

# 간호학 연구에서 효과크기의 사용에 대한 고찰

강현철 · 연구필 · 한상태

호서대학교 응용통계학과

## A Review on the Use of Effect Size in Nursing Research

Kang, Hyuncheol · Yeon, Kyupil · Han, Sang-Tae

Department of Applied Statistics, Hoseo University, Asan, Korea

**Purpose:** The purpose of this study was to introduce the main concepts of statistical testing and effect size and to provide researchers in nursing science with guidance on how to calculate the effect size for the statistical analysis methods mainly used in nursing. **Methods:** For t-test, analysis of variance, correlation analysis, regression analysis which are used frequently in nursing research, the generally accepted definitions of the effect size were explained. **Results:** Some formulae for calculating the effect size are described with several examples in nursing research. Furthermore, the authors present the required minimum sample size for each example utilizing G\*Power 3 software that is the most widely used program for calculating sample size. **Conclusion:** It is noted that statistical significance testing and effect size measurement serve different purposes, and the reliance on only one side may be misleading. Some practical guidelines are recommended for combining statistical significance testing and effect size measure in order to make more balanced decisions in quantitative analyses.

**Key words:** Effect size, Significance probability, Power of test, Statistical significance, Practical significance

### 서 론

간호학 분야에서 특히 경험적 연구를 수행하여 자료를 해석할 때, 대부분 통계적 유의성을 평가하는 방법이 널리 사용되어 왔다. 전통적으로 통계적 가설검정은  $p$ -값(유의확률)을 산출하고 이에 근거하여 결론을 내리는 방식으로 이루어져 왔다. 즉, 검정하려는 가설을 받아들일 수 있는 가능성의 확률이 규범적으로 정한 범위 내에 있는지를 확인하는 것이다. 이러한 과정을 통상 ‘가설검정’ 또는 ‘연구 결과에 대한 통계적 유의성의 검정’이라고 하며, 연구자는 이러한 분석 과정에 의해서 단순히 연구가설이 통계적으로 유의한지의 여부를

제시한다.

그러나 이러한 형식의 가설검정 방법에 대하여 그에 따른 연구 결과의 가치가 매우 한정적이라는 비판이 높게 일고 있다[1-3]. 즉, 단순히 가설검정만을 사용하면 다른 유용한 정보가 손실되는 경우가 많으며, 특히 검정된 통계적 유의성에 대한 효과의 크기나 양의 정도를 고려하지 못하게 되기 때문이다[4]. Sawyer와 Peter [5]가 “우리의 연구가 어떤 사실의 기술이든 혹은 설명이나 예측이든 간에, 연구의 주요 발견 결과는 검증되어서 확증된 연구 효과의 범위와 효과를 설명하는 크기나 양으로 표시되어야 한다”라고 지적한 바와 같이 단순 가설검정의 방법에 문제가 있음을 공감할 수 있다[2].

주요어: 효과크기, 유의확률, 검정력, 통계적 유의성, 실제적 유의성

Address reprint requests to : Han, Sang-Tae

Department of Applied Statistics, Hoseo University, 20 Hoseo-ro 79 beon-gil, Asan 31499, Korea  
Tel: +82-41-540-5901 Fax: +82-41-540-5908 E-mail: sthan@hoseo.edu

Received: September 2, 2015 Revised: September 3, 2015 Accepted: October 7, 2015

This is an Open Access article distributed under the terms of the Creative Commons Attribution NoDerivs License. (<http://creativecommons.org/licenses/by-nd/4.0>)  
If the original work is properly cited and retained without any modification or reproduction, it can be used and re-distributed in any format and medium.

또한 통계적 유의성은 실제적 유의성과 더불어 검토되어야 할 문제이므로 통계적 유의성만의 지나친 집착과 확대해석은 바람직하지 않으며, 통계적 유의성이 검출되지 않은 연구 결과도 때로는 그 자체로 훌륭한 의미를 지닐 수 있다는 점과 추후 연구에 기여할 수 있다는 점에서 발표될 수 있어야 할 것이다[6]. 여기에서 실제적 유의성이란 실제적인 의미를 부여할 수 있을 만큼의 연구 결과가 도출되었는가 하는 것이다.

통계적 유의성의 기준으로  $p$ -값(유의확률)이 널리 사용되는 것과 같이 실제적 유의성의 기준으로는 효과크기(effect size)가 널리 사용된다. Cohen [7]은 효과크기를 “연구된 결과가 모집단에 실제적으로 존재할 가능성의 정도 혹은 귀무가설의 오류가 일어날 가능성의 정도”라고 표현하였다. 효과크기는 연구 결과의 해석을 용이하게 하고, 연구의 타당성을 보다 높일 수 있으며, 실제적 적용성에 대한 가치도 높일 수 있다[8]. 효과크기를 연구 결과에 함께 보고하는 것이 바람직한 이유는 유의성을 검정하는 것이 연구의 효과성을 측정하는 유일한 지표는 아니기 때문이다. 왜냐하면 통계적 유의성을 검정하는 과정에 관련된 지표들은 표본크기에 의해 많은 영향을 받기 때문이다[2].

본 논문에서는 간호학 연구에서 주로 사용되는  $t$ -검정, 분산분석, 상관분석, 회귀분석 등 주요 통계기법들에 대하여 통상적으로 사용되는 효과크기의 정의를 설명하였다. 또한 간호학의 여러 사례들을 통해 효과크기를 계산하고 이용하는 방법을 설명하였다. 그리고 표본크기를 계산하는 데 가장 널리 사용되는 소프트웨어인 G\*Power 3 (<http://gpower.hhu.de>)[9,10]를 이용하여 각 사례에 대한 표본크기를 제시하였다.

## 통계적 가설검정과 효과크기

이 절에서는 앞으로의 논의를 위하여 통계적 가설검정, 유의수준, 검정력과 효과크기 등의 기본적인 개념에 대하여 간단히 기술한다.

### 1. 가설검정과 통계적 유의성

두 모집단의 평균을 비교하는 경우를 고려하자. 먼저, 두 집단의 모평균을  $\mu_1$ 과  $\mu_2$ 라고 하자. 이때 일반적으로 관심이 되는 귀무가설(null hypothesis)은  $H_0 : \mu_1 = \mu_2$ 이며, 대립가설(alternative hypothesis)은 양측검정(two-tailed test)의 경우  $H_1 : \mu_1 \neq \mu_2$ 이고 단측검정(one-tailed test)의 경우  $H_1 : \mu_1 > \mu_2$  혹은  $H_1 : \mu_1 < \mu_2$ 이다. 연구자의 목적은 귀무가설이 기각(대립가설이 채택)되는지의 여부를 결정하는 데 있으며, 귀무가설은 반증이 되기 전까지는 사실인 것으로 가정된다. 대응표본(paired sample)에 대한 평균 차이의 검정에서

귀무가설의 기각(대립가설이 채택) 여부를 결정하는 데 통상적으로 사용되는 검정통계량은 다음과 같은  $t$ -통계량이다.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s/\sqrt{n}} \quad (1)$$

여기에서  $\bar{x}_1$ 와  $\bar{x}_2$ 는 표본 평균이고  $s$ 는 표본 표준편차이다. 그리고 관측된  $t$ 의 절대값이 기준이 되는 임계치보다 큰 경우, 즉

$$|t| > c_\alpha \quad (2)$$

인 경우 귀무가설을 기각(대립가설을 채택)하게 된다. 식 (2)와 같은 영역을 기각역(rejection region)이라고 한다.

