

가설 검정의 기본 개념

가설 검정(Hypothesis Testing)

- 표본 데이터를 사용하여 모집단에 대한 어떤 주장이나 가설이 통계적으로 유의미한지를 판단하는 추론 통계의 핵심적인 과정
- 우연히 발생한 것인지 아니면 실제로 의미가 있는 것인지 판단
- 재판 비유: "유죄라는 명백한 증거가 있기 전까지는 무죄로 추정한다."

귀무가설 (Null Hypothesis, H_0)

- 연구자가 직접적으로 검정하고자 하는 가설
- "차이가 없다", "효과가 없다", "관계가 없다" 같은 보수적이고 기존에 알려진 사실
- 기각(reject)하고자 하는 대상
- 재판 비유: '무죄 추정'에 해당
- 예시: H_0 : 새로운 약은 효과가 없다 (기존 약과 차이가 없다).

대립가설 (Alternative Hypothesis, H_1 또는 H_a)

- 연구자가 귀무가설에 대립하여 새롭게 주장하고 싶은 가설
- "차이가 있다", "효과가 있다", "관계가 있다"
- 입증하고자 하는 내용
- 재판 비유: '유죄' 주장에 해당
- 예시: H_1 : 새로운 약은 효과가 있다 (기존 약과 차이가 있다).

유의수준 (Significance Level, α)

- 귀무가설이 실제로 맞지만, 그것을 기각할(틀렸다고 판단할) 확률의 최대 허용 한계
- 즉, **1종 오류(Type I Error)**를 범할 확률
- 보통 0.05(5%) 또는 0.01(1%)을 사용하며, 연구자가 직접 설정
- 재판 비유: '무고한 사람에게 유죄 판결을 내릴 위험성'에 해당

p-value (Probability Value)

- 귀무가설이 맞다고 가정했을 때, 우리가 현재 관찰한 데이터(또는 그보다 더 극단적인 데이터)가 나타날 확률
- p-value가 매우 작다는 것은, 귀무가설 하에서는 거의 일어나기 힘든 희귀한 일이 발생했음을 의미
- 판단기준(유의수준)보다 p-value가 작다 → 대립가설이 채택, 귀무가설 기각
- 단측검정: 분포 양쪽 끝 중 한쪽 끝단에 속할 확률을 확인
 - e.g. A매장과 B매장 중 어디의 수요량이 더 큰가?
- 양측검정: 분포의 양쪽 끝단에 속할 확률을 확인
 - e.g. 매장 간 수요량 차이가 있나?

검정통계량 (Test Statistic)

- 표본 데이터로부터 계산되는 값으로, 귀무가설이 맞는지 틀리는지를 판단하는 기준
- 데이터가 귀무가설로부터 얼마나 떨어져 있는지를 나타내는 표준화된 값

- e.g., T-값, Z-값, F-값, 카이제곱 값

제 1종 오류 / 제 2종 오류

통계적 결정 \ 실제 상황	H_0 가 사실	H_1 가 사실
H_0 을 채택 (H_1 기각)	옳은 결정 ($1-\alpha$) True Negative	제 2종 오류 (β) False Negative
H_1 을 채택 (H_0 기각)	제 1종 오류 (α) False Positive	옳은 결정 ($1-\beta$) True Positive

- 제 1종 오류 : 귀무가설이 사실인데, 대립가설을 채택할 확률 (확률= α)
 - e.g. 효과 없는 약을 효과 있다고 잘못 판단
- 제 2종 오류 : 대립가설이 사실인데, 귀무가설을 채택할 확률 (확률= β)
 - e.g. 효과 있는 약을 효과 없다고 잘못 판단
- 유의수준(α)을 낮추면(더 엄격한 기준 적용), 1종 오류를 범할 확률은 줄어들지만, 귀무가설을 기각하기 어려워져 2종 오류(β)를 범할 확률이 높아집니다.
- α 와 β 는 서로 트레이드오프(trade-off) 관계에 있습니다.

검정력 (Power, $1-\beta$)

- 대립가설이 사실일 때, 이를 사실로서 올바르게 탐지해낼(즉, H_0 을 기각할) 확률
- 검정력이 높을수록 좋은 검정 방법이라고 할 수 있습니다.
- 검정력은 표본 크기(n), 유의수준(α), 효과 크기(effect size)에 의해 영향을 받습니다.

가설 검정의 절차

- 가설 설정:** 귀무가설(H_0)과 대립가설(H_1)을 설정한다.
- 유의수준 설정:** 1종 오류의 최대 허용 수준인 유의수준(α)을 결정한다. (e.g., $\alpha = 0.05$)
- 검정통계량 계산:** 데이터에 맞는 검정 방법(t-test, ANOVA 등)을 선택하고, 표본 데이터로부터 검정통계량을 계산한다.
- p-value 계산 또는 기각역 설정:** 계산된 검정통계량에 해당하는 p-value를 구하거나, 유의수준(α)에 해당하는 기각역(critical region)을 설정한다.
- 의사결정:**
 - p-value < 유의수준(α) 이면, 귀무가설을 기각한다. (대립가설 채택)
 - p-value \geq 유의수준(α) 이면, 귀무가설을 기각하지 못한다. (대립가설을 채택할 근거가 부족함)

"귀무가설을 기각하지 못한다"는 것이 "귀무가설이 맞다"는 것을 의미하지는 않습니다. 단지, 귀무가설을 기각할 만큼 충분한 증거를 찾지 못했다는 의미입니다.

모수 검정 vs 비모수 검정

구분	모수 검정 (Parametric Test)	비모수 검정 (Non-parametric Test)
----	-------------------------	------------------------------

구분	모수 검정 (Parametric Test)	비모수 검정 (Non-parametric Test)
가정 (Assumption)	데이터가 정규분포를 따른다고 가정	데이터 분포에 대한 가정이 거의 없음
데이터 유형	등간·비율척도 (연속형 데이터)	서열·순위척도 혹은 비정규 데이터
대표 검정 예시	t-test, ANOVA, Pearson 상관분석 등	Mann–Whitney U, Kruskal–Wallis, Spearman 상관 등
검정력(power)	가정이 충족되면 검정력이 높음	가정이 필요 없지만, 상대적으로 검정력이 낮음
데이터 크기	표본이 충분히 크면 적합	소표본이거나 이상치가 많을 때 적합
평균 비교 방식	실제 평균값 비교	중앙값이나 순위(rank) 비교