

적정 표본크기(sample size) 결정을 위한 제언

박 원 우*
손 승 연**
박 해 신***
박 혜 상****

—〈목 차〉—

I. 도입말	IV. 표본크기 관련 국내 연구 실태
II. 표본의 의미와 표본크기 결정 기준	V. 향후 연구를 위한 제언
III. 최소 표본크기 기준 미 충족 시의 문제점	VI. 맺음말

표본은 모집단을 대표한다는 점에서 다수의 실증연구들은 대표성을 가진 표본 추출을 위해 많은 노력들을 기울여 왔다. 특히 표본의 크기는 통계량을 토대로 모수를 추정하는 통계적 검정력에 유의한 영향을 미친다는 점을 고려할 때, 통계적 결론 타당도가 높은 의미 있는 연구가 되기 위해서는 적정 수준의 표본 수를 확보하는 것이 반드시 필요하다 하겠다. 하지만 대부분의 실증연구들은 연구결과를 논함에 있어서 통계적 분석 결과의 논리적 해석에만 초점을 맞추었을 뿐 연구에 사용된 표본의 크기가 통계적 결론 타당도를 뒷받침 할 수 있는 적정 수준인지에 대한 논의는 간과하여 왔다. 즉, 대부분의 연구자들은 적절한 표본 크기가 충족되었다는 암묵적 가정 하에 연구를 진행하여 왔으며, 학계 또한 이런 암묵적 합의 하에 실증 연구에 있어 표본 크기에 대해서는 별도의 관심을 기울이지 않고 있는 실정이다. 하지만 연구결과 해석의 타당성 및 일반화 정도에 미치는 표본 크기의 영향력을 고려할 때, 연구 목적과 사용되는 통계적 기법에 따른 적정 수의 표본 확보는 양질의 연구를 위해 대단히 중요한 부분이라 할 것이다. 따라서 표본 크기의 중요성에 대한 연구자들의 관심을 불러일으키며, 실제로 연구를 진행함에 있어 참고가 될 수 있는 표본 크기 관련 연구 논문의 필요성이 제기된다 하겠다. 본 연구는 이런 목적 하에 진행되었다. 그에 따라, 먼저 문헌 연구를 토대로 표본 크기 결정과 관련된 주요 원칙들을 살펴보고, 실증 연구에서 사용되는 주요 통계기법별로 요구되는 최소한의 표본 크기를 제시하였다. 또한 이 기준을 토대로 국내의 대

* 서울대학교 경영대학 교수
 ** 서울대학교 대학원 경영학과 박사과정
 *** 서울대학교 대학원 경영학과 석사과정
 **** 한국방송통신대학교 경영학과 연구조교

표적인 인사·조직분야 학술지인 경영학 연구와 인사조직연구에 지난 10년간(1998~2007) 게재된 논문을 대상으로 통계기법별 최소 요구 표본 수의 기준 충족 여부를 살펴보았다. 마지막으로 실증연구 시 통계적 결론 타당도를 높이기 위해 필요한 적정 수준의 표본 확보의 중요성을 강조하면서, 향후 연구를 위한 몇 가지 제언을 다루었다.

I. 도입말

실증연구를 시작함에 있어서 결정해야 할 중요한 문제 중 하나는 표본의 크기를 결정하는 일이다(Kerlinger & Lee, 2000). 물론 연구목적에 따라 전수조사가 바람직한 경우도 있지만 신속성, 경제성 등의 이유로 표본조사가 더 효율적인 경우가 많다(채서일, 2003). 따라서 최소 비용으로 최대의 정보를 얻을 수 있는 최적의 표본을 확보하는 것은 연구자들의 주된 관심 중 하나이다(Festinger & Katz, 1953).

많은 연구자들이 각자의 연구목적에 따라 다양한 표본크기 결정 공식을 제시하고는 있지만 사실상 표본크기를 정하는 문제는 결코 쉽지도, 사소하지도 않은 과정으로써, Williams(1978)는 표본크기 결정의 문제가 응용통계에서 가장 어려운 과제라고 하기도 하였다. 또한, 모집단 크기에 따라 표본 크기의 비율을 결정하라는 등의 많은 주먹구구식 방법들(rules of thumb)이 정당성 없이 사용됨에 따라 연구자들을 혼란스럽게 할 뿐 아니라 연구결과의 타당도를 왜곡하는 일도 발생하게 된다.

표본이란 연구자의 진정한 관심이 되는 모집단에서 특정 절차를 거쳐 선정된 모집단의 전형(typicality)으로써 대표성을 가진 연구 대상을 말한다. 표본은 모집단을 대표한다는 차원에서 연구의 외적 타당도에 영향을 미친다. 따라서 표본이 모집단을 대표할 수 있도록 하는 다양한 표본추출방법이 존재하는 데 그 중에서도 모집단 내 모든 표본단위들이 동일한 추출확률을 가지고 있으며, 이를 무작위로 추출하는 단순 무작위 추출이 가장 기초가 된다 하겠다. 이렇게 추출된 표본으로부터 얻어진 통계량은 모수를 추정하는데 있어 일반화 정도가 높은 대표성을 지닌다 하겠다. 또한 연구를 진행함에 있어서 유의한 연구결과를 얻고자 하는 연구자의 본질적 의도를 고려할 때, 무엇보다도 연구가설이 참일 경우 이것이 통계적으로 유의한 결과에 의한 것이라는 것을 주장하기 위해서는 통계적 검정력이 중요한데, 이러한 통계적 검정력은 곧 표본의 크기에 직접적인 영향을

받는다. 즉 통계적 검정력이 통계량을 가지고 모수를 추론하는 통계적 결론 타당도에 영향을 미친다는 점에서 적절한 표본 크기의 선정은 연구의 시종에 지속적인 영향을 미치는 대단히 중요한 부분이라 하겠다(Cook & Campbell, 1979).

그럼에도 불구하고 국내·외를 불문하고 통계적 결론 타당도를 고려하여 표본 크기의 적절성을 다룬 연구는 거의 없는 실정이다. 즉, 이제까지 연구결과를 논함에 있어서 통계적 분석결과의 논리성에만 초점을 맞추었을 뿐 표본크기의 적절성에 대해서는 이미 충족된 것으로 보았다는 것이다. 하지만 표본크기가 적절하지 않을 경우, 특히 표본의 수가 사용된 통계적 기법에서 요구하는 최소 요구 수준보다 적을 경우 통계적 결론 타당도가 저하되어 연구 결과의 해석 및 일반화가 제한되는 점을 고려할 때, 적정 수준의 표본 크기를 갖추었느냐를 확인하는 것은 연구에 있어 매우 중요한 부분이라 하겠다.

따라서 본 연구에서는 표본크기 결정과 관련된 주요 원칙들을 살펴보고 이 중 가장 일반적으로 실증연구에서 사용되는 통계기법별 최소 표본크기의 기준을 제시하고자 한다. 그리고 표본크기가 이러한 기준을 충족하지 못할 경우의 문제점에 대하여 알아볼 예정이다. 또한 국내 문헌 중 실증연구를 중심으로, 사용된 통계기법과 표본 크기의 적절성 여부를 분석하여 향후 국내 연구를 위한 제언을 하고자 한다.

II. 표본의 의미와 표본크기 결정 기준

1. 표본의 의미

연구를 진행함에 있어서 그 대상이 사람이건, 사건이건, 사물이건 관계없이 대상으로부터 어떤 정보를 얻기 위해서는 그들 중 일부를 선택하여 연구를 진행하게 된다. 이렇게 선택된 일부를 통해 얻은 정보를 바탕으로 우리는 연구 대상 전부의 특성이 어떠한다는 결론에 도달하게 된다(Kerlinger & Lee, 2000). 물론 연구대상을 전부 조사하는 것이 가장 확실한 결론에 도달하게 하는 방법이겠지만, 이렇게 일부를 대상으로 한 연구방법이 전통적으로 현명하다고 평가된다. 이는 전수조사보다 훨씬 적은 시간과 비용을 들여 거의 동일한 정보를 얻기 때문이다. 이 때 연구대상이 되는 전체 대상을 모집단이라고 하며, 이 모집단으로부터 추출한 일부를 표본이라고 한다(Williams, 1978). 즉 표본이

란 모집단의 전형(typicality of a population)이자 모집단의 특성을 대변한다 할 수 있을 것이다(Kerlinger & Lee, 2000).

하지만 모집단에서 추출했다고 해서 모든 표본이 의미가 있는 것은 아니다. 즉, 연구자의 실제 관심은 모집단의 특성인 모수에 있는 것이기 때문에 추출된 표본이 이 모집단의 특성을 대표할 수 있어야 표본으로부터 얻은 통계량을 가지고 모수를 적절하게 추론할 수 있는 것이다. 따라서 어떻게 모집단을 대표할 수 있도록 표본을 추출할 것인가에 대해 연구관심이 집중되어 온 것은 당연하다(Aron & Aron, 2003; Festinger & Katz, 1953; Kerlinger & Lee, 2000).

그에 따라 다양한 표본추출방법이 사용되어 왔는데 전통적으로 표본프레임 내에 있는 표본단위들이 표본으로 추출될 확률이 알려져 있는 경우와 그렇지 않은 경우에 따라 확률 표본추출과 비확률 표본추출로 나누어진다. 진정한 의미에서 모집단을 대표하는 표본을 선정하는 방법은 모집단 내에 있는 모든 요소가 추출될 확률이 알려져 있고, 이 확률이 동일하며, 연구자의 의도 등 기타의 외생변수 없이 무작위로(randomly) 표본을 추출하는 단순 무작위 표본추출(simple random sampling)이라 할 수 있다. 하지만 단순 무작위 표본추출을 정의하기는 쉬워도 실제 연구에서 완벽하게 적용하기는 어려운데, 이는 모집단이 확정된다 하더라도 표본프레임(sample frame)의 확보가 어렵기 때문이다. 표본프레임이란 연구대상이 되는 모집단의 표본단위가 수록된 목록을 말한다. 따라서 확정된 모집단의 표본프레임을 가지고 난수표 등을 이용하여 무작위로 표본을 추출한 경우에만 단순 무작위 표본추출이 이루어진 것이고, 이는 곧 표본을 이용한 모수 추정의 일반화가 가능함을 의미한다. 물론 연구목적 또는 자원의 제한 등으로 비확률 표본추출이 종종 사용되기도 하지만 이런 경우에는 표본으로부터 얻어낸 자료로 표본오차를 추정할 수 없어 결과의 일반화에 한계가 존재한다.

따라서 연구의 외적타당성을 높이기 위한 대표성 있는 표본의 추출을 위해 표본 선정 및 추정에 대한 전반적인 절차를 다루는 표본설계가 제대로 이루어져야 하는 것이다(Festinger & Katz, 1953).

