
Hypothesis I

-Comparison of Means-

Contents

1.

가설검정

2.

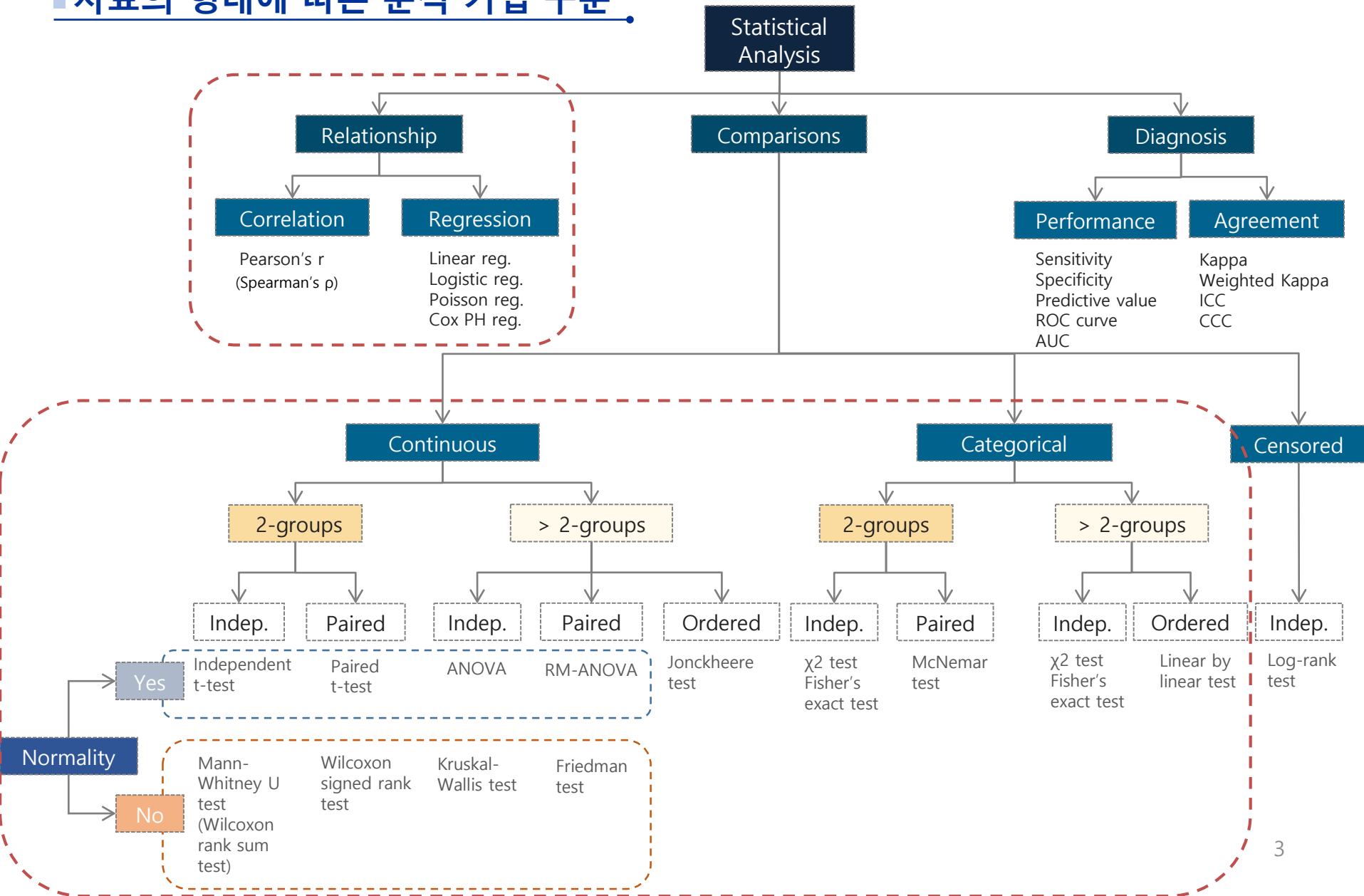
일표본 - 모평균에 대한 추론

3.

