

계량분석

Hypothesis Tests

김현우, PhD¹

¹충북대학교 사회학과 조교수

September 12, 2024



진행 순서

- 1 가설검정의 논리
- 2 단일집단 모평균에 관한 가설검정
- 3 단일집단 모비율에 관한 가설검정
- 4 단일집단 모분산에 관한 가설검정

가설검정의 논리

가설검정의 논리

가설검정은 통계추리의 핵심 논리이므로 명확히 이해하자.

- 확률론 기초, 이론적 확률분포(theoretical probability distribution), 표집분포(sampling distribution), 신뢰구간(confidence interval), 유의수준(significance level) 등의 개념을 복습할 것.
- 추정(estimation)과 가설검정(hypothesis test)의 절차를 복습할 것.
- 반드시 기초통계학 교재를 꼼꼼히 읽고 공부해야 한다. 연습을 통해 자기가 무엇을 하고 있는지 상황을 정확히 파악해야 한다.
- 자신이 남들에게 이 개념들에 관해 설명할 수 있어야 한다(Why?).



가설검정의 논리

모집단이 하나 주어졌을 때 세 종류의 가설검정을 수행할 수 있다.

- ① 모평균에 대한 가설검정(one-sample test of the population mean)
- ② 모비율에 대한 가설검정(one-sample test of the population proportion)
- ③ 모분산에 대한 가설검정(one-sample test of the population variance)



가설검정의 논리

가설검정의 방향에 따라 다시 세 종류로 나뉜다.

- ❶ 좌측검정(left-sided test)
- ❷ 우측검정(right-sided test)
- ❸ 양측검정(two-sided test)



단일집단 모평균에 관한 가설검정

단일집단 모평균에 관한 가설검정

하나의 변수의 평균에 대한 가설을 세우고 이를 검정할 수 있다.

- 그 변수의 모집단이 정규분포(normal distribution)를 따르거나, 표본이 충분히 크거나, 모분산 σ^2 를 안다면 Z -검정을 사용할 수 있다(교재 190페이지).
- 이 경우, **검정통계량(test statistics)**으로 Z 값을 사용한다.

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

- Stata에서 `ztest`를 이용하여 이를 추정할 수 있다. 단 `sd(.)` 옵션을 사용해야 한다 (Why?).



단일집단 모평균에 관한 가설검정

- 표본이 작거나 모분산 σ^2 를 모르면 t -검정을 사용할 수 있다(교재 190페이지).
- 이 경우, **검정통계량(test statistics)**으로 t 값을 사용한다.

$$t = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- Stata에서 `ttest`를 이용하여 이를 추정할 수 있다. 이때는 `sd(.)` 옵션이 없다(Why?).



단일집단 모평균에 관한 가설검정

- 표본의 크기가 작아지면서 **유의확률(p -value)**은 점차 달라진다(Why?).
- 결과 해석에서 좌측검정, 우측검정, 양측검정을 혼동하지 않도록 주의하자.
- 작은 표본(year==88 & age==34)으로 엑셀에서 가설검정을 해보고 Stata로 계산해보자.

	A	B	C	D	E
1	ln_wage		mu	1.67	
2	1.744779		x_bar	1.78158	
3	1.899532		s	0.65169	
4	1.729187		n	54	
5	1.528993		Z	1.25819	
6	1.207312		p-value	0.10416	
7	1.291193				
8	1.115744				
9	2.266703				
10	2.739748				
11	1.207312				
12	1.471526				
13	1.743372				
14	2.186612				



단일집단 모비율에 관한 가설검정

단일집단 모비율에 관한 가설검정

하나의 변수의 비율에 대한 가설을 세우고 이를 검정할 수 있다.

- 그 변수의 모비율 π 에 대한 추정은 특히 사회조사에서 흔하게 이루어진다(Why?).
- $np > 5$ 와 $n(1 - p) > 5$ 라는 두 조건이 충족하면, 다음의 Z -검정을 사용할 수 있다 (교재 193페이지).

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$

- 분모는 추정된 비율의 **표준오차**(standard error)이고 **이항분포**(binomial distribution)에서 도출된다(교재 126페이지).
- Stata에서 `prtest`를 이용하여 이를 추정할 수 있다.



단일집단 모비율에 관한 가설검정

- 여론조사기관에서 발표하는 추정치는 오차범위(margin of error)와 함께 보고된다.
- 그 계산 원리를 정확하게 이해해야 하기 위해 교재 180페이지 8번 문제를 풀어보자 (연습문제).
- 모비율 가설검정의 Z 값과 (가변수의) 모평균 가설검정의 Z 값은 표본 크기가 커지면 점차 같은 값으로 수렴한다(Why?).



단일집단 모분산에 관한 가설검정

단일집단 모분산에 관한 가설검정

하나의 변수의 분산에 대한 가설을 세우고 이를 검정할 수 있다.

- 그 변수의 모분산 σ^2 에 대한 추정은 그다지 사용되지 않지만, 분산분석(analysis of variance; ANOVA)의 기초가 되므로 중요하다.
- 모집단에 대해 정규성 가정(normality assumption)이 성립하면, 다음의 χ^2 값을 검정통계량으로 사용할 수 있다.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

- 양측검정의 귀무가설(null hypothesis)과 s^2/σ^2 의 관계에 대해 생각해보자(교재 196페이지).



단일집단 모분산에 관한 가설검정

- Stata에서 `sdtest`를 이용하여 이를 추정할 수 있다.
- 이때는 분산이 아니라 표준편차를 검정하므로 실수하지 않도록 주의하자.
- χ^2 분포는 대칭적이지 않기 때문에 좌측 유의확률과 우측 유의확률이 일치하지 않는다(Why?).
- 그러나 표본이 충분히 크면 양자는 대체로 일치하게 된다(Why?).
- 이 역시 작은 표본을 가지고 엑셀에서 직접 계산해보자. Stata로 검산하여 자신이 엑셀에서 구한 값과 일치하는지 확인해보자.