2. 표본크기 결정 방법

연구를 시작하려 할 때 우선적으로 고려해야 할 것 중 하나가 표본크기를 어느 정도로

할지를 정하는 문제인데(Comrey & Lee, 1992), 실제 연구자가 관심을 갖는 대상은 모집단이기 때문에 전수조사가 가장 적합한 방법일 수 있겠지만, 여러 가지 제한적 여건(시간, 비용 등)으로 인해 표본조사가 불가피한 경우가 허다하다. 따라서 연구목적에 맞는 적정수준의 표본크기를 정하는 것은 중요한데, 이는 표본크기가 너무 크면 자원의 과도한 낭비가 될 것이고, 반대로 유의한 결과를 얻을 수 없을 정도로 표본의 크기가 적을 경우에는 노력의 낭비가 되기 때문이다. 따라서 연구자에게 있어서 최소자원으로 최대의 정보를 얻을 수 있도록 연구목적에 맞는 표본크기를 정하는 것은 중요한 단계라 하겠다. 물론 표본크기가 얼마나 되어야 하는지에 대한 해답은 결코 단순하지 않다. 하지만 일반적으로 인정되고 있는 규칙을 중심으로 몇 가지 표본 크기 결정 방안을 살펴보고자 한다.

1) 중심극한 정리

모집단의 특성을 나타내는 모수는 표본의 특성을 나타내는 통계량을 통해 추정하게 되는데, 문제는 모수는 고정된 상수 값이지만 통계량은 표본마다 상이하다는 데에 있다. 즉, 모든 표본이 동일한 통계량을 가지고 있지 않는다는 것이다. 그런데 모집단에서 일정 수의 표본을 무작위로 추출 가능한 모든 경우의 수를 고려하여 추출한 후 각 표본의 평균값의 분포를 살펴보면 표본의 크기가 커질수록 표본 평균의 분포가 점점 정규 분포화 되어 가는 현상이 나타난다. 이러한 표본 평균 분포의 특성을 나타낸 것이 중심극한 정리이다(Kurnow, Glasser, & Ottman, 1959). 표본 평균 분포의 통계량과 모수를 비교하면 다음과 같다.

$$\mu_{\bar{x}} = \mu$$

$$\delta_{\bar{x}} = \frac{\delta}{\sqrt{n}}$$

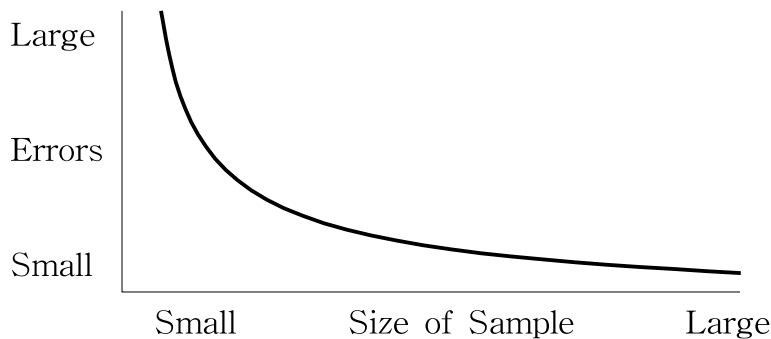
μ	: 모집단 평균
$\mu_{\bar{x}}$: 표본 평균의 평균
δ	: 모집단의 표준편차
$\delta_{\bar{x}}$: 표본 평균의 표준편차
n	: 표본 크기

위의 공식에 의하면 표본 평균의 평균은 모집단의 평균과 같다는 것과 표본의 수가 커질수록 표본 평균의 표준편차가 0에 가까워짐을 알 수 있다.

중심극한 정리는 통계학에서 중요한 의미를 지닌다. 이는 현재의 대부분의 통계적 방법이 정규분포에 기반을 두고 있는데, 모집단의 분포가 정규분포인지의 여부와 관계없이

표본의 크기가 커질수록, 특히 30이상이 되면, 표본 평균의 분포가 정규분포를 이루기 때문에 통계적 추정과 관련된 모든 문제해결을 위해 정규분포가 갖는 특성들을 적용시킬 수 있기 때문이다. 즉, 통계적 분석의 유용성이 보장된다는 점에서 중심극한 정리는 대단히 중요하다(Williams, 1978).

이와 관련하여 중심극한정리와 유사한 표본크기 선정 원칙이 있는데 이를 대수의 법칙이라 한다. 대수의 법칙이란 할 수만 있다면 가급적 많은 표본을 사용하는 것이 좋다는 것이다(Kerlinger & Lee, 2000). 물론 연구에 따라 소수의 표본을 필요로 하는 경우도 존재하기 때문에 가급적 큰 표본을 사용하라는 법칙이 모든 연구에 적용되는 황금률은 아니다. 하지만 일반적으로 표본의 크기는 모수와 통계량의 차이인 오차(sampling error)와 직접적인 연관성을 가진다. 즉, 표본의 크기가 작을수록 오차는 커지는 반면, 표본의 크기가 커질수록 모수와 통계량은 점차 가까워지는 반비례 관계를 보인다는 것이다(Well, Pollatsek, & Boyce, 1990). 이런 표본크기와 오차 간의 관계를 <그림 1>과 같이 표현할 수 있다.



자료원: Kerlinger & Lee, 2000, p. 175.

<그림 1> 표본크기와 오차와의 관계

Wunsch(1986)는 연구 진행에 있어서 항상 발생하는 문제점 두 가지를 표본크기를 정할 때 표본추출 오차(sampling error)를 무시하는 것과 응답 및 무응답 편향을 무시하는 것이라고 지적하였다. 따라서 연구자의 진정한 관심인 모수를 추정함에 있어서 표본에 의한 통계량에 의존할 수밖에 없는 연구인 경우에는 표본의 크기를 30이상으로 하여 정규분포를 이용한 통계적 유용성을 확보할 뿐만 아니라 오차를 최소화시킬 수 있도

록 표본의 크기를 가급적 크게 하는 요구된다 하겠다.

2) Type I 오류를 이용한 방법

이 방법은 표본의 크기와 모집단의 분산 정도가 신뢰구간을 결정한다는 것을 이용한 방법으로, 모집단의 표준편차를 알고 있다고 가정하면 원하는 정확도의 정도를 결정하여 그에 따라 적정 표본크기를 정할 수 있다는 내용이다(Kerlinger & Lee, 2000). 본 방법에 따르면, 우선 연구자는 실제 모집단의 표준편차 또는 그것의 추정치가 필요하다. 이는 과거 자료나 연구를 통해 얻는데, 실제로는 모수의 획득이 거의 불가능하다. 따라서 일반적으로 표본에서 얻은 최대값과 최소값의 차이를 4 또는 6으로 나누는 방법으로 추정하게 된다(Mendenhall & Beaver, 1994; Williams, 1978). 다음으로 허용 가능한 오차의 양을 정한다. 이 오차의 양이란 연구자가 최대한으로 허용할 수 있는 모집단의 평균과 표본의 평균의 차를 말한다. 마지막으로 신뢰수준을 정한다. 이는 통상 알파 오류(또는 Type I 오류)로 알려져 있다. 알파 오류란 귀무가설이 맞지만 이를 틀리다고 할 확률을 말한다. 표본크기 산정 공식은 다음과 같다.

$$n = \frac{Z^2 \times \delta^2}{d^2}$$

<p>Z : 신뢰수준에 따른 Z 값 δ : 모집단의 표준편차 d : 모집단 및 표본 평균값 간 차이</p>

예를 들어, 연구자가 모집단의 표준편차가 0.6이라는 것을 알았으며(혹은 표본의 최대값과 최소값의 차이를 4 또는 6으로 나눈 추정치), α를 0.05로 정했으며(Z값은 1.96이 됨), 또한 연구자가 평균 간 오차를 0.2 정도로 허용할 수 있다고 한다면, 위의 공식에 따라 표본크기 n은 $(1.96)^2 \times (0.6)^2 / (0.2)^2 = 34.6$, 즉 35가 되는 것이다.

이 방법은 여러 학자들에 의해 수용되고 있는데, 예를 들면, Bartlett, Kotrlik, & Higgins(2001)는 알파 오류를 0.01, 0.05, 0.10 나누고 모집단 크기에 따른 적정 표본크기 기준표를 제시하였다. 이 때 사용한 허용오차는 0.3으로 연속형 변수일 경우 일반적으로 수용되는 값이며(Krejcie & Morgan, 1970), 모집단의 표준편차는 7점 척도 설문지를 기준으로 6으로 나눈 값을 사용하였다. 물론, 연구 상황 및 가용 정보에 따라 허용 오차와 예상 모집단 표준편차가 변하면 위의 공식에 대입하는 값이 변함에 따라 모집단에 따른 적정 표본 크기가 달라짐은 당연하다. 하지만 이처럼 알파오류를 이용한 방

법은 많은 학자들이 인정하는 표본크기 결정 공식의 하나이다.

3) Type I 및 II 오류를 이용한 방법

이 방법은 알파 오류를 활용한 방법과 기본적인 논리는 동일하나, 베타 오류(또는 Type II)까지 활용한 방법으로써, 두 오류의 최대 허용치를 정하고 이를 통해 표본의 수를 결정하는 방법이다(Gulifort & Fruchter, 1978). Type II 오류란 귀무가설이 틀리지만 맞다고 결론내릴 확률을 말한다. 절차를 살펴보면, Type I 오류를 이용한 방법과 동일하나 마지막 신뢰수준을 정하는 단계에서 귀무가설 및 대체가설을 토대로 Type I 및 II 오류를 정하여 그에 맞는 Z 값을 이용하는 것이다. 표본 크기 산정 공식은 다음과 같다.

$$n = \frac{(Z_{\beta} - Z_{\alpha})^2 \times \delta^2}{d^2}$$

Z_{α} : α 오류에 따른 Z 값
Z_{β} : β 오류에 따른 Z 값
δ : 모집단의 표준편차
d : 모집단 및 표본 평균값 간 차이

예를 들어 Type I 오류를 이용한 방법의 사례를 적용하였을 경우, α 를 0.05, β 를 0.10으로 정했다면 Z값은 각각 -1.28, 1.645가 된다. 따라서 공식에 대입하면 표본 크기 n 은 $(1.645 - (-1.28))^2 \times (0.6)^2 / (0.2)^2 = 77$, 즉 77이 되는 것이다. 이는 단측 검정의 예를 든 것으로 만약 양측 검정을 할 경우에는 Z_{α} 만 바뀌면 된다(사례의 경우 $Z_{\alpha} = 1.96$). 이것을 다시 공식에 대입하면 $(1.645 - (-1.96))^2 \times (0.6)^2 / (0.2)^2 = 116.96$, 즉 117이 된다.

