제 18 장 유의성 검정

1. 가설검정의 예시
2. 귀무가설과 대립가설
3. 검정통계량과 유의수준
4. 제1종 오류와 제2종 오류
5. 유의성 검정절차
6. 0과 1의 숫자로만 이루어진 상자: 추출 횟수가 많은 경우
7. 0과 1의 숫자로만 이루어진 상자: 추출 횟수가 적은 경우
8. t-검정

유의성 검정

‘Obvious' is the most dangerous word in mathematics.

-E. Bell, 스코트랜드, 1883~1960

1. 가설검정의 원리

가설검정의 예시

새로운 조세제도 제안하는 국회의원의 주장

- “제안된 세법은 조세수입을 증가시키지도 감소시키지도 않는다.”

주장의 타당성 확인

- 모집단: 10만개의 납세자료를 표본틀(sampling framework)로 사용
- 표본: 표본틀로부터 100개 자료를 무작위로 추출
- 계산: 납세액의 차이=(새로운 세법 하 세금)–(종전 세법 하 세금)
  - 납세액 차이의 표본평균 : -21만 9천원
  - 납세액 차이의 표본표준편차 : 73만원

납세자 1인당 평균 21만 9천원 납세액이 감소하는 것으로 관찰된 것은 우연일까 아니면 실질적일까?

1. 가설검정의 원리

가설검정의 예시

상자모형

- 10만장의 카드가 든 상자에서 100장 무작위 추출
- 각각의 카드에 적힌 숫자는 납세자별 납세액의 차이
- 일단 새로운 세법을 제안한 국회위원의 주장이 옳다고 치자. 즉 잠정적으로 상자의 평균이 0이라고 가정해보자.

가설검정

- 이제 표본평균을 가지고 위 주장의 타당성을 검정해보자.
- (표본평균의 표준오차) = 73만원/$\sqrt{100}$ = 7만 3천원
- (자료와 주장의 표준화된 차이)=(-21.9만원-0)/(7.3만원)=-3
- 참으로 모평균이 0이었다면 위의 표준화된 차이가 -3 또는 이보다 더 극단적인 값으로 얻어졌을 가능성은 0.1%에 불과

2. 귀무가설과 대립가설

귀무가설과 대립가설

귀무가설, null hypothesis, $\boldsymbol{H_0}$:
차이는 우연이다. 상자의 평균은 0이다.

대립가설, alternative hypothesis, $\boldsymbol{H_1}$:
차이는 실질적이다. 상자의 평균은 음수이다.

귀무가설이든 대립가설이든 하나의 가설은 상자모형에 대한 하나의 진술(statement)로 표현된다.

3. 검정통계량과 유의수준

검정통계량

z-통계량 또는 t-통계량이 널리 사용됨

이들 통계량은 자료에서 얻은 통계치와 귀무가설 수치의 표준화된 차이

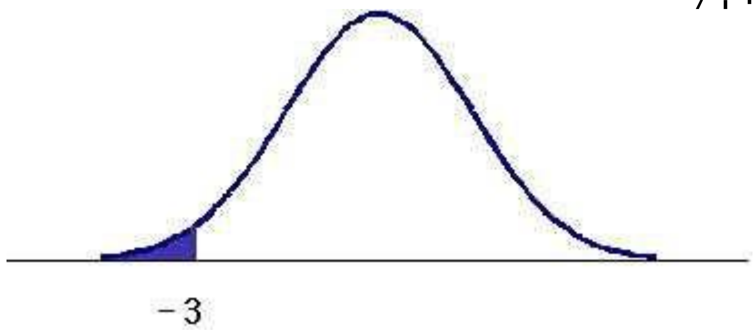
귀무가설 하에서 계산

$$z\text{-통계량 or } t\text{-통계량} = \frac{\text{통계량} - \text{기대값}}{\text{통계량의 표준오차}}$$

3. 검정통계량과 유의수준

p-값은 관측된 유의수준

$$\text{앞의 예에서 계산한 } z\text{- 값} = \frac{-\ 21\text{만}9\text{천원} - 0\text{원}}{7\text{만}3\text{천원}} = -3$$



귀무가설이 옳다면 관측된 차이와 이론적 차이 0의 차이가 관측된 차이의 표준오차 단위로 -3보다 더 작을 확률은 0.1% 에 불과함

p-값은 귀무가설이 옳다는 가정 하에서 실제 관측된 값 또는 그 이상 극단적인 검정통계량 값을 얻을 확률을 의미한다. 앞의 예에서는 0.1%의 확률이 p-값이다.

p-값을 관측된 유의수준이라고도 부른다.

p-값이 작아질수록 귀무가설에 대항하는 반대의 근거는 강해진다.