이표본 - 모평균 차이에 대한 추론

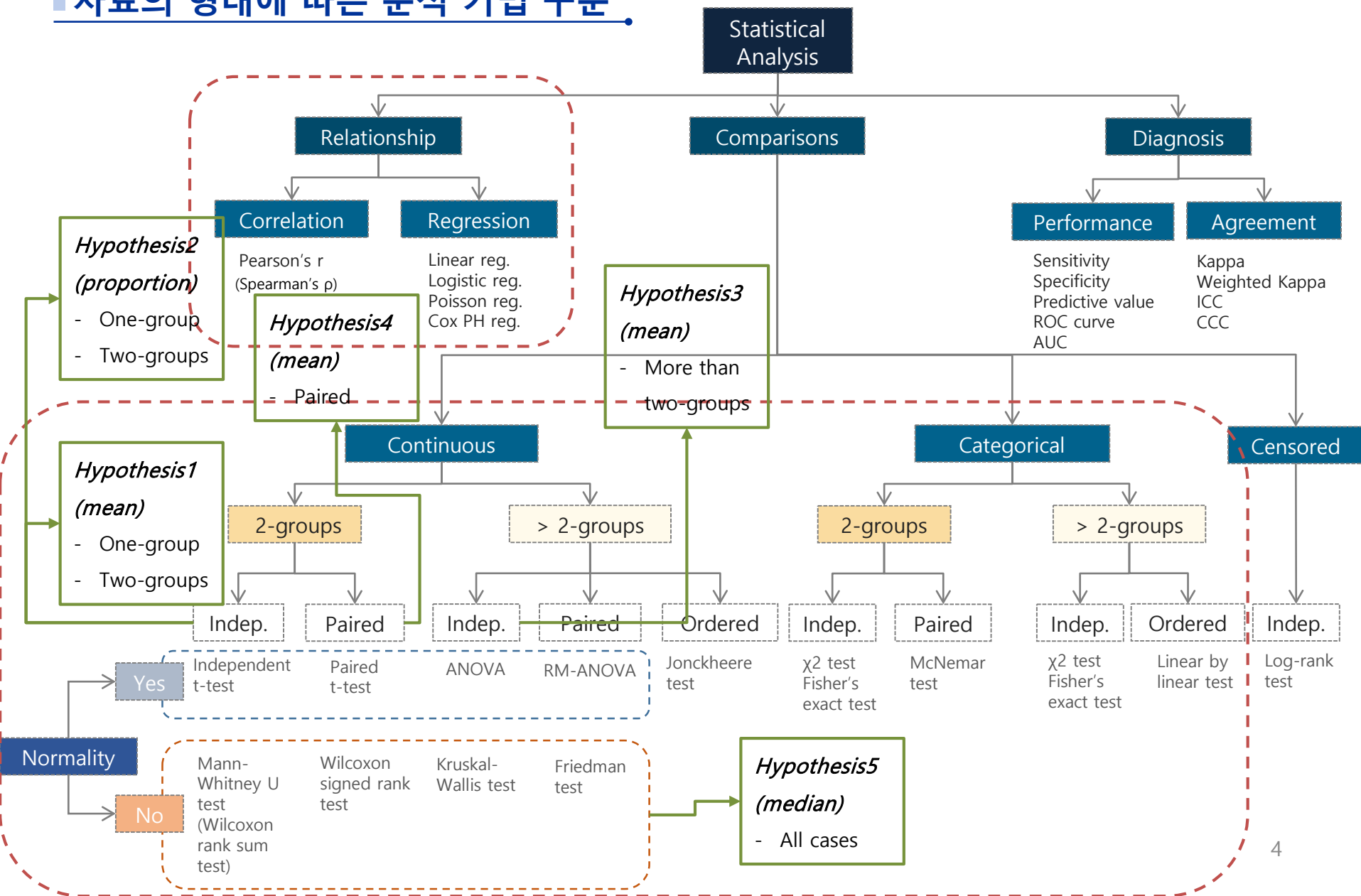
Testing & Modeling

자료의 형태에 따른 분석 기법 구분



Testing & Modeling

자료의 형태에 따른 분석 기법 구분



Project

- 개인 프로젝트 (individual project)

의학자료 분석 프로젝트는 다음과 같은 내용을 기반으로 발표자료를 작성하시면 됩니다.

- 도입(introduction) - 10%
 - ✓ 연구목적과 동기
- 자료 설명 - 20%
 - ✓ <https://vincentarelbundock.github.io/Rdatasets/datasets.html>
 - ✓ <https://www.kaggle.com/>
 - ✓ <https://data.world/datasets/health>
 - ✓ <https://medium.com/@ODSC/15-open-datasets-for-healthcare-830b19980d9>
 - ✓ 대상자 및 변수 설명 등
- 방법론 및 결과 도출 - 50%
 - ✓ Table1 (demographic table) 구성
 - ✓ Model 구축 과정 및 결과
 - ✓ 연구목적에 맞는 결과 해석
 - ✓ 결론 (concluding remark)
- Presentation (10분 이내) + Q&A time (2분 이내) - 20%

Project

- 관심 있는 주제의 의학 저널을 3~5부씩 검색하여 통계부분 및 결과도출 부분을 살펴보자.
 - 다음의 사이트를 비롯한 여러 사이트에서 의학논문을 다운로드 받을 수 있다.
 - ✓ 국외: NCBI의 Pub Med (<http://www.ncbi.nlm.nih.gov/PubMed>)
 - ✓ 국내: 의학연구정보센터 (<http://www.medric.or.kr>)