여기서 동일한 사례를 이용한 Type I 오류를 이용한 방법과 Type I 및 II 오류를 이용한 방법에 의해 구한 표본크기 n 이 다르게 나온 것을 어떻게 이해해야 하는지에 대한 의문이 생긴다. 이는 Type I 및 II 오류의 의미를 되짚어 본다면 간단하다. Type I 오류가 귀무가설이 맞으나 이를 틀리다고 할 확률인 반면 Type II 오류는 귀무가설이 틀리나 이를 맞다고 할 확률이므로, Type I 오류를 이용한 방법에서 구한 표본 크기 35는 단지 귀무가설을 기각하는데 필요한 표본 크기인 반면, Type I 및 II 오류를 이용한 방법에서 도출된 표본 크기 77은(단측 검정의 경우) 귀무가설의 기각 뿐 아니라 연구자가 원하는 충분한 통계적 검정력까지 제공한다는 면에서 보다 신중한 표본의 크기

라 할 수 있는 것이다.

하지만 이상에서 살펴본 Type I 오류, Type I 및 II 오류를 이용한 표본크기 산정 방법은 고려된 변수가 하나라는 것이다. 하지만 실제 연구에서는 다수의 변수들을 동시에 고려하는 경우가 많다. 그에 따라 실제 연구에서는 가장 중요한 변수 또는 표준편차가 가장 클 것으로 예상되는 변수를 정하여 그 변수를 대상으로 표본의 수를 결정하는 방법도 가능하리라 판단된다(이근희, 2001). 또한 통계적 방법에 의한 표본 크기의 산정은 실제 연구를 진행함에 있어서 소요되는 인적, 물적 자원의 제한을 고려하지 않는 한계점도 가진다. 하지만 연구목적상 본 논문에서는 이런 자원의 가용정도가 표본크기 산정에 주는 영향은 별도로 고려하지 않았다.

4) 타당도를 고려한 표본크기

표본크기와 관련된 타당도에는 통계적 결론 타당도와 외적 타당도가 있다. 통계적 결론 타당도란 연구논문의 질을 따지는 개념으로 통계량으로부터 모수를 추정함에 있어 추론이 과연 얼마나 타당한가를 말한다(Cook & Campbell, 1979). 이는 설계된 대로 연구가 되었으며 통계적으로 해석이 타당하게 되었는지를 따지는 것으로, 만약 통계적 결론 타당도가 없다면 아무리 좋은 결과가 나온다 할지라도 그 결과에 따른 모수 추론은 의미 없는 것이 된다. 이 통계적 결론 타당도에 중요한 영향을 미치는 것이 바로 표본크기이다. 통계적 결론 타당도를 높이기 위한 표본크기는 위에서 언급한 중심극한 정리, 즉 표본의 크기가 30이상이 되어야 한다는 것이 기본이 된다. 그래야 표본평균의 정규분포 가정이 성립되어 정규분포를 이용한 통계적 추론 방식의 정당성이 확보되기 때문이다. 부가하여 요인 또는 변수의 수보다 표본의 수가 8 내지 10배 이상 커야 하며, 셀(cell)수가 종속변수의 수보다 커야 보다 높은 통계적 결론 타당도를 확보할 수 있다(박원우, 2003).

또한 표본크기와 관련된 타당도로 외적 타당도가 있는데 이는 연구결과의 일반화 가능성을 말하는 것이다. 즉 연구를 통해 도출된 결과를 기타 조건(시간, 장소, 대상 등)으로 확장해석 할 수 있느냐의 문제이다. 이는 표본이 모집단을 대변하는 대표성을 가지고 있느냐를 논하는 것으로 이때에는 단순 무작위 표출이 반드시 전제가 되어야 한다(Hurlburt, 2003). 이는 모든 표본단위의 표출 확률이 동일한 경우로, 통상 난수표를 이용하여 표본을 선정하게 된다(Gravetter & Wallnau, 2000). 이렇게 단순 무작위

표출의 조건이 충족된 경우, 표본의 크기가 전체 모집단의 5%이상이면 대표성이 있다고 볼 수 있다(박원우, 2003). 지금까지 살펴본 통계적 결론 타당도 및 외적 타당도는 연구 설계에 있어 중요한 의미를 가지고 있기 때문에 연구의 타당도를 고려한 적정 표본크기의 선정은 바른 연구 설계의 기초이며, 연구 결과의 가치를 높이는 핵심이 된다 하겠다.

3. 통계기법별 최소 표본크기

모든 상황에서 보편적으로 적용 가능한 명확한 표본 크기는 존재하지 않는다. 많은 연구자들이 나름대로의 해답을 제시하고는 있지만, 대부분 특정 연구목적에 맞는 규칙일 뿐더러 그 또한 완벽하고 정확한 방법이라 할 수 없다. 그에 따라 많은 주먹구구식 원칙들(rules of thumb)이 정당성 없이 사용되고 있는 경우가 허다하다(Kerlinger & Lee, 2000).

따라서 연구목적에 따라 유의수준, 효과크기를 고려한 통계적 검정력을 기준으로 하여 적정 표본크기를 결정하는 것이 바람직하다. 물론 표본의 크기가 너무 큰 경우도 반드시 바람직하다고 할 수는 없겠지만, 연구자의 현실적 여건, 즉 가용한 시간 및 경제적 자원의 제한성을 고려할 때 표본크기가 과도히 많아서 문제가 되는 경우는 극히 드물 것으로 예상된다. 오히려 최소 표본크기를 충족시키지 못해 연구결과를 왜곡시킬 가능성이 많을 것으로 여겨진다. 또한 대부분의 연구자들이 효과크기를 최대한 높이고 싶어 하지만 이는 실험변수의 조작이 가능한 실험연구에서 가능한 일일 뿐, 설문지를 이용한 현장조사가 대부분인 인사조직, 경영학 분야에서는 모집단이 보유한 특성을 표본을 통해 간접적으로 획득하는 것이기 때문에 원하는 크기의 효과를 원하는 수준만큼 획득한다는 것은 매우 어려운 일이다. 그에 따라 대부분 효과크기를 산정함에 있어 이전 자료를 이용하거나 합리적 결정, 또는 예비 조사 및 문헌 연구를 통해 주관적 판단의 타당성을 확보해야 한다(Eng, 2003). 따라서 실제 연구에 있어서 연구자는 효과크기 및 유의수준을 고려한 통계적 검정력을 기준으로 이에 적합한 최소 표본크기를 정하게 된다. 이 때 대부분의 학자들에게 수용되는 통계적 검정력은, 비록 연구 분야마다 선호되는 수준은 다소 다르지만, 일반적으로 80%이다.¹⁾ 또한 학자들마다 위의 통계적 검정력 수준 및 결과의

1) 예를 들어, 임상연구의 경우에는 연구결과의 중대성을 감안하여 0.9 이상의 통계적 검정력을 선호하는 편이다. 물론 통계적 검정력과 표본크기와의 관계가 trade-off임을 고려할 때 통계적 검정력을

일반화 가능성을 고려하여 통계기법별로 다양한 표본 크기를 제시하였는데(Hair, Black, Babin, Anderson, & Tatham, 2006), 본 연구에서는 실증연구에서 가장 많이 활용되는 회귀분석, (다변량)분산분석, 구조방정식 모델(SEM)을 중심으로 표본크기에 대한 학자들의 일반적 견해를 종합적으로 알아보고자 한다.

첫째, 회귀분석의 경우 단순 회귀분석은 표본 크기가 20인 경우에도 유의한 효과가 나올 가능성이 충분하다고 본다. 독립 및 종속변수가 각각 하나씩인 단순 회귀분석의 경우에는 변수의 수 대 표본크기의 비율을 1 : 10만 유지해도(물론 대표성이 있는 경우에) 유의한 결과가 나올 수 있다는 것이다. 하지만 독립변수가 둘 이상인 다중 회귀분석의 경우에는 통계적 검정력을 0.8이상을 유지하려면 최소 50이상이 되어야 한다고 본다. 하지만 50이라는 표본크기도 최소의 크기일 뿐 대부분의 연구 상황에서는 표본크기가 100이상이 되어야 안정적이라고 보는 보수적인 견해가 지배적이다.

또한 표본 크기는 통계적 검정력 뿐 아니라 결과의 일반화에도 영향을 미치기 때문에 통계적 결론 타당성을 위한 적정 표본크기를 독립변수의 수와 표본의 수의 비율을 최소 1 : 5로 봐야 한다는 입장이 있다. 반면 전통적으로는 1 : 10의 비율이 최소 비율이라는 주장도 존재한다(Halinski & Feldt, 1970; Miller & Kuncce, 1973). 하지만 실제 연구에서는 1 : 15 ~ 1 : 20이 바람직한 수준이라는 의견 또한 만만치 않다. 본 연구에서는 최소 표본 수를 50 이상으로 또한 독립변수의 수와 표본의 수의 비율을 전통적인 관점에 따라 최소 1 : 10으로 보고자 한다.

둘째, (다변량)분산분석의 경우에도 다른 다변량 기법과 마찬가지로 표본크기에 영향을 받게 되는데, 다른 기법들과의 차이점은 표본크기와 관련된 요구조건이 전체적인 표본크기 그 자체보다도 개별 그룹(또는 셀)의 크기와 관련되어 있다는 것이다. 먼저 그룹당 표본의 수가 반드시 종속변수의 수보다 커야 한다. 이는 그룹 간 비교를 원칙으로 하는 분산분석의 특성상 당연하다 하겠다. 또한 각 그룹별로 최소 20개 이상의 표본이 할당되어야 한다. 마지막으로 그룹별 표본의 크기는 가급적 일치시키는 것이 좋다. 이는 그룹 중 최소 표본크기에 의해 분석의 효과성이 좌우되기 때문이다. 본 연구에서는 그룹별 최소 표본크기를 20으로 보고자 한다. 따라서 2×2 요인설계 일 경우 최소 표본크기는 80이 되어야 하는 것이다.

크게 할수록 표본크기가 커지는 것은 불가피한 일이라 하겠다(Eng, 2003).

마지막으로 구조방정식모델의 경우를 살펴보면, 구조방정식이란 다수의 변수 간 관계를 설명하기 위한 통계적 모델의 묶음(family)으로 일련의 방정식에 나타난 상호연관성의 구조를 평가하는데, 이 방정식들은 분석에 포함된 모든 구성개념(즉, 독립 및 종속 변수)간의 관계를 나타낸다. 따라서 구조방정식에서는 모델의 복잡성과 측정 특성에 따라 요구되는 표본크기가 첫째, 모델이 다섯 개 이하의 구성개념을 포함하고 있고, 각 개념이 3개 이상의 항목(item)으로 측정이 되었으며, 측정도구의 신뢰도가 0.6이상이면 100 ~ 150의 표본 수가 적합하고 본다. 둘째, 만약 어느 하나의 신뢰도가 0.45 ~ 0.55이거나 항목이 3개 이하인 구성개념을 포함하고 있으면 200개 이상이 요구된다. 셋째, 만약 신뢰도가 더 낮거나 많은 구성개념의 항목이 3이하인 경우(즉, 아직 구체적으로 확인 되지 않은 구성개념이라고 판단), 최소 300이상이어야 한다고 여겨진다. 넷째, 만약 요인이 6개 이상이고, 이 중 일부가 예측변수로서 3개 이하의 항목을 가지고 측정했으며, 다수의 낮은 신뢰도가 존재한다면, 표본 크기는 500을 넘겨야 한다는 주장이 있다. 한편 표본크기를 미지수와 의 비율로 제시한 연구자들도 있는데, 예를 들어 Bentler & Chou(1987)의 경우에는 미지수의 5 ~ 10배 정도의 표본크기가 필요하다고 하였다. 또한 MacCallum, Brown, & Sugawara(1996)는 통계적 검정력의 수준이 0.8이 되는데 필요한 표본수를 결정할 수 있는 방법으로 검정력 분석 방법을 제안하기도 하였다. 이처럼 구조방정식의 특성 상 조건에 따라, 또한 학자에 따라 다양한 표본크기가 요구되지만, 일반적으로 통계적 검정력을 판단함에 있어서 구조방정식의 경우 표본크기가 200이상이면 바람직하다고 인정되고 있다(Kline, 2005).

