KNU 강원대학교
KANGWON NATIONAL UNIVERSITY

6. 가설 검정 I

정치와 데이터분석

박상훈 (sh.park.poli@gmail.com)
강원대학교

오늘의 목표

10:05-10:45

가설 검정의 논리에 대한 개요를 학습 I.

가설검정의 절차 및 오류와 검정력

10:55-11:40

가설 검정의 논리에 대한 개요를 학습 II.

사례 연구와 실습

11:55-12:40

실습과제 해설 및 질의응답

중간시험 설명

Part I. 통계적 추론과 가설검정

가설 검정 I.

통계적 추론이란?

모집단에 대한 **참 값(모수)**을 알고 싶지만 전체 조사 어려움 $\rightarrow$ **표본** 데이터로 추론

- 표본 통계치(평균, 비율 등)를 이용해 모수에 대해 **추정**하거나 **가설을 검정**함.
- 표본에는 **무작위 오차**가 존재 $\rightarrow$ 추론에는 항상 불확실성이 따름.
- 통계적 추론은 이러한 **불확실성을 계량화**하여 결론을 내리는 과정

가설 검정 I.

가설 검정: 통계적 추론의 한 도구

가설 검정 (hypothesis testing): 모집단에 대한 주장(가설)이 **데이터로 뒷받침되는지** 평가하는 절차

- "새 정책이 투표율을 높인다"는 주장 → 데이터를 통해 사실인지 검정
- 가설 검정은 **표본 오차를 고려**하여 관찰된 효과가 **우연인지 아닌지** 판단하도록 해줌.
- 통계적 추론의 또 다른 도구로 **신뢰구간**이 있음(효과의 범위 추정).

가설 검정 I.

가설 검정: 통계적 추론의 한 도구

여론조사 결과 해석

- 한 선거 여론조사에서 후보 A 지지율 52%, B 지지율 48%로 **A가 앞섬**.
- **의문:** 이 4%포인트 차이가 **우연한 표본 오차**일까, 아니면 **실제로 A가 더 인기**일까?
- 표본 크기에 따라 다르지만, 4% 차이는 미미해서 **통계적으로 유의한 차이인지 불확실**
- 이러한 의문에 답하려면 **통계적 가설검정**이 필요

가설 검정 I.

귀무가설 vs. 대립가설

귀무가설 (H_0): "효과/차이 없다"는 기본 가설 (예: 후보 A와 B의 실제 지지율이 같다)

- 영가설이라고도 하며, 효과/차이가 없다는 무위효과(null effect)

대립가설 (H_A): 연구자가 보여주고 싶은 주장 (예: 후보 A의 지지율이 더 높다)

- 연구가설, 대안가설이라고도 함.

가설검정은 우선 **귀무가설이 사실이라는 전제**로 출발

표본 데이터가 귀무가설과 크게 어긋나면, 귀무가설을 **기각**하고 대립가설을 주장해볼 수 있는 경험적인 근거를 확보한다고 볼 수 있음.

가설 검정 I.

귀무가설 vs. 대립가설

Fisher의 유의성 검정

가설 검정 I.

귀무가설 vs. 대립가설

Fisher의 유의성 검정

R.A. Fisher: **대립가설 설정을 하지 않고** 귀무가설만 고려

p-값 개념을 도입

- 귀무가설 하에서 관찰된 데이터가 극단적일 **확률**로, **증거의 척도**로 사용
- **임의로** 정한 유의수준 0.05로 기계적으로 기각/채택 **반대**
- 통계적 검정은 최종 결정이 아니라 **귀납적 추론 과정**의 한 단계로 보았음.
- 데이터로부터 서서히 귀무가설을 **반증**해가는 과정

가설 검정 I.

귀무가설 vs. 대립가설

Neyman-Pearson의 가설검정

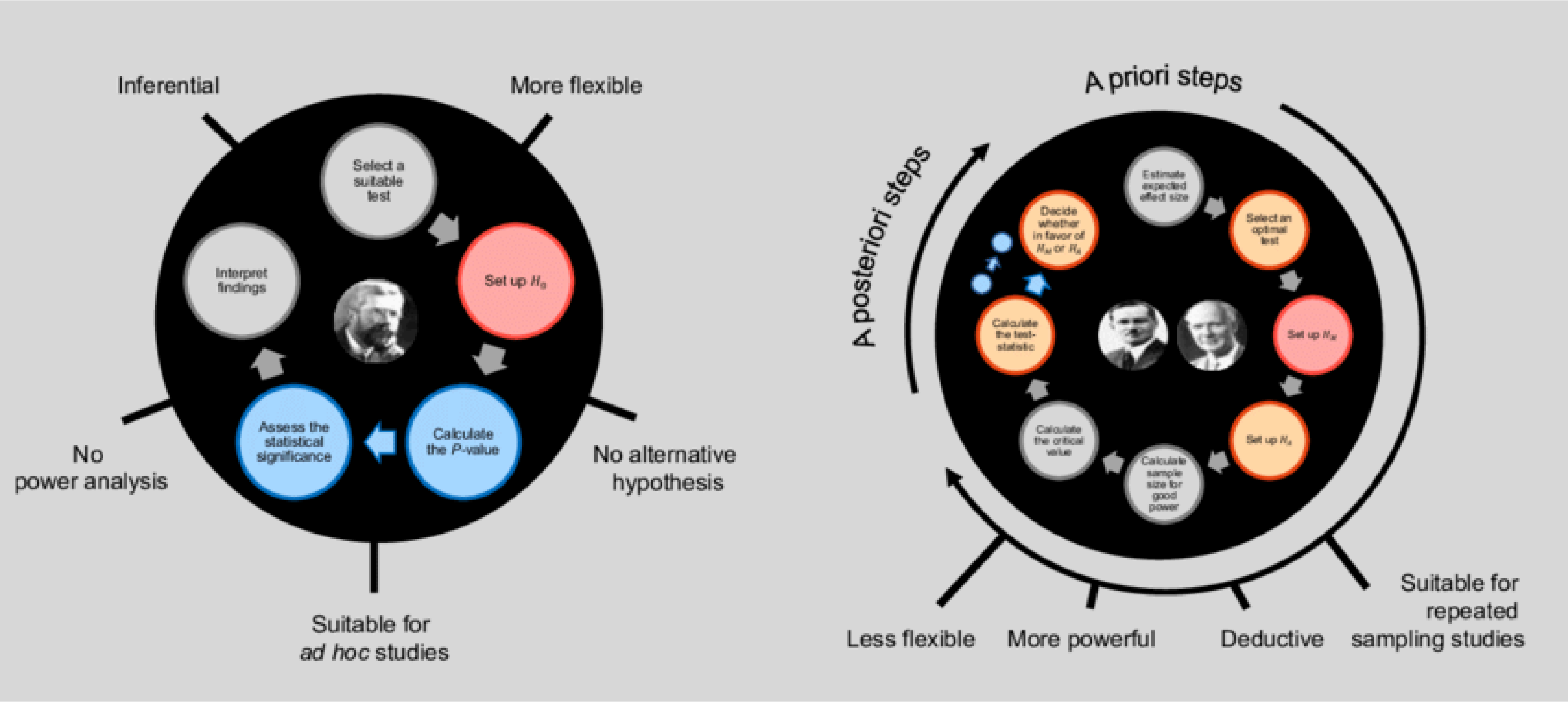
Jerzy Neyman & E.S. Pearson: **두 가지 가설** (H_0, H_A) 을 명시적으로 설정