1. 가설검정

가설검정

■ Definitions

통계적 가설검정(hypothesis testing)

- 기존의 통념을 나타내는 귀무가설(H_0)에 반하는 새로운 주장을 나타내는 대립가설(H_1)을 설정하고, 관측된 데이터를 바탕으로 귀무가설의 채택 또는 기각 여부를 결정하는 통계적 절차

- ✓ 가설검정의 결과는 대립가설이 아닌 귀무가설을 위주로 표현하며, 귀무가설을 '채택(accept)'하거나 '기각(reject)'하거나 둘 중의 하나로 결론을 내린다.
- ✓ 예)

H_0 : 흡연그룹과 비흡연그룹 사이에는 신체활동량에서 차이가 없을 것이다.

H_1 : 흡연그룹과 비흡연그룹 사이에는 신체활동량에서 차이가 있을 것이다.

가설검정

Definitions

제1종 오류(type 1 error)

- 귀무가설이 맞는데도 불구하고 이를 기각하는 오류

제2종 오류(type 2 error)

- 대립가설이 맞는데도 불구하고 귀무가설을 채택하는 오류

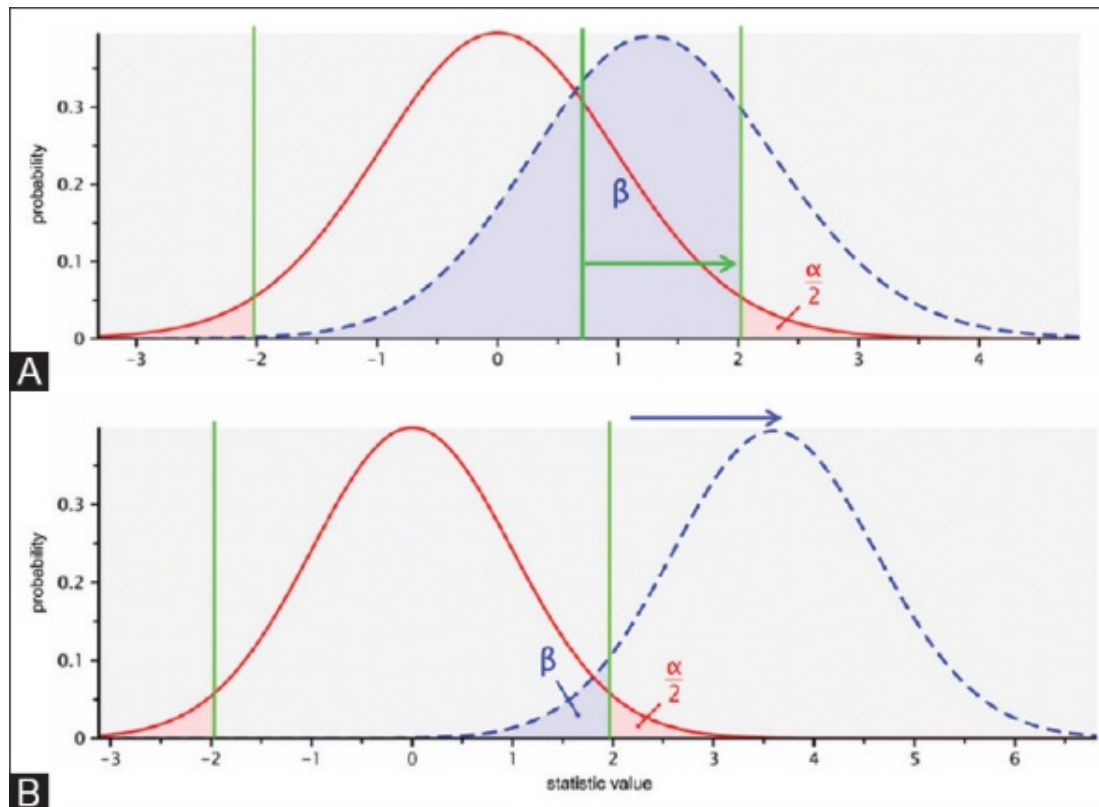
- ✓ 새로운 주장이 틀린데도 받아들이는 오류가 제1종 오류, 새로운 주장이 맞는데도 인정하지 않는 오류가 제2종 오류가 된다.
- ✓ 통계학에서는 제1종 오류를 더 심각하게 받아들여, 가설검정에서 제1종 오류가 발생할 확률을 유의수준(significance level) α 이하로 유지하도록 검정 방법을 설계한다. 유의수준을 만족하면서 제2종 오류가 발생할 확률 β 가 더 작은 가설검정 방법이 더 좋은 것으로 인정된다.