Ⅲ. 최소 표본크기 기준 미 충족 시의 문제점

연구를 진행함에 있어서 연구자가 가장 관심을 기울이는 부분이 귀무가설이 기각되어야 할 때 기각되는, 옳은 결정을 하는 것으로 이를 통계적 검정력이라 한다. 이 때 통계적 검정력 분석을 위해 기본적으로 이용하는 3가지 개념이 있는데, 이는 표본크기(sample size), 유의수준(significance criteria), 효과크기(effect size)이다(Cohen, 1992). 어떤 통계적 방법에도 이들은 서로 관련이 되어 있다. 즉, 연구를 계획함에 있어서 원하는 유의수준과 효과크기, 그리고 통계적 검정력이 주어지면 필요한 표본크기를

결정할 수 있다는 것이다. 이를 다시 말하면 표본의 크기가 유의수준, 통계적 검정력, 효과크기에 영향을 미칠 수 있음을 의미한다. 유의수준이란 앞에서 설명한 것처럼 귀무가설을 기각하게 되는 확률의 최대 위험수준을 말하며, 통상 0.05 또는 0.01을 사용한다. 표본크기와 유의수준과의 관계는 Type I 오류를 이용한 표본크기 결정 방법에서 이미 설명하였기 때문에 이하에서는 표본크기가 통계적 검정력 및 효과크기에 미치는 영향에 대해 알아보고자 한다.

1. 통계적 검정력에 미치는 영향

모집단을 대표하는 표본의 선정이 연구의 외적 타당도와 관련된 개념이라면, 표본에서 추출한 자료를 가지고 모집단의 특성인 모수를 추정함에 있어서 추론의 타당성을 논하는 것은 통계적 결론 타당도라 한다(Cook & Campbell, 1979). 예를 들어 실험연구에서 실험집단과 통제집단 간 실험변수 처리에 의한 실제적인 영향력이 없는데도 두 집단 간 차이가 있다고 할 가능성이 존재하는데, 이런 경우에는 통계적 결론 타당도가 떨어지는 것이다. 통계적 결론 타당도를 위협하는 요소 중 하나로 낮은 통계적 검정력이 있다. 통계적 검정력이란, 효과크기가 0이 아니고, 귀무가설이 기각되어야 할 때 기각되지 못하는 오류가 발생할 확률을 β 라고 한다면, 귀무가설이 기각되어야 할 때 기각되는 옳은 결정을 할 확률을 말한다($1-\beta$)(Aron & Aron, 2003). 혹은 통계적 검정력을 실제 유의한 차이가 존재할 경우 그것을 정확히 짚어 주는 민감도(sensitivity)라고 할 수도 있다고 하였다(Eng, 2003).

표본의 크기가 통계적 검정력에 미치는 영향을 살펴보면, 표본이 커질수록 표본 평균 분포의 표준편차가 줄어들게 되고, 표준편차가 줄어들면 분포의 폭이 좁아지며, 이는 곧 귀무가설과 연구가설에 의한 분포 사이의 중첩이 줄어들어 드는 것을 의미한다. 즉 베타오류가 점차 작아짐에 따라 통계적 검정력인 $1-\beta$ 가 커진다는 것이다. 이는 대수의 법칙을 말하는 것으로 표본의 크기가 커질수록 통계량이 모수에서 벗어난 정도, 즉 오차의 폭이 줄어들음을 의미하는 것이다.

물론 유의수준도 통계적 검정력에 영향을 준다. 만약 다른 조건이 동일하다면 알파를 크게 할수록 상대적으로 베타는 작아져 그 결과 통계적 검정력($1-\beta$)은 높아지기 때문이다. 하지만 통계적 검정력을 크게 하려고 알파를 높일 수는 없다. 왜냐하면 알파, 베타

모두 가능한 한 줄여야 할 오류이기 때문이다. 따라서 일반적으로 연구자들은 알파를 늘리기보다는 연구조건을 통제하여 분산을 줄인다든지 아니면 표본의 크기를 크게 한다는 지 하는 방법을 택하는 것이 좋다고 한다(박정식·윤영선, 2002; 채서일, 2003). 그런데 또 한 가지 유의할 점은 일반적으로 통계적 검정력은 $0.8(\beta=0.2)$ 을 사용하는데, 만약 베타오류를 줄여 통계적 검정력을 0.8이상으로 높이려면 필요한 표본크기가 급격하게 늘어난다는 것이다(Cohen, 1992).

연구자의 관심이 모집단의 특성인 모수의 보다 정확한 추론에 있다고 할 때, 바른 추론이 가능하도록 하기 위해서는 가급적 큰 표본의 수를 확보해야 함은 당연하다 하겠다. 물론 이런 경우에도 대표성 있는 표본의 선정은 필수적이다. 하지만 통계적 검정력을 높이기 위해서 표본의 무작정 크게 한다는 것은 어렵다. 이는 시간과 비용이 과도하게 소요되기 때문이다. 하지만 표본 크기는 연구자가 비교적 쉽게 통제할 수 있는 요인이라는 점에서, 또한 통계적 검정력을 충족시켜 연구목적의 달성해야 한다는 당위성 차원에서, 모집단의 일부로써 관심 있는 값에 일치하는 정보를 제공하는 대표성 있는 적정 수준의 표본 선정을 위한 표본설계가 잘 이루어져야 한다는 점은 거듭 강조되어도 지나치지 않다 하겠다(Festinger & Katz, 1953).

2. 효과크기에 미치는 영향

효과크기란 귀무가설이 기각되는 정도가 얼마나 큰지를 말한다. 예를 들어 분산분석에서 두 개 이상의 집단의 평균을 비교하여 모든 집단의 평균이 같다는 귀무가설이 기각되면, 이는 곧 집단 간 평균이 다름을 의미하는데 여기서 과연 집단 간 평균이 어느 정도 다른지 그 크기에 의문을 갖게 된다. 이 때 각 집단평균의 서로 다른 정도를 측정하는 척도가 바로 효과크기이다(Cohen, 1992). 각 통계기법은 각각의 효과크기 지수(index)를 가지고 있는데, Cohen이 제시한 일부 통계기법별 효과크기는 다음 <표 1>과 같다.

효과크기는 비교되는 값을 표준화 시킨 것으로, 예를 들어 두 집단의 평균차 검증인 t-검정인 경우 효과크기가 0.5라는 의미는 두 집단 간 평균값의 차이가 표준편차의 1/2이라는 의미이다. 또한 상관관계 r의 경우에는 실제 상관관계 값 자체가 효과크기가 된다. 따라서 만약 상관관계가 없다는 귀무가설이 참인 경우에는 효과크기는 0($r=0$)이 되는 것이다. 또한 일원 분산분석의 경우 효과크기는 집단 평균의 표준편차를 집단 공통표준

〈표 1〉 통계기법별 효과크기

통 계 기 법	효 과 크 기		
	작 음	중 간	크 음
t - 검정	0.20	0.50	0.80
상관계수 r 유의성 검정	0.10	0.30	0.50
일원 분산분석	0.10	0.25	0.40
중다회귀분석	0.02	0.15	0.35

자료원: Cohen(1992) 수정인용.

편차로 나눈 값을 말하는 것으로 이는 에타(η)와 같다. 따라서 집단 간 평균의 차가 없다는 귀무가설이 참인 경우 집단 평균의 표준편차가 0이므로 효과크기는 영이 된다.

그런데 효과크기는 작을 때보다 클 경우 연구에 의해 발견될 가능성이 더 높다. 따라서 효과크기가 클수록 통계적 검정력에 더 큰 영향을 주게 되는 것이다. 효과크기와 표본크기와의 관계를 알아보면, 효과크기가 커질수록 유의한 차이, 즉 통계적 검정력이 커지기 때문에 비교적 표본크기가 작아도 됨을 의미한다. 반대로 효과크기가 작을수록 유의한 효과를 얻기 위해서는 큰 표본크기를 요구하게 되는 것이다. 이를 환언하면 표본크기가 커지면 커질수록 작은 효과크기도 확인이 가능함을 의미한다. 따라서 대단히 큰 표본크기를 가지고 연구를 진행하면 검증하지 못할 효과가 없는 것이다. 반대로 표본의 크기가 굉장히 작으면 통계적으로 유의한 결과가 나올 확률이 영에 가까워지게 된다. 이는 다시 위에서 살펴본 통계적 검정력과 연결된다. 즉, 통계적 검정력을 높일 경우 표본의 크기가 급격히 증가된다는 점과 일맥상통하는 것이다.

지금까지 살펴본 통계적 검정력 분석과 관련된 유의수준, 표본크기, 효과크기 간의 관계는 사실 매우 복잡하다. 하지만 Cohen(1992)이 통계적 검정력의 수용수준을 80%로 제시한 이래, 대부분의 연구에서 이를 적용하고 있다(Kerlinger & Lee, 2000; Hair et al., 2006). 따라서 통계적 검정력 80%를 유지하기 위해서는 유의수준, 효과크기, 그리고 표본크기가 동시에 고려되어야 하는 것이다. 유의수준, 효과크기, 표본크기, 검정력이 유기적으로 영향력을 미친다는 사실을 이용해서, 위의 4가지 중 3가지가 결정되면 나머지 하나의 값을 산출할 수 있다. 이러한 계산을 위한 통계적 검정력 프로그램인 GPower 3.0.10.을 이용해서 표본크기가 변함에 따라 달라지는 검정력의 크기를 정리해 보았다. Cohen(1992)의 통계기법별 효과크기를 기준으로 유의수준과 효과크기가

주어졌을 때 표본크기가 변함에 따라 통계적 검정력이 어떻게 변화하는지, 각 통계기법 별로 정리해 보았다.

