output parameters에 나오는 값들은 Table 3에 있는 공식으로도 계산이 가능하다.

계산과 동시에 화면 위 편에는 귀무가설과 대립가설의 분포와, α , β 오류 값, 그리고 표본수 선정에 대한 결정을 내리는 통계값(critical F)이 나오는 그림을 출력한다(Fig. 2). 이 도표는 마우스의 오른쪽을 클릭해서 복사, 저장, 인쇄가

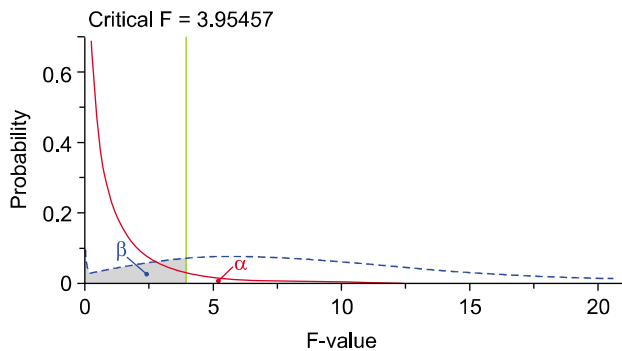


Fig. 2. Figure output window.

가능하다. 이 도표에서 표시된 $F = 3.95457$ 이 Output parameters의 Critical F에도 표기됨을 확인할 수 있다. 또한 Protocol of power analysis 탭을 누르면, 계산 후 입력 및 출력 값이 자동적으로 표기된 것을 확인할 수 있다. 이 값들 또한 동일한 방식으로 저장, 인쇄가 가능하다. X-Y plot for a range of value 단추를 누른 후, Plot창내의 Graph 탭을 누르면 X축, Y축의 변수, X축의 범위, 변수 선택 값에 따른 그래프를 그릴 수 있으며, Table 탭을 누르면 이에 대한 표를 출력할 수 있다.

다음으로 군내비교에서의 표본수 산출을 시행해 보도록 하겠다. Test > Means > Repeated measures, Within factors, ANOVA approach로 들어가면, 군간 비교와 비슷한 화면이 나온다. 군간 비교의 화면과 다른 점은 Nonsphericity correction ϵ 에 해당되는 칸이 생긴다는 점이다. 이 수치는 앞서 언급한 상관인자(correlation factor)이다. 본 예는 단변량 분석을 위한 예제로, 구형성 가정을 한다고 생각할 것이므로 Nonsphericity correction ϵ 에 1을 삽입하고, 나머지 변수들은 군간에서의 계산과 동일하게 적용하면 Total sample size가 28명으로 계산됨을 알 수 있다.

교호작용에서의 표본수 산정은 Test > Means > Repeated measures, Within-between interactions, ANOVA approach로 들어가서, 군내비교에서와 동일한 변수를 삽입하면, Total sample size가 28명으로 계산된다.

2. 효과크기가 정해져 있지 않은 경우

다음으로 위와 비슷한 연구 디자인에서 효과크기가 정해지지 않은 경우를 생각해 보도록 하겠다. 실제로 대부분의 연구에서 효과크기가 정해져 있지 않은 경우가 많으며 효과크기가 연구에 적용될 수 있는지에 대한 이론이 많다. 따라서 임의로 효과크기를 정할 경우 이 값에 대한 논리적 근거를 제시할 수 있어야 한다.

만일 선별해내고자 하는 효과크기가 알려져 있지 않다면,

Table 4. The Results of Pilot Test (Univariate Case)

	2 hr	4 hr	6 hr	$\mu_{i \cdot}$	n_i
Group A	50	40	30	40	10
Group B	40	35	30	35	10
$\mu_{\cdot j}$	45	37.5	30	37.5	20

Values are population mean of VAS for each group and time point. $\mu_{i \cdot}$: means for each group, $\mu_{\cdot j}$: means for each time point, n_i : sample size of each group.

$$SR = \begin{pmatrix} 14 & 0.50 & 0.50 \\ 0.50 & 14 & 0.50 \\ 0.50 & 0.50 & 14 \end{pmatrix}$$

The symmetric matrices SR specify two different covariance structures between measurements taken at different times. The main diagonal contains the SDs of the measurements at each time, and the off-diagonal elements contain the correlations between pairs of measurements taken at different times.

