# 계량분석
## Comparing Means/Proportions

김현우, PhD[1]

[1] 충북대학교 사회학과 조교수

October 18, 2021

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 진행 순서

1. 단일모집단의 평균에 관한 가설검정
2. 두 모집단 평균에 관한 가설검정
3. 두 모집단 비율에 관한 가설검정
4. t 검정의 실제 활용

# 단일모집단의 평균에 관한 가설검정

# 단일모집단의 평균에 관한 가설검정

평균비교(mean comparison)는 관례상 t 검정(t test)라고 불리우는 기법을 통해 수행된다.

- t 검정은 t 분포에 기반하여 일반적으로 "두 표본(또는 두 변수)의 차는 x인가? 혹은 x 보다 작은가? 혹은 x보다 큰가?"에 관한 가설을 확인하는 통계적 방법이다.
- 약 120년 전에 개발된 t 분포는 본래 표준정규분포(standard normal distribution) 가 부적합할만큼 사례가 적고 모집단의 분산에 대해 모르는 경우에 대응하도록 설계된 것이다.
- 오늘날 대규모 데이터가 워낙 흔해지면서 더이상 t 분포에 집착할 필요가 없고 표준정규분포(=Z 분포)를 그냥 사용해도 되는데 (특히 평균비교 맥락에서는) 여전히 t 분포를 사용하는 관행이 지속되고 있다.
- 앞으로 배울 t 분포나 F 분포 그리고 전에 배운 $\chi^2$ 분포 모두 Z 분포와 가설검정의 논리를 충실히 배워야 쉽게 이해할 수 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

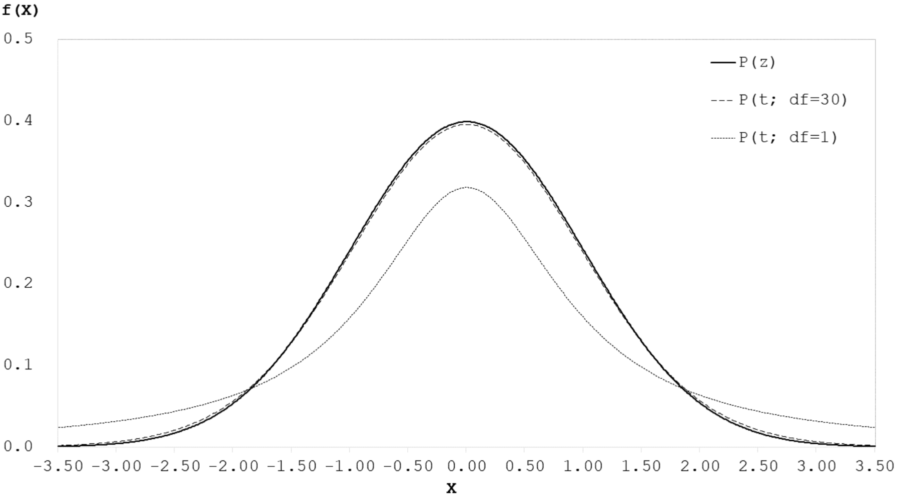
# 단일모집단의 평균에 관한 가설검정

t 분포는 표준정규분포(=Z 분포)와 사실 꽤 비슷하다.

- t 분포(t-distribution)는 자유도(degree of freedom; df)에 따라 그 형태가 정의된다.
- t 검정의 맥락에서 자유도는 N-1로 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 신뢰구간(예컨대 90%, 95%, 99%)을 설정할 때 좀 더 큰 통계량(statistic)을 얻어야만 가설을 기각시킬 수 있게 된다(보수적 추정).

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 단일모집단의 평균에 관한 가설검정



# 단일모집단의 평균에 관한 가설검정

보다 구체적으로 세 종류의 t 검정이 있다고 흔히 이야기된다.

- 단일표본 t 검정(one-sample t test)
- 독립표본 t 검정(t test for independent samples)
- 쌍체표본 t 검정(t test for paired samples)

어떤 교과서는 단일표본 t 검정을 다루지 않는다.