3. 검정통계량과 유의수준

관측된 유의수준인 *p*-값이 미리 설정한 기준보다 낮으면 귀무가설을 기각

가설 검정의 원리는 ‘모순에 의한 논증법’(argument by contradiction)

- 귀무가설의 주장이 옳다면 어떻게 자료 상으로 그렇게 극단적인 차이가 나타날 수 있을까...... (“귀무가설이 틀리지 않고서야 어찌 이런 일이......”)

- α: 미리 설정된 유의수준(significance level)으로서, 관측된 유의수준인 p-값 판정의 기준으로 사용된다. α의 값으로는 통상 5% 내지 1%의 값이 사용됨

- p-값 $\leq \alpha$: 유의수준 α에서 H_0 기각 (통계적으로 유의하다.)
- p-값 $> \alpha$: 유의수준 α에서 H_0 기각 못함 (통계적으로 유의하지 않다.)

3. 검정통계량과 유의수준

단측검정과 양측검정

대립가설의 주장이 방향성을 가지면 단측검정(one-sided test)이고
방향성을 갖지 않으면 양측검정(two-sided test)임

예시)

$H_o : p = \frac{1}{2}$ vs. $H_1 : p \rangle \frac{1}{2}$ 단측검정

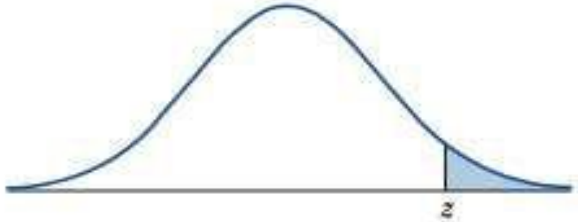
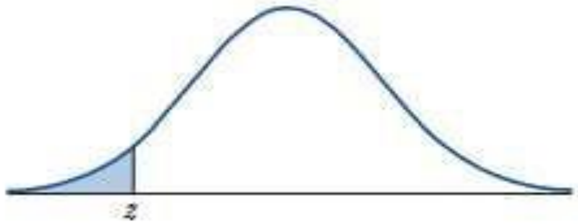
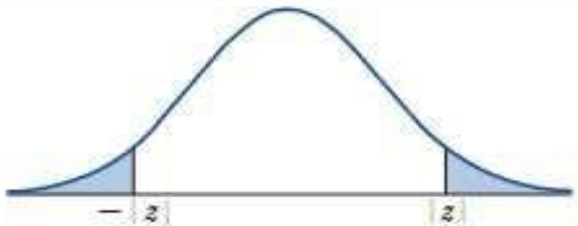
$H_o : p = \frac{1}{2}$ vs. $H_1 : p \langle \frac{1}{2}$ 단측검정

$H_o : p = \frac{1}{2}$ vs. $H_1 : p \neq \frac{1}{2}$ 양측검정

3. 검정통계량과 유의수준

단측검정과 양측검정

표 18－1 대립가설의 형태에 따른 p-값

검정의 종류	귀무가설과 대립가설	p-값 (색칠한 부분)
단측검정	$H_0: \mu=\mu_0$ 대 $H_1: \mu>\mu_0$	
	$H_0: \mu=\mu_0$ 대 $H_1: \mu<\mu_0$	
양측검정	$H_0: \mu=\mu_0$ 대 $H_1: \mu\neq\mu_0$	

4. 제1종 오류와 제2종 오류

생이별의 오류와 착각의 오류

4. 제1종 오류와 제2종 오류

두 종류의 오류를 지적한 우리 속담

- “빈대 잡으려다 초가삼간 태운다” –섣부른 행동의 오류 지적
- “구더기 무서워 장 못 담글까” –미적거림의 오류 지적

4. 제1종 오류와 제2종 오류

가설검정에서의 오류

실제상황 / 검정결과	귀무가설의 상황	대립가설의 상황
귀무가설을 기각 안 함	옳은 결정	제2종 오류
귀무가설을 기각함	제1종 오류	옳은 결정

제1종 오류="생이별의 오류": 애인이 변심하지 않았는데도 변심했다고 잘못 판단하여 생이별하는 오류

제2종 오류="착각의 오류": 애인이 변심했는데도 여전히 자기를 좋아하는 줄 착각하고 자르지 못하는 오류

검정력(power) = 1- 제2종 오류의 가능성.
귀무가설을 기각해야 하는 상황에서 제대로 기각하는 결정을 내릴 확률인 '자를 때 자르는 힘'이 바로 검정력(power)

