

7. 가설 검정

7.1 가설 검정의 이해

통계적 가설검정

표본으로부터의 정보를 이용해 궁금한 사항에 대하여

확률 분포와 확률 계산을 이용한 의사결정을 의미

7.1 가설 검정의 이해

가설의 종류

1. 귀무가설 (Null Hypothesis, H_0)

: 검정하고자 하는 모수에 대해 이미 알려져 있는 **사실 또는 현재의 사실** 등을 나타내는 가설

2. 대립가설 (Alternative Hypothesis, H_1)

: 새로운 연구 또는 조사를 통해 얻은 결과로부터 **기존의 모수와 다르다고 주장**하고자 하는 가설

7.1 가설 검정의 이해

검정 통계량 (Test Statistic, T_0)

- 모수에 대한 가설 검정에서 의사결정의 도구로서 사용되는 통계량을 의미
- 표본으로부터 얻어진 정보, 즉, 표본들의 함수 (표본에 따라 변하는 확률변수)
- 검정 통계량을 T_0 으로 표시하는 이유는 귀무가설 기준의 모수의 값을 따르기 때문
- 귀무가설이 옳다는 가정 하에서 검정 통계량 값을 구하고 이렇게 구한 확률이 크면 귀무가설을 채택, 나타날 가능성이 적으면, 즉 확률이 낮으면 귀무가설을 기각하는 의사결정을 내리게 됨

7.1 가설 검정의 이해

유의수준과 기각역

확률이 크고 적음에 따라 귀무가설을 채택 또는 기각한다고 했는데, 어느 수준보다 크고 적은지에 대한 비교 기준이 필요함

이러한 기준을 유의 수준이라 하고 이는 분석자가 판단하여 결정함

7.1 가설 검정의 이해

유의수준 & 신뢰수준

1. 유의수준 (Significance Level, α)

: 귀무가설이 사실일 때, 이를 기각하는 오류를 범할 최대 허용 확률

2. 신뢰수준 (Confidence Level, $1 - \alpha$)

: 유의수준에서 허용하는 오류를 제외한 확률, 즉 (귀무가설을) 신뢰할만한 수준

: 신뢰수준 = $1 - \text{유의수준} = 1 - \alpha$

7.1 가설 검정의 이해

기각역 (Critical Region, C_α)

기준이 되는 유의수준 α 하에서 귀무가설을 기각하는 검정 통계량 값의 범위

기각역의 범위는 유의 수준의 크기에 따라 결정

(신뢰수준이 99%, 95%, 90%이면 이에 해당하는 유의수준은 1%, 5%, 10%)

7.1 가설 검정의 이해

유의 확률 (p-value)

검정 통계량 값에 대해 귀무가설을 기각할 수 있는 최소의 유의수준

유의 확률이 유의 수준보다 작으면 귀무가설을 기각

$p\text{-value} < \alpha$ 이면 귀무가설 기각

7.1 가설 검정의 이해

가설 검정 방법

1. 기각역을 이용

: 분포의 x축의 값의 크기를 통해 비교하는 방법

2. 유의 확률을 이용

: 검정 통계량 값이 가지는 확률의 크기를 비교해 결과를 판단하는 방법

7.1 가설 검정의 이해

검정 통계량과 유의 확률



검정 통계량 값은 말 그대로 하나의 값을 의미

유의확률은 빨간 빗금의 넓이를 의미

검정 통계량의 값이 커질수록 유의확률은 작아짐

유의확률이 작아지면 귀무가설은 기각될 확률이 높아짐

따라서 **검정 통계량 값이 커질수록 귀무가설은 기각될 확률이 높아짐**

7.2 한 개 집단에 대한 가설 설정

모평균 (μ) 의 검정 절차

어떤 모집단의 평균인 모평균 μ 가 특정 값인가에 대한 가설 검정 방식

가설의 설정	우측 단측 검정 : $H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu > \mu_0$ 좌측 단측 검정 : $H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu < \mu_0$ 양측 검정 : $H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu \neq \mu_0$
유의수준의 설정	일반적으로 1%, 5%, 10% 가 주로 사용
표본의 관측 및 검정 통계량 계산	특정 분포를 따르는 검정 통계량 값을 계산
기각역의 설정	분석자가 정한 유의 수준에 따라 기각역을 설정, 가설의 형태에 따라 분류
의사결정	검정 통계량 값이 유의 수준 하에서 기각역에 속하면 귀무가설을 기각

7.3 두 개 집단에 대한 가설 설정

모평균의 동일성 검정 절차

어떤 모집단의 평균인 모평균 μ 가 특정값인가에 대한 가설 검정 방식

가설의 설정	동일성 검정1 : $H_0 : \mu_x = \mu_y \text{ vs } H_1 : \mu_x > \mu_y$ 동일성 검정2 : $H_0 : \mu_x = \mu_y \text{ vs } H_1 : \mu_x < \mu_y$ 동일성 검정3 : $H_0 : \mu_x = \mu_y \text{ vs } H_1 : \mu_x \neq \mu_y$
유의수준의 설정	일반적으로 1%, 5%, 10% 가 주로 사용
표본의 관측 및 검정 통계량 계산	모분산을 알고 있는 경우와 그렇지 않은 경우 두가지로 구분 알고 있는 경우는 그대로 사용, 모르는 경우는 표본에서 추정하여 사용
기각역의 설정	분석자가 정한 유의 수준에 따라 기각역을 설정, 가설의 형태에 따라 분류
의사결정	검정 통계량 값이 유의 수준 하에서 기각역에 속하면 귀무가설을 기각