조사 및 통계분석

IV. 가설검정

1. 가설검정의 의의
2. 가설이란?
3. 가설의 구비조건
4. 가설의 유형
5. 양측검정과 단측검정
6. 가설검정의 절차
7. 통계적 유의성
8. 유의 수준
9. 유의 확률
10. 단측 검정의 예
11. 양측 검정의 예
12. 가설검정의 오류
13. 적용: 가설검정의 오류

가설검정의 의의

가설검정(hypothesis test)이란? 모수(the parameter)에 대하여 가정되었거나 예상되는 값의 진위를 표본통계량(a statistic)를 이용하여 판정하는 통계적 추론방법이다.

먼저 제기된 논쟁을 기초로 대립가설(alternative hypothesis)을 설정하고, 이에 상반되는 귀무가설(null hypothesis)이 통계적으로 유의한가를 판단한다. 귀무가설은 영가설이라고도 하며 보통 H_0로 표시한다. 귀무가설은 실제로 검정대상이 되며, 대립가설은 검정대상이 아니라, 귀무가설이 기각될 때 자동적으로 수락되는 것이다.

결국 귀무가설이 옳다는 가정아래 표본분포를 설정하고,이 분포를 기반으로 표본조사로부터 추정된 표본통계량이 통계적 유의한가 (우연이라고 보기에는 표본통계량이 너무 크거나 또는 너무 작을 경우)를 평가하여, 유의하다면 귀무가설을 기각하고, 반대로 유의하지 못하면 귀무가설을 채택하게 된다.

가설이란?

Hypothesis ?

가설(hypothesis)이란? "연구 문제에 대한 최선의 예측 또는 잠정적인 해답"으로 정의한다.

하지만 가설은 예측(prognose)과는 다르다. 예측도 이제까지 알려지지 않은 사태에 대한 진술이라는 점에서는 같지만, 그 어떤 설명적인 기능도 가지지 않는다. 따라서 예측은 논리적 추론의 전제에 해당한다기보다는 오히려 그 결론이라 할 수 있다.

예 1. 커피를 마시는 모든 사람들은 인스턴트 커피보다 원두커피를 좋아한다.

예 2. 통학시간이 길수록 그 학생의 학점은 낮아진다.

예 3. 100원짜리 동전의 앞과 뒤가 균형이 맞지 않는다.

가설의 구비조건

1) 가설은 간단 명료해야한다.
2) 최소한 두 변수간의 관계를 확인하는 진술이어야 한다.
3) 양적으로 표현이 가능해야한다.
4) 상관관계 혹은 인과관계 여부가 명확하여야 한다.
5) 긍정적 영향인지 부정적 영향인지 방향성 역시 포함되어야 한다.
6) 진위여부가 계량분석 방법을 통해 입증이 가능하여야 한다.

가설의 유형
Null Hypothesis vs. Alternative Hypothesis

통계적 가설은 귀무가설(null hypothesis)과 대립가설(alternative hypothesis)로 나누인다. 귀무가설은 영가설, 대립가설은연구가설이라고도 불리운다.

대립가설(H_1)은 "어떤 것이 있다, 존재한다, 나타난다, 다르다" 라는 가설이며, **귀무가설**(H_0)은 "어떤 것이 없다, 존재하지 않는다, 나타나지 않는다, 동일하다 " 라는 가설이다.

일반적으로 대립가설은 연구문제에 대한 잠정적인 해답으로 연구자의 주장 또는 논쟁을 나타낸다. 이 두 가설은 하나의 연구 주제에서 논리적으로 한쪽을 채택하면 다른 한쪽을 반드시 기각하는 관계로 짜여진다.

예: 대립가설– " 행정학과 학생들과 경영학과 학생들의 토익 성적은 차이가 있다."
귀무가설 – " 행정학과 학생들과 경영학과 학생들의 토익 성적은 차이가 없다."

양측검정과 단측검정

Both-side Test vs. One-side Test

예를 들어, 행정학과와 경영학과 학생들의 토익 점수가 차이가 나는가를 계량분석방법을 통해 검정하려고 한다. 이 때 2가지 가설이 가능하다.

1) “행정학과와 경영학과 학생들의 토익 점수가 차이가 있다”.