- 첫번째 것은 단일표본(one sample)을 분석하고 나머지 두 개는 2표본(two samples)을 분석한다.
- 첫번째 것은 t 검정의 원리를 따르지만 사실 평균비교(mean comparison)를 위한 기법은 아니다. 다만 기초로서 매우 중요한 부분으로 이것을 이해하지 못하면 나머지 두 개를 전혀 이해할 수 없다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 단일모집단의 평균에 관한 가설검정

가장 먼저 단일표본(one sample)과 2표본(two samples)의 차이를 살펴보자.

- 먼저, 표본이 “하나” 주어졌다고 하자. 이것의 평균의 (가상적인) 표집분포가 t 분포를 따른다고 전제할 때, “$\mu = \mathrm{x}$”라는 영가설을 이 표본이 지지하는가를 살펴볼 수 있다. 이것이 단일표본 t 검정(one-sample t test)이다.
- 다음으로 표본이 “두 개” 주어졌다고 하자. 이 두 표본의 평균들의 차($\mu_1 - \mu_2$)를 확률변수로 보고 그 (가상적인) 표집분포가 t 분포를 따른다고 가정할 때, “$\mu_1 - \mu_2 = \mathrm{x}$” 라는 영가설을 이 두 표본이 지지하는가를 살펴볼 수 있다. 데이터의 구조에 따라 이것은 독립표본 t 검정(t test for independent samples) 또는 쌍체표본 t 검정(t test for paired samples)이 된다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 단일모집단의 평균에 관한 가설검정

단일표본 t-검정의 가설구조를 살펴보자.

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

$$\mathrm{H_0} : \mu = \mathrm{x}, \quad \mathrm{H_a} : \mu \neq \mathrm{x}$$

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

$$\mathrm{H_0} : \mu \geq \mathrm{x}, \quad \mathrm{H_a} : \mu < \mathrm{x}$$

또는

$$\mathrm{H_0} : \mu \leq \mathrm{x}, \quad \mathrm{H_a} : \mu > \mathrm{x}$$

# 단일모집단의 평균에 관한 가설검정

단일표본 t 검정은 평균비교는 아니지만 평균비교의 기초가 된다.

- 1,000명의 학생들에 대해 일괄적으로 시험을 실시하였다. 그리고 30명의 학생을 임의 표본(random sample)으로 추출하였다.
- (가상적인) 표집분포가 t 분포를 따른다는 전제 아래 "그 평균은 60점이다($\mu = 60$)" 라는 영가설을 세웠다. t 분포의 모양은 단순히 자유도(df)에 의해 결정된다. 여기서는 30-1, 즉 df=29다.
- 표본에서 관찰된 평균이 70점, 표준편차가 20점이라고 하자. 이 경우 t 값은 약 2.739이다($= (70 - 60)/(20/\sqrt{29})$). 5% 유의수준(=95% 신뢰수준)을 기준일 때, Stata에서는 **t(29, -2.739) + (1-t(29, 2.739))**를 계산하여 유의확률(p-value)을 구할 수 있다(Why?). 답은 약 0.0104이다.
- p-value가 0.5보다 작으므로 영가설을 5% 유의수준에서 기각한다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 단일모집단의 평균에 관한 가설검정

Stata를 켜고 **ttest** 명령어를 사용해 단일표본 t 검정을 연습해보자(**[Stata 코드]** 참고).

- 먼저 social.dta 파일을 열고 socialself 변수의 요약통계량을 살펴보자.
- (가상적인) 표집분포가 t 분포를 따른다는 전제 아래 "사회적 자아 점수의 평균은 29.9 이다" 라는 영가설을 세웠다. 이를 테스트해보자.
- "사회적 자아 점수의 평균은 29.9보다 작거나 같다" 라는 영가설을 세웠다. 이를 테스트해보자.
- "사회적 자아 점수의 평균은 29.9보다 크거나 같다" 라는 영가설을 세웠다. 이를 테스트해보자.

결과표를 꼼꼼히 들여다보자.

- 평균(mean)과 표준오차(standard error)는 어떻게 계산되었는가?
- t 값과 자유도(degree of freedom)은 어떻게 계산되었는가?
- 대립가설 별로 유의확률(p-value)은 어떻게 계산되었는가?

