# 계량분석

## Comparing Means/Proportions

김현우, PhD[1]

[1] 충북대학교 사회학과 조교수

October 10, 2024

# 진행 순서

# 모평균에 관한 가설검정

# 모평균에 관한 가설검정

평균에 관한 가설검정은 관례상 $t$ 검정으로 수행한다.

- 표본 크기가 충분히 크면($n \geq 30$) 평범하게 표준정규분포, 즉 $Z$ 분포를 사용해서 가설검정할 수 있다.
- 표본 크기가 너무 작으면($n < 30$) $t$ 분포를 사용해서 가설검정할 수 있다.
- 표본 크기가 커졌을 때 $t$ 분포는 점차 표준정규분포(=$Z$ 분포)에 수렴한다.
- 어떤 분포를 사용할지는 모집단의 분포에 대한 사전지식, $\sigma$에 대한 사전지식, 표본크기 $n$에 따라 결정한다(교재 166).

# 모평균에 관한 가설검정

$t$ 분포를 사용하여 ($Z$ 분포와 비슷하게) 가설검정을 수행할 수 있다.

- 오늘 연습할 $t$ 검정($t$ test)에는 세 종류가 있다.
- 첫째, 살펴보려는 모집단이 하나 있을 때 그 평균에 관해 가설검정할 수 있다.
- 먼저 모집단에 관한 귀무가설을 세운다.
- (이것이 옳다는 전제 아래) 그 표본평균의 표집분포가 $t$ 분포를 따른다고 할 때, 나의 표본에 근거하여 귀무가설을 기각할 수 있는가 살펴본다.
- 이것이 (1) 단일표본 $t$ 검정(one-sample $t$ test)이다.

# 모평균에 관한 가설검정

- 둘째, 살펴보려는 모집단이 두 개 있을 때 그 평균의 차이에 관해 가설검정할 수 있다.
- 이것이 이른바 평균비교(mean comparison)이다.
- 각각에서 무한히 표본을 뽑고 두 평균의 차 $\bar{x}_1 - \bar{x}_2$를 확률변수(random variable)로 보고 표집분포를 그릴 수 있다.
- 그 표집분포는 대표본일 때 정규분포한다. 물론 소표본이라면 $t$ 분포한다.
- 이제 귀무가설을 세우고 적절하게 가설검정을 수행한다.
- 다만 자료구조에 따라 (2) 독립표본 $t$ 검정($t$ test for independent samples) 또는 (3) 쌍체표본 $t$ 검정($t$ test for paired samples)을 사용해야 한다.

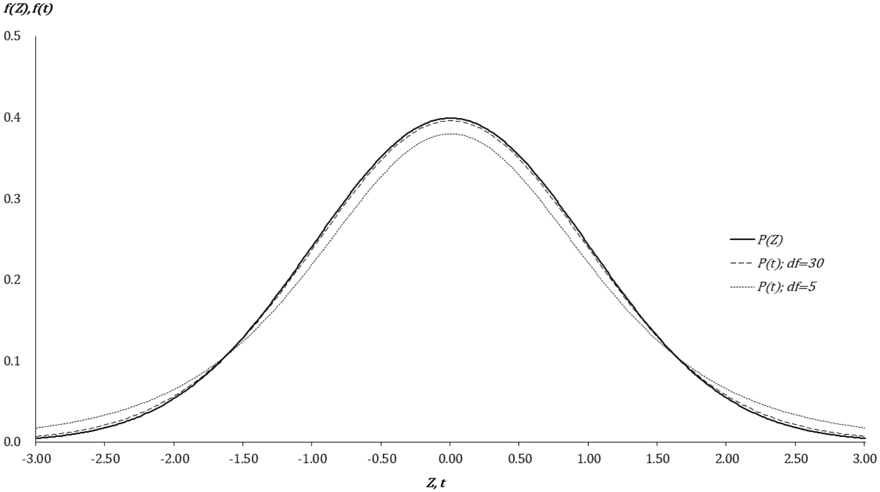
# 단일모집단의 평균에 관한 가설검정

# 단일모집단의 평균에 관한 가설검정

$t$ 분포는 표준정규분포와 사실 꽤 비슷하다.