   $H_{!}: \mu_{"} \neq \mu_{\#}$ 양측검정(both-side test)

2) “행정학과 학생들의 성적이 경영학과 학생들보다 높다”.

   $H_{!}: \mu_{"} > \mu_{\#}$ 단측검정(one-side test)

이 때 귀무가설은 " 두 학과 학생들의 성적은 차이가 없다 ($H_{\$}: \mu_{"} = \mu_{\#}$) ”로 동일하다.

단 2가지 가설의 형태에 따라 양측 모두로 검정대상으로 할 것인가, 아니면 한 측만 대상으로 할 것인가를 결정한다.

가설검정의 절차

1)가설의 설정: 대립 가설을 결정한다. 이 때 단측 검정으로 할 것인가 양측 검정으로 할 것인가를 정한다.

2) 검정방법의 결정: 가설 검정에 활용할 통계적 분포를 결정한다. 즉, t분포, 정규분포, F분포, 카이제곱 분포 등.

3)유의 수준의 결정: 유의수준(α)이란 가설검정에서 용납될 수 있는 오류의 정도를 의미한다.

4)검정통계량과 유의확률 계산: 검정통계량은 가설의 검정을 위하여 활용하는 표본통계량을 말한다. 표본분포를 설정하고, 이를 기반으로 유의 확률(P-value), 즉 검정통계량이 귀무가설을 지지하는 정도를 계산한다.

5)가설 채택 여부 결정: 해당 검정통계량에 해당하는 유의확률이 유의수준보다 작으면 귀무가설을 기각하고, 반면에 크면 채택(혹은 지지)하게 된다.

통계적 유의성
Statistical Significance

통계적 유의성은 모수에 대한 가설이 통계적으로 의미가 있는 정도를 말한다. 다시 말해서, 어떤 실험 또는 조사 결과를 두고 "통계적으로 유의하다."라고 하는 것은 확률적으로 봐서 단순한 우연이라고 생각되지 않을 정도로 그 결과가 의미가 있다는 뜻이다. 반대로 "통계적으로 유의하지 않다."라고 하는 것은 그 결과가 단순한 우연일 수도 있다는 뜻이다.

통계적으로 유의하다라는 말은 단순히 연구자의 결론, 혹은 주장이 합리적이다, 타당하다라는 것을 의미하지는 않는다. 이는 통계적 기법을 적용하여 어떤 가설을 평가했을 때, 연구자의 생각이 단순히 우연이라고 판단되지 않을 정도로 옳을 가능성이 높다는 것을 의미한다.

즉, 통계적으로 유의하다는 말은 통계적으로 의미가 있다는 것으로 "우연이라고 보기에는 어려운 유의미한 차이가 있다"는 것을 말한다.

유의 수준
Level of Significance

유의수준(level of significance)은 통계적인 가설 검정에서 사용하는 기준 값이다. 가설 검증을 할 때, 표본에서 얻은 표본 통계량이 일정한 기각역(rejection area)에 들어갈 확률, 즉 오차 가능성을 얼마나 허용할 것인가의 기준을 말한다. 일반적 시회과학연구에서는 유의 확률(P-값)이 0.05 또는 0.01을 넘지 않는 수준으로 유의 수준(α)을 정한다. 유의확률은 귀무가설이 실제로 참라고 할 때, 이 귀무가설을 잘못 기각할 확률을 말한다.

예를 들어, 유의확률이 0.04라면 귀무가설이 옳음에도 불구하고 100번 중 4번은 그르다고 판단할 확률을 나타낸다는 뜻이다. 유의수준을 5%로 정한다면, 즉, 약 95% 확신하는 수준에서 올바른 의사결정을 하겠다는 뜻이 된다. 만약 95%에 미치지 못할 경우, 그 가설은 "유의수준 0.05에서 기각되었다"라고 말하는데, 이것은 곧 우리가 잘못된 의사결정을 내릴 확률이 0.05보다 크다라는 것을 의미한다.

