▧연구논문

표본의 수와 검정력 분석을 위한 통계팩키지+

이관제
동국대학교 통계학과

Statistical Package for Sample Size and Power Determination

Kwan Jeh Lee
Dept. of Statistics, Dongguk University

Abstract

In application, sample size determination is one of the important problems in designing an experiment. A large amount of literature has been published on the problem of determining sample size and power for various statistical models. In practice, however, it is not easy to calculate sample size and/or power because the formula and other results derived from statistical model are scattered in various textbooks and journal articles. This paper describes some previously published theories that have practical relevance for sample size and power determination in various statistical problems, including life-testing problems with censored cases and introduces a statistical package which calculates sample size and power according to the results described. The screens and numerical results made by the package are demonstrated.

+ 본 연구는 2000년도 동국대학교 전문학술지 논문게재연구비 지원으로 이루어졌음.

1. 서론

실험계획에 의한 통계연구에서 적절한 표본의 수와 이미 얻어진 표본의 수에 대한 검정력 분석은 통계학자나 통계적 방법을 적용하는 연구자에게 매우 중요한 문제이다. 특히 통계적 품질관리(statistical quality control)나 신뢰수명검정(life-testing), 또는 신뢰성 분석(reliability analysis)에서는 실험의 설계단계에서 필히 거쳐야 하는 단계이며, 이때 결정된 표본의 수가 실험후의 통계적 의사결정에 결정적인 영향을 미친다고 하여도 과언이 아니다. 즉, 실험의 결과에 대한 통계적 의사결정에 있어서 일정한 수준의 검정력을 지키는 것이 요구되는데, 실험시작 전에 적절한 표본의 수를 통계적 절차에 의하여 고려되지 못한 실험 중에는 실제로 유의적인 결과가 있다하여도 검정력이 작아서 이 유의적인 결과를 찾아내지 못하는 경우가 발생할 수 있다.

신뢰수명검정(life-testing)이나 신뢰성 이론(reliability theory)에서는 관측되는 확률변수의 측정값이 시험대상의 수명시간(기간)이거나 어느 사건이 일어날 때까지의 시간(기간)인 경우이며, 특히 실험대상이 인간을 포함한 생명체일 경우에는 생물통계학(biostatistics)의 한 통계적 방법인 생존분석(survival analysis)의 범주에 속한다. 이러한 부류의 통계적 방법에는 통상적인 통계적 실험과는 다른 여러 가지 점들이 추가적으로 고려되어야 한다. 먼저 실험 대상들이 모두 한 시점에서 실험에 투여되지 못하는 경우가 발생할 수 있으며, 특히 일부 실험 대상들은 계획된 실험기간 내에 수명시간이 관측되지 않음으로써 발생하게 되는 여러 가지 형태의 중도절단(censored)된 자료를 제공한다. 이 중도에 절단된 자료가 포함된 실험에서는 통계모형(statistical model)의 분석 시에 추가적인 주의가 필요할 뿐만 아니라, 실험전에 표본의 수를 결정할 때에도 통상적인 실험에서 보다 복잡한 과정을 추가하여야 한다. 결과적으로 중도절단으로 인하여 표본의 수가 증가하여야 하며, 같은 표본의 수에서는 중도절단이 없는 경우보다 검정력이 현저히 떨어진다. 이 분야에 대한 좋은 참고서는 Lawless(1982)와 Bain(1978)등이 있다.

통상적인 실험뿐만이 아니라, 신뢰수명검정이나 신뢰성 이론 및 생존분석에 대한 통계모형과 방법론은 여러 가지 교과서나 논문에 산재해 있으므로 매 번의 실험계획 단계에서 각각의 통계모형에 따라 표본의 수와 검정력을 계산하는데 불편함이 따른다. 컴퓨터 프로그램에 의한 표본의 수와 검정력 계산에 관한 비교는 Goldstein(1989)에 소개되어 있다. 상기한 바와 같이 표본의 수는 통계적 의사결정에 직접적인 영향을 줌으로 잘못 계산된 표본의 수는 실험결과 전체를 무의미하게 할 수도 있는 위험이 있다. 그러므로 단순한 작업으로 여겨질 수 있는 표본의 수를 결정하는 작업은 통계학자나 통계관련 연구가에게는 많은 주의를 필요로 하는 작업이다.

본 연구에서는 통상적인 통계실험에서 나타나는 통계 모형에서 뿐만이 아니라, 신뢰수명검정이나 신뢰성 이론에서 다루어지는 중도절단 자료를 포함하는 통계모형에서

의 표본의 수와 검정력을 계산하는 각각의 프로그램을 만들고, 이를 사용자에 편리하게 통합 시스템으로 만들었다. 프로그램에 사용되는 표본계산공식은 Desu와 Ragavarao(1990)을 많이 참조하였고, 신뢰수명검정이나 신뢰성 이론 및 생존분석에 대한 통계모형과 방법론은 Lawless(1982)와 Bain(1978)을 참조하였다. 이 시스템은 SAS 6.12의 AF 모듈을 이용하였다.

2. 표본의 수와 검정력

통계적 가설검정을 수행할 때 제 1종 과오를 범할 확률 (α)는 실험자에 의하여 미리 정하여 진다. 또한 실제로 유의적인 차이가 있는 경우 이것을 찾아내는 확률, 즉 검정력 $(1-\beta)$는 전체 표본의 수에 직접적인 영향을 받는다. 통계적으로 유의적인 결과에 도달하지 못한 경우에는 실제로 유의적인 결과가 없는 경우도 있으나 또한 검정력이 작아서 유의적인 결과를 탐색하지 못한 경우도 있는 것이다. 본 논문에서는 표본의 수와 검정력에 관한 근간이 되는 관계식을 설정하고 각각의 통계문제에서는 이 기본이 되는 관계식에서 변화되는 부분만을 바꿈으로서 표본의 수와 검정력을 계산하는 공식을 유도한다.

표본의 수와 검정력의 관계는 다음과 같이 설명할 수 있다. 먼저 귀무가설 (H_0)하에서 통계량 X는 $N(\mu_0, \Sigma_0^2)$인 분포를 따르고, 대립가설 (H_1)하에서는 $N(\mu_1, \Sigma_1^2)$인 분포를 따른다고 한다. 제 1종 과오를 범할 확률을 α라 하고, 제 2종 과오를 범할 확률을 β라 한다. 이 경우의 검정통계량 $Z = (X-\mu_0)/\Sigma_0$ 이며 귀무가설 하에서 $Z \sim N(0,1)$이다. 그러면 필요한 표본의 수는 귀무가설이 옳은 경우에는 $\Pr(Z > Z_\alpha) = \alpha$와 대립가설이 옳은 경우에는 $\Pr(Z > Z_\alpha) = 1-\beta$를 동시에 만족하여야 한다. 이 두 조건을 만족하는 식은

$$|\mu_1 - \mu_0| = Z_\alpha \Sigma_0 + Z_\beta \Sigma_1 \tag{1}$$

이다. 그러면 표본의 수를 계산하는데 관련이 있는 모수(parameter)는 α, β, μ_0, μ_1, ${\sigma_0}^2$과 ${\sigma_1}^2$이다. 만일 양측검정일 경우에는 Z_α대신 $Z_{\alpha/2}$를 사용한다. 이 관계식에 있는 두 모수간의 거리 $|\mu_1 - \mu_0|$는 $|X_\alpha - \mu_0| = Z_\alpha \Sigma_0$와, $|\mu_1 - X_\alpha| = Z_\beta \Sigma_1$의 합으로부터 구하여 진다. 이 관계식으로부터 다음과 같은 세 가지 통계 문제를 해결할 수 있다.

(1) 검정력 $(1-\beta)$를 확보하기 위한 표본의 수는 얼마인가?

(2) 주어진 표본의 수에 대한 검정력은 얼마인가?

(3) 주어진 표본의 수와 검정력 하에서 얼마만큼의 모수간의 차이를 검정할 수 있는가?

일반적으로 문제(1)는 실험을 계획하는 단계에서 다루어지고, 문제(2)는 실험을 평가하는 단계에서 다루어진다. 문제(3)는 두 단계 어느 곳에서나 고려될 수 있다. 일반적으로 통계량 X의 분산 Σ은 모집단 분산 σ^2와 표본의 수 N의 함수이다. 예를 들면, 통계량 X가 표본 평균일 경우에는 Σ^2는 σ^2/N이다. 그러므로 본 논문에서는 통계량 X의 분산을 $\Sigma^2=\sigma^2/N$으로 가정한다.
그러면 식(1)은 다음과 같이 쓸 수 있다.

$$|\mu_1-\mu_0|=(Z_\alpha\sigma_0/\sqrt{N})+(Z_\beta\sigma_1/\sqrt{N}). \tag{2}$$

실제로 이 식을 표본의 수를 계산하는 문제에 응용할 때에는 $|\mu_1-\mu_0|$는 검정하고자 하는 모수간의 최소의 차이(minimal difference to be detected)로 이해하면 된다. 이 (2)식을 N에 관하여 풀면,

$$N=\left[\frac{Z_\alpha\sigma_0+Z_\beta\sigma_1}{\mu_1-\mu_0}\right]^2 \tag{3}$$

이 되며, 문제(2)를 해결하기 위하여 Z_β에 관하여 풀면

$$Z_\beta=\frac{\sqrt{N}|\mu_1-\mu_0|-Z_\alpha\sigma_0}{\sigma_1} \tag{4}$$

가 된다. 검정력 $(1-\beta)$는 표준정규분포의 확률표(probability table)에서 Z_β에 대응하는 β값을 찾아서 계산한다. 모수간의 검정할 수 있는 유의적인 차이에 관한 질문인 문제(3)는 마찬가지로 식(2)를 이용하여 해결할 수 있다. 문제(3)을 해결하기 위해서는 반복적인 방법이 적용되어야 한다. 그리하여 본 연구에서는 문제(1)과 문제(2)를 위해서는 프로그램을 batch style로 만들었고, 문제(3)을 해결하는 방법으로 프로그램을 interactive style로 만들었다. 소개되는 시스템에서 고려된 점들은 다음과 같다.