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 두 모집단 평균에 관한 가설검정

t 검정은 두 표본의 평균의 차를 가설검정의 대상으로 삼는다.

- 앞서 단일표본 t 검정에서는 “하나의 모집단의 평균($\mu$)이 어떤 특정 값(x)인가”를 테스트하였다.
- 하지만 t 검정이라고 할 때는 보통 두 표본이 주어져 있고 “두 표본의 평균의 차 ($\mu_1 - \mu_2$)가 통계적으로 유의하게 0과 다른가/큰가/작은가”를 분석한다. 이에 해당하는 가설구조는 각각 다음과 같다:

$$\mathrm{H_0} : \mu_1 - \mu_2 = 0, \quad \mathrm{H_a} : \mu_1 - \mu_2 \neq 0$$

$$\mathrm{H_0} : \mu_1 - \mu_2 \geq 0, \quad \mathrm{H_a} : \mu_1 - \mu_2 < 0$$

$$\mathrm{H_0} : \mu_1 - \mu_2 \leq 0, \quad \mathrm{H_a} : \mu_1 - \mu_2 > 0$$

# 두 모집단 평균에 관한 가설검정

그러면 독립표본 t 검정과 쌍체표본 t 검정은 어떻게 다른가?

- 이것은 데이터가 어떻게 생겼는가로 쉽게 이해할 수 있다. 왼쪽(independent samples)은 처방(treatment) 실시 여부를 말해주는 더미변수(dummy variable)가 있는 반면, 오른쪽(paired samples)은 같은 사람에 대해 처방 전후(before and after)로 기록이 짝지어(paired) 있다.

| PID | 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 |

| PID | REC_BEFORE | REC_AFTER |
|---|---|---|
| 1 | 35 | 27 |
| 2 | 31 | 39 |
| 3 | 46 | 33 |
| 4 | 39 | 40 |
| 5 | 31 | 31 |

충북대학교 CHUNGBUK NATIONAL UNIVERSITY

# 두 모집단 평균에 관한 가설검정

먼저 독립표본 t 검정을 연습해보자(**[Stata 코드]** 참고).

- 다시 social_independent.dta 파일을 열고 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 "두 변수 사이의 평균 차이는 없다" 라는 영가설을 세웠다. 이를 테스트해보자.
- "두 변수 사이의 평균 차이는 0보다 작다" 라는 영가설을 세웠다. 이를 테스트해보자.
- "두 변수 사이의 평균 차이는 0보다 크다" 라는 영가설을 세웠다. 이를 테스트해보자.

결과표를 꼼꼼히 들여다보자.

- 평균(mean)의 차(diff)는 어떻게 계산되었는가?
- t 값과 자유도(degree of freedom)은 어떻게 계산되었는가?
- 대립가설 별로 유의확률(p-value)은 어떻게 계산되었는가?

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

다음으로 쌍체표본 t 검정을 연습해보자(**[Stata 코드]** 참고).

- 다시 social_paired.dta 파일을 열고 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 "두 변수 사이의 평균 차이는 없다" 라는 영가설을 세웠다. 이를 테스트해보자.
- "두 변수 사이의 평균 차이는 0보다 작다" 라는 영가설을 세웠다. 이를 테스트해보자.
- "두 변수 사이의 평균 차이는 0보다 크다" 라는 영가설을 세웠다. 이를 테스트해보자.

결과표를 꼼꼼히 들여다보자.

- 평균(mean)의 차(diff)는 어떻게 계산되었는가?
- t 값과 자유도(degree of freedom)은 어떻게 계산되었는가?
- 대립가설 별로 유의확률(p-value)은 어떻게 계산되었는가?

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 두 모집단 평균에 관한 가설검정

t 검정은 등분산(homogeneity of variance)이 가정되고 있다(**[Stata 코드]** 참고).

