사회통계
임계값을 이용한 가설검정

김현우, PhD[1]

[1] 충북대학교 사회학과 조교수

진행 순서

1. 가설검정의 기본 논리
2. 임계값을 이용한 가설검정

가설검정의 기본 논리

가설검정의 기본 논리

오늘 수업까지 포함하여 통계적 추론의 기초가 끝난다.

- 사실 이 부분이야말로 학부 사회통계학의 중급 레벨로 나아가기 위한 관문이다. 가장 어렵고 지루한 파트라고 할 수 있다.
- 5주차에서 "확률이론"의 기초를 배웠다.
- 6주차에서 "이론적 확률분포"로 통계적 추론의 기초 전반부를 배웠다.
- 7주차에서 "표집분포와 추정"으로 통계적 추론의 기초 후반부와 통계적 추론 첫번째 기법인 추정(estimation)을 배웠다.
- 9주차인 오늘은 통계적 추론의 두번째 기법인 가설검정(hypothesis test)을 배운다.

가설검정의 기본 논리

가설이란 모수에 대한 잠정적 주장이다.

- 가설(hypothesis)이 세워지는 대상은 (모집단의 특성인) 모수(parameter)이지 (표본의 특성인) 통계량(statistic)이 아니다.
- 통계량(e.g., 표본평균)에 대해서는 가설 따위가 필요없다(Why?). 우리는 '잠정적 주장'을 내세울 필요없이 직접 그 자리에서 확인하면 되기 때문이다.
- 한편 모수(e.g., 모평균)는 곧바로 알 수 없다. 그러므로 "아마도 모평균 μ는 μ_0가 아닐까?" 하고 잠정적인 주장을 내세울 필요가 있는 것이다.

가설검정의 기본 논리

먼저 귀무가설과 대립가설을 바르게 세울 수 있어야 한다.

- 연구에 앞서 모평균에 대해 귀무가설(null hypothesis)을 세운다.
- 예컨대 "올해 중간시험의 평균 점수는 50점이다"라는 귀무가설은 다음과 같이 표현한다.

$$H_0 : \mu = 50$$

- 다시 한 번 여기서 주목해야 하는 부분은 귀무가설이 "표본평균 $\bar{x}$에 관한 것"이 아니라 "모집단 μ에 관한 것"이라는 점이다!

가설검정의 기본 논리

- 대립가설(alternative hypothesis)은 귀무가설을 기각하면 “자동적으로” 받아들이게 되는 가설이다.
- 예컨대 “올해 중간시험의 평균 점수는 50점이 아니다”라는 귀무가설은 다음과 같이 표현한다.

$$H_a : \mu \neq 50$$

- 귀무가설과 대립가설 사이에서 놓치는 사건이 없어야 한다.
- 예컨대 “올해 중간시험의 평균 점수는 50이다”라고 귀무가설을 세웠는데 대립가설은 “올해 중간시험의 평균 점수는 50보다 크다”라고 하면 잘못된 가설 구조가 된다 (Why?).

가설검정의 기본 논리

가설검정 방식에는 양측검정과 단측검정 두 가지가 있다.

- 양측검정(two-tailed test)에서 귀무가설과 대립가설의 구조는 다음과 같다.

$$H_0 : \mu = \mu_0 \qquad H_a : \mu \neq \mu_0$$

- 이와는 달리 단측검정(one-tailed test)은 둘 중 하나의 구조를 갖는다.

$$H_0 : \mu \geq \mu_0 \qquad H_a : \mu < \mu_0$$

또는

$$H_0 : \mu \leq \mu_0 \qquad H_a : \mu > \mu_0$$

가설검정의 기본 논리

(이름이 암시하듯) 귀무가설은 기각하기 위한 목적으로 세워진다.

- 귀무가설(歸無假說)은 “무로 되돌린다”는 의미를 가지고 있다.
- 내심 내리고 싶은 결론은 오히려 대체가설 쪽이고, 귀무가설은 짐짓 틀리면 좋겠는데 하는 의도를 담고 있다.
- 종종 대립가설을 연구가설(research hypothesis)라고 부르는 이유도 여기에 있다!
- 그러므로 “귀무가설을 기각하지 못한다면” 연구자 입장에서는 아쉬운 일이다.
- 다만 양측검정의 경우에는 언제나 가설 구조가 다음과 같다(Why?).

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

예제 1. 어느 방향제 제조회사에서 생산하는 방향제의 유효시간은 평균 75 일이였다. 이 회사의 전속 조향사인 새우는 자신이 개발한 새로운 첨가제를 넣으면 평균 유효시간이 늘어날 것이라고 기대하고 있다. 새우의 첨가제를 넣은 방향제의 평균 유효시간에 대해 가설검정을 수행할 때, 적절한 귀무가설과 대립가설을 나타내시오.