미리 정한 **유의수준** α에 따라 귀무가설 기각 여부를 판단 (의사결정 규칙)

- **1종 오류** (α), **2종 오류** (β) 개념 도입: 검정의 오류율을 통제하면서 판단
- 과학적 검정을 **최종적인 의사결정**으로 간주하여, 특히 품질관리 등에서 활용

가설 검정 I.

Fisher의 유의성 검정 vs. Neyman-Pearson의 가설검정



가설 검정 I.

Fisher의 유의성 검정 vs. Neyman-Pearson의 가설검정

두 접근은 **철학과 절차**에서 차이가 있었으나, 현대 통계 실무에서는 **절충**되어 사용됨.

- Fisher: $p-$값 자체를 **증거의 연속적 척도**로 해석
- Neyman-Pearson: **유의수준** α를 기준으로 **채택/기각 결정**

오늘날의 가설검정 = $p-$값으로 증거를 평가하되, $\alpha = 0.05$ 등 기준에 따라 **"유의미하다/아니다"** 판단하는 형태

- Fisher는 연구자의 배경지식 없이 기계적 검정에만 의존하는 것을 경계하기도 했음.

가설 검정 I.

가설 검정의 단계

1. **가설 설정:** 귀무가설 (H_0) 과 대립가설 (H_A) 수립
2. **검정통계량 선택:** 가설을 평가할 적절한 통계량 계산 방법 결정
3. **표집분포 고려:** 귀무가설이 참일 때 통계량 분포(분포 모양, 표준오차 등) 파악
4. $p-$**값 계산:** 관찰된 통계량이 귀무가설 하에서 얼마나 극단적인지 확률 산출
5. **결론 내리기:** 미리 정한 유의수준과 비교하여 귀무가설 기각 또는 기각 실패 결정, 결과 해석

가설 검정 I.

가설 검정의 단계

전제 조건: 표집 방법 및 데이터 조건

무작위표본/무작위할당

- 표본이 모집단을 대표하고 실험은 무작위 설계여야 함.
- 추론의 타당성 전제

독립성: 각 관측은 서로 독립이라고 가정(한 데이터가 다른 데이터에 영향 없음)

충분한 표본 크기: 표본이 너무 작으면 검정력이 낮아지고, 통계적 가정도 어긋날 수 있음.

이러한 조건이 충족되어야 통계적 검정 결과를 **신뢰**할 수 있음.

가설 검정 I.

가설 검정의 단계

전제 조건: 분포 가정

모평균 검정

- 모집단 분포가 **정규분포**를 따른다고 가정
- 혹은 표본 크기가 충분히 커서 중심극한정리에 의해 표본평균 분포가 정규 근사됨.

모비율 검정

- 표본 수가 충분히 커서 $(np, n(1-p) \geq 5$ 정도) **이항분포** $\approx$ **정규분포** 근사 가능해야 함.
- **카이제곱 검정**: 기대 빈도가 너무 작으면 (예: < 5) 정확 검정 필요, 충분한 데이터 수 필요

위 가정들을 **검정 전에 점검**해야 결과 해석이 타당함.

가설 검정 I.

가설 검정의 단계

1단계: 가설 설정 (H_0 vs. H_A)

귀무가설: "영향 없음/차이 없음" 등 **중립적 주장** (통상 모수 = 특정 값)

대립가설: 연구자의 **주장/기대** (효과 존재 또는 차이 있음)

- "정책 효과 = 0" (H_0) vs "효과 $\neq$ 0" (H_A)

가설 검정 I.

가설 검정의 단계

1단계: 가설 설정 (H_0 vs. H_A)

양측검정 가설 예시

- **연구질문**: 민주주의 국가와 권위주의 국가의 **평균 GDP**에 차이가 있는가?
- **귀무가설**: 두 집단 모평균이 **동일**하다($\mu_{민주주의} = \mu_{권위주의}$)
- **대립가설**: 두 모평균이 **다르다**($\mu_{민주주의} \neq \mu_{권위주의}$)
- 양측검정: 방향을 특정하지 않았으므로, **차이가 있으면** 어느 쪽이든 검정에 포착됨(큰 차이 여부만 관심).

가설 검정 I.

가설 검정의 단계

1단계: 가설 설정 (H_0 vs. H_A)

단측검정 가설 예시

- **연구질문**: 투표 독려 캠페인이 시행된 집단의 **투표율이 더 높아졌는가?**
- **귀무가설 H_0**: 캠페인 효과 **없음** (캠페인 집단 투표율 = 통제집단 투표율)
- **대립가설 H_a**: 캠페인 집단 투표율이 **더 높다** (캠페인 > 통제)
- 단측검정: 이 경우는 한쪽(우측) 검정으로 연구자가 증가 방향을 기대하므로, **한쪽만** 검정하여 효과 유무를 판단

가설 검정 I.