〈표 2〉 독립표본 t검정 시 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량

표본당 표본크기	유의수준 0.05			유의수준 0.01		
	효과크기			효과크기		
	작음 (0.20)	중간 (0.50)	큼 (0.80)	작음 (0.20)	중간 (0.50)	큼 (0.80)
20	0.095	0.338	0.693	0.025	0.144	0.438
40	0.143	0.598	0.942	0.045	0.349	0.823
60	0.192	0.775	0.991	0.067	0.549	0.959
80	0.242	0.882	0.999	0.093	0.710	0.992
100	0.291	0.940	1.000	0.120	0.824	0.999
120	0.339	0.971	1.000	0.150	0.898	1.000
140	0.385	0.986	1.000	0.181	0.943	1.000
160	0.430	0.994	1.000	0.213	0.969	1.000
180	0.473	0.997	1.000	0.246	0.984	1.000
200	0.514	0.999	1.000	0.280	0.992	1.000

위의 표의 결과를 간략히 해석해 보면 유의수준이 5%이고 효과크기가 작음(0.20)인 경우에는 한 집단의 표본크기가 약 200이상이 되어도 통계적 검정력 80%이상을 확보할 수 없지만, 효과크기가 중간(0.50)인 경우에는 한 집단의 표본크기가 약 80이상, 즉 두 집단을 합친 총 표본크기가 160이 넘어야 하며, 큼(0.80)인 경우 한 집단의 표본크기가 약 40이상, 총 표본크기는 80이상이어야 총 통계적 검정력이 80% 이상이 됨을 알 수 있다. 또한 유의수준이 1%이고 효과 크기가 작은(0.20) 경우에는 한 집단당 200 이상의 표본을 보유해도 80% 이상의 통계적 검정력을 확보할 수 없지만, 효과크기가 중간(0.50)인 경우에는 총 표본크기가 200이상, 효과크기가 큰(0.80) 경우에는 총 표본크기가 80이상일 경우 80% 이상의 통계적 검정력을 확보할 수 있는 것으로 나타났다. 즉 이를 표본크기 입장에서 표현하면 표본크기가 커질수록 작은 효과크기도 유의하게 나타남에 따라 결과적으로 통계적 검정력이 증가한다고 볼 수 있다. 따라서 표본의 크기는 연구 설계 및 결과해석에 있어 연구자로 하여금 보다 많은 정보를 얻을 수 있게

해주는 중요한 결정요소가 된다. 이와 같은 결론은 t검정 외에 본 연구에서 살펴볼 통계 기법인 일원분산분석, 다원분산분석, 중다회귀분석을 시행할 경우에도 비슷한 결과를 도출할 수 있다.

〈표 3〉 일원분산분석 시 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량

전체 표본크기 (집단 수: 3개)	유의수준 0.05			유의수준 0.01		
	효과크기			효과크기		
	작음 (0.10)	중간 (0.25)	큼 (0.40)	작음 (0.10)	중간 (0.25)	큼 (0.40)
60	0.095	0.374	0.776	0.025	0.168	0.544
90	0.121	0.540	0.926	0.035	0.297	0.791
120	0.148	0.676	0.979	0.046	0.432	0.919
150	0.176	0.780	0.995	0.059	0.559	0.972
180	0.204	0.855	0.999	0.072	0.669	0.991
210	0.232	0.907	1.000	0.087	0.759	0.998
240	0.261	0.941	1.000	0.103	0.829	0.999
270	0.290	0.964	1.000	0.120	0.881	1.000
300	0.319	0.978	1.000	0.137	0.919	1.000

〈표 3〉은 독립변수가 3개의 집단으로 나누어지고 3개의 집단에 동일한 표본수가 분산되어있다는 가정을 충족한 일원분산분석을 실시하는 경우를 나타낸다. 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량을 살펴보면 유의수준이 5%이고 효과크기가 중간(0.25)일 경우에는 총 표본의 수가 180이상, 효과크기가 클(0.40) 경우 총 표본의 수가 90이상이어야 80%이상의 통계적 검정력을 확보할 수 있는 것을 알 수 있다. 또한 유의수준이 1%이며 효과크기가 중간(0.25)일 경우 총 표본수가 240이상, 효과크기가 클(0.40) 경우에는 총 표본수가 120이상이면 80%이상의 통계적 검정력을 확보할 수 있다는 결론을 내릴 수 있다.

〈표 4〉는 독립변수가 2개의 요인에 의하여 각 3개의 집단으로(총 9개의 집단) 나누어진 상태에서 집단 간 차이를 분석하기 위해 다원분산분석을 실시할 때의 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량을 나타낸다. 유의수준이 5%이고 효과크기가 중간(0.25)일 경우 총 표본의 수가 270, 효과크기가 클(0.40) 경우 총 표본수가

〈표 4〉 다원분산분석 시 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량

총 표본크기 (총 셀수: 3, 분모자유도: 4일때)	유의수준 0.05			유의수준 0.01		
	효과크기			효과크기		
	작음 (0.10)	중간 (0.25)	큼 (0.40)	작음 (0.10)	중간 (0.25)	큼 (0.40)
90	0.095	0.419	0.855	0.025	0.200	0.661
180	0.154	0.760	0.995	0.049	0.534	0.975
270	0.218	0.922	1.000	0.079	0.790	0.999
360	0.286	0.979	1.000	0.117	0.922	1.000
450	0.355	0.995	1.000	0.160	0.975	1.000
540	0.423	0.999	1.000	0.208	0.993	1.000
630	0.488	1.000	1.000	0.259	0.998	1.000
720	0.550	1.000	1.000	0.313	1.000	1.000
810	0.608	1.000	1.000	0.368	1.000	1.000
900	0.661	1.000	1.000	0.423	1.000	1.000

90이상이면 80%이상의 통계적 검정력을 확보할 수 있지만 유의수준이 1%이고 효과크기가 중간(0.25)일 경우 총 표본수가 360이상, 효과크기가 클(0.40) 경우에는 총 표본수가 180이상이면 80%이상의 통계적 검정력을 확보할 수 있다.

〈표 5〉는 독립변수가 2개일 때 중다회귀분석을 실시할 때의 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량을 나타낸다. 유의수준이 5%이고 효과크기가 중간(0.02)일 경우 총 표본의 수가 80이상, 효과크기가 클(0.35) 경우 총 표본수가 40이상이면 80%이상의 통계적 검정력을 확보할 수 있지만, 유의수준이 1%이고 효과크기가 중간(0.15)일 경우에는 총 표본수가 100이상, 효과크기가 클(0.35) 경우에는 총 표본수가 60이상이면 80%이상의 통계적 검정력을 확보할 수 있다.

〈표 6〉은 두 변수의 Pearson의 상관관계수를 이용해서 상관관계를 분석할 때 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량을 나타낸다. 유의수준이 5%이고 효과크기가 중간(0.30)일 경우 변수당 표본의 수가 180이상, 효과크기가 클(0.50) 경우 변수당 표본수가 90이상이면 80%이상의 통계적 검정력을 확보할 수 있다. 유의수준이 1%일 때 효과크기가 중간(0.30)일 경우 변수당 표본수가 270이상, 효과크기가 클(0.50) 경우에는 변수당 표본수가 120이상이면 80%이상의 통계적 검정력을 확보할 수 있다.

〈표 5〉 중다회귀분석 시 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량

총 표본크기 (독립변수: 2)	유의수준 0.05			유의수준 0.01		
	효과크기			효과크기		
	작음 (0.02)	중간 (0.15)	큼 (0.35)	작음 (0.02)	중간 (0.15)	큼 (0.35)
20	0.076	0.275	0.573	0.018	0.100	0.297
40	0.110	0.547	0.906	0.030	0.294	0.736
60	0.145	0.747	0.985	0.045	0.507	0.933
80	0.182	0.870	0.998	0.061	0.686	0.987
100	0.220	0.937	1.000	0.080	0.815	0.998
120	0.258	0.971	1.000	0.100	0.897	1.000
140	0.297	0.987	1.000	0.123	0.946	1.000
160	0.335	0.995	1.000	0.146	0.973	1.000
180	0.373	0.998	1.000	0.172	0.987	1.000
200	0.410	0.999	1.000	0.198	0.994	1.000

〈표 6〉 상관관계분석 시 표본크기, 유의수준, 효과수준에 따른 통계적 검정력 변화량

변수 당 표본크기	유의수준 0.05			유의수준 0.01		
	효과크기			효과크기		
	작음 (0.10)	중간 (0.30)	큼 (0.50)	작음 (0.10)	중간 (0.30)	큼 (0.50)
30	0.066	0.197	0.451	0.015	0.070	0.230
60	0.083	0.360	0.761	0.022	0.165	0.537
90	0.101	0.507	0.910	0.028	0.275	0.765
120	0.119	0.631	0.969	0.035	0.389	0.894
150	0.138	0.730	0.990	0.043	0.498	0.956
180	0.156	0.806	0.997	0.051	0.597	0.983
210	0.174	0.863	0.999	0.060	0.683	0.994
240	0.193	0.904	1.000	0.069	0.755	0.998
270	0.211	0.934	1.000	0.078	0.813	0.999
300	0.230	0.955	1.000	0.087	0.860	1.000

IV. 표본크기 관련 국내 연구 실태

1. 실증연구 분석

모든 연구문헌을 대상으로 적정 표본크기의 적합성 여부를 분석하는 것은 바람직 할 것이다. 하지만 본 연구의 목적이 모든 문헌 하나하나에 대한 표본크기 적절성 여부를 분석하는 것이 아니라 실증연구에 있어서 표본크기의 중요성을 재고시키는 데 있기 때문에 표본크기가 미 충족된 국내 문헌이 존재하는가를 밝히는 것만으로도 연구 목적을 달성하기에 충분하다 하겠다. 그에 따라 통계기법 중 가장 많이 사용되는 몇 가지 기법을 선정하여 그 기법을 이용한 논문을 중심으로 표본크기 적절성 여부를 확인하는 방식을 택하고자 한다. 또한 사용되는 통계기법의 빈도 또는 비율이 얼마 이상이 되어야 많이 사용된다는 것을 객관적으로 인정해주는 기준은 없지만, 본 연구에서는 전체 실증연구 중 특정 통계기법들을 이용한 논문의 수가 70%를 넘을 경우 해당 통계기법의 다용성을 수용하여 연구를 지속하고자 한다. 그에 따라 본 연구에서는 인사·조직 분야의 대표적 학술지로서 속성상 설문지를 이용한 현장조사 연구가 많은 인사·조직 연구 및 경영학 연구를 채택하였으며, 실증연구에서 가장 많이 사용되는 통계기법 중 일부인 다중회귀분석, (다변량)분산분석, 구조방정식 모형을 대상 통계기법으로 선정하고 인사·조직 및 경영학 연구에 1998년부터 2007까지 10년 간 게재된 논문을 대상으로 사용된 통계기법을 분석하였다. 그 결과는 아래와 같다.