효과크기는 대립 가설 하에 모집단의 특성을 나타내는 변수들로부터 구할 수 밖에 없다. 이 때 연구자들은 변수에 대한 정보를 얻어야만 되는데, 방법에는 1) 이전 연구 (previous study)에서 자료를 얻는 법, 2) 예비연구(pilot study)를 시행하는 법, 3) 경험에 의해 얻어지는 방법 등이 있다. 이 방법들 중 이전 연구에서 자료를 얻는 경우는, 문헌에서 표본수 산정을 위한 충분한 변수를 얻을 수 없는 경우가 많으므로, 연구자들이 이전에 시행한 연구가 아니라면, 이전 연구에 대한 문헌의 저자들과 접촉하여 변수에 대한 정보를 구할 수 밖에 없는 경우가 많다.

본 예에서는 이 전에 시행된 연구가 없었으므로 연구진들은 표본수의 산정을 위해 두 약에서 10명씩을 대상으로 예비연구를 시행하였으며, 결과는 Table 4와 같았다고 가정한다. 단변량 분석의 예이므로 구형성 가정을 만족시키기 위하여 반복 측정된 모든 자료의 분산이 동일하고 반복 측정된 자료들의 쌍들 사이의 상관관계가 동일하다고 가정한다.

이를 바탕으로 군간의 차이를 선별하기 위한 표본수를 산정하도록 하겠다.

Test > Means > Repeated measures, Between factors, ANOVA-approach로 들어가면 Fig. 1의 화면이 나오게 된다. 이 화면에서 Determine 단추를 누르면, Effect size drawer가 열린다(Fig. 3). 여기에서 평균을 알고 있으므로, select procedure에서 Effect size from means로 설정하고, 군이 A, B로 2개이므로 Numbers of groups를 2로, 표준편차가 14로 동일하므로, SD σ within each group에 14를 A군과 B군의 평균이 40과 35이고, 10명씩을 대상으로 하였으므로 Mean에는 40과 35을 Size에는 10을 넣는다. 이후 calculate and transfer to main windows를 click하면 effect size 0.1786이 계

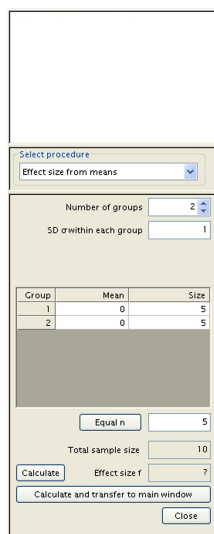


Fig. 3. Effect size drawer.

산되어 main 화면으로 이동되게 된다. 이 화면에서 a err prob에 0.05, Power에 0.8를, Number of groups에 2, Number of measurement에 3, Table 4에서 각 시간 사이에 시각통증 점수들 사이의 상관관계는 0.5였으므로, Corr among rep measures에 0.5를 넣고 calculate 버튼을 누르면, Total sample size 168이 계산되어 나오게 된다.

Effect size drawer에서 select procedure에서 Effect size from variance를 이용하는 방법은 다음과 같다. A군과 B군의 평균은 40과 35이며 전체의 평균은 37.5이므로 군간의 분산은 다음과 같다.

$$\sigma_{\mu}^2 = \frac{(40 - 37.5)^2 + (35 - 37.5)^2}{2} = 6.25$$

여기에서 나온 값인 6.25를 Variance explained by effect에 넣고, Variance within groups에는 주어진 표준편차가 14이므로 $14^2 = 196$ 을 삽입한다. 여기에서 calculate and transfer to main windows 버튼을 누르면 Effect size from mean에서와 동일한 effect size 0.1786이 계산되어 main 화면으로 이동하게 되는 것을 확인 할 수 있다.

다음으로 군내의 차이를 선별하기 위한 표본수를 산정하도록 하겠다.

Test > Means > Repeated measures, Within factors, ANOVA approach로 들어가서, Type of power analysis를 A priori: Compute required sample size-given α , power, and effect size로 설정하고, Determine 단추를 누르면, Effect size drawer가 열린다. 여기에서는 From variances와 Direct를 선택할 수 있는데, 이 경우 분산을 알 수 있으므로 From variances를 선택한다.

2 hr, 4 hr, 6 hr에서의 평균은 45, 37.5, 30이며, 전체의 평균은 37.5이므로 군내의 분산은 다음과 같다.

$$\sigma_{\mu}^2 = \frac{(45 - 37.5)^2 + (37.5 - 37.5)^2 + (30 - 37.5)^2}{3} = 37.5$$

여기에서 산출된 37.5를 Variance explained by effect에 넣고 Variance within group에는 주어진 표준편차가 14이므로 $14^2 = 196$ 을 넣는다. 여기에서 calculate and transfer to main windows의 단추를 누르면 effect size 0.4374이 계산되어 main 화면으로 이동하게 되는 것을 확인할 수 있다. a err prob에 0.05, Power 0.8, Number of groups에 2, Number of measurement 3, Corr among rep measures에 0.5를, ϵ 에 1을 삽입하고 calculate 버튼을 누르면, Total sample size 12가 나오게 된다.