- $t$ 분포는 자유도(degree of freedom; df)에 따라 그 형태가 정의된다.
- $t$ 검정의 맥락에서 자유도는 $N-1$로 결정된다. 자유도는 흔히 데이터에서 자유롭게 선택될 수 있는 관찰값(observations)의 수로 설명된다.
- 자유도가 커질수록 $t$ 분포는 표준정규분포(=Z 분포)와 점점 더 닮아간다.
- 자유도가 작을때 $t$ 분포는 표준정규분포에 비해 꼬리가 좀 더 두껍다.
- 이로 인해 $t$ 분포를 사용하면 같은 신뢰수준이라도 좀 더 큰 검정통계량(test statistic)을 얻어야만 가설을 기각시킬 수 있게 된다(보수적 추정 원리).

# 단일모집단의 평균에 관한 가설검정



# 단일모집단의 평균에 관한 가설검정

단일표본 $t$ 검정의 가설구조를 살펴보자.

- 다른 검정 기법의 가설구조와 마찬가지로 양측검정과 단측검정으로 나뉜다.
- 양측검정에서 귀무가설과 대립가설의 구조는 다음과 같다.

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

- 단측검정은 두 가지 형태 중 하나의 귀무가설과 대립가설의 구조를 갖는다.

$$H_0 : \mu \geq \mu_0, \quad H_a : \mu < \mu_0$$

또는

$$H_0 : \mu \leq \mu_0, \quad H_a : \mu > \mu_0$$

# 단일모집단의 평균에 관한 가설검정

단일표본 $t$ 검정은 평균비교의 기초가 된다.

- nlswork 자료에서 ln_wage 변수의 요약통계량을 살펴보자.
- “모평균은 1.67점이다($\mu_0 = 1.67$)” 라는 귀무가설을 세우기로 한다.
- 표본평균의 표집분포가 $t$ 분포를 따른다고 전제한다. $t$ 분포의 모양은 자유도(df)에 의해 결정되며, 여기서는 $28534 - 1 = 28533$이다.

# 단일모집단의 평균에 관한 가설검정

- 만약 표본에서 $\bar{X} = 1.674$이고 $s = 0.478$이라면, $t$ 값은 약 26.47이다.

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{1.674 - 1.67}{0.478/\sqrt{28534}} = 1.733$$

- 이론적 확률분포의 누적분포함수(cdf)를 직접 계산하여 $t$ 값과 유의확률($p$-value)을 구할 수 있다.
- $p < 0.001$이므로 99.9% 신뢰수준에서 통계적으로 유의하게 귀무가설을 기각할 수 있다.
- Stata에서는 ttest 명령어로 단일표본 $t$ 검정을 수행할 수 있다.

# 두 모평균에 관한 가설검정

# 두 모평균에 관한 가설검정

평균비교는 두 모집단의 평균의 차이를 가설검정의 대상으로 삼는다.

- 평균비교는 양측검정과 단측검정으로 수행될 수 있다. 다만 양측검정이 압도적으로 많이 사용된다(Why?).

$$
\begin{array}{lll}
H_0 : \mu_1 - \mu_2 = 0 & H_0 : \mu_1 - \mu_2 \geq 0 & H_0 : \mu_1 - \mu_2 \leq 0 \\
H_a : \mu_1 - \mu_2 \neq 0 & H_a : \mu_1 - \mu_2 < 0 & H_a : \mu_1 - \mu_2 > 0
\end{array}
$$

- 사실 0이 아닌 숫자로 가설검정을 할 수 있다. 하지만 0이 실무나 연구에서 많이 쓰이므로 이걸로 충분하다.

# 두 모평균에 관한 가설검정

- 단일표본 $t$ 검정은 현실적으로 그다지 사용되지 않는다고 생각하기 쉽다(Why?). 하지만 나중에 정말 중요하게 사용된다!
- 물론 당장은 그룹 간의 평균비교를 수행하는 경우가 훨씬 많다.