유의 확률

Significance Probability

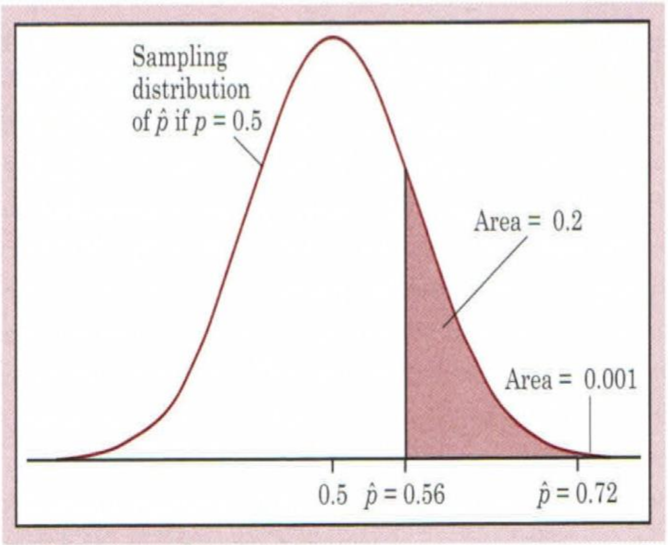


Figure 22.2 The sampling distribution of the proportion of 50 coffee drinkers who prefer fresh-brewed coffee. This distribution would hold if the truth about all coffee drinkers is that 50% prefer fresh coffee. The shaded area is the probability that the sample proportion is 56% or greater.

유의 확률**(**significance probability) 또는 **P-값**(*P*-value)은 귀무 가설이 맞다고 가정할 때, 추정된 표본통계량보다 같거나 더 극단적인 통계량이 실제로 관측될 확률이다(옆의 그림에서 그늘진 부분).

P-값이 작을수록 그 정도가 약하다고 보며, 특정 값(유의수준)보다 작을 경우 귀무가설을 기각하는 것이 관례이다.

즉 P-값이 유의수준보다 작다면, 귀무가설을 기각하고, 대립가설을 채택하게 된다.

양측 검정의 예

Buffon은 동전을 던져 동전의 앞 면이 몇 번 나오는가를 실험하였다. 그 결과 4040번 던져서 2048번이 앞면이 나오는 결과를 얻었다. 표본 통계량인 앞 면이 나온 비율은 $\hat{p} = 2048/4040 = 0.507$ 로 추정되었다. 이 결과를 보고 동전이 앞뒤 균형이 맞지 않는다라는 주장이 있다.

1단계: 가설 설정 귀무가설 $H_0: p = 0.5$, 대립가설 $H_1: p \neq 0.5$

2단계: 표본 분포 귀무가설에 근거한 표본 분포는 평균 $p = 0.5$, 표준편차 $\sqrt{p(1-p)/n} = \sqrt{(0.5)(0.95)/4040} = 0.00787$인 정규분포로 가정한다.

3단계: 유의 확률(P-값) 산출 앞 면이 나온 비율 0.507일 때의 표준점수 $z = (0.507 - 0.5)/0.00787 = 0.88945$ (≈ 0.9)이며, 앞 면이 나올 비율이0.5임에도 불구하고, 0.507보다 크거나, 0.493보다 작을 확률은 약 0.37이 된다.

4단계: 결론 동전의 앞뒤가 균형이 잡혀 있음에도 불구하고 표본 통계량이 0.507보다 크거나, 0.493보다 작을 확률은 0.37로 귀무가설을 기각할 수 없다. 왜냐하면, 균형이 잡혀 있음에도 100번 중 37번은 이러한 결과를 가져다 줄 수 있기 때문이다. 결국 귀무가설을 기각하기에는 충분한 통계적 유의성을 갖고 있다고 할 수 없다.