〈표 7〉 1998 ~ 2007년간 게재된 인사조직 연구 논문 수

구 분	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07	계	
총 논문 수	19	10	12	19	16	15	25	17	19	23	175	
실증연구 수	11	8	11	17	12	13	19	13	16	16	136	
주요 통계 기법	다중회귀분석	5	3	5	9	8	7	13	8	11	6	75
	(M)ANOVA	2	0	3	3	1	1	0	0	1	1	12
	SEM	3	2	1	2	0	3	3	3	1	4	22
	소 계	10	5	9	14	9	11	16	11	13	11	109

〈표 8〉 1998 ~ 2007년간 게재된 경영학 연구 논문 수

구 분	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07	계	
총 논문 수	55	48	51	54	78	71	70	73	73	68	641	
실증연구 수	21	37	47	48	75	61	63	65	62	56	535	
주요 통계 기법	다중회귀분석	13	9	19	19	33	31	24	30	25	23	226
	(M)ANOVA	4	7	4	6	10	2	6	7	7	2	55
	SEM	2	9	11	7	14	13	18	3	16	14	107
	소 계	19	25	34	32	57	46	48	40	48	39	388

분석결과, 인사조직 연구의 경우 실증연구는 136편으로 총 175편 논문 대비 77.7%였다. 이 중 다중회귀분석, (M)ANOVA, SEM을 이용한 연구 수는 총 109편으로(각 75, 12, 22편) 실증연구 수 대비 80.1%를 차지하였다(각 55.1%, 8.8%, 16.2%). 경영학 연구는 총 641편 중 실증연구는 535편으로 83.5%였으며, 상기 3가지 주요통계기법을 이용한 연구 수는 388편으로(각 226, 55, 107편) 실증연구 대비 72.5%를 차지하였다(각 42.2%, 10.3%, 20%). 위에서 살펴본 바와 같이 실증연구 대비 위의 3가지 주요통계기법을 이용한 연구 문헌의 수가 각 학술지 별로 70%를 넘는다는 것을 알 수 있다. 그에 따라 위의 3가지 통계기법을 이용한 연구 문헌을 대상으로 연구를 진행함에 무리가 없을 것으로 판단된다.

위에서 살펴본 통계기법별 최소 표본크기 내용 중 본서에서 적용할 기준을 정리하면 다음과 같다. 이하에서는 본 기준을 토대로 국내 연구문헌에 대한 표본크기의 적절성 여부를 분석해 보고자 한다. 분석에 앞서 연구 목적에 따라 모든 연구의 표본은 단순무작위 표본추출방법에 의한 것임을 가정한다.

〈표 9〉 통계기법별 최소 표본크기 기준

구 분	표본 크기	비 고
회귀분석	1 : 10이상	독립변수 당 표본 수의 비율
(다변량)분산분석	20이상	그룹(또는 셀)당 최소 표본 수
구조방정식 모델	200이상	

통계기법별 최소 표본크기 기준에 따라 회귀분석, (다변량)분산분석, 구조방정식 모델

을 이용하여 논문을 분석한 결과, 인사조직 연구의 경우 회귀분석을 사용한 총 75편의 논문 중 10편이 최소 표본크기를 충족시키지 못하였으며, (다변량)분산분석은 총 12편 대비 4편이, 구조방정식 모델은 총 22편 중 7편이 최소 표본기준을 넘지 못하였다. 경영학 연구의 경우를 살펴보면 총 226편의 회귀분석 사용 논문 중 31편이, (다변량)분산분석 기법을 사용한 논문은 총 55편 중 16편이, 구조방정식 모델의 경우에는 총 107편 중 25편이 최소 표본기준을 넘지 못하는 것으로 나타났다. 학술지별로 통계기법에 따른 최소 표본크기 미 충족 논문 실태는 다음과 같다.

〈표 10〉 학술지별 통계기법에 따른 최소 표본크기 미 충족 논문 현황

학술지	통계기법	총 논문 수	미 충족 논문 수	비율 ^a	총 비율
인사조직 연구	회귀분석	75	10	13.3	19.3
	(M)ANOVA	12	4	33.3	
	SEM	22	7	31.8	
경영학 연구	회귀분석	226	31	13.7	18.6
	(M)ANOVA	55	16	29.1	
	SEM	107	25	23.4	

주) ^a: 기법별 최소 표본크기 미 충족 논문 수 / 통계기법별 논문 수 × 100%

이 중 통계기법별 최소 표본크기 미 충족 논문의 일부를 아래와 같이 예시하였다. 아래에 예시한 논문은 특별한 기준에 의해 선정된 것은 아니며, 단지 표본크기에 대한 연구자들의 인식 부족이 얼마나 만연된 문제인가를 상기시키기 위한 것으로써 비판의 의도는 전혀 없음을 밝혀둔다(따라서 참고문헌의 목록에서도 제외하였음을 양해바람).

이상에서 표본 크기가 미 충족된 논문들을 살펴보면 대체로 표본의 단위가 집단 또는 조직수준인 경우가 많았다. 이는 개인 수준을 넘는 표본단위를 대상으로 하는 연구에서는 모집단 자체가 제한적이어서 가용 표본의 수 또한 제약을 받는다는 점과 개인을 대상으로 하는 연구에 비해 집단 이상 수준의 표본 확보는 비용이 급증하기 때문일 것이다. 따라서 연구자들은 표본의 크기를 증대시킴에 따라 얻어지는 정밀도 향상과 그 같은 대규모 표본에 수반되는 비용의 증대를 고려하여, 통계적 결론 타당도가 높은 그러나 현실적인 경제적 문제 또한 고려되는 최소한의 표본으로 최대의 통계적 검정력을 확보하는 것이 필요하리라 생각된다(Stone, 1978).

〈표 11〉 통계기법별 최소 표본크기 미 충족 사례 (「인사·조직연구」)

년 도	논 문 제 목	통계 기법	변수(셀) 수 ^a	표본 크기 ^b	기준 (최소표본)
1998	조직정의와 조직몰입에 관한 연구: 다중 몰입을 중심으로	SEM (LISREL)	4	116	200
1999	조직변수와 산업변수가 경영성과전환에 미치는 효과	다중 회귀분석	8	70	82
2000	면접구조화가 면접의 신뢰성과 타당성에 미치는 영향에 관한 연구	ANOVA	2	5 (총 표본 수 11)	셀 당 20
2001	업무환경이 교육훈련의 전이에 미치는 영향	다중 회귀분석	8	58	80
2002	CEO 보상에 있어서 재무적 측정치와 비 재무적 측정치의 비중에 대한 연구	SEM (LISREL)	4	43	200
2003	인적자원관리와 조직성과에 대한 실증 연구	ANOVA	4	10 (총 표본 수 100)	셀 당 20
2004	중소기업의 기술학습과 기업성과: 우리나라 전자부품 산업 1990~1995	다중 회귀분석	18	108	180
2005	인지 및 동기요인이 지식근로자의 팀 성과와 혁신행동에 미치는 영향	SEM (AMOS)	6	121	200
2006	경영전략과 인적자원관리의 적합성에 따른 조직성과 차이	ANOVA	4	8 (총 표본 수 124)	셀 당 20
2007	집단성과 결정에 있어서 인적자본과 사회적 자본의 효과	다중 회귀분석	7	60	70

주) ^a: 회귀분석은 독립변수, (M)ANOVA는 셀, SEM은 측정변수의 수입

^b: (M)ANOVA의 경우 셀 중 최소 표본크기를 나타냄

두 번째로, 표본크기 부족에 대한 한계를 인정한 논문이 비교적 적다는 것이다. 또한 적정표본크기의 부족을 인정한 논문도 결과의 일반화가 제한된다는 측면만 언급했을 뿐, 부족한 표본크기에 의한 통계적 검정력 미흡 가능성 등에 대해서는 거의 관심을 기울인 연구가 없다. 이런 연구자들의 관심 미흡은 연구결과의 일반화 이전에 연구결과의 신뢰성 자체에 의구심을 품게 만들 수 있다. 예를 들어, 적정 표본크기가 되었다면 충분히

〈표 12〉 통계기법별 최소 표본크기 미 충족 사례 (「경영학연구」)

년 도	논 문 제 목	통계 기법	변수(셀) 수	표본 크기	기준 (최소표본)
1998	창업주의 사망과 기업가치의 관련성	다중 회귀분석	6	38	60
1999	EDI 도입 성과 결정요인 및 문제해결 방안에 관한 연구	ANOVA	3	10 (총 표본 수 46)	셀 당 20
2000	EDI성파에 영향을 미치는 요인에 관한 연 구: EDI 이용수준의 매개역할을 중심으로	다중 회귀분석	8	70	80
2001	국내기업간 전략적 제휴의 핵심성공요인 에 관한 연구: 상호호혜적 조정전략을 중심으로	SEM (LISREL)	6	156	200
2002	인사조직혁신에 대한 노조 간부들의 태 도에 영향을 미치는 요인: 사회적 네트 웍 관점을 중심으로	SEM (AMOS)	4	122	200
2003	국제다각화, 제품다각화 및 조직적 학습 이 기업의 가치에 미치는 영향	다중 회귀분석	9	58	90
2004	An application of hybrid technology acceptance model to object-oriented computing	SEM (LISREL)	5	109	200
2005	비즈니스 전략, 아웃소싱 전략 및 기업 성과: 구성적 관점의 탐색적 연구	ANOVA	3	11 (총 표본 수 57)	셀 당 20
2006	본-자회사간 혁신 노하우의 국제적 이전 을 결정하는 요인에 관한 연구	다중 회귀분석	15	78	150
2007	서비스 대기시간과 품질평가의 관계에 있어서 영향을 주는 요인들에 관한 연구	SEM (LISREL)	5	101	200

유의한 효과를 확인할 수 있었음에도 불구하고 적은 표본으로 인해 실제 효과를 없는 것으로 제시한 연구가 존재할 가능성이 있다는 것이다. 이는 연구결과의 일반화 보다 더 근본적인 문제라 하겠다.

V. 향후 연구를 위한 제언

국내 연구문헌을 대상으로 통계적 기법에 따른 최소표본 크기 기준 충족 여부 분석결과, 표본크기에 대한 연구자들의 관심이 이론적 수준에만 국한되어 있을 뿐, 실제 연구를 진행함에 있어서는 그리 중요하게 여기고 있지 않다는 것을 보여준다. 하지만 실제 존재하는 효과의 유의성 정도를 나타내고, 연구결과의 일반화를 가능케 하는 적정 표본 크기에 대한 요구는 아무리 강조해도 지나치지 않다. 그러나 문제는 연구를 설계함에 있어서 가장 먼저 부딪히는 질문 중 하나가 표본 크기를 얼마나 크게 해야 하는가이지만, 정확한 답변도 어려울 뿐 아니라, 혹 해결책이 있다 하더라도 연구자마다 연구목적에 따라 다양한 표본크기 선정공식을 사용함에 따라 연구자들의 혼란만 가중시키고 있는 실정이다. 본 연구에서도 구체적인 표본크기 선정 공식의 제시로 이런 혼돈을 가중시키고자 하지 않는다. 구체적인 표본크기 선정을 위한 방법은 개별 연구자들의 영역에서 해결되어야 할 문제이기 때문이다. 하지만 연구목적에 따라 표본의 크기는 다를 수 있어도, 모든 연구에서 공통적으로 충족되어야 할 최소한의 원칙은 존재하기 마련이다. 이하에서는 그런 표본 크기 결정과 관련하여 많은 학자들이 수용하는 일반적 틀을 제시하여 향후 연구를 돕고자 한다.