가설 검정의 단계

1단계: 가설 설정 (H_0 vs. H_A)

단측검정 vs 양측검정 선택

단측검정은 효과의 **방향을 미리 확신**하거나, 반대 방향 결과에는 관심이 없을 때 사용

- 장점: 예상한 방향에서 **검출력(power)**이 높아짐(같은 효과에도 p값 더 작게 나옴).
- 단점: 효과가 반대 방향으로 나타나면 알아채지 못하고, 잘못된 방향에 대해서는 검정 불가

양측검정은 효과 방향에 확신이 없거나, **어느 쪽이든 차이**가 있어도 중요할 때 사용(보통 기본 선택)

- 연구윤리상 애매하면 **양측검정**하는 것이 일반적(단측은 명확한 이론 근거가 있을 때만)

가설 검정 I.

가설 검정의 단계

퀴즈

*"국민의 50%가 정부를 지지한다"*는 주장을 검증하려 한다. 귀무가설과 대립가설을 어떻게 세우면 될까?

- **힌트**: 모집단 지지율 p 를 사용. 귀무가설은 $p = 0.5$, 대립가설은 상황에 따라 $p \neq 0.5$ (양측) 또는 방향 설정.

가설 검정 I.

가설 검정의 단계

2단계: 검정통계량 선택

검정통계량(test statistic): 표본 데이터로부터 계산한 **요약치**로, 귀무가설 하에서 그 분포를 아는 통계량

- 보통 **"관찰된 효과 - 귀무가설 값"**을 **표준오차로 나눈 값**을 사용 → 표준화된 차이
- 검정통계량은 **크게 벗어날수록** 귀무가설과 데이터의 차이가 크다는 의미
- 가설검정에서는 적절한 검정통계량을 선택하고 계산하여, 이를 기준 분포와 비교함.

가설 검정 I.

가설 검정의 단계

2단계: 검정통계량 선택

검정통계량의 일반 형태

일반적으로:

$$t = \frac{(\text{표본 추정치}) - (\text{귀무가설 값})}{\text{표준오차(SE)}}$$

- 추정치와 귀무값의 차이를 **표준 오차 단위**로 나타낸 값
- 관찰된 표본평균이 귀무가설 평균보다 2배의 표준오차만큼 크다면, $t \approx 2$
- $t-$값이 **0에 가까우면** 귀무가설과 잘 맞고, **크게 벌어질수록** 귀무가설과 거리가 멀다는 얘기

가설 검정 I.

가설 검정의 단계

2단계: 검정통계량 선택

검정통계량의 일반 형태: 평균에 대한 $t-$통계량

모평균에 대한 검정 (모분산 모름)

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

- 여기서 $\bar{y}$는 표본평균, s는 표본 표준편차
- 표본평균이 귀무가설 하 모평균 μ_0에서 **몇 표준오차만큼 벗어났는지** 나타냄.
- 귀무가설이 참이면, 이 통계량 t는 자유도 $n-1$인 **t 분포**를 따름(표본 크기 충분하면 근사적으로 정규분포).

가설 검정 I.

가설 검정의 단계

2단계: 검정통계량 선택

검정통계량의 일반 형태: 비율에 대한 z 통계량

모비율에 대한 검정

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

- 여기서 $\hat{p}$는 표본비율, p_0는 귀무가설의 모비율 값
- 표본비율과 귀무가설 비율의 차이를 표준오차로 나눈 값
- 본 크기가 충분히 크다면, 귀무가설 하에서 z 통계량은 **표준 정규분포** $(N(0, 1))$을 따름.

가설 검정 I.

가설 검정의 단계

2단계: 검정통계량 선택

두 집단 평균 차이 통계량

두 독립표본 평균 차이 검정 (등분산 가정 없이):

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

- 두 집단 간 표본평균 차이를 그 **표준오차**로 나눈 값
- 귀무가설(H_0: 두 모평균 같음) 하에서 해당 $t-$통계량은 근사적으로 특정 자유도의 t 분포를 따름(보통 Welch의 $t-$검정 사용).
- 두 표본의 분산, 크기가 다를 때도 적용 가능

가설 검정 I.

가설 검정의 단계

2단계: 검정통계량 선택

범주형에 대한 카이제곱 통계량

범주형 자료의 적합도 검정이나 독립성 검정에 사용되는 통계량:

$$\chi^2 = \sum \frac{(\text{관측도수} - \text{기대도수})^2}{\text{기대도수}}$$

- 관측된 빈도분포가 귀무가설(예: 분포가 같다, 독립이다)과 **얼마나 차이** 나는지 합산 측정
- 귀무가설이 참이면, χ^2 통계량은 자유도에 따른 **카이제곱 분포**를 따름.
- 값이 클수록 관측이 귀무 가정에서 벗어난 정도가 크다는 뜻

가설 검정 I.

가설 검정의 단계

3단계: 귀무가설 하의 표집분포

검정통계량의 **표집분포**란, **귀무가설이 참일 때** 그 통계량이 취할 수 있는 값들의 분포

- 예를 들어, 귀무가설이 "동전은 공평하다"이면, 100번 던졌을 때 앞면 개수의 분포(평균 50, 분산 25) 등을 계산 가능
- 이 분포를 알아야 관측된 통계량이 **평범한지 드문지** 판단할 수 있음.
- 각 검정통계량마다 귀무가설 하 분포가 정해져 있음 (z, t, χ^2 등)

가설 검정 I.