가설검정의 기본 논리

- 새우가 수행해야 하는 실험은 다음의 가설 구조를 갖는다.
  (1) 귀무가설은 "방향제의 유효시간은 75일이거나 그보다 작다."
  (2) 대립가설은 "방향제의 유효시간은 75일보다 길다."
- 새우가 내심 원하는 방향은 기본적으로 방향제의 유효시간이 늘어나는 것이다 (Why?). 대립가설이 그 방향에 따라 서술한다.

$$H_0 : \mu \leq 75$$
$$H_a : \mu > 75$$

가설검정이란 귀무가설이 표본에 의해 지지되는가를 테스트하는 것이다.

- "중간시험의 평균 점수는 50이다" 또는 $H_0 : \mu = 50$ 이라는 가설을 세웠다.
- 중간시험 전체 응시자 모집단으로부터 충분히 큰($n > 30$) 임의표본(random sample)을 추출하였는데, 이 표본에서 구한 표본평균 $\bar{x}$이 52점이였다.
- 이렇게 표본평균 $\bar{x}$이 "귀무가설에서 설정된" 모평균($H_0 : \mu = 50$)과 제법 비슷하다면 이는 무엇을 시사할까?
- "귀무가설로 세운 모집단의 속성은 표본에 의해서 지지되므로 귀무가설은 아마 옳다" 라고 할 수 있을 것이다.
- 또는 (귀무가설은 기각하기 위한 목적으로 세워지기 때문에) "귀무가설을 아마 기각하지 못한다" 라고 할 수 있다.

가설검정의 기본 논리

- 그런데 표본평균 $\bar{x}$을 다시 계산해보니 50점이 아니라 사실 25점이였다고 해보자.
- 만일 이렇게 표본에서 구한 평균이 "귀무가설에서 설정된" 모평균과 크게 동떨어져 있다면 무엇을 시사할까?
- "귀무가설로 세운 모집단의 속성이 표본에 의해서 지지되지 않으므로 귀무가설은 <u>아마</u> 틀리다" 라고 할 수 있을 것이다.
- 또는 (귀무가설은 기각하기 위한 목적으로 세워지기 때문에) "귀무가설을 <u>아마</u> 기각할 수 있다" 라고 할 수 있다.

왜 자꾸 "아마" 라는 단서가 붙어있을까?

- 하나의 표본평균 $\bar{x}$가 귀무가설로 세워진 모평균 μ과 동떨어져 있더라도 "우연히" 그런 표본이 뽑혔을 가능성을 완전히 배제할 수는 없기 때문이다.
- 표본평균이 귀무가설로 세워진 모평균과 가까우면 가까울수록 아무래도 귀무가설을 지지한다고 볼 수 있다. 그런 값이 뽑힐 확률이 정말로 높아지기 때문이다!
- 반대로 표본의 평균이 귀무가설 평균과 다르면 다를수록 귀무가설은 점점 의심받아 마땅하다.

가설검정의 기본 논리

이것이 오늘 수업의 가장 중요한 논리적 핵심이다.

- 표본의 속성(e.g., 표본평균)이 귀무가설로 제시한 속성(e.g., 모평균)과 제법 가깝다면, (1) 귀무가설이 옳다는 가정 아래 (2) 그런 표본이 나올 확률이 제법 높으니 (3) 귀무가설이 틀리지 않음을 의미한다.
- 반대로 표본의 속성(e.g., 표본평균)이 귀무가설로 제시한 속성(e.g., 모평균)으로부터 너무 동떨어져 있다면, (1) 귀무가설이 옳다는 가정 아래 (2) 그런 표본이 나올 확률이 너무 낮으니 (3) 귀무가설이 틀림을 의미한다.
- 지금 위 두 문장은 너무나 중요하기 때문에 다른 곳에 적어두고 꼭 기억해두자!

가설검정의 기본 논리

어떻게 하면 "아마" 라는 꼬리표를 떼어낼 수 있을까?

- 두 가지 방법이 있다:

  (1) 임계값(critical value)을 이용한 가설검정
  (2) 유의확률(p-value)을 이용한 가설검정

- 둘 다 근본 원리는 똑같다. 모두 표집분포 아이디어와 중심극한정리를 사용하여 통계적으로 "아마"의 확률의 계산한다.

임계값을 이용한 가설검정

임계값을 이용한 가설검정

95% 신뢰구간 안에 포함되는지 여부를 따져 "아마"의 확률을 계산해보자.