4. 제1종 오류와 제2종 오류

제 1종 오류와 제 2종 오류

제1종 오류를 범할 확률: 미리 설정된 유의수준보다 작거나 같음. 즉, 미리 설정한 유의수준=제1종 오류를 범할 확률의 상한

유의수준을 미리 일정 정도로 설정함으로써 제1종 오류를 범할 확률을 통제한 채 가설검정을 수행함. 일반적으로 1%, 5%, 10%가 주로 사용됨

제1종 오류를 범할 가능성과 제2종 오류를 범할 가능성간에는 상충관계가 존재함

4. 제1종 오류와 제2종 오류

제 1종 오류와 제 2종 오류

형사사건 재판의 경우

- 변호사는 피의자의 무죄 주장 (귀무가설)
- 검사는 피의자의 유죄 주장 (대립가설)

제1종 오류는 피의자가 무죄임에도 유죄라고 잘못 판정하는 오류이고 제2종 오류는 피의자가 유죄임에도 무죄라고 잘못 판정하는 오류

위 두 종류의 오류 가운데 생사람 잡는 제1종 오류의 심각성이 범죄자를 풀어주는 제2종 오류의 위험보다 크기 때문에 미리 유의수준을 극히 작은 값으로 설정하는 게 타당함

- 구체적으로 유죄의 입증책임(burden of proof)을 검사 측에 지움

4. 제1종 오류와 제2종 오류

제 1종 오류와 제 2종 오류

암 진단의 경우: 의사는 초기 검진을 통해 발암 가능성 판정

- 암세포 없음 (귀무가설)
- 암세포 가능성. 정밀검진 필요 (대립가설)

제1종 오류는 암세포가 없는데도 불구하고 정밀검사를 받게 함으로써 병원비를 낭비하게 하는 오류이고 제2종 오류는 발암 가능성이 있는데도 불구하고 정밀검사를 의뢰하지 않고 집으로 돌려보내는 오류임

두 종류 오류 가운데 상대적으로 제2종 오류의 심각성이 훨씬 크기 때문에 의사는 발암여부를 판정할 때 유의수준을 매우 높게 설정 (과도한 정밀검사 의뢰로 의료비 상승에 기여하는 측면도 있음)

5. 유의성 검정절차

유의성 검정절차

상자모형을 만들고 귀무가설을 세운다.

유의수준 α 를 정한다.

일단 귀무가설이 옳다고 가정한다.

자료(관측치)와 이론(귀무가설)의 차이를 측정할 검정통계량의 값을 구한다.

관측된 유의수준인 p-값을 계산한다.

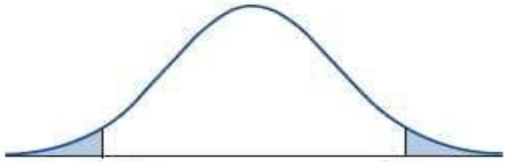
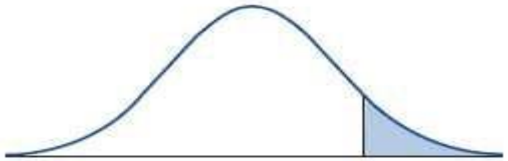
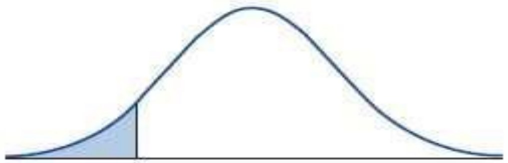
관측된 유의수준인 p-값과 미리 설정된 유의 수준인 α 를 비교한다.

5. 유의성 검정절차

기각역

p-값이 유의수준 α보다 같거나 작아지게 하는 검정통계량 값의 영역

표 18－2 대립가설의 형태에 따른 기각역

검정의 종류	귀무가설과 대립가설	기각역 (색칠한 부분의 가로축 좌표)
양측검정	$H_0: \mu = \mu_0$ 대 $H_1: \mu \neq \mu_0$	 $-z_{\alpha/2}$ $z_{\alpha/2}$ $\lvert z \rvert \geq z_{\alpha/2}$
단측검정	$H_0: \mu = \mu_0$ 대 $H_1: \mu > \mu_0$	 z_α $z \geq z_\alpha$
	$H_0: \mu = \mu_0$ 대 $H_1: \mu < \mu_0$	 $-z_\alpha$ $z \leq -z_\alpha$

6. 0과 1의 숫자로만 이루어진 상자: 추출횟수가 많은 경우

0-1상자와 가설검정: 표본크기가 큰 경우

종합주가지수의 등락을 잘 맞춘다고 소문난 사람이 있었으니... 그의 이름은 '칼'

100거래일 동안 매일매일 주가지수의 등락을 예측하는 게임을 진행

100거래일 중 65거래일의 주가지수 등락을 옳게 예측

무능력자가 주가지수의 등락을 무작위로 찍을 때 맞출 것으로 기대하는 50일보다 15일을 더 맞춘 것은 우연인가, 아니면 '칼'의 칼같은 능력인가?