(1) 남성의 연 평균소득과 여성의 연 평균소득의 차이

(2) 가출경험이 있는 중학생이 가진 비행경험 친구의 수와 가출경험이 없는 중학생이 가진 비행경험 친구의 수의 차이

(3) 집중강화 트레이닝을 받기 전 2군 선수들의 이전 평균기록과 이후 평균기록의 차이

(4) 처방약을 복용한 환자의 혈압과 위약을 복용한 환자의 혈압의 차이

## 독립표본 $t$ 검정과 쌍체표본 $t$ 검정은 어떻게 다른가?

- 왼쪽은 쌍체표본(paired samples)으로 같은 사람에 대해 처방 전후(before and after)로 기록이 짝지어(paired) 있는 반면, 오른쪽은 독립표본(independent samples)은 처방(treatment) 실시 여부를 말해주는 더미변수(dummy variable)가 있다.

| ID | BEFORE | AFTER |
|---|---|---|
| 1 | 35 | 27 |
| 2 | 31 | 39 |
| 3 | 46 | 33 |
| 4 | 39 | 40 |
| 5 | 31 | 31 |

| ID | TREATED | RECORD |
|---|---|---|
| 1 | 0 | 35 |
| 2 | 0 | 31 |
| 3 | 0 | 46 |
| 4 | 0 | 39 |
| 5 | 0 | 31 |
| 1 | 1 | 27 |
| 2 | 1 | 39 |
| 3 | 1 | 33 |
| 4 | 1 | 40 |
| 5 | 1 | 31 |

# 두 모평균에 관한 가설검정

먼저 독립표본 $t$ 검정을 연습해보자.

- social_independent.dta에서 wave 별로 socialself 변수의 요약통계량을 살펴보자.
- 1년 전후로 사회적 자아 두 변수의 평균 차이($\mu_1 - \mu_2$)의 (가상적인) 표집분포가 $t$ 분포를 따른다는 전제 아래, 다음의 귀무가설을 검정해보자.
- (1) “두 변수 사이의 평균 차이는 없다.”
- (2) “두 변수 사이의 평균 차이는 0보다 작다.”
- (3) “두 변수 사이의 평균 차이는 0보다 크다.”

# 두 모평균에 관한 가설검정

- 표본평균의 차(difference)는 어떻게 계산되었는가?
- $t$ 값과 자유도(df)은 어떻게 계산되었는가?
- 대립가설 별로 유의확률($p$-value)은 어떻게 계산되었는가?

# 두 모평균에 관한 가설검정

다음으로 쌍체표본 $t$ 검정을 연습해보자.

- 이번에는 social_paired.dta에서 socialself1과 socialself2 변수의 요약통계량을 살펴보자.
- 1년 전후로 사회적 자아 두 변수의 평균 차이($\mu_1 - \mu_2$)의 (가상적인) 표집분포가 $t$ 분포를 따른다는 전제 아래, 다음의 귀무가설을 검정해보자.
- (1) "두 변수 사이의 평균 차이는 없다."
- (2) "두 변수 사이의 평균 차이는 0보다 작다."
- (3) "두 변수 사이의 평균 차이는 0보다 크다."

# 두 모평균에 관한 가설검정

- 표본평균의 차(difference)는 어떻게 계산되었는가?
- $t$ 값과 자유도(df)은 어떻게 계산되었는가?
- 대립가설 별로 유의확률($p$-value)은 어떻게 계산되었는가?
- 앞서 독립표본으로 계산했던 $t$ 값과 어떻게 다른지 비교해보자.

# 두 모평균에 관한 가설검정

$t$ 검정의 결과를 해석할 때 그 가정에 대해 살짝 주의해야 한다.

- 두 모집단의 분산이 같다는 가정을 등분산(homogeneity of variance)이라고 부른다.
- 결과표를 들여다 보면 타이틀에 Two-sample t test with equal variances라고 적혀있다. 다시 말해, "두 표본이 나온 모집단들의 분산이 같다"고 전제된 것이다.
- 만일 실험 전후로 같은 사람을 짝지워 평균을 비교하는 경우라면 이 가정은 (상대적으로) 크게 문제되지 않을 수 있다. 가령 같은 초등학생을 1년 전후로 추적하여(before vs after) 사회적 자아를 조사한다면 이 가정의 타당성이 별로 의심스럽지는 않다(Why?).
- 하지만 (쌍체표본 $t$ 검정과는 달리) 독립표본 $t$ 검정을 사용한다면 반드시 같은 사람끼리 짝지어 평균을 비교하지 않을 수도 있다!
- 다른 사람이라면 이 가정의 타당성이 특히 의문스럽고 이 가정은 아예 완화되어야 할 필요가 있다.