실제 모집단의 모평균 차이가 없는데도 불구하고 실험이나 관측을 통해서 표본으로부터 얻어진 평균 차이가 크게 나타날 수가 있다. 만약, 표본으로부터 얻은 평균 차이가 우연히 크게 나타났다면 연구자는 사실과 다르게 귀무가설을 기각(대립가설을 채택)하는 잘못을 범하게 된다. 이러한 종류의 오류를 제1종 오류(Type I error)라고 한다. 그리고 이러한 오류가 일어날 확률의 최대 허용치를 유의수준(significance level)이라고 하며 통상  $\alpha$ 로 표기한다. 식 (2)에서 임계치  $c_\alpha$ 는 주어진 유의수준에 의존한다. 통상 유의수준  $\alpha$ 는 매우 낮은 수준으로 정해지도록 요구되며, 많은 경우 0.01, 0.05, 0.10 정도가 적용된다. 예를 들어,  $\alpha$ 를 0.05로 정하고  $|t| > c_\alpha$ 인 것으로 나타났을 때 “유의수준 5% 하에서 귀무가설이 기각(대립가설이 채택)된다” 또는 “유의수준 5% 하에서 평균 차이가 통계적으로 유의하다”고 표현한다.

반면에 대립가설이 옳는데도 불구하고 표본으로부터 관측된 평균 차이가 우연히 매우 작을 수도 있다. 이때 연구자는 이러한 결과에 근거하여 귀무가설을 기각하지 않는 잘못을 범하게 되는데, 이러한 오류를 제2종 오류(Type II error)라고 하며 이러한 오류가 일어날 확률을  $\beta$ 로 표기한다. 대립가설이 사실일 때 대립가설을 채택(귀무가설을 기각)하게 될 확률은  $1 - \beta$ 가 되는데 이를 검정력(power of test)이라고 한다. 검정력  $1 - \beta$ 는 0.80에서 0.95 사이에서 정하는 것이 보통이다.

### 2. 효과크기와 실제적 유의성

앞에서의 귀무가설과 대립가설을 보다 일반화하여  $H_0 : \mu_1 - \mu_2 = \delta$ 와  $H_1 : \mu_1 - \mu_2 \neq \delta$ 를 고려하자. 단측검정의 경우에는 대립가설이  $H_1 : \mu_1 - \mu_2 > \delta$  또는  $H_1 : \mu_1 - \mu_2 < \delta$ 이다. 여기에서  $\delta$ 는 연구분야에서 실제적으로 의미 있다고 간주할 수 있는 평균 차이를 의미한다. 예를 들어, 새로운 간호중재 방법이 기존의 방법에 비하여 소요되는 비용과 노력을 고려할 때 어떤 신체 특성치에 있어서  $\delta = 10$  이상의 차이를 나타내어야 새로운 간호중재 방법을 도입하는 의미가 있다면,  $H_1 : \mu_1 - \mu_2 > 10$ 이라는 대립가설이 채택될 수 있어야 한다. 이때

$\delta$ 를 효과크기(effect size)라고 한다.

두 모집단 평균을 비교하는 경우 보다 일반적으로는 효과크기를 다음과 같이 정의한다.

$$d = \frac{|\mu_1 - \mu_2|}{\sigma} \quad (3)$$

식 (3)의 효과크기  $d$ 는 다음과 같은 특징이 있다. 첫째, 측정단위에 의존하지 않는다. 따라서,  $d$ 를  $\delta$  ( $=\mu_1 - \mu_2$ )와 대비하여 표준화된 효과크기라고 한다. 예를 들어, 어떤 간호중재 방법에 따른 신체 특성치의 변화에 있어서 측정단위를 센티미터(cm)로 하였으나 미터(m)로 하였느냐에 따라서  $\delta$ 는 100배만큼 다르게 되지만  $d$ 는 측정단위에 무관하게 동일한 값으로 계산된다. 둘째, 표본크기에 의존하지 않는다. 통계적 유의성을 나타내는 식 (1)의 검정통계량의 경우에는 표본크기가 커짐에 따라서 이 값도 커지게 된다. 반면에 실제적 유의성을 나타내는 효과크기  $d$ 는 표본크기에 의존하지 않는다.

실제 연구에서 효과크기가 가지는 주요 의미는 다음과 같다. 첫째, 어떤 연구의 결과로서 보고되는 효과크기는 실제적 유의성을 나타낸다. 즉, 효과크기  $d$ (또는  $\delta$ )를 보고함으로써 그 연구에서 어느 정도의(큰 또는 작은) 평균 차이가 나타났는지를 제시할 수 있다. 둘째, 실험 또는 관측 연구를 계획하는 단계에서 효과크기는 표본크기를 정하는 것과 밀접한 관련이 있다. 앞에서 설명하였듯이 표본크기는 유의수준  $\alpha$ , 검정력  $1 - \beta$ , 효과크기  $\delta$ 와 관련이 있다. 통상 유의수준  $\alpha$ 는 0.05 정도의 값으로 고정된다는 점을 고려할 때, 효과크기가 큰 경우에는 작은 표본크기로도 높은 검정력을 기대할 수 있지만, 효과크기가 작은 경우에는 높은 검정력을 달성하기 위해서 많은 표본크기가 필요하다. 검정력이 낮다는 것은, 연구자의 주장인 대립가설이 실제 사실임에도 불구하고, 애써 수행한 실험에서 대립가설이 사실이라는 충분한(통계적으로 유의한) 근거를 보일 수 없게 된다는 것을 의미한다.