가설 검정의 단계

3단계: 귀무가설 하의 표집분포

동전 던지기: H_0: 공정동전($p = 0.5$)

- 100번 중 앞면 개수 $X \sim$ **이항분포**$(n = 100, p = 0.5)$
- 평균 50, 표준편차 5; 정규근사하면 $N(50, 5^2)$ 정도
  - 만약 실제 실험에서 앞면이 60번 나왔다면, 이 분포에서 **평균보다 +10(2 σ)**
    $\rightarrow$ 다소 드문 사건
- **평균의 분포:** 모집단 분산이 σ^2 일 때 표본평균 $\bar{Y} \sim N(\mu, \sigma/\sqrt{n})$ (n이 큰 경우)
  - 표본평균은 모평균 주변에 이런 분포로 흩어짐. 표집분포의 표준편차 = 표준오차(SE)

가설 검정 I.

가설 검정의 단계

3단계: 귀무가설 하의 표집분포

표준오차 (Standard Error, SE)

표준오차 = 추정치(통계량) **표집분포**의 **표준편차**

- 표본 하나의 통계량이 얼마나 **불확실**한지, 추정의 **정밀도**를 나타냄.
- 표준오차가 작을수록 (즉 분포가 좁을수록) 표본 통계치가 모수 주변에 촘촘히 분포 $\rightarrow$ **정밀한 추정**
- 표준오차는 주로 **표본 크기(n)**의 제곱근에 반비례 $(n \uparrow \rightarrow SE \downarrow)$ 하고, 모분산(변동성)에 비례

가설 검정 I.

가설 검정의 단계

3단계: 귀무가설 하의 표집분포

표준오차 (Standard Error, SE)

모평균 추정치의 SE:

$$SE(\bar{Y}) = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

- 모표준편차 σ 모르면 표본표준편차 s 사용

모비율 추정치의 SE (큰 표본):

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

가설 검정 I.

가설 검정의 단계

3단계: 귀무가설 하의 표집분포

표준오차 (Standard Error, SE)

두 표본 평균차의 SE: $(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}})$

표본 크기 증가 $\rightarrow$ SE가 감소 (추정 더 정확); 모집단 변동성 클수록 SE 증가

가설 검정 I.

가설 검정의 단계

퀴즈

퀴즈: 표본의 크기를 4배로 늘리면 표준오차(SE)는 몇 배로 줄어들까?

정답: **1/2 배.** SE는 대략 $\sqrt{n}$에 반비례하므로, n이 4배 되면 SE는 $1/\sqrt{4} = 1/2$로 감소

가설 검정 I.

가설 검정의 단계

4단계: $p-$값 계산

$p-$**값 (유의수준)**: **귀무가설이 참**이라고 가정할 때, **지금 얻은 데이터 정도로 극단적인 결과**가 나올 확률

- 한마디로 **데이터가 귀무가설과 얼마나 안 맞는지**를 확률로 표현한 것
- $p-$값이 **작을수록** 현재 데이터는 귀무가설 하에서 보기 **힘든** 일 $\rightarrow$ 귀무가설에 대한 **반증 증거**가 강함.
- $p-$값이 **크면** 이런 데이터는 귀무가설 하에서 흔히 일어날 수 있음. $\rightarrow$ 특별한 증거가 아님.
- $p-$값은 0부터 1 사이의 값이며, **연속적인 증거의 척도**로 활용(0.001 매우 강한 증거 vs 0.1 약한 증거). 단, 이 기준은 임의의 기준임.

가설 검정 I.

가설 검정의 단계

4단계: $p-$값 계산

양측검정 vs. 단측검정의 $p-$값

양측검정 (H_A :$\neq$)의 $p-$값: 관찰된 통계량 이상으로 **양쪽 꼬리**에서 극단적일 확률의 합

- $z = +2$ 관찰 (오른쪽 꼬리), $p-$값 = 오른쪽 꼬리확률 $\times 2$(양쪽 대칭이므로 왼쪽도 동일)

단측검정 (H_A :$>$ 또는 $<$)의 $p-$값: 해당 **한쪽 꼬리**에서 더 극단적일 확률

- H_A: "크다" 상황에서 $z = +2$라면 $p-$값 = $\Pr(Z \geq 2)$(오른쪽 꼬리만)
- 같은 통계량 값이라도, 단측검정 $p-$값은 양측의 절반 정도가 됨(검정력이 높아짐)

가설 검정 I.

가설 검정의 단계

4단계: $p-$값 계산

통계량과 p값 계산 예시

예를 들어, **표준정규 통계량** $z = 2.0$ 이 나왔다면:

- 양측검정의 $p-$값 = $\Pr(Z \leq -2$ 또는 $Z \geq 2) \approx 0.045$ (약 4.5%)
- 단측검정 (오른쪽 꼬리)의 $p-$값 = $\Pr(Z \leq 2) \approx 0.0225$ (약 2.3%)
- **두 꼬리의 p값 시각화:** 관측된 z = 2에서 양쪽 꼬리의 면적 합이 p값을 나타냄(약 4.5%).

가설 검정 I.

가설 검정의 단계

4단계: $p-$값 계산

통계량과 p값 계산 예시

$z = 3.0$ 이라면:

- 양측 $p-$ 값 ≈ 0.0027 (0.27%): 매우 작음.
- $z = 1.0$이라면 양측 $p \approx 0.32$ (32%): 매우 큼(거의 증거 없음).

👉 $p-$값은 통계량과 귀무가설 분포로부터 계산되며, 소프트웨어가 자동으로 계산 및 제공

가설 검정 I.

가설 검정의 단계

4단계: $p-$값 계산

$p-$값의 의미와 해석

$p-$값이 작다 (예: 0.01): "귀무가설이 참인데 이런 데이터가 나올 확률이 1%뿐" $\rightarrow$ **귀무가설에 대한 강한 의심**

$p-$값이 크다 (예: 0.5): "귀무가설 하에서도 흔히 일어날 결과" $\rightarrow$ 데이터는 귀무가설과 잘 맞음(특별한 효과 증거 없음).

가설 검정 I.

가설 검정의 단계

4단계: $p-$값 계산

$p-$값의 의미와 해석

주의

- $p-$값은 "귀무가설이 **참일 확률**"이 아님!
- 귀무가설이 참/거짓인지 확률로 표현할 수 없음.
- $p = 0.05$라고 해서 "귀무가설이 5% 확률로 참"이라는 뜻이 **아니라**, "귀무가설 하에서 이런 이상치가 나올 확률이 5%"

가설 검정 I.