- 결과표를 들여다 보면 타이틀에 Two-sample t test with equal variances라고 적혀있다. 다시 말해, "두 표본이 나온 모집단들의 분산이 같다"고 전제된 것이다.
- 사실 이 가정은 오늘 조금 있다가 배울 일원분산분석(one-way ANOVA)에서도 똑같이 적용된다.
- 만일 실험 전후로 같은 사람을 짝지워 평균을 비교하는 경우라면 이 가정은 (어디까지나 상대적으로) 크게 문제가 되지 않을수 있다. 우리가 사용한 데이터는 같은 초등학생을 1년 전후로 추적하여(before vs after) 사회적 자아를 조사한 것이므로 문제가 적을 것 같기는 하다.

# 두 모집단 평균에 관한 가설검정

다행히 이 가정은 완화될 수 있다(**[Stata 코드]** 참고).

- 독립표본 t 검정은 반드시 같은 사람을 짝지워 평균을 비교하지 않을 수도 있다!
- 예컨대 같은 데이터를 가지고 이번엔 성별에 따라 평균비교를 수행해보자. 보통 사람의 성별이 1년 전후로 바뀌지 않으므로 이것은 명백히 같은 사람을 짝지은 것이 아니다. 그래도 독립표본 t 검정에 의해 평균비교는 할 수 있다. 쌍체표본 t 검정과는 이 점에서 크게 다르다.
- 어쨋든 이렇게 사람까지 달라지면 등분산 가정의 타당성은 더욱 의문스럽다.
- 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.
- 당연히 이런 옵션은 단일표본(one-sample)의 경우나 쌍체표본(paired tests)에서서는 아무런 의미도 없다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 두 모집단 평균에 관한 가설검정

정반대로 쌍체표본의 형태로 데이터가 주어졌지만 같은 사람이 아닌 경우도 있다([Stata **코드**] 참고).

- 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.
- 이런 경우도 대응하기 쉽다. **ttest** 명령어의 옵션으로 **unpair**를 붙이면 된다. 그 결과를 독립표본 t 검정의 결과와 비교해 보면 완전히 똑같다.
- 물론 필요에 따라 데이터를 리코딩하여 얼마든지 독립표본의 형태로 바꾸어 다시 분석할 수도 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 두 모집단 비율에 관한 가설검정

# 두 모집단 비율에 관한 가설검정

아까는 평균비교(mean comparison)를 살펴보았고 이번엔 비율비교 (proportion comparison)을 살펴보자.

- 앞서 살펴보았듯 평균비교는 두 표본의 평균을 비교한다. 그 가설은 “두 변수 간 평균의 차이는 없다” 또는 “두 변수 간 평균의 차이는 0보다 크다/작다” 하는 식으로 설정된다.
- 비율비교는 두 표본의 비율을 비교한다. 예컨대 샘플에서 어떤 지역재해사건 전후로 환경 문제가 가장 심각한 문제(Most Important Problem; MIP)라고 인식하는 사람의 비율이 달라지는가를 확인한다고 하자. 이 경우 영가설은 “두 표본 간 비율의 차이는 없다” 또는 “두 표본 간 비율의 차이는 0보다 크다”, “0보다 작다” 하는 식으로 설정된다.

충북대학교 CHUNGBUK NATIONAL UNIVERSITY

# 두 모집단 비율에 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

Stata에서는 **prtest** 명령어로 비율비교를 수행할 수 있다(**[Stata 코드]** 참고).

- **prtest**로도 단일표본과 쌍체표본, 독립표본을 모두 분석할 수 있다는 점에서 평균비교의 **ttest**와 동일하다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# 두 모집단 비율에 관한 가설검정

(보건의료 통계학과는 별개로) 사회과학 통계학에서는 그다지 쓰임새가 없는 편이긴 하다.