다음으로 군간-군내의 교호작용에 있어 차이를 선별하기 위한 표본수를 산정하도록 하겠다.

Test > Means > Repeated measures, Within-between interactions, ANOVA approach로 들어가서, Type of power analysis를 A priori: Compute required sample size-given α , power, and effect size로 설정하고, Determine 단추를 누르면, Effect size drawer가 열린다. 여기에서는 From variances와 Direct를 선택할 수 있는데, 이 경우 분산을 알 수 있으므로 From variances를 선택한다.

여기에서 $\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot}$ 의 분산을 계산하면

$$\sigma_{\epsilon}^2 = \frac{(50 - 45 - 40 + 37.5)^2 + (40 - 37.5 - 40 + 37.5)^2 + \dots + (30 - 30 - 35 + 37.5)^2}{6} = 4.16667$$

이므로 Variance explained by effects에 4.16667을 넣으면 된다. 나머지는 군내에서 표본수를 산정하는 것과 동일한 과정을 거치면 Total sample size가 8로 나오는 것을 확인할 수 있다.

다변량 분석

G Power는 몇 가지 다변량 분석에 대한 표본수 산정을 할 수 있다. 이 중 Hotelling T^2 tests는 단변량 t test를 중속 변수가 2회 이상 측정된 다변량 분석으로 확장시킨 형태로, 두 개의 평균을 비교하는 대신 평균을 포함한 벡터가 비교되며, 평균을 비교하는 대신, 분산-공분산 행렬을 비교한다. 이에 대한 설명은 생략한다. 이 분석에 대한 자세한 내용은 *G Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences*를 [8] 참고하기 바란다.

G Power는 단변량 분석에서와 마찬가지로, 다변량 분석에서도 군간, 군내, 군내-군간의 교호작용에 대한 표본수 산정 및 검정력 계산을 할 수 있는 매우 유사한 각각의 분석에 대한 별도의 방법 및 과정을 제공한다.

이 경우 또한, 이전에 시행된 연구가 없었으므로 연구진

들은 표본수의 산정을 위해 두 약에서 10명씩을 대상으로 예비연구를 시행하였으며, 연구 결과는 Table 5과 같았다고 한다. 이 경우는 다변량 분석의 예로, 구형성 가정을 만족시키지 못하고 있으며, 반복 측정된 자료의 분산이 다르고 반복 측정된 자료들의 쌍들 사이의 상관관계 또한 다르다.

먼저 군간의 차이를 선별하기 위한 표본수를 산정해보도록 하겠다. Test > Means > Repeated measures, between factors, MANOVA approach로 들어가서, Type of power analysis를 A priori: Compute required sample size-given α , power, and effect size로 설정하고, Option을 Click하여 Choose Approximation Method창을 연다. 여기에서 Effect size calculation에서 use mean correlation in effect size calculation을 해제하고, Approximation에서 Pillai V를 선택하고, Algorithm에서 O'Brien and Shieh를 선택한 후 OK를 click한다. a err prob에 0.05, Power 0.8, Number of groups에 2, Number of measurement에 3을 넣는다. Determine을 click하면, Effect size drawer가 열린다. 여기에서 Select procedure에서 Effect size from mean and cov를 선택한다. 우리가 표준편차와 상관계수를 알고 있으므로, Input method에서 SD and correlation matrix를 선택한 후 Specify/Edit matrices 단추를 누른다. Means tab과 Cov sigma tab에서 Table 2의 수치를 입력한다. Cov Sigma tab에서는 대각선을 중심으로 대칭이므로 상관계수는 음영이 표시되지 않은 한 쪽에 입력하면 음영이 표시된 반대 편까지 입력이 되는 것을 확인할 수 있다. OK 단추를 눌러서 Effect size drawer로 돌아온 후, Calculate 단추를 누르면 Effect size f(v)가 0.2257로 계산되어 나옴을 확인할 수 있다. 여기에서 Calculate and transfer to main window 단추를 누른 후 Close 단추를 누른다. 이 후 Calculate 단추를 누르면 Total sample size가 158이 나오는 것을 확인할 수 있다.