(1) 각 그룹간의 표본의 수가 다른 경우를 고려 각 그룹의 표본수 비율을 Q_1과 Q_2라 하면, $Q_1+Q_2=1$이고 $n_1=Q_1N$, $n_2=Q_2N$이다.

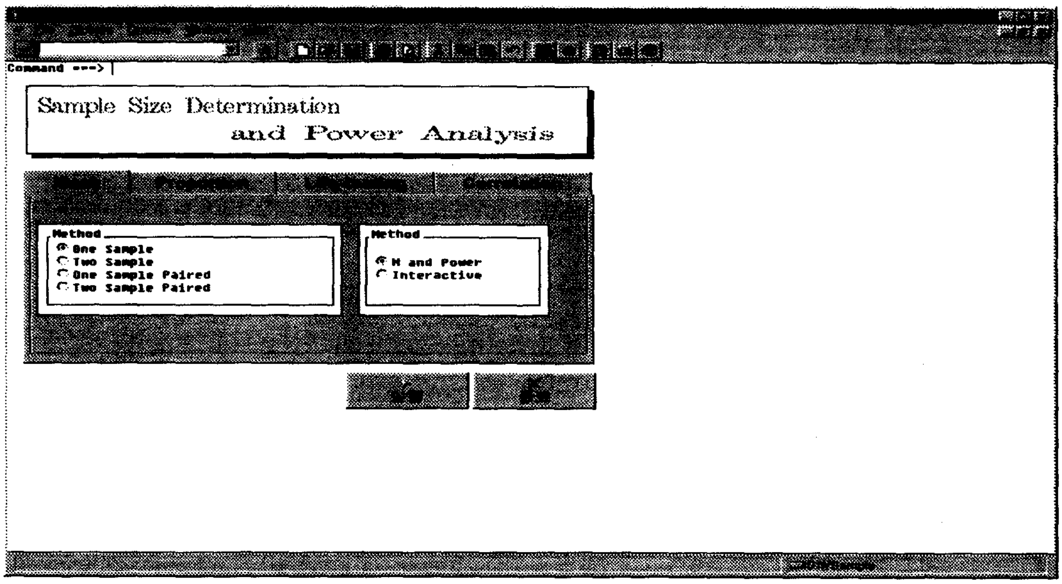
(2) 표본의 수의 결정은 가능한 자원과 연구의 여러 가지 목적과 타협의 과정이다. 그러므로 타협에 따르는 조건의 변화에 맞추어 여러 차례의 반복된 작업이 이루어 져야 하는데, batch style의 시스템 Output에는 α=0.1, 0.05, 0.01 세 경우와 20여개의 검정력을 분할표로 하여 표로 작성하여 제공한다. 그러므로 한번의 수행으로 여러 조건에 맞는 표본의 수나 검정력표를 제공한다. (<표 1>과 <표 2> 참조)

(3) 신뢰수명검정이나 신뢰성 분석 및 생존분석에서 나타나는 중도절단에 대한 문제를 다룬다.

(4) 생존분석에서 나타날 수 있는 중도탈락의 비율을 고려한다.
(5) 프로그램은 batch style과 interactive style를 모두 가능하게 구현한다. [화면 2]부터 [화면 17]중에서 짝수번호는 batch style의 화면이고, 홀수번호의 화면은 interactive style의 화면이다.

먼저 시스템을 시작하면 아래와 같은 화면이 만들어지고, 각각의 통계문제에 따라 화면의 메뉴를 선택한다.

[화면 1] 초기 프로그램 화면



3. 평균을 이용한 표본의 수와 검정력

3.1 일 표본 문제

모평균에 대한 검정 시에 모분산(population variance)은 미리 주어지거나 미지의 값으로 간주된다. 분산이 주어진 경우에 Z 검정 통계량이 사용되며, 분산이 미지인 경우의 검정통계량은 t 분포를 따르는 스튜던트 T 검정통계량이 사용된다. 일 표본 검정 시에는 귀무가설과 대립가설에서의 분산이 같다는 가정($\sigma_0{}^2 = \sigma_1{}^2 = \sigma^2$)을 하는 것이 일반적이다. t 분포도 자유도가 크면 표본정규분포에 점근하므로, 분산이 미지인 경우는 표본분산 s^2으로 σ^2을 추정하여 식(3)과 (4)를 이용한다. 귀무가설과 대립가설에서의 분산이 같다는 가정 하에서 표본의 수와 검정력을 구하는 식은 다음과 같다.

$$N=\left(\frac{Z_{\alpha}+Z_{\beta}}{\mu_1-\mu_0}\right)^2\sigma^2, \tag{5}$$

$$Z_{\beta}=\frac{\sqrt{N}|\mu_1-\mu_0|-Z_{\alpha}\sigma}{\sigma} \tag{6}$$

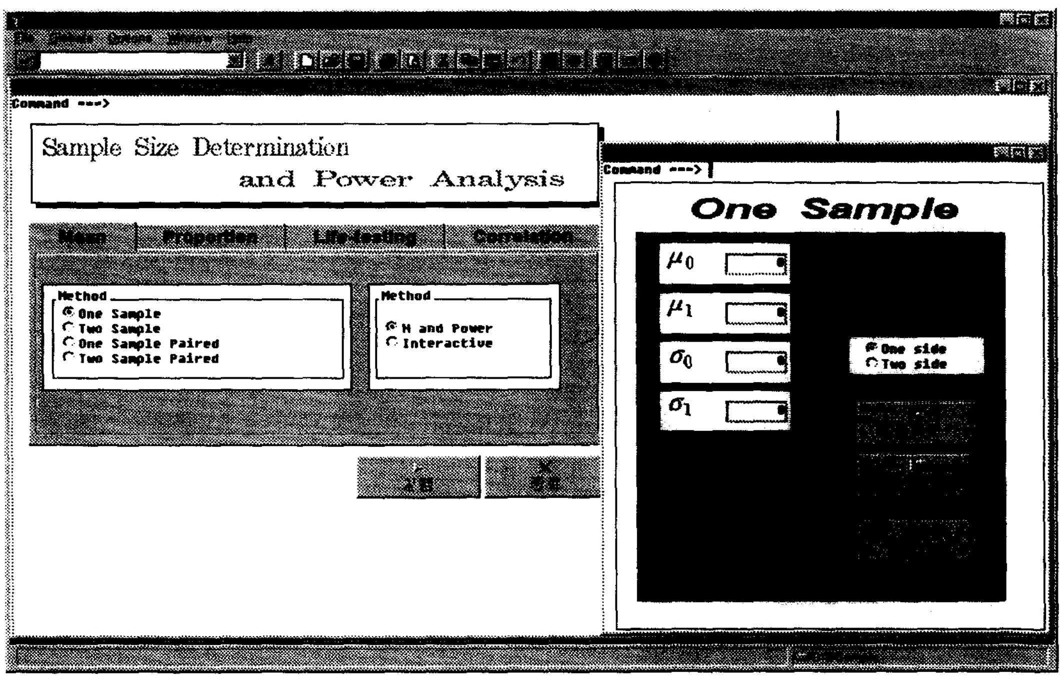
화면에 입력하여야 하는 모수 값들은 귀무가설 하에서의 모평균값, 대립가설 하에서의 모평균값, 귀무가설 하에서의 모표준편차, 및 대립가설 하에서의 모표준편차 등이다. 두 모표준편차가 같으면 σ_0와 σ_1에 같은 값을 입력한다. 공식 (3)과 (4)나, 공식 (5)와 (6)에 있는 $|\mu_1-\mu_0|$는 검정하고자 하는 모수의 차이이므로 귀무가설과 대립가설이 구체적으로 결정이 되지 않은 상태에서 표본의 수를 계산하고자 할 때는 μ_0에는 0을 입력하고 μ_1에는 검정하고자 하는 모수의 차이를 입력한다.

3.2 이 표본 문제

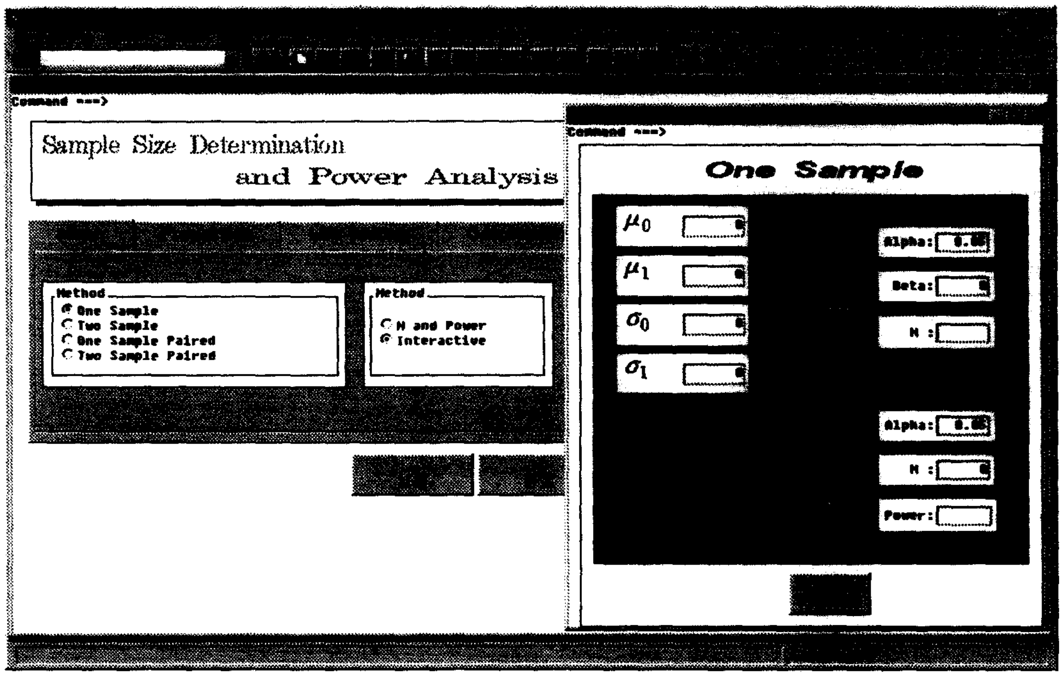
두 모집단의 모평균을 검정하는 문제로서 두 모집단의 분포를 각각 $N(\mu_0, \sigma_0^2)$과 $N(\mu_1, \sigma_1^2)$으로 가정하고, 이 때의 귀무가설은 $H_0: \mu_0=\mu_1$ 이다.