- 예를 들어 어떤 모집단에 대해 $H_0 : \mu = 25$라는 귀무가설을 세웠다고 하자(당연히 대립가설은 $H_a : \mu \neq 25$가 된다).
- 이 모집단에서 $n \geq 30$인 표본을 무한히 계속해서 뽑아 그 표본들의 평균을 계산하고, 다시 그 평균의 확률분포(=평균의 표집분포)를 그려보면 (중심극한정리 덕택에) 이것은 정규분포한다는 것을 알 수 있다.
- 표본($n = 30$)에서 표본평균은 $\bar{x} = 52$이고 표본표준편차는 $s = 15$라고 하자.
- 이제 "이 귀무가설이 옳다고 가정하고" 평균의 표집분포의 95% 신뢰구간을 (지금까지 배웠던대로) 계산한다.

임계값을 이용한 가설검정

- 먼저 원점수의 표집분포를 Z-점수로 표준화하여 표준정규분포를 만들고 "최대한 확률이 높은 부분으로만" 곡선 아래 95% 면적을 색칠해보자.
- 표본평균 $\bar{x}$를 Z-점수로 다음과 같이 바꿀 수 있다.

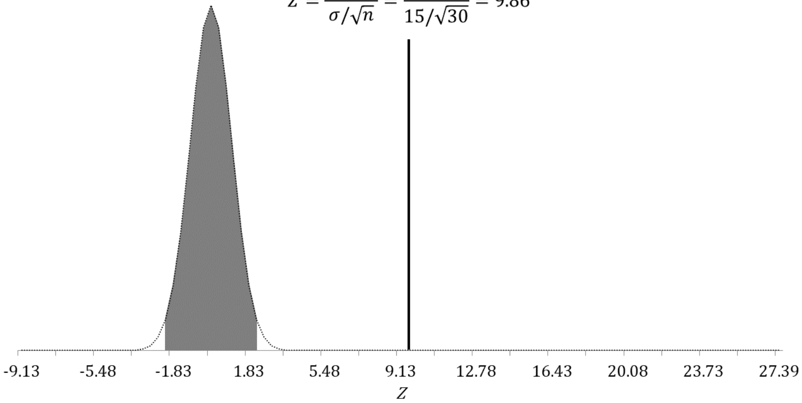
$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{52 - 25}{15/\sqrt{30}} = 9.86$$

- 표준정규분포상에서 $Z_{0.025}$와 $Z_{0.975}$는 NORM.INV(0.025, 0, 1)와 NORM.INV(0.975, 0, 1)을 통해 각각 ± 1.96임을 알 수 있다.
- (Z-점수로 나타낸) 표준정규분포에서 $[-1.96, 1.96]$ 사이의 "곡선 아래 95% 면적"은 수식으로 이렇게 표현된다.

$$P(Z_{0.025} < Z < Z_{0.975}) = 0.95$$

임계값을 이용한 가설검정

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{52 - 25}{15/\sqrt{30}} = 9.86$$



임계값을 이용한 가설검정

이것이 실질적으로 무엇을 의미할까?

- 만일 이 귀무가설이 옳다면 95%의 확률로 (표준화된) 표본평균이 ± 1.96 사이에 있어야 했다(Why?).
- 참고로 이 Z-값을 원점수로 환원해보면 19.6점과 30.4점이다(꼭 확인해보자).
- 그런데 표본에서 계산된 평균 $\bar{x}$은 52점이고 Z-값은 9.86이다. 이것은 1.96보다 훨씬 크다. 그러므로 (지금까지 논리를 뒤집어) 95%의 확률로 귀무가설을 기각한다 (reject).
- 19.6점과 30.4점은 각각 좌우 임계값(critical values)이 된다.
- 즉 표본평균이 좌우 임계값을 벗어나 있다면 귀무가설을 기각할 수 있다.

임계값을 이용한 가설검정

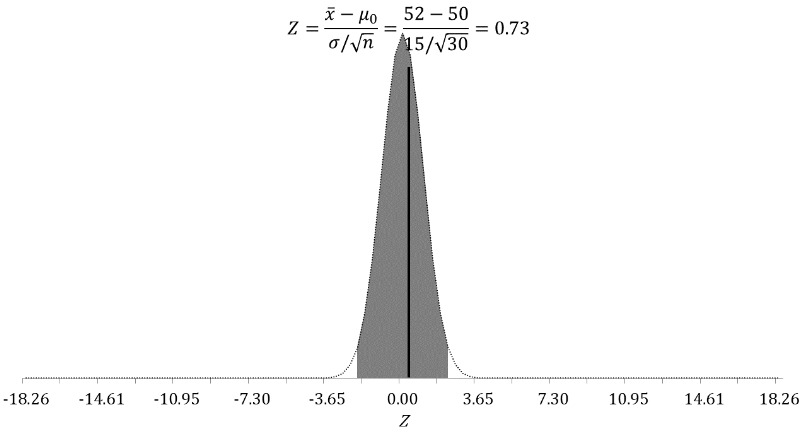
다른 귀무가설을 한 번만 더 테스트해보자.