가설 검정의 단계

퀴즈

퀴즈: $p = 0.04$와 $p = 0.004$ 두 경우가 있을 때, **어느 경우가 귀무가설에 대해 더 강한 반증의 증거일까?**

정답

- $p = 0.004$ 가 훨씬 강한 증거
- 두 값 모두 0.05보다 작아 유의하지만, 0.004는 0.04보다 데이터가 귀무가설과 맞지 않는 정도가 훨씬 크다는 뜻

가설 검정 I.

가설 검정의 단계

5단계: 유의수준 (significance level, α)

유의수준 α = 기각 결정의 기준 확률 (관례적으로 0.05, 즉 5%)

- 연구자가 **1종 오류**(귀무가설이 참인데 잘못 기각)로 허용할 최대 확률을 의미
- 보통 사회과학에서는 $\alpha = 0.05$를 많이 사용 (5%의 위험은 감수)
- 연구 상황에 따라 0.01 (1%)처럼 더 엄격하게 쓰기도 하고, 탐색적 연구에선 0.10 (10%)도 쓰기도 함.
- α는 **미리 결정**해야 하며, 결과를 본 후 임의 변경하면 안 됨(p-hacking 피해야).

가설 검정 I.

가설 검정의 단계

5단계: 유의수준 (significance level, α)

유의수준과 판단 기준

통계분석 결과 p−**값** $\leq \alpha$ 이면: 결과가 **유의수준보다 충분히 드물게** 나왔으므로 **귀무가설을 기각**

- 이때 "통계적으로 유의미한 결과"라고 표현

가설 검정 I.

가설 검정의 단계

5단계: 유의수준 (significance level, α)

유의수준과 판단 기준

$p-$값 $> \alpha$ 이면: 결과가 그다지 드물지 않으므로 **귀무가설을 기각하지 못함.**

- 통계적으로 유의하지 않은 결과. 증거 부족으로 결론 유보
  - $\alpha = 0.05, p = 0.03 \rightarrow 0.03 < 0.05$ 이므로 귀무가설 기각 (유의)
  - $\alpha = 0.05, p = 0.15 \rightarrow 0.15 > 0.05$ 이므로 기각 못함 (유의하지 않음)

가설 검정 I.

가설 검정의 단계

5단계: 유의수준 (significance level, α)

유의수준과 판단 기준

"통계적으로 유의미함"이란?

- 귀무가설 하에서 **보기 힘든 극단적인** 데이터를 얻었다는 것 $\rightarrow$ 우연이라고 보기 어려움.
    - 즉, **효과나 차이가 있다고 볼 만하다**는 의미
- 통계적으로 유의하다는 것이 **실용적으로나 이론적으로 중요한 효과**인지를 뜻하지는 않음.
    - 표본이 엄청 크면 아주 미묘한 차이도 $p < 0.05$ 될 수 있음.
    - 유의미하다는 것은 일단 **통계적 증거의 유무**를 말할 뿐

가설 검정 I.

가설 검정의 단계

5단계: 유의수준 (significance level, α)

귀무가설 기각 실패 시 주의

$p-$값이 유의수준보다 크면 **귀무가설을 기각하지 않음**. 하지만 이것을 **"귀무가설을 채택"**했다고 표현하지 않음.

- **기각 실패** = 현재 데이터로는 귀무가설을 반증할 증거가 부족할 뿐

실제로 귀무가설이 참일 수도, 아니면 단지 자료가 부족해서 못 밝힌 것일 수도 있음.

- 즉, 귀무가설이 옳다고 **증명된 것은 아님**. *"무죄 추정"*에 비유 가능
- 피고인이 무죄 평결을 받아도, 그것이 **피고인이 실제 결백함을 증명**하는 건 아니고 증거 불충분일 뿐(유죄로 입증될 정도 증거가 없음)

가설 검정 I.

가설 검정의 단계

5단계: 유의수준 (significance level, α)

비유: 형사재판에서의 무죄추정

귀무가설 = 피고인 무죄, 대립가설 = 피고인 유죄로 비유 가능

검사는 대립가설(유죄)을 입증할 증거를 제시해야 함.

- **합리적 의심 이상의 증거**($p-$값이 매우 작음)에 도달해야 유죄 판결 (귀무가설 기각)
- 증거 부족 시 무죄 판결(귀무가설 기각 실패) $\rightarrow$ 그러나 이는 피고가 실제로 결백하다는 **증명은 아님.**

가설 검정 I.

가설 검정의 단계

퀴즈: 유의수준 $\alpha = 0.05$로 설정했다고 하자. 분석 결과 $p = 0.08$이었다면, 귀무가설에 대해 어떻게 해야 할까?

정답: 귀무가설을 기각하지 못함.

- $p-$값이 $0.08 > 0.05$이므로 통계적으로 유의하지 않은 결과

Part II. 오류와 검정력, 신뢰구간

가설 검정 I.

오류와 검정력

1종 오류 (Type I Error)

1종 오류 = 실제로 **귀무가설이 참인데** 잘못 기각하는 오류 (false positive)

- "없는데 있다고 잘못 판단"하는 오류
- 효과 없는 신약을 효과 있다고 결론내림, 무죄인 사람을 유죄 판결함.
- 1종 오류가 발생할 확률 = $alpha$ **(유의수준)** 로 연구자가 통제
  - $\alpha = 0.05$로 정했다면, *"귀무가설이 참인 상황에서 5%는 기각으로 오판 가능"*을 감수한다는 의미

가설 검정 I.