6. 0과 1의 숫자로만 이루어진 상자: 추출횟수가 많은 경우

0-1상자와 가설검정: 표본크기가 큰 경우

귀무가설: 칼에게는 특별한 능력이 없다.

귀무가설 하에서 설정된 상자모형:

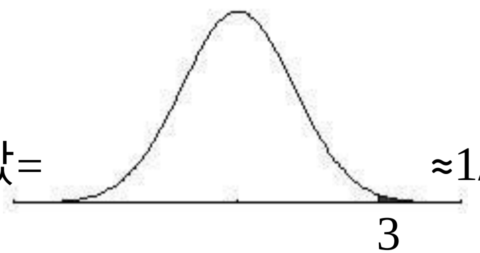
1	0

"귀무가설 하에서 100 거래일 동안의 실험은 좌측 상자처럼 0과 1이 각각 절반씩 들어 있는 상자로부터 100장의 카드를 무작위 복원추출로 뽑은 것과 같다."

상자의 표준편차= $\sqrt{0.5 \times 0.5} = 0.5$

개수의 표준오차= $\sqrt{100} \times 0.5 = 5$

$$z = \frac{65 - 50}{5} = 3,$$

p-값= 

≈1/1000

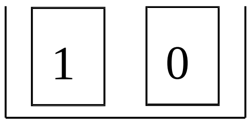
7. 0과 1의 숫자로만 이루어진 상자: 추출횟수가 적은 경우

0-1상자와 가설검정: 표본크기가 작은 경우

SI Kim**의 데이트 신청**

- 이전에는 HR Lee가 2 번 신청 시 한번 꼴로 만나줌
- 최근엔 9 번 데이트 신청에 단 한번만 만나줌

- 과연 HR Lee는 SI Kim에 대해 변심한 것일까.
- 귀무가설: HR Lee는 변심하지 않았다. (“old faithful”)
- 대립가설**:** HR Lee는 변심했다. (“gone yesterday”)



귀무가설 하에서 최근 9번 전화해서 얻은 자료는 마치 좌측 상자에서 무작위 복원추출로 뽑은 9장의 카드와 같다.

$$Z = \frac{1-4.5}{0.5\times\sqrt{9}} \approx -2.33,$$ p-값= 1% ???

7. 0과 1의 숫자로만 이루어진 상자: 추출횟수가 적은 경우

0-1상자와 가설검정: 표본크기가 작은 경우

앞에서 추출횟수가 적기 때문에 정규분포를 이용하기는 무리다.

귀무가설 하에서 데이트 성공 횟수(X)는 매번 데이트 신청의 응답이 독립이면 다음의 이항분포를 따른다. 이 사실을 이용하여 검정하는 게 보다 낫다.

이처럼 표본크기가 작은 경우 이항분포를 이용하여 수행하는 검정을 부호검정(sign test)이라고 부른다.

$$X \sim B(9, 1/2)$$

$$\Pr(X \leq 1 \mid H_0) = {}_9C_0\left(\frac{1}{2}\right)^9 + {}_9C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^8 \approx 2\%$$

8. t-검정

t-검정

과속을 단속하는 고속도로의 속도측정기는 달리는 자동차의 속도를 체계적으로 과대 측정하는 경향이 있는가? (운전자들의 항의에 따른 가설검정)

- 귀무가설: 속도측정에 있어 체계적인 편의(bias)는 없다.
- 고속도로에 설치된 5개 속도측정기를 무작위로 선택
- 시속 100 km로 설정하고 달리는 자동차의 속도를 측정시킴:
  각 측정기는 108.0, 113.0, 99.4, 102.0, 116.6의 속도로 측정함
- 측정치들의 평균 = 107.8, 측정치들의 표준편차 = 7.22
- 표본평균의 표준오차 = $7.22 / \sqrt{5} = 3.23$

$$z = \frac{107.8 - 100}{3.23} \approx 2.4$$

8. t-검정

t-검정

표준정규분포곡선을 이용하여 z=2.4 우측의 면적을 구하면 *p*-값 < 1%

하지만 측정횟수가 적으므로 정규근사를 이용하기 어렵다. 표준정규분포곡선 대신 자유도 4(=관측치수-1)의 t 분포곡선을 이용하자.

t 분포곡선은 표준정규분포곡선과 유사하나 꼬리부분이 좀더 두꺼워 p-값이 좀더 크게 얻어진다. 지금의 예에서 자유도 4인 t 분포곡선에 따라 p-값을 구하면 p-값은 5%로 크게 얻어진다. 자유도가 작아질수록 차이가 커진다.