1. 표본 설계의 중요성

표본 설계는 모집단의 확정, 표본프레임의 결정, 표본추출방법의 결정, 표본크기의 결정, 실제 표본추출에 이르기까지 전체적인 연구 설계에 있어 표본과 관련된 제반 문제를 다루는 부분이다. 그런데 전수조사가 아닌 표본을 통해 모집단의 특성을 추론하는 점에서 많은 오류 발생 가능성이 존재한다. 따라서 모집단을 대표할 수 있는 표본을 선정할 수 있도록 하는 좋은 표본 설계는 연구 설계에 있어 가장 중요한 부분 중 하나라 할 수 있다.

표본 설계는 먼저 정보를 얻으려는 대상 집단인 모집단을 확정하는 것에서부터 시작된다. 의사결정을 함에 있어서 유용한 정보를 얻기 위해서는 가능한 모집단을 명확히 하는 것이 중요하다. 모집단을 정밀하게 규정하지 못하면 이후 표본 선정 자체가 애매해지며, 표본을 선정하여 연구가 진행되었다 하더라도 연구결과의 신뢰성과 타당성의 확보가 불

투명해지기 때문이다. 두 번째로, 일단 모집단이 확정되면 표본프레임(sampling frame)을 선정해야 한다. 표본프레임이란 연구대상이나 표본단위가 수록된 목록을 말하는데, 이로부터 최종적인 표본이 추출된다. 예를 들어 모집단이 특정학교 학생이라면 학생명부가, 특정 학술지 논문이 모집단이라면 그 학술지에 게재된 논문제목 리스트가 표본프레임이 된다. 물론 대부분의 연구에서 모든 모집단의 구성단위를 수록한 표본프레임을 확보하기란 현실적으로 불가능하다. 따라서 모집단과 표본프레임간 불일치로 발생하는 표본프레임 오차가 존재하기 마련이다. 그렇다고 오차를 미리 고려해서 표본프레임 없이 연구를 진행하라는 의미는 아니다. 중요한 것은 가급적 표본프레임 오차를 줄이기 위해 모집단과 표본프레임을 일치시키는 노력이 필요하다는 것과, 부적절한 통계적 추론을 하지 않기 위해서는 표본프레임 오차를 연구자가 분명히 인식하여 추론결과의 일반화 과정에 주의를 기울여야 한다는 것이다. 이렇게 표본프레임이 확보되면, 표본프레임 내에 있는 연구대상 중 일부를 어떻게 결정할 것인지를 정하게 되는데 이를 표본추출방법의 결정이라 한다. 표본추출방법은 크게 표본 추출확률이 알려져 있고 무작위로 추출하는 확률표본 추출방법과 이러한 특성을 갖지 않는 비확률표본 추출방법으로 나누어지는데, 연구자가 모집단으로부터 얻은 정보를 바탕으로 어떤 방법으로 표본을 추출할 것인지를 결정한다. 그 다음으로 표본크기 결정방식에 따라 실제 표본크기를 결정하여 실제 표본을 추출하면 되는 것이다.

지금까지 살펴본 표본 설계의 절차는 모두 모집단을 대표하는 표본의 선정이라는 목표를 지향한다. 비록 표본을 이용한 연구가 통계적 추론을 위해 효율적인 의사결정 정보를 제공해 주지만, 표본이 모집단을 대표하지 못하면 표본을 통해 획득한 정보는 아무런 가치가 없을 뿐 아니라 오히려 연구결과를 왜곡시킬 수 있기 때문이다. 따라서 이전 연구에서 활용 가능한 충분한 정보가 있거나 사전조사를 통해 정보를 얻은 경우를 제외하고는 연구초기에 표본크기를 정하는 것은 피하고 위에서 제시한 표본 설계 절차를 따르는 것이 바람직하다(Williams, 1978).

2. 통계적 검정력 및 효과크기를 고려한 표본 크기 결정

표본크기가 통계적 검정력, 유의수준, 효과크기와 관련되어 있다는 것은 앞에서 이미 살펴보았다. 연구자는 연구를 진행함에 있어서 원하는 수준의 통계적 검정력 또는 모수

추정의 정확성 달성 가능 여부(혹은 둘 다)에 기반을 두고 연구를 계획해야 하는데, 이때 우선고려 대상이 표본크기이다. 또한 효과크기에 따라 표본크기가 달라지는 점도 간과할 수 없다. 따라서 연구자는 통계적 검정력과 효과크기를 고려한 표본크기를 결정해야 한다.

연구자는 먼저 효과크기를 정하는 것이 좋다. 비록 효과크기가 통상 잘 알려져 있지 않고 또한 추정하기도 어려워 표본크기 계획에 있어 가장 신경쓰이는(troublesome) 부분이지만, 통계적 검정력을 확보하여 유의한 연구가 되기 위해서는 효과크기를 구체적으로 정할 필요가 있다. Polit & Sherman(1990)은 효과크기를 결정하는 방법으로 다음의 네 가지를 제시하였다. 첫째, 자신의 연구와 비슷한 선행연구를 통해 효과크기를 추정하는 것이다. 특히 관련 연구가 다수 일 경우에는 메타분석(meta-analysis)를 사용해서 추정하는 것이 더욱 바람직하다 하겠다. 둘째, 관련 연구가 거의 없는 경우에는 예비연구(pilot study)를 통해 효과크기를 추정하는 것도 좋은 방법이다. 셋째, 예비연구가 불가능할 경우에는 더미표(dummy table)을 만들어 실증적으로나 이론적으로나 가치가 있다고 간주될 수 있는 가장 작은 효과크기를 선정한다. 넷째, 과거 자신의 연구경험을 토대로 효과크기의 대(large), 중(medium), 소(small)을 결정하고 그에 따라 정해진 추정치를 사용하는 것이다(Cohen, 1988).

이렇게 정해진 효과크기와 일반적으로 수용되는 80%의 통계적 검정력 및 유의수준(5% 또는 1%)을 가지고 표본크기를 정하면 되는 것이다. 이 때, 검정력 분석을 위한 소프트웨어 프로그램을 이용하면 통계적 검정력과 효과크기, 그리고 유의수준을 고려한 적정 표본크기를 구할 수 있다.²⁾

그 다음으로 중요한 것은 연구자가 정한 통계적 검정력과 효과크기, 그리고 이들을 고려한 표본크기의 적절성 여부를 연구 논문에 명시하는 것이다. 많은 학술지에서 투고 논문의 수락 여부를 결정하는 암묵적 전제조건이 통계적 검정력이라 할 수 있는데(Maxwell, Kelley, & Rausch, 2008), 그 어떤 논문도 자신의 연구가 얼마만큼의 통계적 검정력을 가지는지, 어느 정도의 효과크기를 예상하는지, 그에 따라 적정 표본의 크기가 얼마인지를 명시한 경우가 없다. 비록 일반적으로 통계적 검정력을 80%로 한다 하더라도 과연 연구에서 사용된 표본의 크기가 그 수준을 충족시킬 수 있는 정도의 크기인지에 대

2) Kerlinger & Lee(2000)는 통계적 검정력 분석을 위한 윈도우 및 DOS용 프로그램을 제공하는 웹사이트 정보를 일부 제공하였다. (p. 294).

한 진위 여부는 연구자의 양심에 맞기거나 기본적으로 주어진(given) 것으로 여기는 분위기가 지배적이다. 이는 합리성, 객관성, 논리의 타당성을 중시하는 사회과학의 특성과는 다소 거리가 먼 현상이다.

더군다나 일반적으로 정확한 모수 추정을 위해 필요한 표본의 크기는 실제 존재하는 효과의 유의성을 나타내 주기에 필요한 표본의 크기보다 더 큰 경우가 많다. 이는 통계적 유의성 검증을 위해 요구되는 표본보다 연구결과의 일반화를 위해 필요한 표본의 크기가 더 커야 함을 의미한다(Maxwell et al., 2008). 이렇게 결과의 일반화 가능성을 높이는 표본크기를 가지고 진행된 연구는 통계적 결론 타당도가 높다. 연구자들은 일반적으로 단지 자신의 가설을 지지하는 통계적 검정력만 원하는 것이 아니라 일반화 가능성도 높은 연구결과도 추구한다. 따라서 실제 연구에 필요한 표본의 크기는 통계적 검정력 및 효과크기를 고려한 수준보다 더 큰 크기를 필요로 하게 된다. 즉, 통계적 검정력과 효과크기를 고려한 표본크기는 최저 기준이 되는 셈이다. 따라서 연구의 신뢰성을 제공하는 차원에서 통계적 검정력, 효과크기, 표본크기 적절성 여부를 논문에 포함시켜야 하는 것은 당연한 일이라 하겠다.

3. 통계기법별 최소 표본크기 기준 적용

통계기법별 최소 표본크기 기준은 위에서 설명한 통계적 검정력과 일반화를 고려한 보다 구체적인 표본크기의 기준을 의미한다. 우리는 위에서 회귀분석, (다변량)분산분석, 구조방정식 모델을 중심으로 최소 표본크기에 대한 학자들의 일반적 견해를 알아보았다(Hair et al., 2006). 국내 문헌 분석에서 볼 수 있었던 것처럼, 적지 않은 연구들이 통계기법별 최소 표본크기를 미 충족 하였으며, 이로 인해 통계적 결론 타당도에 위협이 초래될 가능성 또한 있음도 언급하였다. 따라서 연구자들은 가설 검정을 위해 자신이 사용하는 통계기법별 최소 표본크기를 명확히 알고 이 기준에 의해 표본 수를 정해야 하며, 그런 과정에 대해서도 논문에 명시하는 것이 바람직하다.

회귀분석, (다변량)분산분석, 구조방정식 모델 이외에도 실증연구에서 활용되는 다수의 통계적 기법에 적합한 최소 표본크기를 간략히 살펴보면, Hair et al.(2006)은 요인분석의 경우에는 표본크기가 최소 100은 넘어야 하거나 또는 설문 문항 수 대비 5배 이상은 되어야 한다고 하였으며, 판별분석의 경우에는 예측변수 대비 표본의 비율이 1 : 20

이 적당하고, 정준상관분석의 경우에는 변수 당 표본의 크기가 10배를 넘어야 한다고 하였다. 이 외에도 다양한 통계기법별 최소 표본크기에 대한 기준이 존재하는데, 여기서 중요한 점은 전통적으로 사용되었던 주먹구구식 방법이 아닌 통계적 검정력 및 연구결과 의 일반화를 고려하여 다수의 학자들이 인정하는 통계기법별 최소 표본크기를 따르는 것이 연구결과의 신뢰성과 타당성을 확보할 수 있는 안전한 경로라는 것이다.