표본크기 결정의 문제에 있어서 Polity와 Sherman [11]은 효과크기를 정하기 위해 다음과 같은 네 가지 방법을 설명하였다[12]. 1) 과거의 비슷한 연구로부터 효과크기를 추정하는 것이다. 관련된 연구가 많이 있을 때는 메타분석(meta analysis)을 사용해서 추정하는 것이 가장 바람직하다고 하겠다. 2) 관련된 연구가 거의 없다면 작은 규모의 시험연구(pilot study)를 실시하여 추정하는 것이 좋다. 3) 시험 연구를 실시할 상황이 아니면 가상표(dummy table)를 작성하여 임상적으로나 이론적으로 가치가 있기에 충분히 크다고 간주할 수 있는 가장 작은 효과크기를 계산한다. 예를 들어, 어떤 금연 요법이 적어도 10%의 금연효과를 가져올 때 효과적이고 가치 있는 것이라고 간주할 수 있다면, 효과크기를 이 값에 근거하여 추정할 수 있다. 4) 이 방법은 앞의 세 가지 방법을 적용할 수 없을 때 사용하는 방법인데, 과거의 경험적 결과로부터 단지 효과크기가 작을지

(small), 적당한지(medium), 큰지(large)만을 선택하여 그에 따라 정해진 추정치를 사용하는 것이다[7]. 이 방법을 사용하는 데 있어서 강조할 점은 새로운 연구 주제에서는 보통 효과크기가 기대보다 작을 수 있다는 것이다.

Cohen [7]은 행동과학 연구에서 주로 사용되는 여러 가지 통계분석 방법들에 대하여 효과크기를 계산하는 공식을 제안하였고, 작고 큰 정도를 나름대로 정의하였다. 평균 차이의 비교에 대한 효과크기  $d$ 의 경우에서 효과크기의 의미를 예를 들어 설명한다. 예컨대, 실험군(새로운 간호중재를 받은 환자 그룹)이 대조군(기존의 간호중재를 받은 환자 그룹)에 비하여 어떤 신체 특성치의 증가에 있어  $d$ 만큼의 효과를 가진다고 하자. 만약,  $d=1$ 이면 실험군의 평균  $\mu_1$ 은 귀무가설 분포의 상위 16%에 해당한다. 즉, 실험군의 중간에 해당하는 환자가 대조군의 84%보다 낮다는 것을 의미한다. 이런 식으로 Cohen [7]이 정의한 효과크기의 의미를 설명하면 다음과 같다. 1)  $d=0.2$  (small, 작은 효과크기): 실험군의 평균 수준이 대조군 분포의 상위 42% 정도에 해당함. 2)  $d=0.5$  (medium, 중간의 효과크기): 실험군의 평균 수준이 대조군 분포의 상위 31% 정도에 해당함. 3)  $d=0.8$  (large, 큰 효과크기): 실험군의 평균 수준이 대조군 분포의 상위 21% 정도에 해당함.

## 평균 차이 검정

### 1. 독립표본 t-검정(Independent Sample t-test)

독립표본 t-검정은 두 개의 서로 다른 집단으로부터 독립적으로 추출된 자료의 분석에 사용되며, 두 모집단의 평균 차이를 검정한다. 이때 관심이 되는 귀무가설은  $H_0 : \mu_1 = \mu_2$ 이며, 일반적으로 효과크기는 식 (3)의  $d = |\mu_1 - \mu_2|/\sigma$ 가 사용된다. 이때  $\sigma$ 는 공통 표준편차(common standard deviation)를 의미한다.

$$\begin{aligned} \sigma^2 &= (n_1\sigma_1^2 + n_2\sigma_2^2)/n \\ &= (\sigma_1^2 + \sigma_2^2)/2, \quad n_1 = n_2 \text{ 일 경우} \end{aligned} \quad (4)$$

여기에서  $n_1$ 과  $n_2$ 는 각각 두 집단의 표본크기이다( $n = n_1 + n_2$ ).

(예 3-1A)

양측검정(two-tailed test), 유의수준  $\alpha=0.05$ , 검정력  $1 - \beta = 0.08$ ,  $n_1 = n_2$ 의 조건 하에서, Cohen [7]의 기준에 따라 G\*Power 3를 이용하여 효과크기별로 최소 표본크기를 구하면 다음과 같다.

- $d=0.2$  (small):  $n=788$  ( $n_1 = n_2 = 394$ )
- $d=0.5$  (medium):  $n=128$  ( $n_1 = n_2 = 64$ )
- $d=0.8$  (large):  $n=52$  ( $n_1 = n_2 = 26$ )

(예 3-1B)

Ko와 Lee [13]의 연구에서 종속변수 우울(depression)에 대하여 실험군과 대조군의 표본크기, 평균, 표준편차는 각각 다음과 같이 나타났다.

- 표본크기:  $n_1=34, n_2=34$
- 평균:  $\bar{y}_1 = -3.77, \bar{y}_2=1.23$
- 표준편차:  $s_1=6.50, s_2=6.76$

이 예에서 공통 표준편차를 구하면 다음과 같다.

- 공통 분산:  $s^2=(6.50^2+6.76^2)/2=43.97,$
- 공통 표준편차:  $s=\sqrt{43.97}=6.63$ 이다.

따라서, 효과크기는 다음과 같이 추정된다.

$$\hat{d} = |-3.77 - 1.23| / 6.63 = 5.00 / 6.63 = 0.75$$

### 2. 대응표본 t-검정(Paired Sample t-test)

대응표본 t-검정은 서로 짝지어진 자료의 분석에 사용되며, 두 모평균의 차이를 검정한다. 이때 관심이 되는 귀무가설은  $H_0 : \mu_1 = \mu_2$ 이며, 일반적으로 효과크기는 식 (3)의  $d = |\mu_1 - \mu_2| / \sigma$ 가 사용된다. 대응표본 t-검정의 경우  $\sigma$ 는 짝지어진 두 변수의 차이에 의해 정의되는 변수  $D_y (= Y_1 - Y_2)$ 의 표준편차를 의미한다.

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 \tag{5}$$

여기에서  $\sigma_1$ 과  $\sigma_2$ 는 각각  $Y_1$ 과  $Y_2$ 의 표준편차이며,  $\rho$ 는  $Y_1$ 과  $Y_2$ 의 상관계수이다. 대응표본 t-검정에서 식 (1)의 t-통계량과 효과크기 사이에는  $\hat{d} = t / \sqrt{n}$ 의 관계가 있다.

(예 3-2A)

유의수준  $\alpha=0.05$ , 검정력  $1-\beta=0.80$ 의 조건 하에서, Cohen [7]의 기준에 따라 G\*Power 3를 이용하여 효과크기별로 최소 표본크기를 구하면 다음과 같다.

- $d=0.2$  (small):  $n=199$
- $d=0.5$  (medium):  $n=34$
- $d=0.8$  (large):  $n=15$

(예 3-2B)

Wang과 Kim [14]의 연구에서 종속변수 자기효능감(self-efficacy)의 경우 실험군의 표본크기, 사전 평균 및 사후 평균,  $D_y (= Y_1 - Y_2)$ 의 표준편차는 다음과 같이 나타났다.