양측 검정의 예

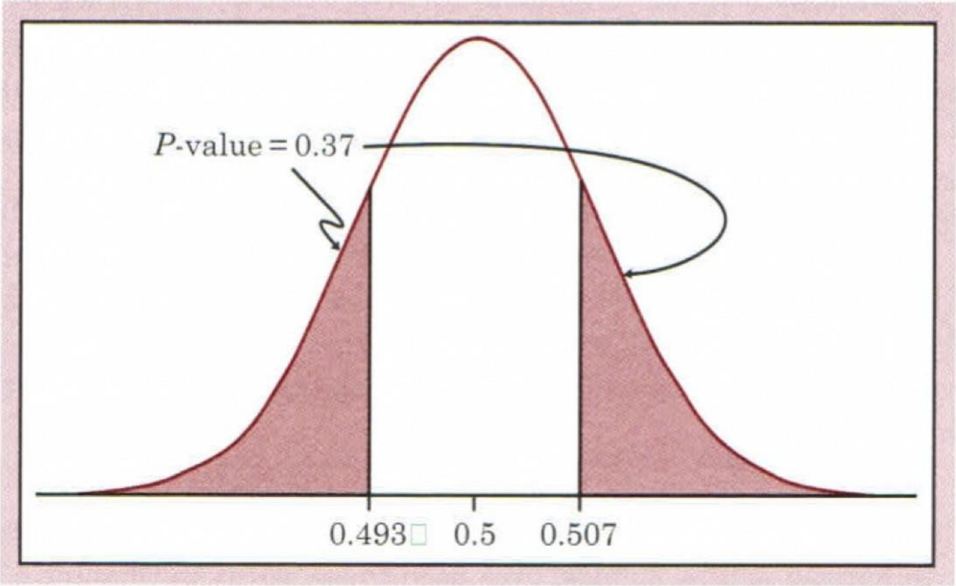


Figure 22.4 The *P*-value for testing whether Count Buffon's coin was balanced. This is the probability, calculated assuming a balanced coin, of a sample proportion as far or farther from 0.5 as Buffon's result of 0.507.

가설 검정의 오류

표본추출을 확률표본추출에 의해 정확히 진행하였다고 하더라도 '표본 통계량(a statistic)'이 모수(the parameter)와 정확히 일치할 수 없다(이를 '표본 변동(sampling variability)'이라고 함). 이렇게 표본 통계량과 모수의 차이를 '표본 오차(sampling error)'라 하는데 가설검정에서 표본 오차는 피할 수 없다, 표본오차가 크면 가설검정의 오류가 발생할 수 있다.

구분	귀무가설이 참	귀무가설이 거짓
귀무가설 채택	옳은 결정	제2종 오류(β)
귀무가설 기각	제1종 오류(α)	옳은 결정

적용: 가설검정의 오류

제1종 오류와 제2종 오류

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(제1종 오류), 해일이 왔음에도 불구하고 대피명령을 내리지 않았다면(제2종 오류) 의사결정에 오류가 발생한 것이다.

보편적 가치 측면에서 보면 제2종 오류가 휠씬 더 치명적이다. 따라서 제2종 오류가 발생할 가능성이 조금이라도 있는 상황이면 제1종 오류가 발생할 가능성을 감안하더라도 귀무가설을 기각하는 것이 현명한 선택이다.