다음으로 군내의 차이를 선별하기 위한 표본수를 산정하도록 하겠다. Test > Means > Repeated measures, Within

Table 5. The Results of Pilot Test (Multivariate Case)

	2 hr	4 hr	6 hr	$\mu_{i\cdot}$	n_i
Group A	50	40	30	40	10
Group B	40	35	30	35	10
$\mu_{\cdot j}$	45	37.5	30	37.5	20

Values are population mean of VAS for each group and time point. $\mu_{i\cdot}$: means for each group, $\mu_{\cdot j}$: means for each time point, n_i : sample size of each group.

$$SR = \begin{pmatrix} 15 & 0.6 & 0.3 \\ 0.6 & 14 & 0.4 \\ 0.3 & 0.4 & 13 \end{pmatrix}$$

The symmetric matrices SR specify two different covariance structures between measurements taken at different times. The main diagonal contains the SDs of the measurements at each time, and the off-diagonal elements contain the correlations between pairs of measurements taken at different times.

factors, MANOVA approach로 들어가서, Type of power analysis를 A priori: Compute required sample size-given α , power, and effect size로 설정하고, Option창과, Effect Size Drawer창에서 동일한 과정을 반복하면, Effect size $f(v)$ 가 0.9096로 계산되어 나오며, Total sample size 16이 나오는 것을 확인할 수 있다.

다음으로 군간-군내의 교호작용에 있어 차이를 선별하기 위한 표본수를 산정하도록 하겠다. Test > Means > Repeated measures, Within-between interactions, MANOVA approach로 들어가서, Type of power analysis를 A priori: Compute required sample size-given α , power, and effect size로 설정하고, Option단추를 눌러서 Choose Approximation Method창을 연다. 여기에서 Approximation에서 Pillai V를 선택하고, Algorithm에서 O'Brien and Shieh를 선택한 후 OK단추를 누른다. a err prob에 0.05, Power 0.8, Number of groups에 2, Number of measurement에 3을 넣는다. Determine을 click하면, Effect size drawer가 열린다. 여기에서 Select procedure에서 Effect size from mean and cov를 선택한다. 우리가 표준편차와 상관계수를 알고 있으므로, Input method에서 SD and correlation matrix를 선택한 후 Specify/Edit matrices단추를 누른다. Means 탭과 Cov sigma 탭에서 Table 2의 수치를 입력한다. OK단추를 눌러서 Effect size drawer로 돌아온 후, Total sample size에 20을 입력한다. Calculate단추를 누르면 Effect size $f(V)$ 에 0.3032가 계산되는 것을 확인할 수 있다. 다음 Calculate and transfer to main window를 Click한 후 Close단추를 누른다 이 후 Calculate단추를 누르면 Total sample size가 108이 나오는 것을 확인할 수 있다.

군간, 군내, 군내-군간 교호작용의 표본수 산정 과정 모두에서 Effect size drawer에서 Input method에 나오는 Variance-covariance matrix를 click하면, Means 탭에서는 SD and correlation matrix에서 입력한 값들이 그대로 있는 것을 확인할 수 있으며, Cov sigma탭을 누르면 좌상귀에서 우하귀에 이르는 대각선에서는 표준편차-상관계수 행렬에서 표준편차를 제공한 분산이 위치하며, 대각선에서 벗어난 부위에는

$$\text{Cov}(X,Y) = \text{SD}(X) \times \text{SD}(Y) \times \text{Corr}(X,Y)$$

(SD: 표준편차, Corr [X,Y]: 상관계수, Cov [X,Y]: 공변량)에서 계산된 공변량이 위치한다는 것을 확인할 수 있다.

Input method에서 SD and correlation pattern을 click하면, 전체 행렬의 상관계수를 지정할 수 있으며 Specify/Edit matrices단추를 누르고, Cov Sigma탭을 누르면 Table 1의 2) 상관계수가 동일한 경우와 같은 행렬을 이루고 있음을 확인할 수 있으며, 1st autocorr을 click한 후 Specify/Edit matrices 단추를 누르고, Cov Sigma탭을 누르면 상관계수의 pattern이

Table 1의 3) 상관계수가 동일한 경우의 1차 자기상관양상을 보임을 확인할 수 있다.

앞서 설명하였듯이 반복측정자료의 분석에서 구형성에 대한 가정을 고려한 분석을 시행하는 것은 대단히 중요하다. 다음의 세 가지 경우를 비교해 보도록 하겠다. 한 경우는 구형성 가정을 통과하여 단변량 분석을 시행한 경우이고, 또 한 경우는 단변량 분석이되 구형성 가정을 위반하여 이를 고려한 상관인자 ϵ 를 적용한 경우이며, 마지막의 경우는 다변량 분석을 실시한 경우이다. 모든 경우에서 Effect size = 0.25, 1종오류 $\alpha = 0.05$, 총 표본수 = 30, 군 수 = 2, 반복측정횟수 = 3, 상관계수 = 0.5으로 지정한 후 Post hoc으로 검정력을 계산하기로 한다.