검정결과 \ 진실	귀무가설이 맞음	대립가설이 맞음
	옳은 결정	제2종 오류
귀무가설 채택		
귀무가설 기각	제1종 오류	옳은 결정

가설검정

■ Definitions

- ✓ 통계학에서는 제1종 오류를 더 심각하게 받아들이며, 가설검정에서 제1종 오류가 발생할 확률을 유의수준(significance level) α 이하로 유지하도록 검정 방법을 설계한다. 유의수준을 만족하면서 제2종 오류가 발생할 확률 β 가 더 작은 가설검정 방법이 더 좋은 것으로 인정된다.



가설검정

■ Definitions

유의수준(significance level)

- 귀무가설이 맞는데도 불구하고 이를 기각할 확률(제1종 오류)의 최대값

유의확률(significance probability, p-value)

- 귀무가설이 맞는데도 불구하고 잘못해서 기각할 실제 확률(제1종 오류의 확률)

- ✓ 검정 시 유의확률이 유의수준보다 작으면 그 검정 결과를 받아들일 수 있다.
- ✓ 유의수준은 일반적으로 고정시키고, 유의확률은 자료 및 검정 방식에 따라 다르게 산출된다.
- ✓ 가장 많이 사용되는 유의수준은 $\alpha = 0.05$ 이다. (또는 $\alpha = 0.1, 0.01$)
- ✓ $P\text{-value} < \alpha$ 이면 귀무가설을 기각
- ✓ $P\text{-value} > \alpha$ 이면 귀무가설을 채택(기각할 수 없음)

가설검정

■ 가설검정(예제).

[재판과정으로 가정한다면]

- 사람은 특별한 증거가 없는 한 무죄이므로 귀무가설=무죄
- 확신할 증거가 나오면 유죄이므로 대립가설=유죄로 정의
- 유죄인 사람을 무죄로 선고하는 것도 심각한 오류이지만, 무죄인 사람을 유죄로 선고하는 것이 보다 치명적인 오류이므로 제1종 오류가 된다.

검정결과 \ 진실	귀무가설이 맞음	대립가설이 맞음
	옳은 결정	제2종 오류 (죄가 있지만 무죄선고)
귀무가설 채택		
귀무가설 기각	제1종 오류 (죄가 없는데 유죄선고)	옳은 결정

가설검정

■ Definitions

검정 통계량(test statistic)

- 가설검정('채택' 또는 '기각' 둘 중의 하나를 선택하는 기준)을 위해 사용되는 통계량

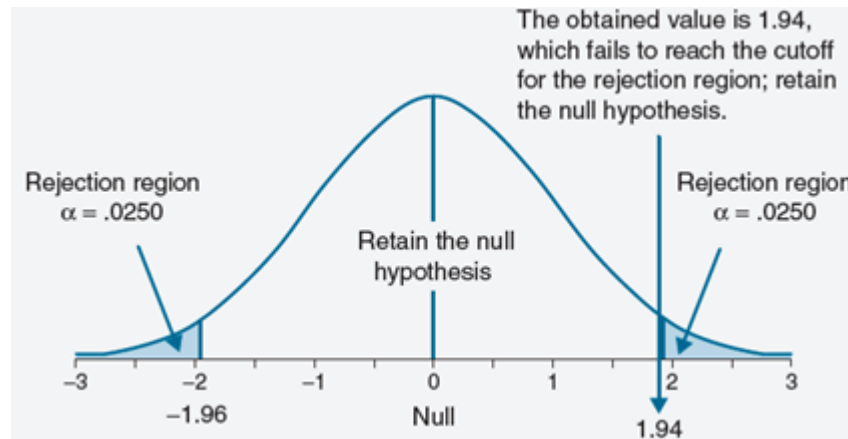
기각역(rejection region)

- 귀무가설을 기각하는 검정통계량의 영역

임계치(critical point)

- 기각역의 경계치

- ✓ 즉, 검정통계량이 임계점을 넘어 기각역에 위치하면 귀무가설을 기각하고, 그렇지 않으면 귀무가설을 채택한다.



가설검정

■ 가설검정 절차

가설검정 절차(testing process) – 앞으로의 모든 검정에 아래 절차대로 진행!

- 귀무가설 설정 $H_0: \theta = \theta_0$
- 대립가설 설정 $H_1: \theta < \theta_0 \text{ or } \theta > \theta_0 \text{ or } \theta \neq \theta_0$
- 유의수준 α 설정
- 검정통계량과 기각역 설정
- 관측된 표본 데이터로부터 검정통계량의 관측치 계산
- 귀무가설(H_0) 채택 또는 기각 결정
- 실제자료를 기반으로 가설에 대한 해석

Thank you!