- $n=30$
- 사전 평균:  $\bar{y}_1=39.70$ , 사후 평균:  $\bar{y}_2=48.83$
- $D_y$ 의 표준편차:  $s=5.74$

따라서, 효과크기는 다음과 같이 추정된다.

$$\hat{d} = |39.70 - 48.83| / 5.74 = 9.13 / 5.74 = 1.59$$

한편, 이 연구에서는  $t=8.72$ 로 보고되어 있는데, 이를 이용하여 다음과 같이 효과크기를 계산할 수도 있다.

$$\hat{d} = 8.72 / \sqrt{30} = 1.59$$

### 3. 일원분류 분산분석(One-Way ANOVA)

분산분석(analysis of variance)은 두 개 이상의 범주를 갖는 독립 변수에 대한 종속변수의 범주별 평균 차이를 검정하는 방법이다. 이 경우 관심이 되는 귀무가설은  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ 이고, 일반적으로 사용되는 효과크기는 다음과 같다.

$$f = \frac{\sigma_m}{\sigma} \tag{6}$$

여기에서  $\sigma_m$ 은 집단 평균들의 표준편차이다.

$$\begin{aligned} \sigma_m^2 &= \sum_{i=1}^k [n_i(\mu_i - \bar{\mu})^2] / n, \quad n = \sum_{i=1}^k n_i, \quad \bar{\mu} = \sum_{i=1}^k (n_i\mu_i) / n \\ &= \sum_{i=1}^k (\mu_i - \bar{\mu})^2 / k, \quad n_1 = n_2 = \dots = n_k \text{ 일 경우} \end{aligned} \tag{7}$$

또한  $\sigma$ 는 공통 표준편차이다.

$$\begin{aligned} \sigma^2 &= \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2 / n = \sum_{i=1}^k (n_i\sigma_i^2) / n \\ &= \sum_{i=1}^k \sigma_i^2 / k, \quad n_1 = n_2 = \dots = n_k \text{ 일 경우} \end{aligned} \tag{8}$$

전체 분산(total variance)

$$\sigma_t^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{\mu})^2 / n \tag{9}$$

을 알고 있는 경우 ( $\sigma_t^2 = \sigma_m^2 + \sigma^2$ ), 또 다른 효과크기의 측도인  $\eta^2 = \sigma_m^2 / \sigma_t^2$ 를 이용할 수도 있는데 식 (6)의  $f$ 와  $f = \sqrt{\eta^2 / (1 - \eta^2)}$ 의 관계가 있다. Cohen [7]은 범위 측도(range measure)  $d = (\mu_{\max} - \mu_{\min}) / \sigma$ 를 이용하여 효과크기  $f$ 를 계산하는 방법도 다루고 있다.

(예 3-3A)

일반적으로 최소 표본크기를 구하기 위해서는 효과크기  $f$ 와 집단의 수가 요구된다. 예를 들어 집단의 수가 3일 때, 유의수준  $\alpha = 0.05$ , 검정력  $1-\beta=0.80$ 의 조건 하에서, Cohen [7]의 기준에 따라 G\*Power 3를 이용하여 효과크기별로 최소 표본크기를 구하면 다음과 같다.

- $f=0.10$  (small):  $n=969$  ( $n_1=n_2=n_3=323$ )
- $f=0.25$  (medium):  $n=159$  ( $n_1=n_2=n_3=53$ )
- $f=0.40$  (large):  $n=66$  ( $n_1=n_2=n_3=22$ )

(예 3-3B)

효과크기 계산에 필요한  $\sigma_m^2$ ,  $\sigma^2$ ,  $\sigma_f^2$ 의 분자에 대응하는 추정치는 SPSS 소프트웨어의 경우 각각 집단-간 제곱합(between sum of squares), 집단-내 제곱합(within sum of squares), 합계 제곱합(total sum of squares)이라는 이름으로 출력된다. 예를 들어, Han 등[15]의 p. 35에서 이들은 각각 다음과 같이 나타났다.

$$\begin{aligned} \cdot \text{제곱합: } & \sum_{i=1}^k n_i(\bar{y}_i - \bar{y})^2 = 1636.5, \quad \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = 2018, \\ & \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = 3654.5 \end{aligned}$$

따라서, 효과크기  $f$ 와  $\eta^2$ 은 각각 다음과 같이 추정된다.

$$\cdot \hat{f} = \sqrt{1636.5/n} / \sqrt{3654.5/n} = 0.90, \quad \eta^2 = \sqrt{1636.5/n} / \sqrt{3654.5/n} = 0.45$$

(예 3-3C)

Lee와 Kweon [16]의 연구에서 종속변수 사회적지지(social support)의 경우 세 집단의 표본크기, 평균, 표준편차는 각각 다음과 같이 나타났다.

- 표본크기:  $n_1=294, n_2=173, n_3=91$
- 평균:  $\bar{y}_1=95.37, \bar{y}_2=92.02, \bar{y}_3=89.56$
- 표준편차:  $s_1=15.37, s_2=15.72, s_3=15.71$

이 예에서 전체 평균, 집단 평균들의 표준편차, 공통 표준편차를 구하면 다음과 같다.

- 전체 평균:  $\bar{y} = \frac{294 \times 95.37 + \dots + 91 \times 89.56}{558} = \frac{52,108}{558} = 93.38$
- 집단 평균들의 분산:  $s_m^2 = \frac{294 \times (95.37 - 93.38)^2 + \dots + 91 \times (89.56 - 93.38)^2}{558} = \frac{2,812}{558} = 5.04$
- 집단 평균들의 표준편차:  $s_m = \sqrt{5.04} = 2.24$
- 공통 분산:  $s^2 = \frac{294 \times 15.37^2 + \dots + 91 \times 15.71^2}{558} = \frac{134,664}{558} = 241.33$
- 공통 표준편차:  $s = \sqrt{241.33} = 15.53$

따라서, 효과크기  $f$ 는 다음과 같이 추정된다.

$$\cdot \hat{f} = 2.24/15.53 = 0.14$$

#### 4. 공분산분석(ANCOVA)

공분산분석(analysis of covariance)은 공변량(covariate)이라고 불리는 연속형 독립변수  $X$ 를 추가로 고려하여 종속변수  $Y$ 의 집단별 평균 차이에 대한 추론을 수행하고자 하는 기법이다. 이 경우 관심이 되는 귀무가설은  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ 이고, 일반적으로 사용되는

효과크기는 앞의 식 (6)과 동일하다. 다만 공분산분석에서는  $\mu_i$ 에 대한 추정치로 집단 평균  $\bar{y}_i$ 가 사용되는 것이 아니라 고려되는 공분산분석 모형에 의하여 추정된다는 점이 다르다. 공변량의 효과를 통제하고 추정된 평균  $\hat{\mu}_i$ 는 SPSS 및 SAS와 같은 소프트웨어에서 추정된 주변평균(estimated marginal mean) 또는 최소제곱평균(least squares mean)이라는 이름으로 출력된다.