첫번째로, 구형성 가정을 통과하여 단변량 분석을 시행할 경우를 보면, Test > Means > Repeated measures, Within factors, ANOVA-approach으로 들어가서 Type of power analysis를 Post hoc으로 선택한다. 화면에서 Effect size 0.25, alpha 0.05, total sample size 30, Number of groups 2, Number of measurement 3, corr 0.5으로 넣은 후 구형성 가정을 통과한다고 하였으므로 Nonsphericity ϵ 를 1로 지정한다. calculate 단추를 누르면 power가 0.8407로 계산되어 나옴을 확인할 수 있다.

두번째로, 구형성 가정을 통과하지 않는 경우를 보면, Test > Means > Repeated measures, Within factors, ANOVA-approach으로 들어가서 Type of power analysis를 Post hoc으로 선택한다. 화면에서 Effect size 0.25, alpha 0.05, total sample size 30, Number of groups 2, Number of measurement 3, corr 0.5으로 넣은 후, 구형성 가정을 통과하지 않았으므로, $\epsilon = \frac{1}{n-1}$ (n = 반복측정회수)에서 반복측정횟수가 3이므로 $\epsilon = \frac{1}{3-1} = \frac{1}{2}$ 이므로 Nonsphericity에 0.5를 넣으면 power가 0.6297로 계산되어 나옴을 확인할 수 있다.

마지막으로, 다변량 분석을 시행하는 경우를 확인하면, Test > Means > Repeated measures, Within factors, MANOVA-approach으로 들어가서 Type of power analysis를 Post hoc으로 선택합니다. 화면에서 Effect size 0.25, alpha 0.05, total sample size 30, Number of groups 2, Number of measurement 3, corr 0.5으로 넣은 후, calculate 버튼을 누르면 Power가 0.8127로 나옴을 확인할 수 있다.

여기에서 검정력의 크기는 구형성 가정을 통과하여 단변량 분석을 시행한 경우 > 다변량 분석을 실시한 경우 > 단변량 분석이되 구형성 가정을 위반하여 상관인자 ϵ 를 적용한 경우의 순임을 알 수 있다.

여기에서 확인할 수 있는 점은 첫 번째로 구형성 가정을 통과할 수 있다면, 단변량 분석을 실시하는 것이 가장 검정력이 크며, 두 번째로 구형성 가정을 위반한 경우에는 다변량 분석을 실시하는 것이 단변량분석에서 상관인자로 보정한 경우보다 검정력을 증가시킬 수 있음을 알 수 있다.

결론적으로 반복측정 자료에 대한 분석은 대상수를 늘리지 않고, 검정력을 늘릴 수 있으며, 시간의 흐름에 따른 변화를 명확하게 파악하게 하고, 혼란변수의 영향을 줄일 수 있는 강력한 방법이지만, 반복측정 자료의 분산과 상관관계를 포함한 구형성에 대한 가정에 대한 고려를 반드시 하여야 한다. G Power는 무료로 GUI로 되어 있어 쉽게 사용할 수 있는 소프트웨어로 반복측정자료의 분석에 대한 다양한 계산 방법을 제공한다. 그리고 연구계획단계에 있어 구형성 가정을 확신할 수 있다면, 단변량 분석을 이용할 것으로 계획하는 것이 검정력을 증가시킬 수 있는 방법이지만, 구형성 가정을 만족할 수 없다고 생각될 때에는 다변량 분석을 사용할 것으로 계획하는 것이 검정력을 증가시킬 수 있는 방법이다.

참 고 문 헌

1. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994; 272: 122-4.
2. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978; 299: 690-4.
3. Guo Y, Logan HL, Glueck DH, Muller KE. Selecting a sample size for studies with repeated measures. *BMC Med Res Methodol* 2013; 13: 100.
4. Liu G, Liang KY. Sample size calculations for studies with correlated observations. *Biometrics* 1997; 53: 937-47.
5. Huynh H, Feldt LS. Conditions under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions. *JASA* 1970; 65: 1582-9.
6. O'Brien RG, Kaiser MK. MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychol Bull* 1985; 97: 316-33.
7. Cohen J. *Statistical power analysis for the behavioral sciences* (2nd ed): Hillsdale, 1988.
8. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007; 39: 175-91.