오류와 검정력

2종 오류 (Type II Error)

2종 오류 = 실제로 **귀무가설이 거짓인데** 기각하지 못하는 오류 (false negative)

- "있는데 없다고 잘못 판단"하는 오류
- 효과 있는 정책을 효과 없다고 놓침, 유죄인 범인을 증거 부족으로 풀어줌.
- 2종 오류 확률 = β 로 표시 (일반적으로 쉽게 계산되지 않지만, 검정력으로 보완 설명)
  - 연구자가 보통 직접 β를 정하진 않지만, **표본크기나 검정방법**으로 영향을 줌 (β 줄이는 방향으로 설계)

가설 검정 I.

오류와 검정력

구분	의미	예시	확률
1종 오류 (Type I)	H_0이 참인데 기각	코로나19에 걸리지 않았는데 양성이라고 판단 (false positive)	α
2종 오류 (Type II)	H_0이 거짓인데 기각하지 않음	코로나19에 걸렸는데 음성이라고 판단 (false negative)	β
검정력 (Power)	H_0이 거짓일 때 기각할 확률	코로나19에 걸렸을 때 양성으로 정확히 판단할 확률	1−β

통계검정 결정에는 두 가지 정답과 두 가지 오류 가능성이 존재

- 1종 오류 (α): H_0 참인데 기각 (False Alarm)
- 2종 오류 (β): H_0 거짓인데 채택(기각 실패) (Miss)

검정력 (power) = $1 - \beta$ = H_0 거짓일 때 올바로 기각할 확률

가설 검정 I.

오류와 검정력

검정력(Power)

검정력 = $1 - \beta$ ="“귀무가설이 거짓일 때 (실제로 효과가 있을 때) 이를 올바르게 밝혀낼 확률"

- 쉽게 말해, 참된 효과를 놓치지 않는 능력
- 검정력이 높을수록, 실제 효과를 검출해낼 가능성이 크고 2종 오류 위험이 낮음.

사회과학에서는 통상 80% 이상의 검정력을 목표로 연구 설계 $(\beta = 0.20)$

- "효과가 있다면 10번 중 8번은 유의하게 발견할 수 있도록"

가설 검정 I.

오류와 검정력

검정력에 영향을 미치는 요인

표본 크기(n) $\uparrow \rightarrow$ 표준오차 $\downarrow \rightarrow$ 작은 효과도 검출 가능 $\rightarrow$ 검정력 증가

효과의 실제 크기 $\uparrow$ (모평균 차이, 모비율 차 등) $\rightarrow$ 통계량 더 크게 나옴 $\rightarrow$ 검정력 증가

- 모집단 변동성 감소 (데이터 잡음 적음) $\rightarrow$ 통계량 분포 좁아짐 $\rightarrow$ 검정력 $\uparrow$
- 유의수준 $\alpha \uparrow$ (기준 완화) $\rightarrow$ 기각 범위 넓어짐 $\rightarrow$ 검정력 증가 (하지만 1종 오류도 증가)

1종 오류와 2종 오류는 상충관계: α를 줄이면 β는 늘어나는 경향

가설 검정 I.

오류와 검정력

실제 연구에서 α와 검정력 결정

연구 분야와 맥락에 따라 α 선택

- 인명에 영향 큰 의학 연구 등은 α 를 매우 작게 (1종 오류 최소화)
- 탐색적 연구나 예비 연구는 α 를 0.10까지 쓰기도 (증거 포착 중시)

표본크기 결정 시, 연구자는 통상 "검정력 0.8" 등을 목표로 설정

- 예상 효과크기, 허용 α 로부터 필요한 n 을 계산 (power analysis)
- 지나치게 작은 표본 $\rightarrow$ 2종 오류 우려 $\rightarrow$ 의미 있는 효과도 못 발견 가능

결국 자원 한계 내에서 최대한 높은 검정력 확보가 중요

가설 검정 I.

오류와 검정력

오류율 균형의 비유

엄격한 기준 ($\alpha \downarrow$): 1종 오류 (거짓 긍정) 거의 안 내지만, 2종 오류(거짓 부정) 늘어남

- 무고한 사람 처벌 최소화 (1종 $\downarrow$) $\leftrightarrow$ 많은 범죄자 놓침 (2종 $\uparrow$)

완화된 기준 ($\alpha \uparrow$): 2종 오류 줄어 검정력 높아지지만, 1종 오류 위험 상승

- 범죄자 대부분 잡아냄 (2종 $\downarrow$) $\leftrightarrow$ 일부 무고한 사람도 잡힘 (1종 $\uparrow$)

이 상충관계를 고려해 분야마다 적절한 α 선택 (예: 법은 무죄추정 $\rightarrow$ 1종 오류 최소화)

가설 검정 I.

오류와 검정력

퀴즈

퀴즈: 유의수준을 0.05에서 0.01로 더 엄격하게 낮추면, 1종 오류와 2종 오류는 각각 어떻게 변할까?

정답

- 1종 오류 발생 확률은 줄어들고, 2종 오류 (놓칠 위험)는 커지게 됨.
- $\alpha \downarrow \rightarrow$ 검출 기준 엄격 $\rightarrow$ 놓치는 경우 $\uparrow$

가설 검정 I.

신뢰구간

자, 이제 다시 신뢰구간으로 돌아와보자.

신뢰구간(confidence interval): 모수(모집단의 참값)에 대한 추정 범위를 제시

- 점추정 + 오차범위 형태 (예: "모비율은 55% ± 4%p")
- 보통 95% 신뢰구간 많이 사용 (90%, 99% 등도 상황에 따라)

신뢰구간은 효과의 크기와 방향을 함께 제시, 단순한 유의성 여부보다 풍부한 정보 제공

- "모수는 이 범위 어딘가에 있을 것으로 확신한다"

가설 검정 I.

신뢰수준

신뢰수준 (Confidence Level)

- 95% 신뢰구간이라 함은, 장기적으로 동일 방법으로 표본 추출 반복 시 구간들이 95% 확률로 참값을 포함함을 뜻함.
- 한 번 계산된 특정 구간에 대해 "95% 확률로 참값 포함"이라고 말하진 않음(참값은 고정, 구간이 랜덤).
- "모평균은 100~120 사이에 있을 것으로 95% 신뢰한다"
- 신뢰수준 높일수록, 구간이 더 넓어져 더 두루뭉술하게 잡지만 안전함(모수 포함 확률 ↑)

가설 검정 I.