N을 두 표본을 합친 전체 표본의 수라하고 각 표본의 비율은 Q_0와 Q_1이라 한다.

[화면 2] 일 표본 표본의 수와 검정력 계산



[화면 3] 일 표본 표본의 수와 검정력 계산 - interactive



즉, $Q_0+Q_1=1$이며 $n_0=NQ_0$와 $n_1=NQ_1$이다. 이때의 두 표본평균의 분포는 각각의 모집단에 따라 $\overline{X} \sim N(\mu_0,\ \Sigma_0{}^2)$과 $\overline{Y} \sim N(\mu_1,\ \Sigma_1{}^2)$이다. 그리고 이 표본 검정시의 또 다른 가정은 등분산 가정 $\sigma_0^2=\sigma_1^2=\sigma^2$인데, 이 σ^2이 주어지지 않은 경우에는 혼합표본분산(pooled sample variance) s_p^2에 의하여 추정된다. 일 표본에서와 같이 모분산이 알려진 경우에는 Z 검정통계량을 사용하고 모분산이 알려지지 않은 경우에는 T 통계량을 사용한다.

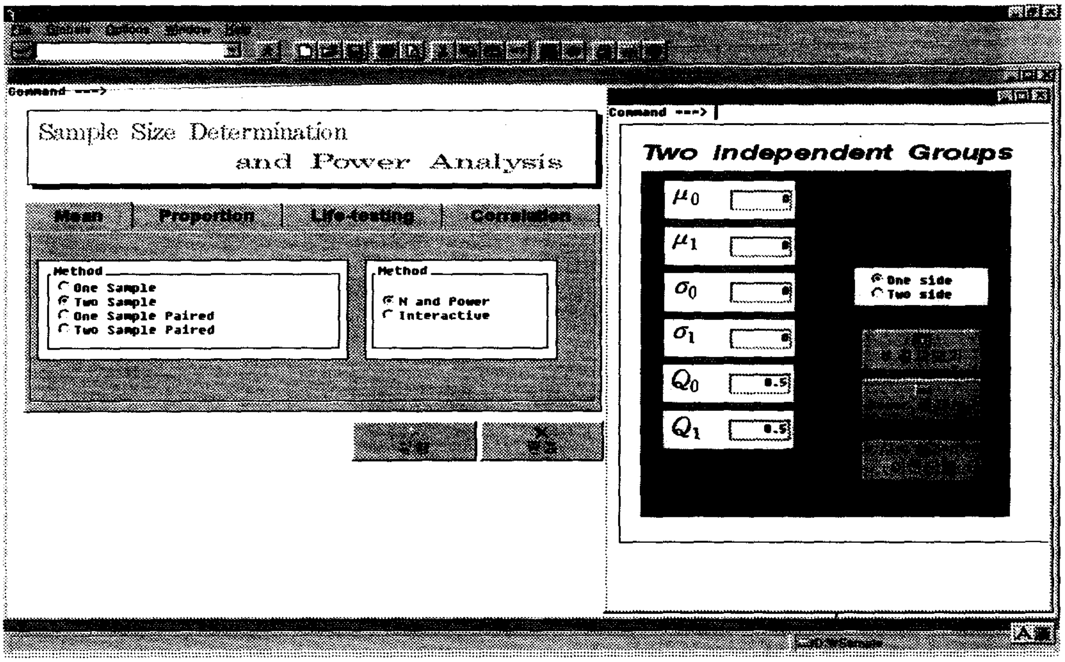
또한 등분산(equal variance)가정에 따라 $\Sigma_0^2=\Sigma_1^2=\dfrac{\sigma^2(\ Q_0{}^{-1}+\ Q_1{}^{-1})}{N}$ 이며, 이 조건들을 포함시킨 표본의 수와 검정력 계산식은 식 (3)과 (4)를 변형하여 아래와 같이 얻어진다.

$$N=\frac{\sigma^2(Q_0^{-1}+Q_1^{-1})(Z_\alpha+Z_\beta)^2}{|\mu_1-\mu_0|^2},$$

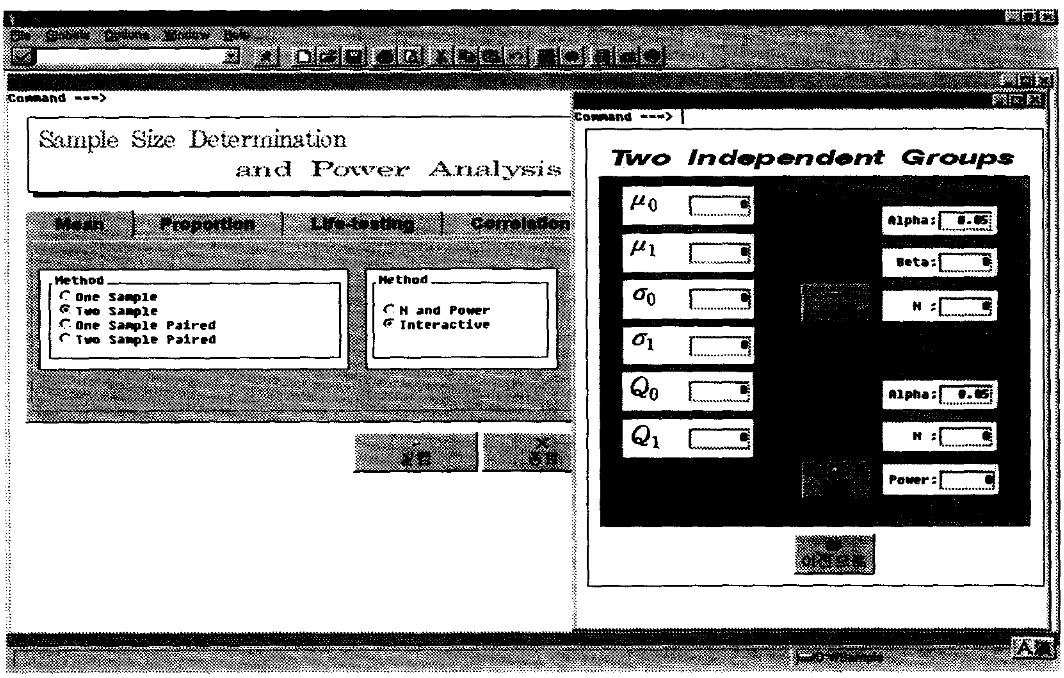
$$Z_\beta=\frac{|\mu_1-\mu_0|\sqrt{N}-Z_\alpha\sigma\sqrt{\ Q_0{}^{-1}+\ Q_1{}^{-1}}}{\sigma\sqrt{Q_0^{-1}+Q_1^{-1}}}.$$

여기서 두 표본의 수가 같은 경우에는 $Q_0{}^{-1}+\ Q_1{}^{-1}=4.0$이다.