(예 3-4A)

일반적으로 최소 표본크기를 구하기 위해서는 효과크기  $f$ , 집단의 수, 고려하는 효과의 자유도, 공변량의 수가 요구된다. 예를 들어 집단의 수가 2, 자유도  $1(=2-1)$ , 공변량의 수가 1일 때, 유의수준  $\alpha = 0.05$ , 검정력  $1-\beta = 0.80$ 의 조건 하에서, Cohen [7]의 기준에 따라 G\*Power 3를 이용하여 효과크기별로 최소 표본크기를 구하면 다음과 같다.

- $f=0.10$  (small):  $n=787$  ( $n_1=n_2 \approx 394$ )
- $f=0.25$  (medium):  $n=128$  ( $n_1=n_2=64$ )
- $f=0.40$  (large):  $n=52$  ( $n_1=n_2=26$ )

(예 3-4B)

효과크기 계산에 필요한  $\sigma_m^2$ 과  $\sigma^2$ 의 분자에 대응하는 추정치는 SPSS 소프트웨어의 경우 각각 집단-간 제곱합(between sum of squares)과 오차 제곱합(error sum of squares)이라는 이름으로 출력된다. 예를 들어, Han 등[15]의 p. 138에서 이들은 각각 다음과 같이 나타났다.

- 제곱합: 집단-간 제곱합=1352.7, 오차 제곱합=8502.5

따라서, 효과크기  $f$ 는 다음과 같이 추정된다.

$$\cdot \hat{f} = \sqrt{1352.7/n} / \sqrt{8502.5/n} = 0.40$$

#### 5. 이원분류 분산분석(Two-Way ANOVA)

이원분류 분산분석은 2개의 독립변수에 대한 종속변수의 범주별 평균 차이를 검정하는 방법이다. 주 효과(main effect)에 대한 검정 및 효과크기의 계산은 앞에서 설명한 바와 유사하므로, 여기에서는 상호작용 효과(interaction effect)에 대한 내용을 설명하도록 한다. 2-요인 상호작용 효과에 대한 귀무가설은  $H_0 : \delta_{ij} = \delta_{ki}$ (모든  $i, j$ 와  $k, l$  쌍에 대하여)이고, 여기에서  $\delta_{ij} = \mu_{ij} - \mu_i - \mu_j$ 이다. 효과크기에 대한 정의는 식 (6)과 동일하며, 이때  $\sigma_m$ 은 상호작용 효과의 표준편차이다.

$$\begin{aligned} \sigma_m^2 &= \sum_{i=1}^k \sum_{j=1}^l n_{ij}(\delta_{ij} - \bar{\delta})^2/n, \quad n = \sum_{i=1}^k \sum_{j=1}^l n_{ij}, \quad \bar{\delta} = \sum_{i=1}^k \sum_{j=1}^l (n_{ij}\delta_{ij})/n \\ &= \sum_{i=1}^k \sum_{j=1}^l (\delta_{ij} - \bar{\delta})^2/(kl), \quad n_{11}=n_{12}=\dots=n_{kl} \text{ 일 경우} \end{aligned} \tag{10}$$

(예 3-5A)

일반적으로 최소 표본크기를 구하기 위해서는 효과크기  $f$ 와 관심이 되는 주 효과 또는 상호작용 효과의 자유도가 요구된다. 예를 들어, A 요인의 수준이 2이고 B 요인의 수가 3일 때 상호작용 효과를 검정하는 경우를 고려하자. 이 경우 상호작용 효과에 대한 자유도는  $(2-1) \times (3-1) = 2$ 이며, 집단의 수는  $2 \times 3 = 6$ 이다. 유의수준  $\alpha = 0.05$ , 검정력  $1 - \beta = 0.80$ 의 조건 하에서, Cohen [7]의 기준에 따라 G\*Power 3를 이용하여 효과크기별로 최소 표본크기를 구하면 다음과 같다.

- $f = 0.10$  (small):  $n = 967$  ( $n_{11} = n_{12} = \dots = n_{23} \approx 162$ )
- $f = 0.25$  (medium):  $n = 158$  ( $n_{11} = n_{12} = \dots = n_{23} \approx 27$ )
- $f = 0.40$  (large):  $n = 64$  ( $n_{11} = n_{12} = \dots = n_{23} \approx 11$ )

(예 3-5B)

Park 등[17]의 연구에서는 실험군과 대조군 각각 3번의 반복측정을 취하고 있다. 이 연구에서 종속변수 임파워먼트(health empowerment)의 경우 집단  $\times$  시간별 평균과 표준편차는 다음과 같이 나타났다(이 예에서는 편의상  $n_{11} = n_{12} = \dots = n_{33}$ 으로 가정하자).

- 평균:  $\bar{y}_{11} = 3.56, \bar{y}_{12} = 3.67, \bar{y}_{13} = 4.00, \bar{y}_{21} = 3.80, \bar{y}_{22} = 3.61, \bar{y}_{23} = 3.52$
- 표준편차:  $s_{11} = 0.62, s_{12} = 0.93, s_{13} = 0.57, s_{21} = 0.98, s_{22} = 0.65, s_{23} = 0.71$

먼저 2요인의 수준별 평균을 구하면 다음과 같다.

- 요인 A (실험군, 대조군):  
 $\bar{y}_1 = (3.56 + 3.67 + 4.00) / 3 = 3.74,$   
 $\bar{y}_2 = (3.80 + 3.61 + 3.52) / 3 = 3.64$
- 요인 B (Baseline, 8<sup>th</sup> week, 24<sup>th</sup> week)  
 $\bar{y}_1 = (3.56 + 3.80) / 2 = 3.68, \bar{y}_2 = (3.67 + 3.61) / 2 = 3.64,$   
 $\bar{y}_3 = (4.00 + 3.52) / 2 = 3.76$

상호작용 효과들의 계수를 구하면 다음과 같다.