- 이번에는 $H_0 : \mu = 50$을 귀무가설을 세워보자(당연히 대립가설은 $H_a : \mu \neq 50$이다).
- 이 원점수의 Z-값은 다음과 같다.

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{52 - 50}{15/\sqrt{30}} = 0.73$$

- (귀무가설이 옳다는 전제 아래) 95% 신뢰구간에서 좌우 임계값을 구해보면 각각 ± 1.96이다.
- 좌우 임계값에 대응하는 원점수는 각각 44.6점와 55.4점이다(직접 구해보자).
- 표준화된 표본평균은 0.73이므로 ± 1.96 안에 포함되어 있다. 그러므로 이 귀무가설은 95% 확률로 기각할 수 없다(fail to reject)."

임계값을 이용한 가설검정



임계값을 이용한 가설검정

가설검정에는 조금 특별한 전문용어가 사용된다.

- 만일 (표준화된) 표본평균이 (귀무가설이 옳다는 전제 아래 그려진 표집분포의) 95% 신뢰구간 바깥에 위치해 있다면,
- 95%의 신뢰수준(confidence level) 또는 5%의 유의수준(significance level)에서 통계적으로 유의(statistically significant)하게 귀무가설을 기각한다.
- (표준화된) 표본통계량(e.g., 표본평균)이 임계값 바깥에 놓였을 때 귀무가설을 기각했다.
- 그렇기 때문에 임계값 바깥의 구간은 기각역(rejection region)이라고 부른다.
- 임계값 안의 구간은 채택역(acceptance region)이라고 부른다.

임계값을 이용한 가설검정

신뢰수준은 반드시 95%여야 하는 것은 아니다.

- 하지만 관습에 따라 흔히 90%, 95%, 99%, 99.9%를 많이 본다.
- 신뢰수준을 높일수록 연구자가 원하는 가설을 기각하기 어려워진다(실제로 확인해보자).
- 가설 기각이 어려워질수록 추정 결과를 보수적으로 받아들이겠다는 의미가 된다 (Why?).

임계값을 이용한 가설검정

예제 2. 고래는 청주시에 소재한 C대학 재학생의 매달 용돈 평균이 100만원이라는 귀무가설을 세웠다. 고래가 이 학교 재학생으로 구성된 100명의 임의표본을 뽑아 확인해보니 표본평균은 97만원이고 표본표준편차는 10만원이였다. 고래의 가설을 서술하고 이를 90% 신뢰수준에서 검정하시오.

임계값을 이용한 가설검정

- 고래의 가설은 다음과 같다.

$$H_0 : \mu = 100$$
$$H_a : \mu \neq 100$$

- 이 귀무가설이 옳다는 전제 아래 $\mu = 100$ 인 모집단에서 무한히 많은 표본평균들을 구해 (가상적인) 표집분포를 그린다고 하자.
- 그런데 고래가 구한 표본평균의 Z-값은 다음과 같다.

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{97 - 100}{10/\sqrt{100}} = -3$$

임계값을 이용한 가설검정

- (귀무가설이 옳다는 전제 아래) 표집분포를 표준화하였을 때, 좌우 임계값은 $Z_{0.05}$와 $Z_{0.95}$이다.
- 이것들의 실제 Z-값은 NORM.INV(0.05, 0, 1)와 NORM.INV(0.95, 0, 1)로 알 수 있다. 각각 ± 1.64이다.
- 명백히 아까 구한 표본평균 $Z = -3$은 90% 신뢰구간 바깥에 놓여있다.
- 그러므로 고래의 귀무가설은 90% 신뢰수준에서 통계적으로 유의하게 기각된다.

임계값을 이용한 가설검정

실제로는 굉장히 쉽게 기각 여부를 판정할 수 있다.

- 어차피 언제나 신뢰수준이 90%라면 표준화된 임계값은 ± 1.64, 신뢰수준이 95%라면 표준화된 임계값은 ± 1.96, 신뢰수준이 99%라면 표준화된 임계값은 ± 2.58이다 (Why?).
- 표본평균을 얻은 Z-값으로 표준화한 뒤, 이것의 절대값이 임계값보다 크면 기각할 수 있다.
- 바로 이렇게 쉽게 계산하기 위해 자꾸 Z-값으로 변환하여 계산하는 연습을 하라고 그동안 말했던 거다!