[화면 4] 이 표본 표본의 수와 검정력 계산

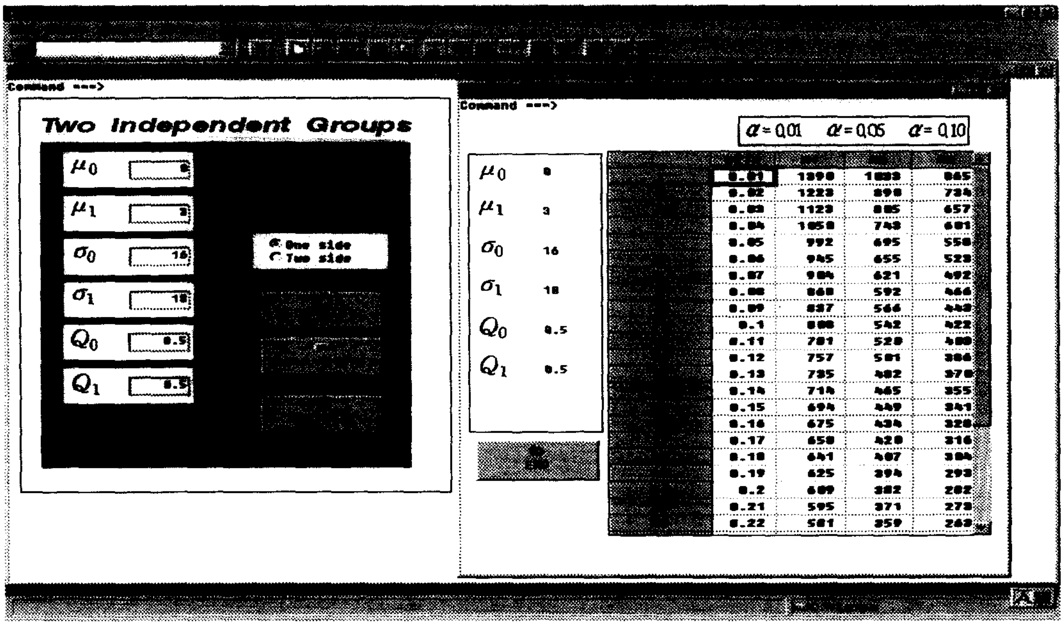


[화면 5] 이 표본 표본의 수와 검정력 계산 - interactive

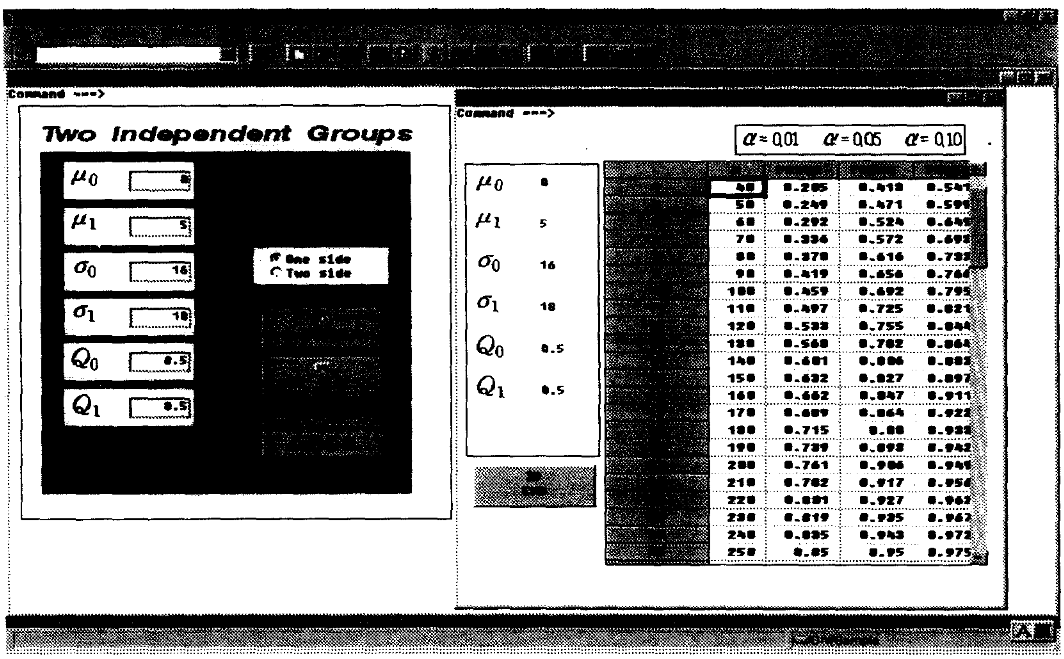


< 표 1 > 이 표본 문제에서의 표본의 수와 검정력

- 표본수 -



- 검정력 -



3.3 쌍(pair)을 이룬 이 표본 문제

두 번의 반복측정 시에 나타나는 경우로 각각의 측정시점에서의 모평균은 μ_0와 μ_1으로 표시한다. 이 경우에는 두 번의 측정이 한 개체 또는 한 사람으로부터 반복적으로 관측됨으로써 두 관측치 사이에 독립성이 결여된다. 그러므로 상기한 이 표본 검정문제에서 사용한 방법을 그대로 사용할 수 없다. 이러한 경우에는 첫 번째 측정값과 두 번째 측정값과의 차이(difference)를 새로운 확률변수로 사용하며, 통계량으로는 두 표본평균의 차 $\overline{d}= \overline{X}_1- \overline{X}_0$를 이용한다. 물론 역으로 두 번째에서 첫 번째를 뺀 값을 사용하여도 무방하다.

ρ를 첫 번째 측정값과 두 번째 측정값과의 상관계수이며 $\sigma_0^2=\sigma_1^2=\sigma^2$이라 하면, $\overline{d}$의 표준오차(standard error)는 $\sigma_d/\sqrt{N}$이고, 이 때 $\sigma_d^2=2\sigma^2(1-\rho)$이다. 그러므로 식 (5)와 (6)을 이용하여 다음과 같은 표본의 수와 검정력 계산식을 얻는다.

$$N=\left[\frac{(Z_\alpha+Z_\beta)\sigma_d}{(\mu_1-\mu_0)}\right]^2,$$

$$Z_\beta=\frac{|\mu_1-\mu_0|\sqrt{N}-Z_\alpha\sigma_d}{\sigma_d}.$$

4. 모비율을 이용한 표본의 수와 검정력

4.1 일 표본 모비율

통계실험에서 관측치가 양적인 변수가 아닌 질적인 수로써 특히 성공과 실패와 같은 이항분포를 따르는 변수인 경우가 많다. 즉 성공의 확률을 π라 하면 관측치 들은 모수가 N과 π인 이항분포를 따른다.

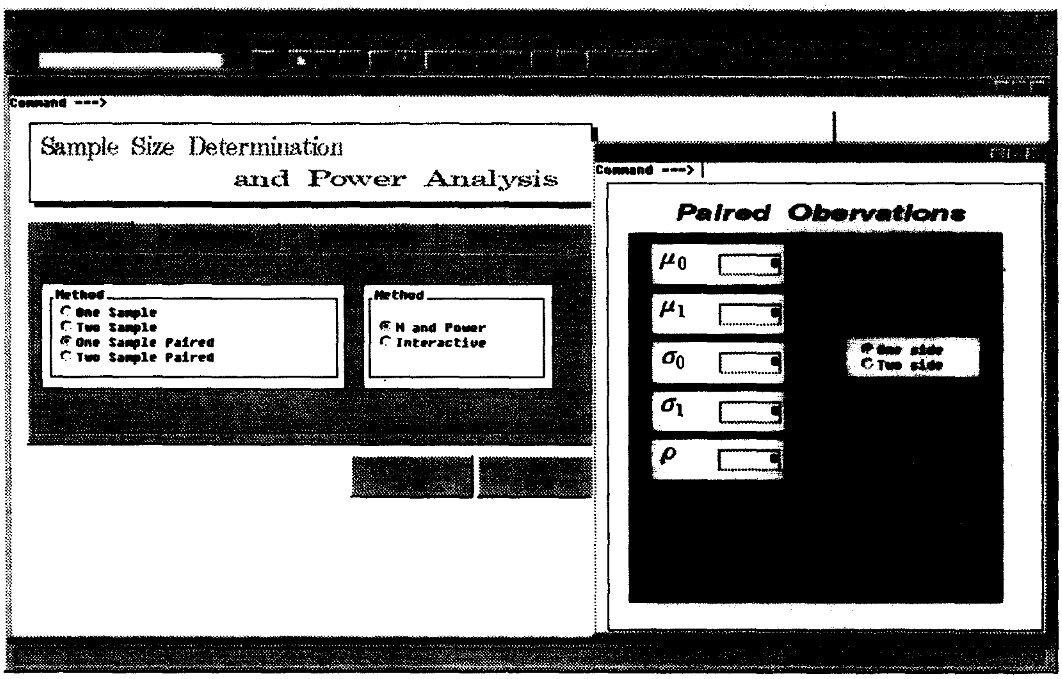
일 표본 모비율에 관한 검정인 경우에 귀무가설은 $H_0: \pi=\pi_0$ 이고 대립가설은 $H_1: \pi=\pi_1$이다. 또한 $\sigma_0^2=\pi_0(1-\pi_0)$와 $\sigma_1^2=\pi_1(1-\pi_1)$이다. 표본의 수와 검정력 계산식은 다음과 같다.

$$N=\left[\frac{Z_\alpha\sqrt{\pi_0(1-\pi_0)}+Z_\beta\sqrt{\pi_1(1-\pi_1)}}{\pi_1-\pi_0}\right]^2 \tag{7}$$

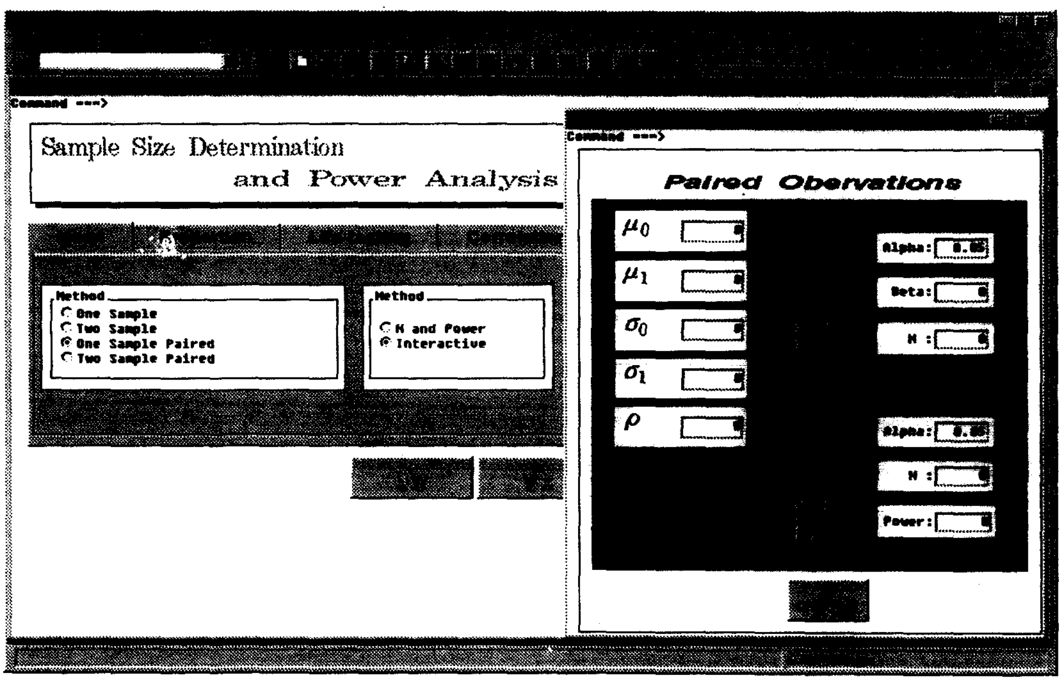
과

$$Z_\beta=\frac{\sqrt{N}|\pi_1-\pi_0|-Z_\alpha\sqrt{\pi_0(1-\pi_0)}}{\sqrt{\pi_1(1-\pi_1)}}. \tag{8}$$