신뢰수준: 계산

일반 형태: 추정치 $\pm$ (임계값) $\times$ 표준오차

95% 신뢰구간: 추정치 $\pm$ 1.96 $\times$ SE (정규 근사, 양측 95%)

- 정확히는: 추정치 $\pm\, z_{\alpha/2} \times$ SE (혹은 $t-$분포 임계값)
- 표본평균 $\bar{y} = 50$, $s = 10$, $n = 100$ 일 때: $SE = 10/\sqrt{100} = 1$
  - 95% CI = $50 \pm 1.96 \times 1$ = (48.04, 51.96)

신뢰구간은 유의수준 $\alpha = 0.05$의 양측 검정과 직접 연결됨.

가설 검정 I.

신뢰구간과 검정의 연결

귀무가설 값(효과 = 0)이 신뢰구간 내에 포함된다면, 해당 값이 가능한 값으로 남아 있다는 뜻으로 동일한 수준에서 귀무가설을 기각할 수 없다(유의하지 않다)는 것을 의미

귀무가설 값이 신뢰구간 밖에 있다면, 그 값은 배제되었다는 뜻으로 귀무가설을 기각(통계적으로 유의함)

- 95% 신뢰구간이 (0.1, 5.2)%이면 0% 효과는 구간 밖이므로 귀무가설 기각 ($p < 0.05$)
- 99% 신뢰구간이면 $\alpha = 0.01$ 검정과 대응. 신뢰수준과 유의수준 합은 100% 대응 관계

가설 검정 I.

신뢰구간과 검정의 연결

퀴즈

퀴즈: 한 연구에서 두 처리집단 간 효과의 95% 신뢰구간이 -0.5 ~ 1.2 (%p) 였다고 하자. "효과가 없다"는 귀무가설(효과 = 0)을 이 결과로 기각할 수 있을까?

정답: 기각할 수 없음. (0이 신뢰구간에 포함되므로 통계적으로 유의하지 않음.)

가설 검정 I.

한계 1: 오해하기 쉬운 $p-$값의 의미

p값은 흔히 오해되기 쉬운 통계량

- " $p-$ 값 = 0.05이면 귀무가설이 95% 확률로 거짓이다" ❌ (틀림)
- " $p-$ 값 = 0.05이면 결과는 우연일 가능성이 5%다" ✔ (맞는 해석)

$p-$값은 데이터가 귀무가설 하에 나올 확률이지, 귀무가설이 맞을 확률이 아님.

또한 $p-$값은 사후 확률이 아니라 가정 하의 가능성이라, "신뢰도"와 혼동하면 안 됨.

가설 검정 I.

한계 2: 통계적 유의성과 실질적 중요성

유의미한 차이 $\neq$ 큰/중요한 차이

- 큰 표본으로 연구하면 아주 작은 효과도 $p < 0.05$ 나올 수 있음.
- 실질적으로는 의미 없는 차이라도 통계적으로 유의 가능
  - n=1,000,000에서 평균 차이 0.1도 유의하게 검출될 수 있음.
- 반대로 표본이 작으면 꽤 큰 효과도 유의하지 않게 나와 놓칠 수 있음(2종 오류).

따라서 p값만 볼 게 아니라, **효과의 크기(추정치)**와 신뢰구간도 함께 봐야 함.

가설 검정 I.

한계 3: 다중검정 문제와 $p-$hacking

만약 여러 가설을 동시에 테스트하면, 우연히 유의한 결과가 나올 확률이 크게 증가

- 유효한 효과는 없는데 20가지 변수 시험 $\rightarrow \alpha = 0.05$ 기준 평균 1개는 가짜로 유의 나올 수 있음.
- 이를 악용/실수하는 것이 $p-$hacking: 유의한 결과가 나올 때까지 분석을 반복하거나 여러 것 중 유의한 것만 강조

교정법(Bonferroni 등)으로 다중비교 보정하거나, 탐색적 분석 결과는 별도로 검증하는 것이 필요

항상 사전 계획된 가설을 검정하고, 결과를 선택적으로 보고하지 않도록 주의해야 함.

Part III. 중간시험 Prep. 및 질의응답

다음 강의에는 중간시험 리뷰 및 가설검정의 논리에 대해 다시 한 번 살펴볼 것

중간시험

일시: 2025년 10월 28일 화요일 10:00 - 11:45 (105분)

총 15개 문항: 객관식 + 약술형

범위

과학적 연구와 이론, 인과관계와 연구설계

개념과 측정

단변량과 양변량 분포

확률과 통계적 추론

정규분포와 신뢰구간

가설검정의 기초

중간시험

인과관계 성립을 위한 필수 조건

관련 개념: 인과관계 수립의 4가지 관문 (공변, 시간적 선행, 허위 관계 배제 등)

참고 슬라이드: `02-slides.pdf`

- 상관관계가 인과관계가 아님.
- '인과관계 수립을 위한 관문' 4가지: (A) 공변규칙성, (B) 역인과관계 (시간적 선행 관련), (C) 허위적 관계 (제3변인 통제) 등 확인

중간시험

측정척도와 중심경향값

관련 개념: 측정 수준(명목형), 중심경향성(평균, 중앙값, 최빈값)

참고 슬라이드: `03-slides.pdf` 및 `04-slides.pdf`

$p-$값의 올바른 해석

관련 개념: p-value의 정의, 가설검정

참고 슬라이드: `06-slides.pdf` 및 `07-slides.pdf`

- $p-$값의 의미는 "'귀무가설이 참'(효과가 실제로 없다)일 때 이런 데이터가 나올 확률"
- $p-$값은 '귀무가설이 참일 확률'이 아님!"

중간시험

측정의 신뢰도와 타당도

관련 개념: 신뢰도(Reliability), 타당도(Validity)

참고 슬라이드: `03-slides.pdf`

상관계수의 올바른 해석

관련 개념: 상관계수(Correlation Coefficient)의 강도와 방향

참고 슬라이드: `04-slides.pdf`

- '상관계수'는 -1(완벽한 음)부터 +1(완벽한 양) 사이의 값을 가짐.
- "상관관계는 인과관계가 아님!"