- 이는 대규모 표본을 주로 분석하는 사회과학 분야에서 이항분포의 정규근사(normal approximation to the binomial)가 자연스럽게 활용될 수 있기 때문이다.
- 앞서 범주형(categorical) 척도인 인종(1=백인; 2=흑인; ...)이나 종교(0=없음; 1=기독교; 2=불교; ...)의 평균에는 의미가 없다고 말했다. 하지만 이런 범주형 척도도 일단 더미변수로 만들면(dummy coding) 비율의 의미가 생겨난다. 예컨대 "더미변수로서 백인"의 평균은 곧 백인의 비율이다(더미 변수 논의를 참고할 것).
- N=500 정도 되는 표본에서 **prtest**와 **ttest**의 결과표를 비교해보자.
- 방법론이나 보건의료 분야에 관심이 있다면 이항분포와 더불어 수학적으로도 반드시 더 공부해야 한다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# t 검정의 실제 활용

# t 검정의 실제 활용

경험적 연구논문에서 t 검정은 크게 두 부분에서 주로 활용된다.

- 첫번째는 표본에 관한 기술통계(descriptive statistics)를 제시하는 부분이고, 두번째는 회귀분석(regression analysis)에서 계수(coefficients)의 유의성 검정 (significance test) 부분이다.
- 먼저 연구자는 자신의 표본 안의 “핵심이 되는 이분형(dichotomous) 관심변수 내지 종속변수에 따라” 다른 여러 변수들이 어떻게 달라지는지 t 검정을 통해 살펴볼 수 있다. 그 차이는 기술통계(descriptive statistics)의 일부로 보고할 수 있다. 이때 이분형 변수는 예/아니오 척도로 측정된 것이며 예컨대 성별, 대졸 여부 등을 생각해 볼 수 있다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# t 검정의 실제 활용

첫번째 맥락으로 활용된 한 논문의 기술통계 파트에서 t 검정이 실제로 어떻게 활용되는지 살펴보자

- 김경희 외 (2007)의 ⟨표4⟩를 꼼꼼히 살펴보자. 이 표는 가출경험(Runaway experience) 여부(None vs Have)에 따라 7개의 관심 변수가 어떻게 다른가를 보여준다.
- 학교생활 만족감(Content with school life) 점수는 가출경험 별로 어떤 차이가 있나? 그 차이는 90% 수준에서 통계적으로 유의(statistically significant)한가?
- 친구에 대한 애착도(Affection of friend)는 가출경험 별로 어떤 차이가 있나? 그 차이는 95% 수준에서 통계적으로 유의한가?
- 비행경험 친구의 수(Number of delinquency friend)는 가출경험 별로 어떤 차이가 있나? 그 차이는 99% 수준에서 통계적으로 유의한가?

김경희 · 김희영 · 김수강. 2007. “중학생 가출경험에 영향을 미치는 예측요인.” 지역사회간호학회지 18(4): 662–672.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

# t 검정의 실제 활용

t 검정은 두번째 맥락인 회귀분석 계수의 유의성 검정(significance test)에서도 사용된다.

- 그런데 앞서 설명하였듯 대규모 데이터라면 t 검정이 큰 의미를 갖지 않고 평범하게 표준정규분포를 사용해도 된다고 하였다.
- 전통적으로 회귀분석 계수의 유의성 검정에서 대표본이더라도 t 검정을 사용해왔다.
- 이 맥락에서 영가설은 "k번째 회귀계수가 0이다($H_0 : b_k = 0$)"로, 다시 말해 해당 독립변수는 종속변수를 설명하는데 의미가 없다는 뜻이다. 물론 연구자는 이 영가설을 기각하고 싶기 마련이다($H_a : b_k \neq 0$).
- 이에 관해 더 자세한 내용은 몇 주 뒤에 다루게 된다.

충북대학교
CHUNGBUK NATIONAL UNIVERSITY

기술통계와 밀접하게 잇닿아있는 주제로 t 검정은 또다른 중요한 용도를 갖고 있다.

- 이른바 성향점수(propensity scores)를 활용한 인과관계 분석에서는 통제집단(control group)과 처치집단(treatment group) 사이에 사전적으로 밸런스(balance)가 맞는지, 그리고 매칭(matching), 가중치 부여(weighting), 또는 계층화(stratification) 이후에 "관찰가능한 변수들에 대한 조건부(conditional on observables)"로 밸런스가 맞추어졌는지를 확인하기 위해서 쓰인다.
- 다만 이것은 이 수업의 수준을 한참 벗어난 것이므로 더이상 다루지 않는다.