- $\delta_{11} = 3.56 - 3.74 - 3.68 = -3.86, \dots, \delta_{23} = 3.52 - 3.64 - 3.76 = -3.89$   
 즉,  $\delta_i = (-3.86, -3.71, -3.50, -3.52, -3.67, -3.88), \bar{\delta} = -3.69$

상호작용 효과들의 표준편차는 다음과 같이 계산된다.

$$s_{\delta}^2 = \frac{(-3.86 + 3.69)^2 + \dots + (-3.88 + 3.69)^2}{2 \times 3} = \frac{0.1318}{6} = 0.0218$$

$$s_m = \sqrt{0.0218} = 0.148$$

공통 표준편차는 다음과 같이 계산된다.

$$s^2 = \frac{0.62^2 + 0.93^2 + \dots + 0.71^2}{2 \times 3} = \frac{4.460}{6} = 0.743, s = \sqrt{0.743} = 0.862$$

따라서, 효과크기는 다음과 같이 추정된다.

$$\hat{f} = 0.148 / 0.862 = 0.172$$

## 비율 차이 검정

### 1. 독립표본 비율 차이 검정

두 개의 서로 다른 집단으로부터 독립적으로 추출된 자료를 통해 두 모집단의 비율 차이를 검정하는 경우를 고려하자. 이때 관심이 되는 귀무가설은  $H_0 : p_1 = p_2$ 이며, 일반적으로 사용되는 효과크기는 다음과 같다.

$$h = |\phi_1 - \phi_2|, \phi_i = 2 \arcsin \sqrt{p_i}, i = 1, 2 \quad (11)$$

Cohen [7]은 이와 같은 효과크기에 대하여  $h = '0.2: \text{small}'$ ,  $'0.5: \text{medium}'$ ,  $'0.8: \text{large}'$ 와 같은 기준을 제시하였다.

(예 4-1A)

예를 들어,  $p_1 = 0.5$ 이고  $p_2 = 0.8$ 인 경우를 고려하자. 이때  $\phi_1 = 1.571$ 이고  $\phi_2 = 2.214$ 가 된다(Cohen [7]은  $p_i$ 에 따른  $\phi_i$ 의 변환표를 제시하고 있다. 또는 EXCEL의 ASIN과 SQRT 함수를 이용하여 쉽게 구할 수 있다). 따라서, 효과크기는  $h = |1.571 - 2.214| = 0.643$ 이 된다. 그리고 이와 같은 경우에 대하여, 양측검정(two-tailed test), 유의수준  $\alpha = 0.05$ , 검정력  $1 - \beta = 0.80$ ,  $n_1 = n_2$ 의 조건 하에서, G\*Power 3를 이용하여 최소 표본크기를 구하면  $n = 88$ ( $n_1 = n_2 = 44$ )이 얻어진다(G\*Power 3에서는  $p_1$ 과  $p_2$ 의 입력을 요구한다).

(예 4-1B)

Yoo [18]의 연구에서 시뮬레이션 교육에 대하여 실험군과 대조군의 표본크기 및 '매우 만족'의 빈도가 각각 다음과 같이 나타났다.

- 표본크기:  $n_1 = 25, n_2 = 24$
- '매우 만족'의 빈도:  $y_1 = 18, y_2 = 10$
- '매우 만족'의 비율:  $\hat{p}_1 = 18/25 = 72.0\%, \hat{p}_2 = 10/24 = 41.7\%$

따라서, 효과크기는 다음과 같이 추정된다.

$$\hat{\phi}_1 = 2 \arcsin \sqrt{0.720} = 2.026, \hat{\phi}_2 = 2 \arcsin \sqrt{0.417} = 1.404$$

$$\hat{h} = |2.026 - 1.404| = 0.622$$

### 2. 대응표본 비율 차이 검정

두 변수  $Y_1$ 과  $Y_2$ 가 모두 0, 1의 값을 가지는 이항변수일 때 ( $Y_1 = 1, Y_2 = 1$ ), ( $Y_1 = 1, Y_2 = 0$ ), ( $Y_1 = 0, Y_2 = 1$ ), ( $Y_1 = 0, Y_2 = 0$ )에 해당하는 쌍들의 비율을 각각  $\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00}$ 라고 하자( $\pi_{11} + \pi_{10} + \pi_{01} + \pi_{00} = 1$ ). 이때 관심이 되는 귀무가설은  $H_0 : \pi_{10} = \pi_{01}$ (또는  $H_0 :$

$\pi_{11} + \pi_{10} = \pi_{11} + \pi_{01}$ 이며, 이러한 검정을 통상 맥니머(McNemar) 검정이라고 한다. 이 경우 일반적으로 사용되는 효과크기는 다음과 같은 오즈비(Odds ratio)이다.

$$OR = \pi_{10} / \pi_{01} \quad (12)$$

이러한 오즈비는 0~∞ 사이의 값을 가지며 OR=1이면 비율 차이가 없음을 의미한다.

(예 4-2A)

예를 들어,  $\pi_{11}=0.54$ ,  $\pi_{10}=0.08$ ,  $\pi_{01}=0.32$ ,  $\pi_{00}=0.06$ 인 경우를 가정하자. 이 경우  $Y_1=1$ 의 비율은  $\pi_{1.} = 0.54 + 0.08 = 62\%$ 이고  $Y_2=1$ 의 비율은  $\pi_{.1} = 0.54 + 0.32 = 86\%$ 이다. 따라서, 효과크기는  $OR = 0.08 / 0.32 = 0.25$ 이다. 그리고 이와 같은 경우에 대하여, 양측검정(two-tailed test), 유의수준  $\alpha=0.05$ , 검정력  $1-\beta=0.80$ 의 조건 하에서, G\*Power 3를 이용하여 최소 표본크기를 구하면  $n=50$ 이 얻어진다(G\*Power 3에서는 불일치 쌍의 비율(proportion of discordant pairs)  $\pi_D = \pi_{10} + \pi_{01}$ 의 입력이 요구된다).

### 3. 교차분석

교차분석은 두 범주형 변수에 의해서 생성되는  $r \times c$  교차표(contingency table)에 대하여 두 범주형 변수의 연관성을 분석하는 방법이다. 두 변수의 범주별 조합에 의한 셀의 수가  $m (=r \times c)$ 개이고, 대립가설과 귀무가설 하에서 각 셀의 비율을  $P_{i.}$ 와  $P_{.i}$ 라고 하자 ( $i=1, \dots, m$ ). 이 경우 일반적으로 사용되는 효과크기는 다음과 같다.

$$w = \sqrt{\frac{\sum_{i=1}^m \frac{(P_{i.} - P_{.i})^2}{P_{.i}}}{P_{.i}}} \quad (12)$$

(예 4-3A)

행의 수가 2이고 열의 수가 3인 교차표를 고려하자. 이때 자유도는  $(2-1) \times (3-1) = 2$ 이다. 유의수준  $\alpha=0.05$ , 검정력  $1-\beta=0.80$ 의 조건 하에서, Cohen [7]의 기준에 따라 G\*Power 3를 이용하여 효과크기별로 최소 표본크기를 구하면 다음과 같다.