[화면 6] 쌍을 이룬 이 표본 표본의 수와 검정력 계산

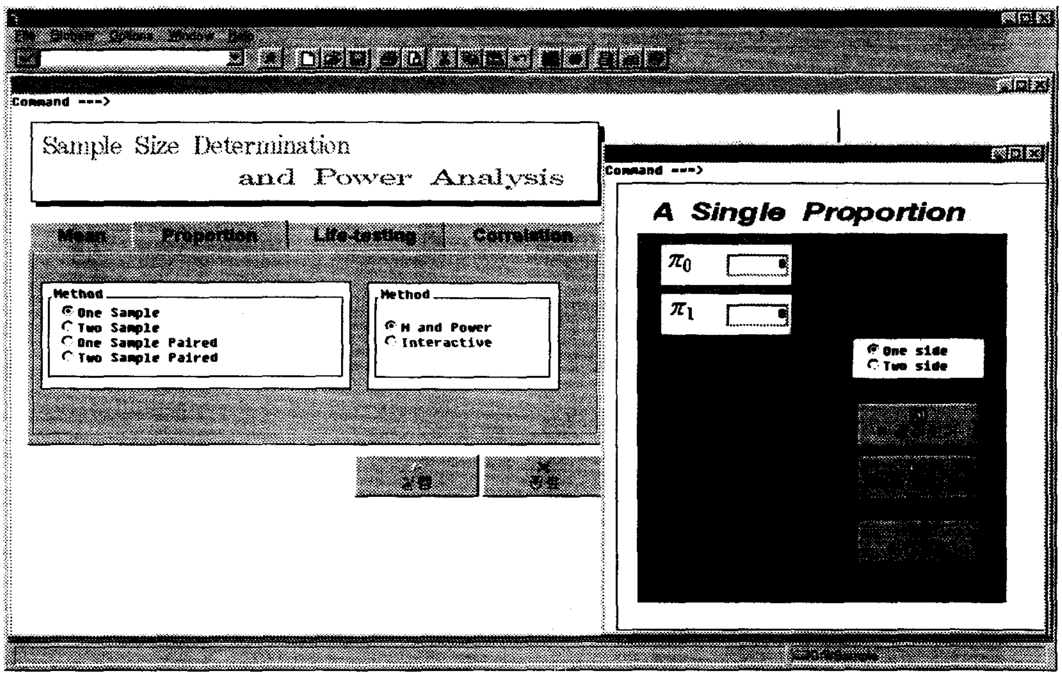


[화면 7] 쌍을 이룬 이 표본 표본의 수와 검정력 계산 - interactive



화면에 입력할 때는 공식 (7)과 (8)에 있는 $|\pi_1 - \pi_0|$는 검정하고자 하는 모수의 차이이므로 귀무가설과 대립가설이 구체적으로 결정이 되지 않은 상태에서 표본의 수를 계산하고자 할 때는 π_0에는 0을 입력하고 π_1에는 검정하고자 하는 모비율의 차이를 입력한다.

[화면 8] 일 표본 모비율에서의 표본의 수와 검정력 계산



4.2 이 표본 모비율

각각의 표본의 수 $n_0 = Q_0N$, $n_1 = Q_1N$ 인 모비율 π_0와 π_1을 검정하는 문제에서는 귀무가설을 $H_0 : \pi_1 = \pi_0$으로 놓고, 검정통계량은 $Z = \frac{(p_1 - p_0)}{s}$ 을 사용한다. p_0와 p_1은 표본비율이며 s^2은 $s^2 = (n_0^{-1} + n_1^{-1})\bar{p}(1-\bar{p})$로 정의하며 $\bar{p} = Q_0p_0 + Q_1p_1$으로 계산한다. 귀무가설 하에서 $Z \rightarrow N(0, 1)$으로 접근한다. 표본의 수와 검정력을 계산하기 위해서는 먼저 σ_0^2과 σ_1^2을 결정하여야 하는데, 귀무가설 하에서는 $\bar{\pi} = Q_0\pi_0 + Q_1\pi_1$으로 정의하면 $\sigma_0^2 = (Q_0^{-1} + Q_1^{-1})\bar{\pi}(1-\bar{\pi})$이고, 대립가설 하에서는 $\sigma_1^2 = [\pi_0(1-\pi_0)Q_0^{-1} + \pi_1(1-\pi_1)Q_1^{-1}]$이다. 그러면 기본공식 (1)로 부터 다음과 같은 공식을 얻을 수 있다.

$$\sqrt{N}|\pi_1 - \pi_0| = Z_\alpha\sqrt{\bar{\pi}(1-\bar{\pi})(Q_0^{-1} + Q_1^{-1})} + Z_\beta\sqrt{\pi_0(1-\pi_0)Q_0^{-1} + \pi_1(1-\pi_1)Q_1^{-1}}\,.$$

이 공식을 N과 Z_β에 대하여 풀면 표본의 수와 검정력을 구할 수 있다.

$$N = \frac{\left(Z_\alpha \sqrt{\bar{\pi}(1-\bar{\pi})(Q_0^{-1}+Q_1^{-1})} + Z_\beta \sqrt{\pi_0(1-\pi_0)Q_0^{-1} + \pi_1(1-\pi_1)Q_1^{-1}}\,\right)^2}{(\pi_1-\pi_0)^2},$$

$$Z_\beta = \frac{\sqrt{N}|\pi_1-\pi_0| - Z_\alpha \sqrt{\bar{\pi}(1-\bar{\pi})(Q_0^{-1}+Q_1^{-1})}}{\sqrt{\pi_0(1-\pi_0)Q_0^{-1} + \pi_1(1-\pi_1)Q_1^{-1}}}.$$

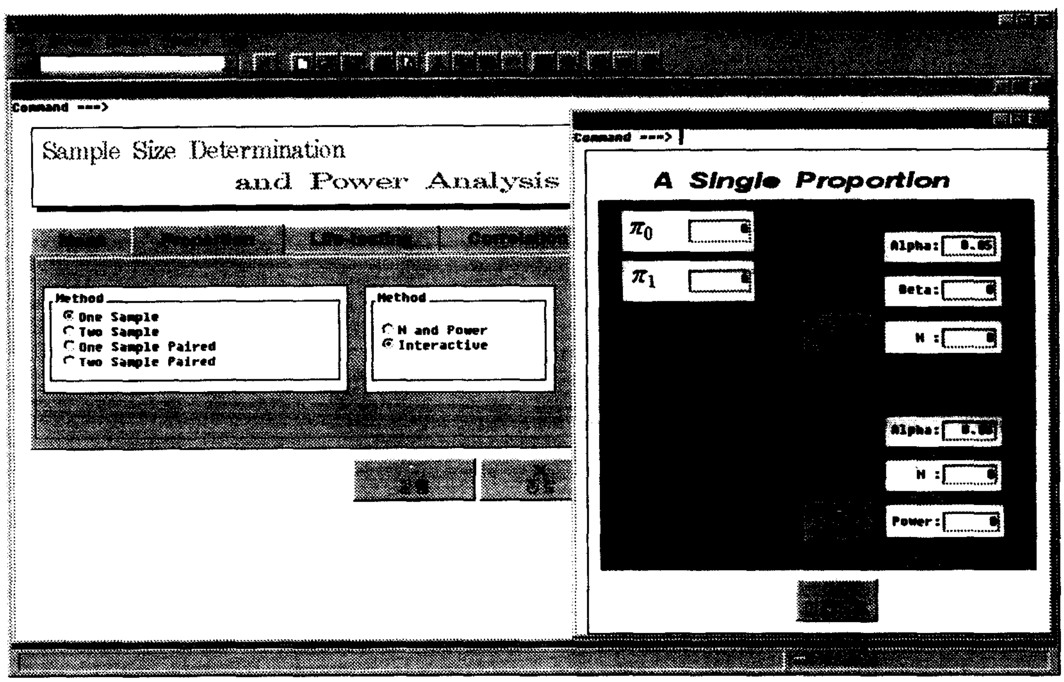
표본의 수가 같은 경우에는 $\sigma_0^2 = 4\bar{\pi}(1-\bar{\pi}) \geq \sigma_1^2 = 2\pi_0(1-\pi_0) + 2\pi_1(1-\pi_1)$이 성립하므로 다음과 같은 간편한 공식을 얻을 수 있다. 즉,

$$N = \frac{(Z_\alpha + Z_\beta)^2\, 4\bar{\pi}(1-\bar{\pi})}{(\pi_1-\pi_0)^2},$$

$$Z_\beta = \frac{\sqrt{N}|\pi_1-\pi_0|}{2\sqrt{\bar{\pi}(1-\bar{\pi})}} - Z_\alpha.$$

를 얻을 수 있다.

[화면 9] 일 표본 모비율에서의 표본의 수와 검정력 계산 - interactive



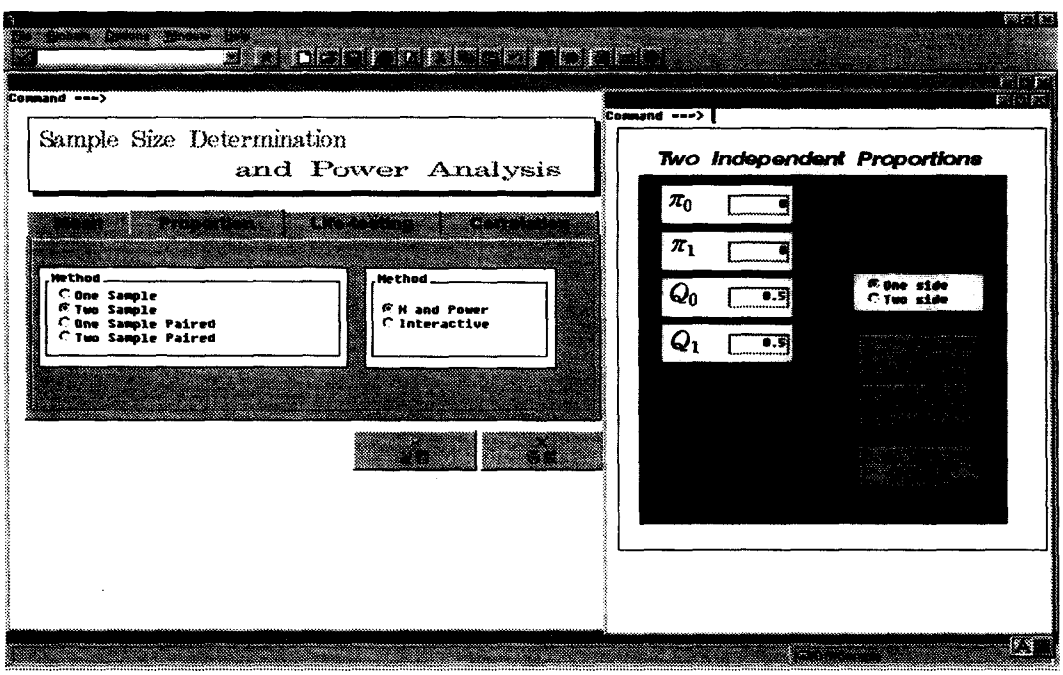
5. 신뢰수명검정

신뢰수명분석에서 사용되는 변수들의 예로는 기계 부품이 고장나는 시간이라든지, 생명체가 죽는 시간 또는 어떠한 기술을 배우는데 걸리는 시간 등의 시간변수이거나 시간의 개념이 들어있는 변수들이다. 서론에서 언급된 바와 같이 신뢰수명분석이 다른 분석과 다른 점은 중도절단(censored)이 일어나는 것이다. 중도절단에는 여러 가지가 있으며 일반적으로 관측대상이 연구종료 시까지 고장나지 않거나, 또는 생명체인 경우에는 연구종료 시까지 살아있는 경우이다. 이외에도 연구종료에 의한 중도절단외에도 실험의 규정에 어긋나서 계속 관찰을 할 수 없는 경우에도 발생하는데, 특히 생존분석에서와 같이 환자의 치료에 관한 임상 실험인 경우에는 환자가 다른 곳으로 옮겼거나 또는 실험을 계속할 수 없을 만큼의 부작용 등이 있는 경우도 중도절단의 예이다.