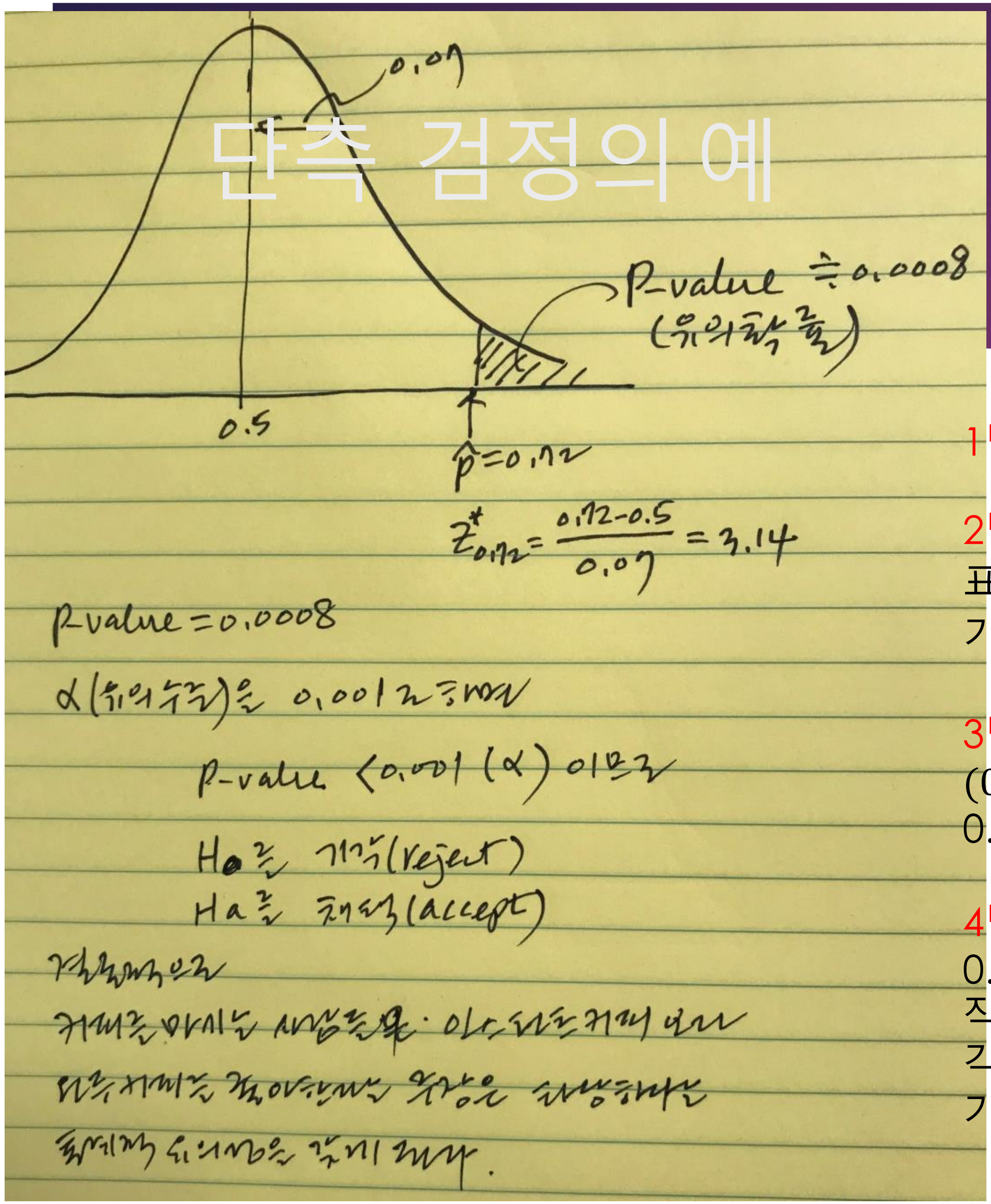
단측 검정의 예

“커피를 마시는 모든 사람들은 원두커피를 좋아한다.”라는 주장(claim)이 있다. 이 주장을 맞는가를 판단하기 위해 다음과 같은 실험을 하였다.

커피를 마시는 무작위로 50명을 선택하여 동일한 컵에 동일한 온도, 유사한 맛을 내는 인스턴트 커피와 원두 커피를 맛보게 한 뒤 선호도를 조사하였다. 그 결과, 36명이 원두 커피를 선호하는 것으로 조사되♘다.

즉, 원두 커피 선호율 ^ = 36/50 = 0.72 로 추정되♘다.

단측 검정의 예



1단계: 가설 설정 귀무가설 $H_0: p = 0.5$, 대립가설 $H_a: p > 0.5$

2단계: 표본분포 설정 귀무가설에 근거한 표본 분포는 평균 $p = 0.5$, 표준편차 $\sqrt{p(1-p)/n} = \sqrt{(0.5)(0.95)/50}$ = 0.07인 정규분포로 가정한다.

3단계: 유의확률(P-값) 산출 선호율 0.72일 때의 표준점수 $z = (0.72 - 0.5)/0.07$ =3.14이며, 선호율이 0.5임에도 불구하고, 0.72보다 클 확률은 0.001보다 작다.

4단계: 결론 모집단의 선호율이 0.5임에도 불구하고 표본통계량이 0.72가 나올 확률은 유의 수준을 0.001로 정했을 경우, 이보다 작으므로(1000번에 1번 정도도 나올 수 없음), 귀무가설을 기각한다. 결국 연구자의 주장, 즉 원두커피를 좋아한다는 대립가설을 채택한다.