- $w=0.1$  (small):  $n=964$
- $w=0.3$  (medium):  $n=108$
- $w=0.5$  (large):  $n=39$

(예 4-3B)

Kang 등[19]의 p. 295에서 귀무가설과 대립가설 하에서의 비율들은 다음과 같이 나타났다.

- $P(H_0)$ :  $p_{11}=0.019$ ,  $p_{12}=0.116$ ,  $p_{13}=0.449$ ,  $p_{21}=0.014$ ,  $p_{22}=0.083$ ,  $p_{23}=0.319$

- $P(H_1)$ :  $p_{11}=0.017$ ,  $p_{12}=0.117$ ,  $p_{13}=0.450$ ,  $p_{21}=0.016$ ,  $p_{22}=0.082$ ,  $p_{23}=0.318$

따라서, 효과크기는 다음과 같이 추정된다.

- $w=0.02$

## 상관분석 및 회귀분석

### 1. 상관분석

상관분석(correlation analysis)은 두 양적변수의 선형적 연관성을 분석하는 방법이다. 이때 관심이 되는 귀무가설은  $H_0 : \rho=0$ 이며, 일반적으로 효과크기는 피어슨의 적률상관계수(Pearson's moment correlation coefficient)  $\rho$ 가 사용된다.

(예 5-1A)

양측검정(two-tailed test), 유의수준  $\alpha=0.05$ , 검정력  $1-\beta=0.80$ 의 조건 하에서, Cohen [7]의 기준에 따라 G\*Power 3를 이용하여 효과크기별로 최소 표본크기를 구하면 다음과 같다.

- $\rho=0.1$  (small):  $n=782$
- $\rho=0.3$  (medium):  $n=84$
- $\rho=0.5$  (large):  $n=29$

### 2. 회귀분석

회귀분석(regression analysis)은 하나 이상의 설명변수  $X_1, X_2, \dots, X_p$ 가 양적 반응변수  $Y$ 에 어떠한 영향을 미치는지를 분석하는 방법이다. 이때 설명변수들에 의해 반응변수가 설명되는 정도를 평가하는 측도 중 하나는 결정계수(coefficient of determination)이며, 이는 통상  $R^2$  (R-square)로 표기된다. 이때 관심이 되는 귀무가설 중 하나는  $H_0 : R^2=0$ 이며, 일반적으로 사용되는 효과크기는 다음과 같다.

$$f^2 = \frac{R^2}{1 - R^2} \quad (13)$$

여기서  $R^2 = \sigma_R^2 / \sigma_Y^2$ 으로서 전체 분산( $\sigma_Y^2$ ) 중에서 회귀식으로 설명되는 분산( $\sigma_R^2$ )의 비율을 의미한다. 또한  $R^2 = f^2 / (1 + f^2)$ 이 성립한다.

(예 5-2A)

일반적으로 최소 표본크기를 구하기 위해서는 효과크기  $f^2$ 과 설명변수의 수가 요구된다. 예를 들어 고려 대상인 설명변수의 수가 5일 때, 유의수준  $\alpha=0.05$ , 검정력  $1-\beta=0.80$ 의 조건 하에서, Cohen [7]의 기준에 따라 G\*Power 3를 이용하여 효과크기별로 최소 표본

크기를 구하면 다음과 같다.

- $f^2=0.02$ ,  $R^2=0.02$  (small):  $n=647$
- $f^2=0.15$ ,  $R^2=0.13$  (medium):  $n=92$
- $f^2=0.35$ ,  $R^2=0.26$  (large):  $n=43$

(예 5-2B)

Lee 등[20]의 연구에서 6개의 설명변수에 의한 회귀식의 결정계수가  $R^2=0.23$ 으로 나타났다. 따라서, 효과크기는 다음과 같이 추정된다.

$$\hat{f}^2=0.23/(1-0.23)=0.23/0.77=0.30$$

### 3. 로지스틱 회귀분석

로지스틱 회귀분석(logistic regression analysis)은 하나 이상의 설명변수  $X_1, X_2, \dots, X_p$ 가 이항 반응변수  $Y$ 에 어떠한 영향을 미치는지를 분석하는 방법이다. 설명변수가 하나인 경우 로지스틱 회귀 모형은 다음과 같다.

$$P(Y=1|X) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} \quad (14)$$

이 때 관심이 되는 귀무가설은  $H_0: \beta_1=0$ 이며, 일반적으로 사용되는 효과크기는 다음과 같은 오즈비(Odds ratio)이다.

$$OR = \exp(\beta_1) \quad (13)$$

이러한 오즈비는  $0 \sim \infty$  사이의 값을 가지며  $OR=1$ 이면 반응변수  $Y$ 에 대하여 설명변수  $X$ 의 영향력이 없음을 의미한다.

(예 5-3)

일반적으로 최소 표본크기를 구하기 위해서는 효과크기  $OR$ 과 귀무가설 하에서의 사건의 비율  $p_1=P(Y=1|H_0)$ 이 요구된다. 이것은 통상 사전확률(prior probability)이라고 불린다. 예를 들어, 관심 군의 비율  $p_1$  및 설명변수  $X_1$ 의 분포 등이 각각 다음과 같이 나타났다고 하자.

- 관심 군의 비율:  $\hat{p}_1=0.5$
- 설명변수  $X_1$ 의 분포: 정규분포(Normal distribution),  $\hat{\mu}=0$ ,  $\hat{\sigma}=1$
- 오즈비:  $\widehat{OR}=1.5$

이와 같은 경우에 대하여, 양측검정(two-tailed test), 유의수준  $\alpha=0.05$ , 검정력  $1-\beta=0.80$ 의 조건 하에서, G\*Power 3를 이용하여 최소 표본크기를 구하면  $n=208$ 이 얻어진다.

### 결론 및 논의

본 논문에서는 간호학 연구에서 주로 사용되어지는 주요 통계기법

들에 대하여 통상적으로 사용되는 효과크기의 정의를 설명하였다. 또한 간호학의 여러 사례들을 통해 효과크기를 계산하고 이용하는 방법을 설명하였다. 이러한 내용은 특히 표본크기 결정의 문제에 있어서 적절한 효과크기를 산출해야 하는 간호학 연구자들에게 큰 도움이 될 것으로 기대된다. 참고로 G\*Power 3에서는 필요한 몇 가지 수치들을 입력함으로써 다양한 분석 상황에서 효과크기를 계산할 수 있는데 이에 대하여는 G\*Power 3 manual (<http://gpower.hhu.de>)을 참조하기 바란다.

