

1. 가설검정의 의의

1) 가설과 가설검정의 개념

예를 들어, K대학교 행정학과 학생들을 대상으로 통학 시간과 학점 관계에 대하여 연구를 한다고 가정하자. 이 때, 가설은 “ 통학 시간이 길어지면 길어질수록 학점은 낮아질 것이다.” 라고 제시된다.

가설은 연구자가 분석방법을 채택하여 분석을 본격적으로 하기 전 그 결과를 미리 예측해 보는 것이다.

2) 가설의 조건

가설로서 인정을 받으려면 몇 가지 조건을 만족하여야 한다.

1. 가설은 간단명료해야한다.
2. 최소한 두 변수 간의 관계를 확인하는 진술이어야한다.
3. 양적으로 표현가능하여야 한다.
4. 상관관계 혹은 인과관계 여부가 명확해야한다.
5. 긍정적 영향인지 부정적 영향인지 명확해야한다.

3) 가설의 유형

분석을 위한 가설을 통계적 가설이라 한다. 통계적 가설은 대립가설(연구가설)과 귀무가설(영가설)로 나뉜다. 대립가설은 연구문제에 대한 잠정적인 해답으로 연구자가 기대하는 결과이고 귀무가설은 대립가설과 논리적으로 반대되는 진술이다. 귀무가설이 채택되면 대립가설은 기각되었음을, 귀무가설이 기각되면 대립가설이 채택되었음을 의미한다.

4) 양측검정과 단측검정

행정학과와 경영학과 학생들의 토익점수 차이가 나는지를 검정한다고 가정하자. 이 예에서 귀무가설은 “행정학과와 경영학과의 토익점수는 차이가 없다.”이다. 대립가설은 “행정학과와 경영학과의 토익 점수는 서로 다르다.” 또는 “행정학과의 토익 점수가 경영학과의 토익 점수보다 클 것이다.”이다. 첫 번째 경우의 가설을 수식으로 표현하면 $H_1: \mu_1 \neq \mu_2$ (μ_1 : 행정학과 학생들의 토익점수의 평균, μ_2 : 경영학과 학생들의 토익점수의 평균) 이고, 두 번째 경우의 대립가설을 수식으로 표현하면 $H_1: \mu_1 > \mu_2$ 이다.

2. 모집단과 표본추출

1) 모집단과 표본추출의 의미

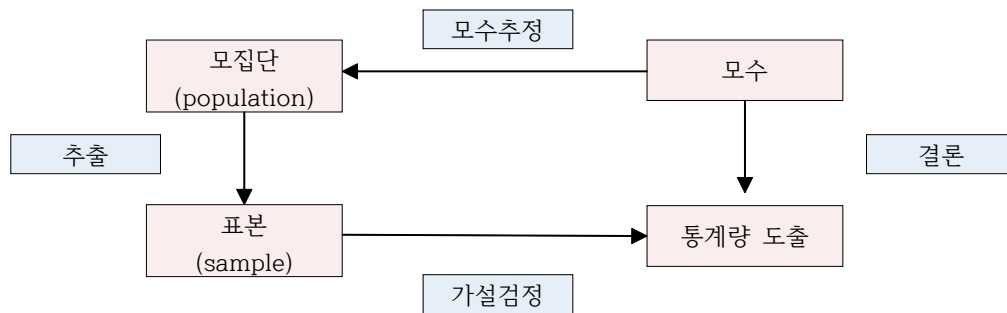


그림1. 분석의 단위

연구의 대상이 되는 모집단 중 일부를 표본으로 추출하여 추출된 표본을 대상으로 연구자는 검정통계량을 도출하게 된다. 검정통계량은 가설검정의 기본이 되는 통계량으로서 이를 바탕으로 연구자는 모수를 추정하고 가설의 채택 여부를 결정하게 된다. 통계량을 도출하여 이를 바탕으로 모집단에 적용하는 과정을 모수추정이라 한다. 모집단 전체를 대상으로 하지 않고 일부만 표본으로 추출하여 연구하는 이유는 첫째, 시간과 비용, 둘째, 표본추출만 잘 이루어지면 표본을 대상으로 진행한 연구가 오히려

더 정확한 정보를 전달해 줄 수 있다. 셋째, 더 많은 양의 정보를 획득할 수 있다.

2) 표본추출의 방법

표본추출의 방법은 크게 확률표본추출과 비확률표본추출로 구분할 수 있다. 확률표본추출은 표본추출 과정에서 연구자의 주관이 개입되지 않고 사례가 표본으로 선택될 확률이 사전에 알 수 없는 경우이고, 비확률표본추출은 연구자의 의도대로 표본을 뽑는 것을 의미한다. 다음은 확률표본추출 방법과 개념이다.

확률표본추출 방법	개념
단순무작위 추출	모든 사례를 포함하는 표본추출을 위한 목록을 작성한 후, 무작위 과정을 거쳐 표본을 추출
계통적 표본추출	표본추출을 위한 목록을 작성한 후, 표본추출 간격 $1/k$ 를 계산, 첫 단위는 무작위로 추출하고 매 k 번째 구성 단위를 추출
층화표본 추출	모집단을 뚜렷한 특징을 가진 소집단별로 표본추출을 위한 목록을 작성, 각 소집단별로 표본을 무작위로 추출한 다음, 이들을 합산
집락표본 추출	각 행정적 또는 조직 체계상 자연스럽게 구분된 집락에 대한 표본추출을 위한 목록을 작성, 집락에 대하여 무작위로 추출, 다단계추출에서는 이 과정을 되풀이함.