독립 사건의 확률

관련 개념: 확률의 종류(결합확률), 사건의 독립

참고 슬라이드: `05-slides.pdf`

- '결합확률 $\Pr(A \cap B)$'과 조건부 확률의 정의를 읽어볼 것
- '독립(Independence)'이란?
    - "어떤 사건의 결과를 알아도 다른 사건의 결과에 정보가 없다면" 두 사건이 독립
    - $\Pr(A \cap B) = P(A) \times P(B)$

중간시험

95% 신뢰구간의 해석

관련 개념: 신뢰구간(Confidence Interval)의 해석, 가설검정과의 관계

참고 슬라이드: `06-slides.pdf` 및 `07-slides.pdf`

- 신뢰구간과 가설검정의 관계를 리뷰할 것
- 귀무가설의 값이 신뢰구간 *안에* 있으면 기각할 수 없고 *밖에* 있으면 기각함.
- 신뢰수준(90%, 95%, 99%)에 따른 구간 폭의 변화 이해

중간시험

이항분포의 기대값과 분산

관련 개념: 이항분포(Binomial distribution)의 $\mathbb{E}[Y]$와 $Var(Y)$

참고 슬라이드: `05-slides.pdf`

- 이항분포의 특성으로 기대값 $\mathbb{E}[Y] = np$

표본 크기(n)와 표준오차(SE)의 관계

관련 개념: 표준오차(Standard Error)의 정의, 중심극한정리

참고 슬라이드: `03-slides.pdf`, `05-slides.pdf`, `06-slides.pdf`, `07-slides.pdf`

- $SE = s/\sqrt{n}$, $SE = \sigma/\sqrt{n}$, n이 커질수록 SE가 감소하며 '정밀도'가 증가
- 표준오차:"추정치 표집분포의 표준편차", "추정의 정밀도"를 나타내고 $\sqrt{n}$에 반비례

중간시험

표본 크기 증가의 '수확체감' 현상

관련 개념: 표준오차, 오차범위, n의 $\sqrt{n}$ 효과

참고 슬라이드: `06-slides.pdf`

- n이 1,000 $\rightarrow$ 2,500 $\rightarrow$ 10,000으로 증가할 때 오차범위가 $\pm 3\%$ $\rightarrow$ $\pm 2\%$ $\rightarrow$ $\pm 1\%$로 줄어드는 예시
- 오차를 절반으로 줄이려면 표본이 4배 필요
- "통계적으로 유의미한 개선에는 큰 표본 증가 필요 (수확체감 발생)"

중간시험

히스토그램 분포 특징 서술

관련 개념: 히스토그램 해석, 분포 형태(왜도), 이탈치

참고 슬라이드: `03-slides.pdf`

산점도 관계 해석 및 인과관계의 한계

관련 개념: 산점도(Scatterplot) 해석, 상관관계와 인과관계, 혼란변수

참고 슬라이드: `02-slides.pdf`, `03-slides.pdf`, `04-slides.pdf`

- 상관관계가 인과관계가 아닌 이유로 '허위적 관계'와 '혼란변인(confounding variable)' 설명 이해

중간시험

R 코드 읽기

```
set.seed(123) # 예시: 난수 생성을 재현 가능하도록 기준점을 123으로 설정
population <- 1:100 # (1) 1부터 100까지의 숫자가 담긴 'population' 벡터를 생성
sample_means <- c() # (2) 표본 평균들을 저장할 'sample_means'라는 이름의 빈 벡터(깡통)를 저장
for (i in 1:100) { # (3) 1부터 100까지, 총 100번의 반복
  s <- sample(population, size = 10, replace = TRUE) # (4) 'population'에서 10개의 숫자를 복원추출
  sample_means[i] <- mean(s) # (5) 's' 벡터의 평균을 계산하여 'sample_means' 벡터의 i번째 위치에 저장
}
mean(sample_means) # (6) 100번 반복하여 'sample_means'에 저장된 100개 평균값들의 ' 평균'을 계산
```

중간시험

일표본 $t-$검정 통계량 계산 및 유의성 판단

관련 개념: 일표본 $t-$검정(One-sample t-test), 검정통계량(t-statistic)

참고 슬라이드: `06-slides.pdf` 및 `07-slides.pdf`

- 주어진 일표본 $t-$검정통계량 공식을 이용하여 계산
- 유의성 판단 기준: $\alpha = 0.05$ (양측)의 임계값(critical value)이 약 ± 1.96임을 확인하고, 계산된 t-값이 이보다 크면 유의 아니면 유의하지 않다고 결론

중간시험

이표본 $t-$검정 통계량 계산 및 유의성 판단

관련 개념: 이표본 $t-$검정(Independent two-sample t-test), 검정통계량

참고 슬라이드: `06-slides.pdf` 및 `07-slides.pdf`

- 주어진 이표본 $t-$통계량 공식을 이용하여 계산
- 유의성 판단 기준: $\alpha = 0.05$ (양측)의 임계값이 약 ± 1.96임을 확인하고, 계산된 t-값이 이보다 크면 유의 아니면 유의하지 않다고 결론

중간시험

제1종 오류(Type I)와 제2종 오류(Type II)

관련 개념: 가설검정의 오류 (Type I Error, Type II Error)

참고 슬라이드: `06-slides.pdf` 및 `07-slides.pdf`

- 1종 오류와 2종 오류에 대한 표를 확인
- 1종 오류 (Type I Error)는 "H_0이 참인데 잘못 기각하는 오류 (false positive)"
- 2종 오류 (Type II Error)는 "H_0이 거짓인데 기각하지 못하는 오류 (false negative)"

감사합니다!

궁금한 것이 있으면 언제든 연락하세요.

강사 연락처

연락처	박상훈
	sh.park.poli@gmail.com
	sanghoon-park.com/
	영상바이오관 405