4. 표본크기 관련 향후 연구 방향

지금까지 연구 진행 상 간과되어 왔던 표본크기에 대한 관심 증대를 언급하였다. 하지만, 본 연구에서 표본크기와 관련하여 모든 방안을 제시한 것은 아니다. 보다 신뢰할만 하고 타당성이 높은 연구를 위해서는 향후 연구를 통해 표본크기와 관련하여 보다 정밀하고 명확한 틀의 제공이 필요하다.

첫째, 본 연구에서는 최소 표본크기에 초점을 두었지만, 표본크기가 커지면 커질수록 효과크기가 아무리 작아도 통계적 검정력이 높아져 이는 유의한 결과를 나타낼 가능성이 높다는 점도 인식해야 한다. 다시 말하면, 표본의 크기가 대단히 커지면 실제적인 관점에서 볼 때 극히 작은 효과크기라도 그 결과는 항상 통계적으로 유의하게 나온다는 것이다(Cohen, 1969). 이는 결과 분석 상 통계적 유의성과 실제적 유의성이 항상 일치하지 않음을 의미한다(Hays, 1973). 따라서 향후 연구는 최소 표본기준에 부가하여 최대 표본크기 기준제시를 위한 연구가 필요하다.

둘째, 연구목적에 따라 많은 학자들이 나름대로의 표본크기 결정 공식을 제시함에 따라 연구자 수만큼 표본크기 결정공식이 존재함으로써 오히려 연구자들을 더욱 혼란스럽게 하고 있는 실정이다. 이로 인해 오히려 근거 없는 주먹구구식 방법이 지속되고 있다는 것을 이미 언급하였다. 따라서 연구 분야별로 유사한 결정공식을 일정 수준이상의 통계적 검정력이 확보되는 수준에서 통합함으로써 실증연구를 위한 실질적인 가이드라인의 제시가 요구된다 하겠다. 이는 연구의 통계적 결론 타당도를 높여 과학적 지식의 축적을 위한 연구자들의 노력에 더욱 박차를 가하는 일이 될 것이다.

VI. 맺음말

지금까지 매우 중요하지만 간과되기 쉬운 개념인 표본크기에 대하여 알아보았다. 표본은 연구자의 관심인 모집단을 대표한다는 면에서 대단히 중요하다. 그에 따라 연구를 설계함에 있어서 가장 먼저 부딪히는 질문 중 하나가 표본의 크기를 결정하는 일인데, 이 질문은 단순해 보이지만, 정확한 답변을 하는 것은 정말 어려운 일이다. 가장 손쉬운 방법은 실증적인 접근법으로써 유사한 연구에서 사용된 표본 크기를 적용하는 것이다. 이 방법은 경험이 부족한 연구자에게 다른 학자들의 공격에 대비한 적절한 표본크기 판단의 기준을 제공한다 할 수 있을 것이다(Sudman, 1976). 하지만 다수의 학자들이 표본크기 결정공식을 우후순식음으로 제시하여 혼란을 가중시킴으로써 오히려 근거 없는 주먹구구식 방법들이 마치 황금률인양 지속되고 있는 것이 현실이다. 그에 따라 본 연구에서는 적정 표본크기 결정 방안과 이를 충족하지 못했을 경우 발생할 수 있는 문제점을 짚어보고, 향후 연구를 위한 조언을 제시함으로써 연구의 신뢰성과 타당성을 높이는데 기여하고자 하였다.

통계적 검정력과 추론의 정확성에 대한 대부분의 논의들이 표본크기를 강조함에 따라 연구자들은 통계적 검정력과 추론의 정확성에 영향을 미치는 요인 중 그들이 통제할 수 있는 유일한 것이 표본크기라고 결론을 내리는 경우가 많다. 사실 표본크기가 중요한 역할을 한다는 것은 분명한 사실이지만, 분명한 것은 통계적 검정력과 추론의 정확성을 높이기 위해 연구자가 통제할 수 있는 기타의 요인들도 존재한다는 것이다. 예를 들어, McClelland(2000)는 추정 회귀계수, 평균오차 제곱, 예측변수의 분산, 표본크기 등을 이용하여 회귀계수에 대한 신뢰구간에 관한 공식을 제시하면서, 표본크기 외에도 신뢰구간의 폭을 줄여 통계적 검정력을 높일 수 있는 다양한 방법이 있다고 하였으며, 구체적으로 어떤 요인들을 연구자가 통제할 수 있는지를 설명한 연구도 존재한다(Shadish, Cook, & Campbell, 2002; West, Biesanz, & Pitts, 2000). 하지만 정확한 모수의 추정과 작은 효과라도 유의하게 나타내 줄 수 있는 통계적 검정력에 필요한 표본 크기의 확보는 종종 연구자들의 가용자원의 초과를 요구하는 경우가 많다. 따라서 현실적인 여건을 고려할 때 연구의 통계적 결론 타당도를 높이기 위해 연구자가 가장 손쉽게 통제할 수 있는 요인이 표본크기라는 것은 부인할 수 없는 사실이라 하겠다.

참 고 문 헌

- 박원우. 2003. 「연구방법론 강의 자료집」. 서울: 서울대학교.
- 박정식 & 윤영선. 2002. 「현대 통계학」(4판). 서울: 다산출판사.
- 이군희. 2007. 「사회과학 연구방법론」(수정판). 서울: 법문사.
- 채서일. 2003. 「사회과학 조사방법론」(2판). 서울: 학현사.
- Aron, A., & Aron, E. N. 2003. *Statistics for psychology*(3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Bartlett, J. E., II, Kotrlik J. W., & Higgins, C. C. 2001. Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19(1): 43-50.
- Bentler, P. M., & Chou, C. 1987. Practical issues in structural modeling. *Sociological Methods and Research*, 16(1): 78-117.
- Comrey, A. L., & Lee, H. B. 1992. *A first course in factor analysis*(2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. 1969. *Statistical power analysis for the behavioral science*. New York: Academic Press.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Cohen, J. 1992. A power primer. *Psychological Bulletin*, 112(1): 155-159.
- Cook, T. D., & Campbell, D. T. 1979. *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA : Houghton Mifflin Company.
- Eng, J. 2003. Sample size estimation: How many individuals should be studied? *Radiology*, 227(2): 309-313.
- Festinger, L., & Katz, D. (Eds.). 1953. *Research methods in the behavioral sciences*. New York: The Dryden Press.
- Gravetter, F. J. & Wallnau, L. B. 2000. *Statistics for the behavioral*

- sciences*(5th ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Guilford, J. P., & Fruchter, B. 1978. *Fundamental statistics in psychology and education*(6th ed.). New York: Mcgraw-Hill.
- Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. 2006. *Multivariate data analysis*(6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Halinski, R. S., & Feldt, L. S. 1970. The selection of variables in multiple regression analyses. *Journal of Educational Measurement*, 7(3), 151-158.
- Hays, W. L. 1973. *Statistics for the social science*(2nd ed). NY: Holt, Rinehart and Winston.
- Hurlburt, R. T. 2003. *Comprehending behavioral statistics*(3rd ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Kerlinger, F. N., & Lee, H. B. 2000. *Foundations of behavioral research* (4th ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Kline, R. B. 2005. *Principles and practice of structural equation modeling*(2nd ed.). New York: Guilford Press.
- Krejcie, R. V., & Morgan, D. W. 1970. Determining sample size for research activities. *Educational and Psychological Measurement*, 30: 607-610.
- Kurnow, E., Glasser, G. J., & Ottman, F. R. 1959. *Statistics for business decisions*. Homewood, IL: Richard D. Irwin.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. 1996. Power analysis and determination of sample size for covariance structural modeling. *Psychological Methods*, 1(2): 130-149.
- Maxwell, S. E., Kelley, K., & Rausch, J. R. 2008. Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59: 537-563.
- McClelland, G. H. 2000. Increasing statistical power without increasing sample size. *Americal Psychologist*, 55(8): 963-964.

- Mendenhall, W., & Beaver, R. J. 1994. *Introduction to probability and statistics*(9th ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Miller, D. E., & Kunce, J. T. 1973. Prediction and statistical overkill revisited. *Measurement and Evaluation in Guidance*, 6(3): 157-163.
- Polit, D. F., & Sherman, R. E. 1990. Statistical power in nursing research. *Nursing Research*, 39(6): 365-370.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Stone, E. F. 1978. *Research methods in organizational behavior*. Oakland, CA: Scott, Foresman and Co.
- Sudman, S. 1976. *Applied sampling*. New York: Academic Press.
- Well, A. D., Pollatsek, A., & boyce, S. J. 1990. Understanding the effects of sample size on the variability of the mean. *Organizational Behavior and Human Decision Process*, 47(2): 289-312.
- West, S. G., Biesanz, J. C., & Pitts, S. C. 2000. Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd(Eds.), *Handbook of research methods in social and personality psychology*: 40-48. New York: Cambridge Univ. Press.
- Williams, B. 1978. *A sampler on sampling*. New York: John Wiley & Sons.
- Wunsch, D. 1986. Survey research: Determining sample size and representative response. *Business Education Forum*, 4(5): 31-34

A proposal on determining appropriate sample size considering statistical conclusion validity

Won-Woo Park*
Seung Yeon Son**
Haeseen Park***
Hae Sang Park****

ABSTRACT

Sampling has been critical issue for empirical studies because the sample used in empirical study represents the population. Considering that sample size is significantly related to statistical power, conducting study with optimal sample size is essential to have statistical conclusion validity of the results. However, most empirical research has overlooked whether the sample size is enough to support statistical conclusion validity of the results, but focused only on logical interpretation. In fact, researchers have assumed that sample size used in research satisfy the proper level to validate the result and hardly paid attention to this dubious assumption.

Considering that sample size significantly influence on validation and generalization of the results, determining appropriate sample size according to the statistical method and the purpose of the study is noteworthy to improve

* Professor, College of Business Administration, Seoul National University

** Doctoral student, College of Business Administration, Seoul National University

*** Master student, College of Business Administration, Seoul National University

**** Research Assistant, Department of Business Administration, Korea National Open University

the quality of research. Therefore, study on determining proper sample size would provide practical information for researchers who conduct empirical study. The purpose of the current study is to extend the knowledge base regarding sample size in organizational study. Specifically, on the basis of the statistical principles, we offered criteria to determine appropriate sample size according to various statistical methods. Using these criteria, we examined all articles published during 1998~2007 in Korean Business Review and Korean Journal of Management, the major Korean journals, whether they are qualified these criteria. Moreover, we provided proposals for future research.

Key words: sample size, statistical conclusion validity