3. 가설검정의 일반적 절차

1) 개요

가설검정은 가설의 설정부터 시작한다. 이후 가설 설정을 위하여 가장 적합한 검정 방법을 결정하고 유의수준과 p 값을 결정하게 된다. 결정된 검정 방법에 따라 검정통계량을 계산하고, 이를 바탕으로 서정한 가설의 채택 여부를 확인하게 된다.

2) 가설의 설정

연구자가 추정하는 결과를 기술한 가설이 대립가설이다. 연구자는 단측검정을 위한 가설을 세울지, 양측검증을 위한 가설을 세울지를 결정하게 된다. 단측 검정이 양측검정에 비해 귀무가설을 기각할 가능성이 높아진다. 이를 검정력이 높아진다고 표현하며 단측검정이 양측검정에 비해 검정력이 높다.

3) 검정 방법의 결정

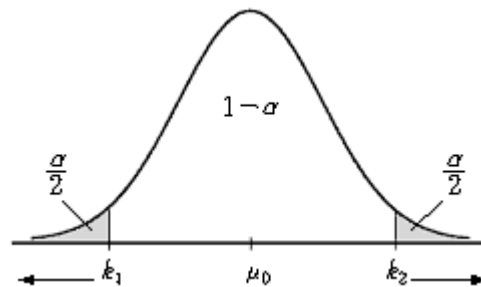
검정방법을 결정한다는 의미는 가설 검정에 활용할 분포를 결정하는 것과 동일한 의미이다.

4) 유의수준과 유의확률의 결정

유의수준이란 가설검정에서 용납될 수 있는 오류의 정도 혹은 가설검정이 오류를 범할 확률을 의미하며 구간추정의 신뢰수준과 같은 의미이다. 다시 말해, 가설검정에서 표본을 분석하여 얻은 통계량이 일정한 기각역에 들어갈 확률을 의미하며 오차 가능성이라 할 수 있다. 일반적으로 유의수준은 1%, 5%, 10%를 사용한다. 10%의 유의수준을 사용하면 5%, 1%의 유의수준을 사용할 때 보다 귀무가설이 기각될 확률이 높아진다. 유의수준은 일반적으로 α 로 표기한다.

p값으로 표현되는 유의확률은 검정통계량이 귀무가설을 지지하는 정도이다. 따라서 이 값이 작아질수록 대립가설이 지지될 가능성이 높아지고, 이 값이 커질수록 대립가설이 기각되고 귀무가설이 지지될 가능성이 높아진다.

다음 그림은 양측 검정일 경우 채택역, 기각역, 임계값을 도식화한 것이다.



5) 검정통계량의 계산

검정통계량이란 모수통계량을 계산할 수 없으므로 가설의 검정을 위하여 활용하는 표본통계량을 의미한다. 유의수준 α 에서 검정통계량을 Z분포를 활용하는 z값, T분포를 활용하는 t값, F분포를 활용하는 F값 등으로 전환한 다음 그 값의 p값이 α 보다 작으면 귀무가설을 기각하고 크면 채택하게 된다.

4. 가설검정의 오류

1) 제1종오류와 제2종오류

표본추출을 확률표본추출에 의하여 정확히 진행하였다고 가정하여 표본통계량이 모수와 정확히 일치할 수는 없다. 표본통계량과 모수의 차이를 표본오차라고 하는데 가설검정에서 표본오차는 피할 수 없다. 가설검정에서 발생할 수 있는 오류는 두가지가 있는데 다음과 같다.

구분	귀무가설이 참	귀무가설이 거짓
귀무가설 채택	옳은 결정	제2종오류(β)
귀무가설 기각	제1종오류(α)	옳은 결정

귀무가설이 참임에도 불구하고 귀무가설을 기각할 때 발생하는 오류를

제1종오류, 귀무가설이 거짓임에도 불구하고 귀무가설을 채택하는 오류를 제2종오류라고 한다. 제1종오류가 작아지면 제2종오류는 커지고, 제1종오류가 커지면 제2종오류는 작아진다.

2) 적용

영화 해운대에서 해일이 도시를 덮쳤는데 대피명령을 내렸다면 매우 적절한 의사결정이다. 또한 해일이 도시를 덮치지 않았는데 대피명령을 내리지 않았다면 이 역시 적절한 의사결정이다. 그런데 해일이 오지 않았음에도 대피명령을 내리거나(제1종오류), 해일이 왔음에도 대피명령을 내리지 않았다면(제2종오류) 의사결정에 오류가 발생한 것이다.

이 상황에서 어떤 오류가 더 치명적일까? 보편적인 가치라는 측면에서 판단해 보면 제2종오류가 제1종오류에 비해 훨씬 치명적이다. 제1종오류는 정부의 신뢰도가 하락한 정도로 무마되고 시간이 어느 정도 지나면 잊혀질 수 있는 오류이지만 제2종오류는 막대한 사상자가 발생할 수 있는 오류이다.