# 두 모평균에 관한 가설검정

다행히 이 가정없이도 평균비교를 수행할 수 있다.

- 똑같은 명령어에 unequal 옵션을 붙여 다시 시행해보자.
- 이번엔 타이틀에 Two-sample t test with unequal variances 라고 나와있는 것을 확인할 수 있다.
- 물론 결과도 조금씩 다르다.

# 두 모평균에 관한 가설검정

정반대로 쌍체표본의 형태로 데이터가 주어졌지만 같은 사람이 아닌 경우도 있다.

- 가령 남녀의 달리기 기록이 두 개의 변수로 제공되었다고 하자. 한 사람이 동시에 남자이자 여자일수는 없는데도 쌍체표본의 형태로 주어졌다.
- 이런 경우는 자료 입력자의 미숙함이 가장 큰 원인일 것 같다.
- 물론 필요에 따라 데이터를 재배열(reshape)하여 얼마든지 독립표본의 형태로 바꾸어 다시 분석할 수도 있다.

# 두 모비율에 관한 가설검정

# 두 모비율에 관한 가설검정

평균 뿐만 아니라 비율도 비교할 수 있다.

- 앞서 살펴보았듯 평균비교는 두 표본의 평균을 비교한다. 귀무가설은 “두 변수 간 평균의 차이는 없다” 또는 “두 변수 간 평균의 차이는 0보다 크다/작다” 같은 식으로 설정된다.
- 비율비교(proportion comparison)는 두 표본의 비율을 비교한다.
- 예컨대 표본에서 어떤 재해사건 전후로 환경 문제가 가장 심각한 문제(Most Important Problem; MIP)라고 인식하는 사람의 비율이 달라지는가를 확인한다고 하자.
- 이 경우 귀무가설은 “두 표본 간 비율의 차이는 없다” 또는 “두 표본 간 비율의 차이는 0보다 크다”, “0보다 작다” 하는 식으로 설정된다.

# 두 모비율에 관한 가설검정

Stata에서는 prtest 명령어로 비율비교를 수행할 수 있다.

- prtest로 단일표본과 쌍체표본, 독립표본을 모두 분석할 수 있다.
- 명령어 사용법은 평균비교의 ttest와 대체로 같다.

# 두 모비율에 관한 가설검정

그건 그렇고 아까 표본평균을 했는데 왜 표본비율은 또 할까?

- 표본평균은 숫자형(numerical) 척도로 측정된 변수에 대해서만 의미를 갖는다. 예컨대 숫자형 척도인 키, 몸무게의 평균은 의미를 갖지만, 범주형(categorical) 척도인 인종(1=백인; 2=흑인; ...)이나 종교(0=없음; 1=기독교; 2=불교; ...)의 평균에는 의미가 없다.
- 반면 범주형(categorical) 척도는 비율이 의미를 갖는다. 예컨대 여성의 비율, 백인의 비율, 기독교의 비율 하는 식으로.
- 수학적으로 볼 때, 평균비교는 $t$ 분포(t distribution)에 직결되어 있으나, 표본비율은 이항분포(binomial distribution)에 직결되어 있다

# 두 모비율에 관한 가설검정

- 범주형 척도인 인종이나 종교의 평균에는 의미가 없지만, 일단 가변수로 만들면 비율의 의미가 생겨난다.
- 그러므로 비율비교 대신 평균비교를 수행할 수도 있다. 대표본에서 prtest와 ttest의 결과표를 비교해보자.
- 결과는 사뭇 유사한데, 이는 대표본일 때 이항분포의 정규근사(normal approximation to the binomial)가 성립하기 때문이다.

# 두 모비율에 관한 가설검정

- 양적방법론, 여론조사, 보건의료 분야에 관심이 있다면 (이항분포와 더불어) 모비율에 관한 가설검정과 신뢰구간도 수학적으로도 반드시 더 공부해야 한다(교재 192–195, 212–215).
- 특히 표준오차(standard error)가 다르게 계산된다는 것에 주의하자.

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