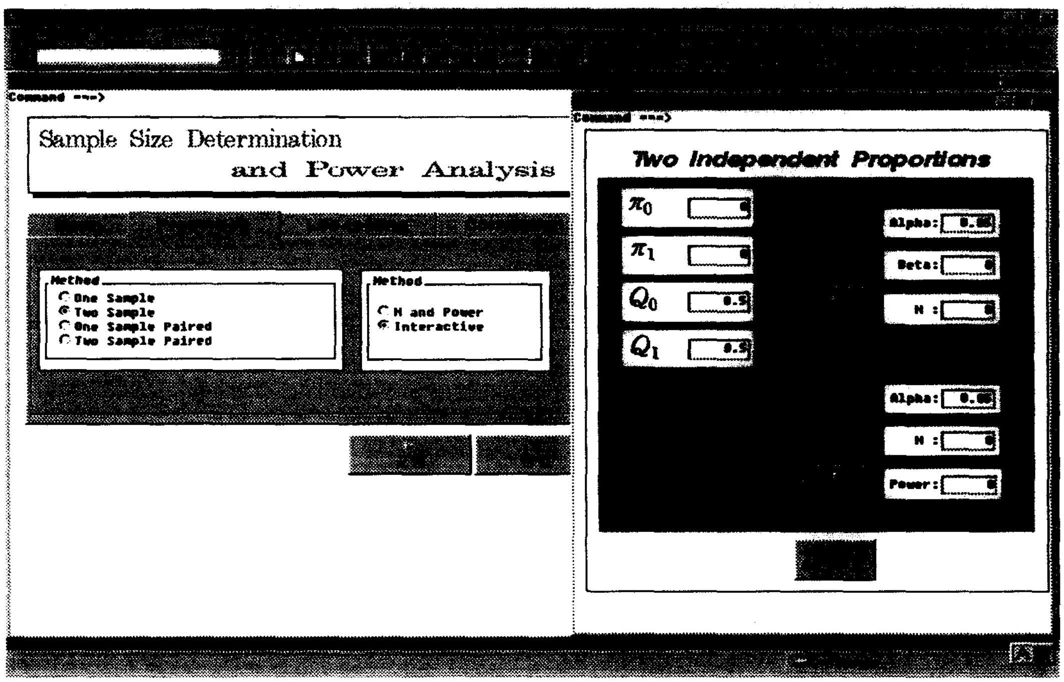
중도절단된 불완전한 자료를 버리고 나머지 중도 절단되지 않은 자료들만으로 신뢰수명분석을 실행할 수도 있으나, 이러한 경우에는 평균 수명기간의 추정값이 작아지는 경우가 방생한다. 그리하여 중도절단될 때까지의 자료에 중도절단 됐다는 정보를 함께 포함하여 분석하는 것이 일반적인 방법이다.

기본적인 신뢰수명 분석에 사용되는 통계기법은 비모수적 방법(nonparametric method)이 사용되나, 표본의 수를 계산하기 위해서는 통계모형에 대한 가정이 필요하다. 본 연구에서는 수명에 대한 분포로 지수분포를 가정한다. 위험률(hazard rate)을

[화면 10] 이 표본 모비율에서의 표본의 수와 검정력 계산



[화면 11] 이 표본 모비율에서의 표본의 수와 검정력 계산 - interactive



λ라고 하면 어느 시점 t에서의 생존함수(survival function) $S(t)=e^{-\lambda t}$이며 평균수명을 M이라 하면 위험률 λ는 $L=M^{-1}$으로 추정되며, L의 분포는 근사적으로 정규분포로 접근한다. 즉 $L \rightarrow N(\lambda,\ \lambda^2/N)$이다. 보다 자세한 내용은 Lawless(1982)와 Bain(1978)에 설명되어 있다.

5.1 이 표본의 비교 (중도절단이 없는 경우)

두 표본의 수를 각각 n_0와 n_1이라 하고 귀무가설 $H_0 : \lambda_1-\lambda_0=0$으로 설정한다. 이때 검정통계량은 $Z=\dfrac{L_1-L_0}{s}$이고, L_0와 L_1은 각각 두 표본의 위험률에 대한 추정량이며 $s=(n_0^{-1}+n_1^{-1})(Q_0L_0+Q_1L_1)$이다. 귀무가설 하에서 Z는 정규분포로 점근한다. 표본의 수와 검정력 분석을 위하여 $\mu_1=|\lambda_1-\lambda_0|$이라 하고, σ_0^2과 σ_1^2은 $\sigma_0^2=\bar{\lambda}^2(Q_0^{-1}+Q_1^{-1})$, $\bar{\lambda}=Q_0\lambda_0+Q_1\lambda_1$, $\sigma_1^2=(\lambda_0^2Q_0^{-1}+\lambda_1^2Q_1^{-1})$로 구하여 진다. 이와 같은 결과를 이용하여 다음과 같은 식을 얻는다.

$$\sqrt{M}|\lambda_1-\lambda_0|=Z_\alpha\sqrt{\bar{\lambda}^2(Q_0^{-1}+Q_1^{-1})}+Z_\beta\sqrt{\lambda_0^2Q_0^{-1}+\lambda_1^2Q_1^{-1}}.$$

이 식을 N과 Z_β에 관하여 풀면 표면의 수와 검정력을 구할 수 있다.

$$N = [\frac{(Z_\alpha \sqrt{\overline{\lambda}^2 (Q_0^{-1} + Q_1^{-1})} + Z_\beta \sqrt{\lambda_0^2 Q_0^{-1} + \lambda_1^2 Q_1^{-1}})}{\lambda_1 - \lambda_0}]^2$$

$$Z_\beta = \frac{\sqrt{N}|\lambda_1 - \lambda_0| - Z_\alpha \sqrt{\overline{\lambda}^2 (Q_0^{-1} + Q_1^{-1})}}{\sqrt{\lambda_0^2 Q_0^{-1} + \lambda_1^2 Q_1^{-1}}}.$$

5.2 중도절단이 있는 경우의 이 표본의 비교

신뢰수명분석에서 중도절단이 없는 경우는 비현실적이라 할 수 있다. 보다 현실적으로 실험이 어느 시점 T에서 종료되며, 실험대상들은 구간 (0, T)에서 균일한 확률을 갖고 실험에 투입된다고 가정한다. 여기서

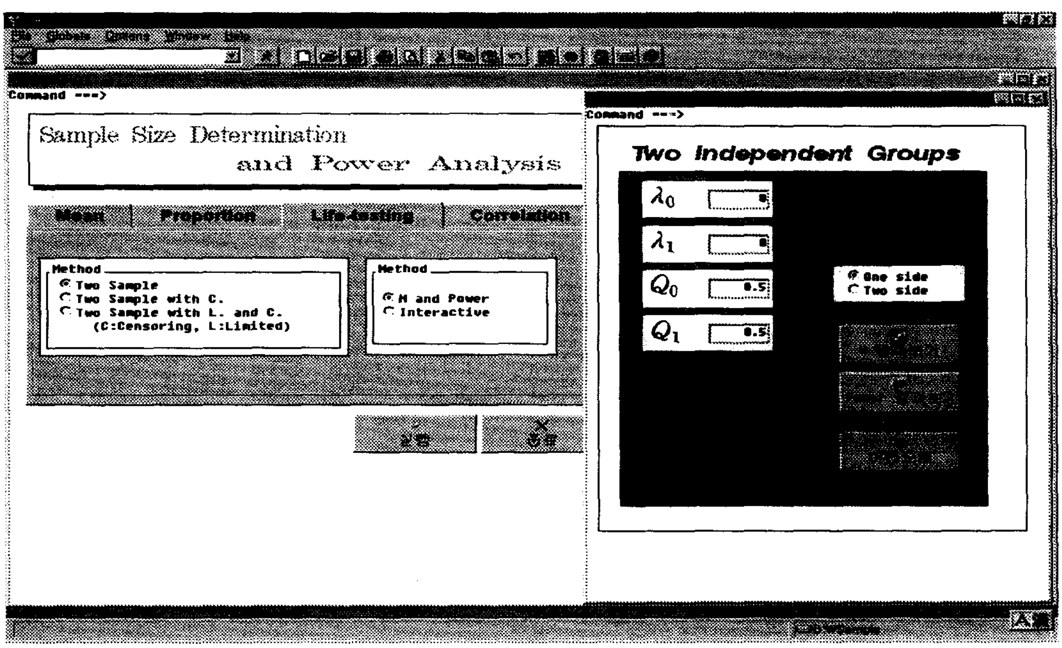
$$\phi(\lambda) = \frac{\lambda^3 T}{(\lambda T - + e^{\lambda T})}$$