최근 많은 분야의 학계에서는 연구자들이 연구 결과를 보고할 때, 효과크기를 함께 보고할 필요가 있음을 주장하고 있다[21-23]. 예를 들어, 미국 심리학 및 교육학 저널에서 가장 널리 이용되는 출판 매뉴얼 중의 하나인 APA [24] Publication Manual은 양적 방법론을 사용했음에도 효과크기를 보고하지 않은 연구는 결점이 있는 것으로 간주한다. APA 방식을 따르는 24개의 저널들은 명시적으로 효과크기를 보고하도록 요구하고 있다[25]. 이러한 저널들은 효과크기가 보고되지 않은 연구는 출판을 불허할 정도로 효과 크기에 대한 관심은 강화되고 있다[26].

국내 간호학 연구에서는 연구 설계 단계에서 표본크기를 결정하거나 메타분석 등의 연구를 위해 기존 연구의 효과크기를 산출하는 경우는 있으나, 연구 결과에 효과크기를 보고하고 이에 근거한 논의를 하는 경우는 거의 없는 것으로 나타났다[27-30]. 앞으로 간호학 연구에서도 효과크기를 적극 활용함으로써 연구 결과에 대한 보고의 질이 더욱 높아지기를 기대하여 본다.

### REFERENCES

1. Im SH. A review on the statistical significance and statistical power analysis. *The Journal of Education Studies*. 2011;48(1):267-281.
2. Na WB, Park CW. A meta analysis on the effect sizes of marketing experiments in Korea. *Korean Marketing Review*. 2006;21(4):189-207.
3. Kwon SD. A review of the significance testing and effect size in empirical research. *Journal of Educational Evaluation*. 2005;18(2):1-16.
4. Peterson RA, Albaum G, Beltramini RF. A meta-analysis of effect sizes in consumer behavior experiments. *Journal of Consumer Research*. 1985;12(1):97-103. <http://dx.doi.org/10.1086/209039>
5. Sawyer AG, Peter JP. The significance of statistical significance tests in marketing research. *Journal of Marketing Research*. 1983;20(2):122-133. <http://dx.doi.org/10.2307/3151679>
6. Kang H. Analysis of statistical method applied in the Journal of Korean Academy of Nursing for recent 3 years (1999-2001). *Journal of Korean Academy of Nursing*. 2002;32(6):929-935.
7. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd



- ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
8. Cooper HM. On the significance of effects and the effects of significance. *Journal of Personality and Social Psychology*. 1981;41(5):1013-1018. <http://dx.doi.org/10.1037//0022-3514.41.5.1013>
  9. Faul F, Erdfelder E, Lang AG, Buchner A. G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*. 2007;39(2):175-191. <http://dx.doi.org/10.3758/BF03193146>
  10. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*. 2009;41(4):1149-1160. <http://dx.doi.org/10.3758/brm.41.4.1149>
  11. Polit DF, Sherman RE. Statistical power in nursing research. *Nursing Research*. 1990;39(6):365-369.
  12. Lee JW, Park MR, Lee JB, Lee SJ, Park ES, Park YJ. A review on the methods of sample size determination in nursing research. *Journal of Korean Academy of Women's Health Nursing*. 1998;4(3):375-387.
  13. Ko JM, Lee JK. Effects of a coaching program on comprehensive lifestyle modification for women with gestational diabetes mellitus. *Journal of Korean Academy of Nursing*. 2014;44(6):672-681. <http://dx.doi.org/10.4040/jkan.2014.44.6.672>
  14. Wang HJ, Kim IO. Effects of a mobile web-based pregnancy health care educational program for mothers at an advanced maternal age. *Journal of Korean Academy of Nursing*. 2015;45(3):337-346. <http://dx.doi.org/10.4040/jkan.2015.45.3.337>
  15. Han ST, Kang H, Huh MH. PASW statistics: SPSS analysis of variance. Seoul: Hannare Publishing Co.; 2009.
  16. Lee JA, Kweon YR. Factors affecting social competence in school-aged children according to alcohol consumption by parents'. *Journal of Korean Academy of Nursing*. 2014;44(5):495-503. <http://dx.doi.org/10.4040/jkan.2014.44.5.495>
  17. Park C, Song M, Cho B, Lim J, Song W, Chang H, et al. Effects of a multi-disciplinary approached, empowerment theory based self-management intervention in older adults with chronic illness. *Journal of Korean Academy of Nursing*. 2015;45(2):192-201. <http://dx.doi.org/10.4040/jkan.2015.45.2.192>
  18. Yoo SY. Development and effects of a simulation-based education program for newborn emergency care. *Journal of Korean Academy of Nursing*. 2013;43(4):468-477. <http://dx.doi.org/10.4040/jkan.2013.43.4.468>
  19. Kang H, Han ST, Choi H. SPSS (PASW Statistics) introduction to data analysis. Paju: Freeacademy; 2010.
  20. Lee H, Cho SH, Kim JH, Kim YK, Choo HI. Influence of self efficacy, social support and sense of community on health-related quality of life for middle-aged and elderly residents living in a rural community. *Journal of Korean Academy of Nursing*. 2014;44(6):608-616. <http://dx.doi.org/10.4040/jkan.2014.44.6.608>
  21. Anderson DR, Burnham KP, Thompson WL. Null hypothesis testing: Problems, prevalence, and an alternative. *The Journal of Wildlife Management*. 2000;64(4):912-923. <http://dx.doi.org/10.2307/3803199>
  22. Breaugh JA. Effect size estimation: Factors to consider and mistakes to avoid. *Journal of Management*. 2003;29(1):79-7. [http://dx.doi.org/10.1016/S0149-2063\(02\)00221-0](http://dx.doi.org/10.1016/S0149-2063(02)00221-0)
  23. Cohen J. The earth is round ( $p < .05$ ). *American Psychologist*. 1994;49(12):997-1003.
  24. American Psychological Association. Publication manual of the American Psychological Association. 5th ed. Washington, DC: Author; 2001.
  25. Thompson B. Foundations of behavioral statistics: An insight-based approach. New York, NY: The Guilford Press; 2006.
  26. Kim M. Quantitative methods in geography education research: Concept and application of effect size. *The Journal of the Korean Association of Geographic and Environmental Education*. 2011;19(2):205-220.
  27. Kim CG, Cho MK. Effects on depression intervention programs of hemodialysis patients in Korea: Meta-analysis. *Journal of the Korean Data Analysis Society*. 2012;14(5):2513-2528.
  28. Jeon HR, Park JS. A meta-analysis about effects of transcutaneous electrical nerve stimulation on postoperative pain. *Journal of the Korean Data Analysis Society*. 2012;14(2):851-865.
  29. Jun JY. A meta-analysis of effects on the warming methods for postoperative patients. *Journal of the Korean Data Analysis Society*. 2009;11(4):1819-1829.
  30. Lee EH. On the introduction and application of effect size for data analysis methods [master' thesis]. Asan: Hoseo University; 2014.