라 놓으면

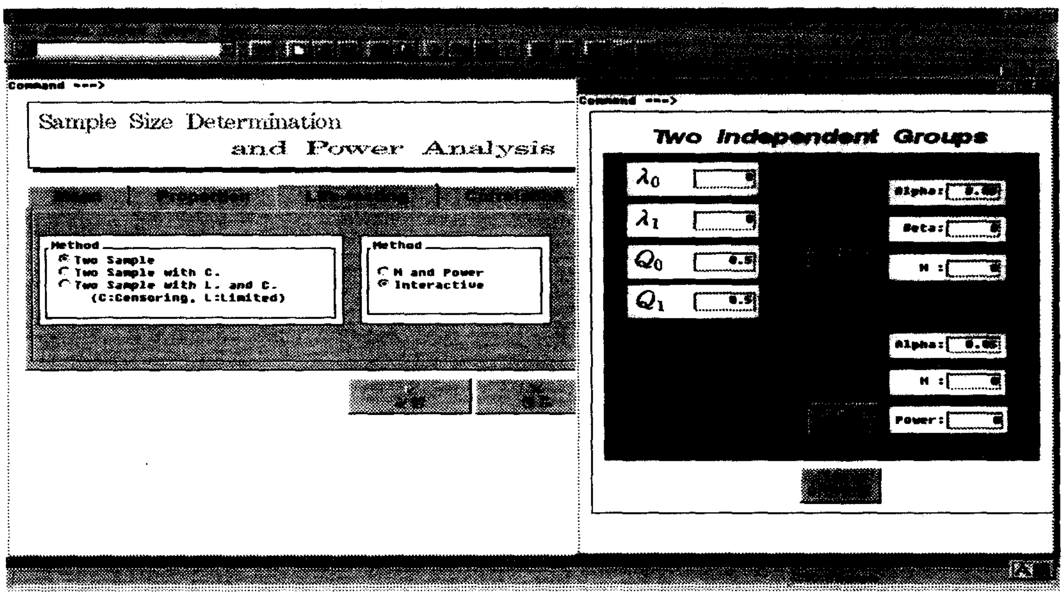
$$\sigma_0^2 = \phi(\overline{\lambda})(Q_0^{-1} + Q_1^{-1}),$$

$$\sigma_1^2 = \phi(\lambda_0) Q_0^{-1} + \phi(\lambda_1) Q_1^{-1}$$

[화면 12] 신뢰수명검정에서의 표본의 수와 검정력 계산



[화면 13] 신뢰수명검정에서의 표본의 수와 검정력 계산 - interactive



로 나타낼 수 있으며 $\bar{\lambda} = Q_0\lambda_0 + Q_1\lambda_1$이다.

그러면 표본의 수와 검정력에 관한 관계식은

$$\sqrt{N}|\lambda_1 - \lambda_0| = Z_\alpha\sqrt{\phi(\bar{\lambda})(Q_0^{-1} + Q_1^{-1})} + Z_\beta\sqrt{\phi(\lambda_0)Q_0^{-1} + \phi(\lambda_1)Q_1^{-1}}$$

와 같이 얻을 수 있으며, 이를 N과 Z_β에 관하여 풀면 표본의 수와 검정력을 얻을 수 있다.

$$N = \frac{\left(Z_\alpha\sqrt{\phi(\bar{\lambda})(Q_0^{-1} + Q_1^{-1})} + Z_\beta\sqrt{\phi(\lambda_0)Q_0^{-1} + \phi(\lambda_1)Q_1^{-1}}\right)^2}{(\lambda_1 - \lambda_0)^2},$$

$$Z_\beta = \frac{\sqrt{N}|\lambda_1 - \lambda_0| - Z_\alpha\sqrt{\phi(\bar{\lambda})(Q_0^{-1} + Q_1^{-1})}}{\sqrt{\phi(\lambda_0)Q_0^{-1} + \phi(\lambda_1)Q_1^{-1}}}.$$

6. 상관계수

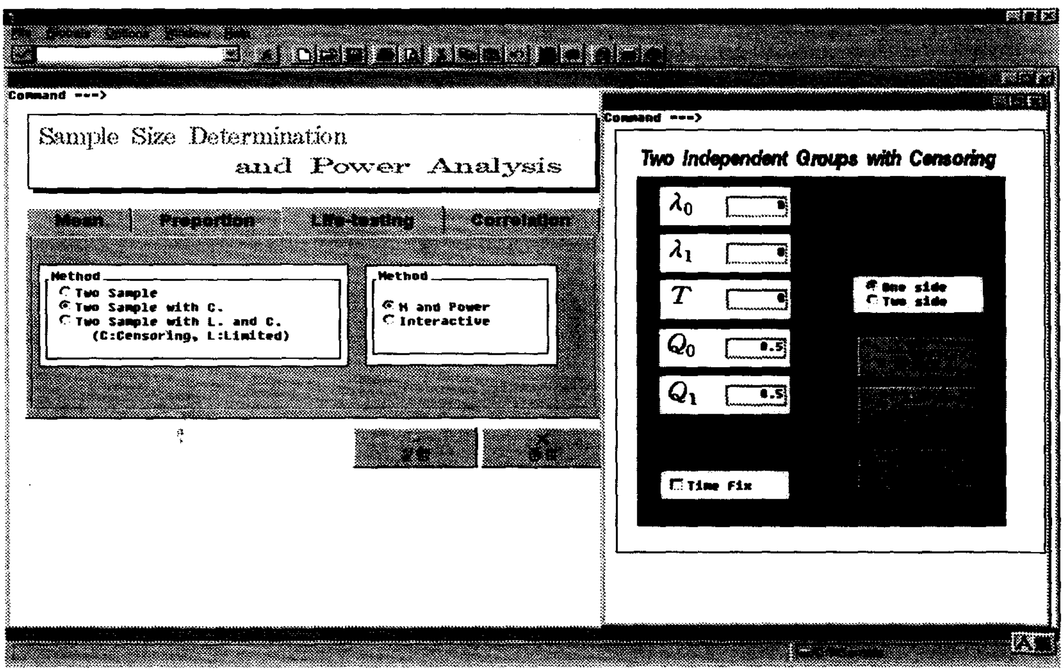
모상관계수를 ρ라 하고, 표본상관계수를 r이라 하면 귀무가설 $H_0 : \rho = 0$도 설정할 수 있다. 모상관계수에 대한 검정문제는 Freiman, et al.(1978)에 소개된 다음 변수

변환을 고려하여 해결할 수 있다.

$$C(r) = \frac{1}{2}\log_e \frac{1+r}{1-r}.$$

표본의 수가 N이 경우에 $C(r)$은 평균이 $C(\rho)$이고 분산이 $1/(N-3)$인 정규분포로 접근한다. $\rho=0$일 때 $C(0)=0$이므로 상기 귀무가설에 대한 검정통계량은 $Z = C(r)\sqrt{N-3}$으로 표준정규분포에 접근한다. 이를 이용한 표본의 수와 검정력에 관한 관계식은 $\sqrt{N-3}\,C(\rho_1) = Z_\alpha + Z_\beta$을 얻을 수 있으며, ρ_1은 대립가설 하에서의 모상관계수의 값이다. 두 모상관계수를 비교하는 문제에서는 귀무가설 $H_0: C(\rho_1) - C(\rho_0) = 0$으로 설정하고, 이때의 검정통계량은 $Z = \dfrac{C(r_1) - C(r_0)}{\Sigma_0}$를 사용한다.

[화면 14] 중도절단이 있는 경우의 표본의 수와 검정력 계산



$\Sigma_0 = N^{-1}(Q_0^{-1} + Q_1^{-1})$이며 $n_0 - 3 = Q_0 N$, $n_1 - 3 = Q_1 N$이다. 또한 귀무가설 하에서 Z는 표준정규분포로 접근한다. 표본의 수와 검정력에 관한 식은 다음과 같다.

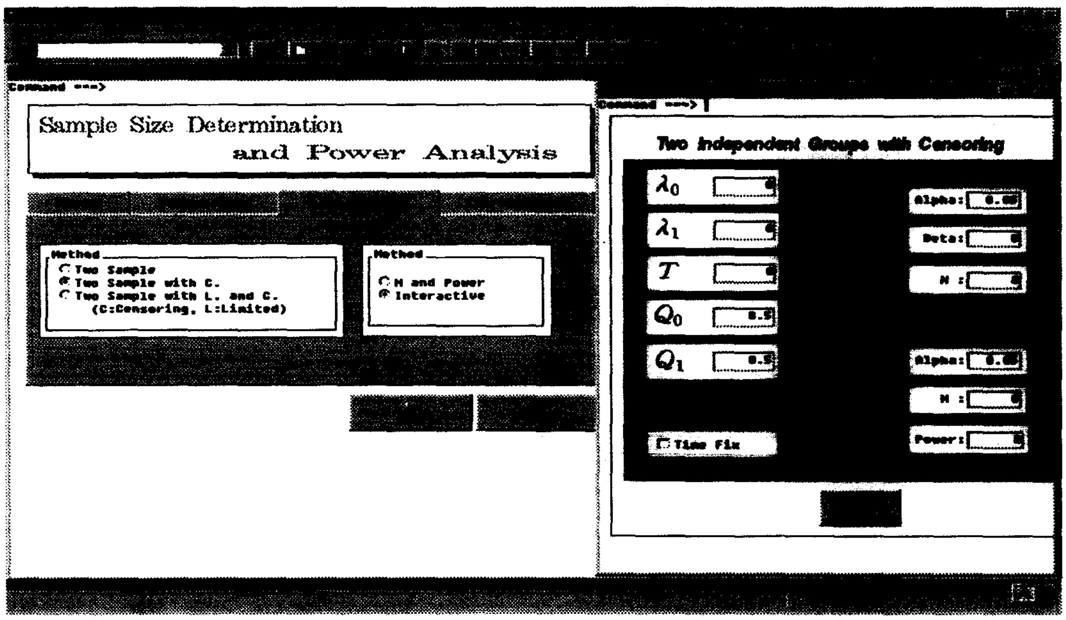
$$\frac{\sqrt{N}\,|C(\rho_1) - C(\rho_0)|}{\sqrt{Q_0^{-1} + Q_1^{-1}}} = Z_\alpha + Z_\beta.$$

이식을 N과 Z_β에 대하여 각각 풀면 표본의 수와 검정력을 구할 수 있다.

$$N=\frac{\{(Z_\alpha+Z_\beta)\sqrt{Q_0^{-1}+Q_1^{-1}}\}^2}{(C(\rho_1)-C(\rho_0))^2},$$

$$Z_\beta=\frac{\sqrt{N}\,|C(\rho_1)-C(\rho_0)|}{\sqrt{Q_0^{-1}+Q_1^{-1}}}-Z_\alpha.$$

[화면 15] 중도절단이 있는 경우의 표본의 수와 검정력 계산 - interactive

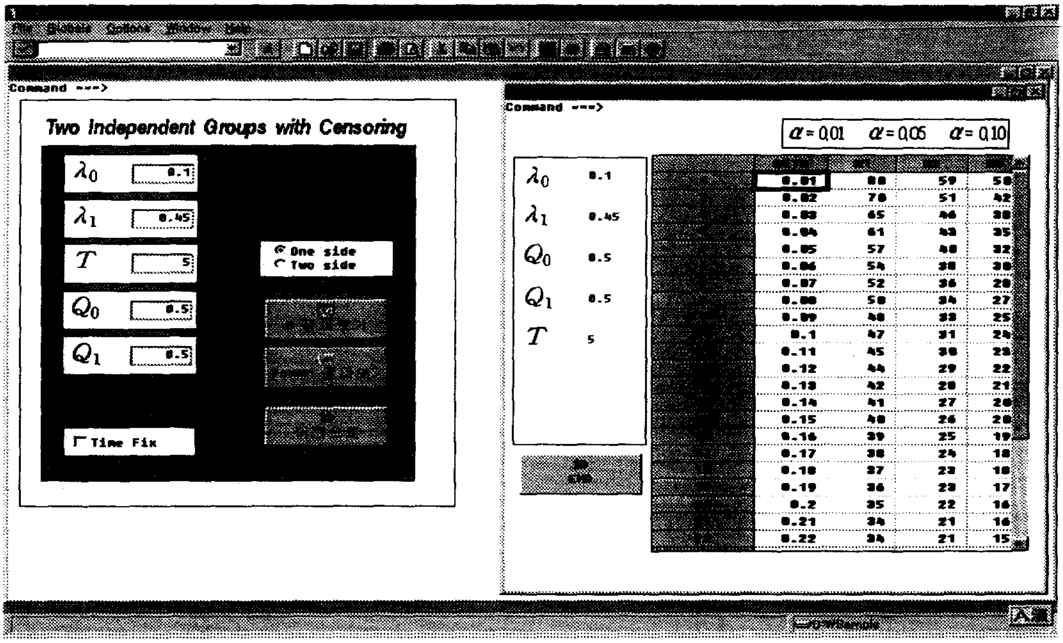


7. 맺는 말

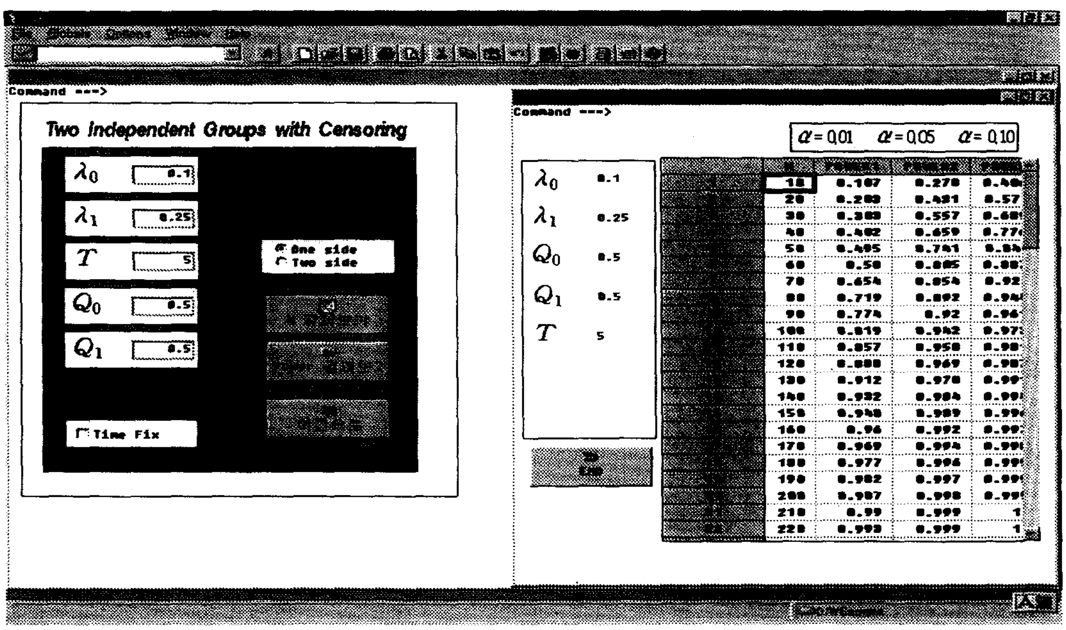
본 연구에서는 각 경우에서의 대표적인 표본의 수와 검정력에 관련된 식을 통하여 표본의 수와 검정력을 찾고자하였다. 어떠한 경우에 본 논문에서 사용한 공식들 외에 보다 효율적이고 다양한 식들이 연구 발표되었다. 향후 이러한 공식을 찾아 소개된 시스템을 보완함이 바람직하며 본 연구에서 다루지 못한 여러 경우의 통계모형을 포함하는 확장작업이 필요하다. 또한 이용자들이 편리하도록 사용된 표본의 수나 필요한 내용을 볼 수 있는 도움말 기능을 첨부함이 바람직하다.

< 표 2 > 신뢰수명분석 문제에서의 표본의 수와 검정력

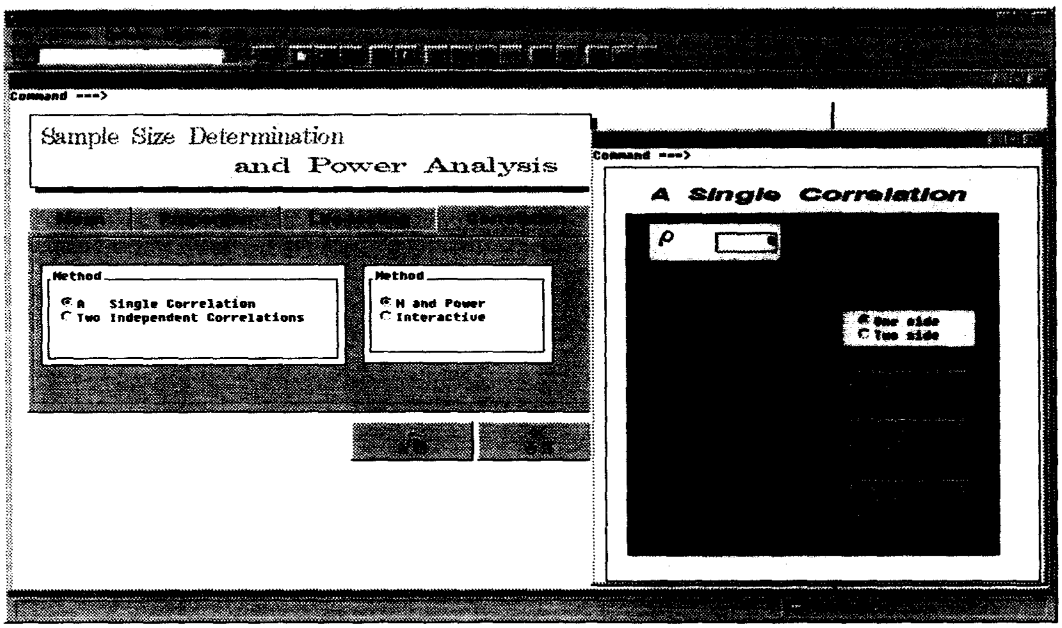
- 표본수 -



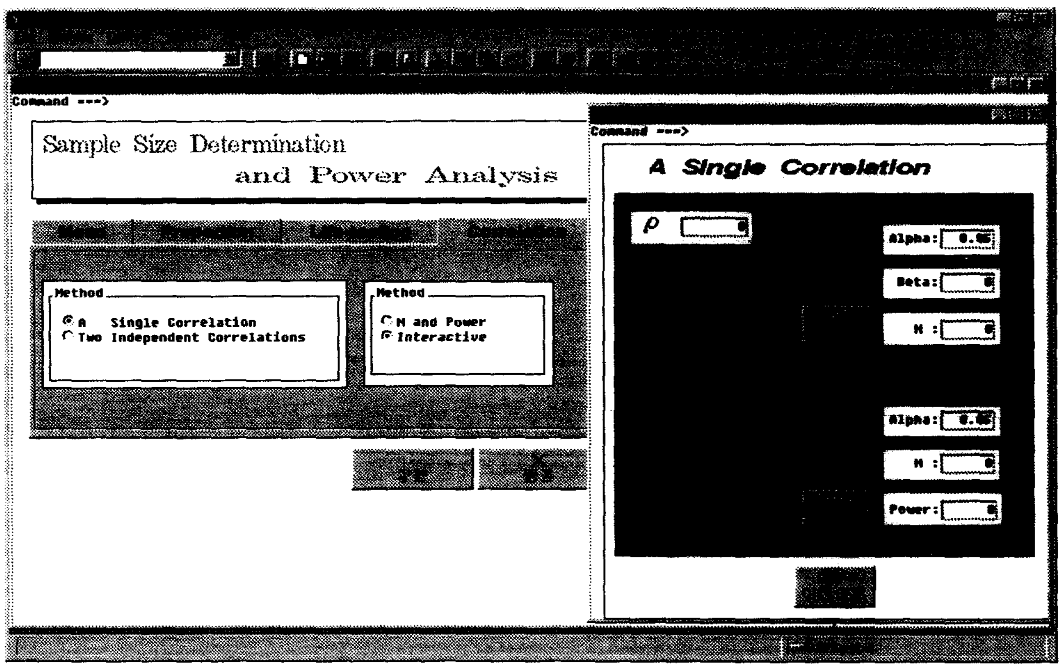
- 검정력 -



[화면 16] 상관분석에서의 표본의 수와 검정력 계산



[화면 17] 상관분석에서의 표본의 수와 검정력 계산 - interactive



참고문헌

[1] Bain, L.J.(1978), *Statistical Analysis of Reliability and Life-Testing Models - Theory and Methods*, Marcel Dekker, inc.

[2] Desu, M.M. & Raghavarao, D.(1990), *Sample Size Methodology*, Academic Press.

[3] Freiman J.A., Chalmers T.C., Smith, H., Kuebler R.(1978), "The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Controlled Trial," New England Journal of Medicine, Vol. 299, pp. 690-694.

[3] Goldstein, R.(1989), "Power and Sample size via MS/PC-DOS Computers," The American Statistician, Vol. 43, No. 4, pp. 253-260.

[4] Lawless, J.F.(1982), *Statistical Models and Methods for Lifetime Data.* Wiley.

[5] SAS Institute Inc.(1996), *SAS/AF Software, Release 6.11*, SAS Institute Inc.