4. 검정의 종류와 과정

가설 검정 과정

가설 설정 → 표본 자료의 관측 → 가설 검정에 사용할 통계량 선택 (검정 통계량) →

관측값 계산 → 기각역/유의 확률 계산 → 검정 결과 해석 및 가설 채택

검정 통계량

가설 검정에 사용되는 통계량

가설 검정의 결과를 결정하는데 이용되는 표본의 함수

$\bar{X}$를 관측하여 그 값으로부터 μ에 대한

가설 검정을 결정할 때 검정통계량으로 사용

검정을 위한 기준

기각역

유의 확률
(P-value)

기각역

$\bar{X}$ 가 취하는 구간 중에서 H_0을 기각하는 구간

$R: \bar{X} \leq c$ 로 표현

$\bar{X}$가 c이하면 H_0을 기각한다고 판단

기각역의 올바른 선택이 검정의 가장 중요한 부분

바람직한 기각역은 두 오류를 범할 확률을 최소화하는 것

유의 확률

- 표본 자료가 대립 가설을 지지하는 정도를 0과 1사이의 숫자로 나타낸 최소의 유의 수준 값
- P-value 라고 부르기도 함
- 표준 정규 분포표를 이용해 P값을 구해야 함

유의 수준과 P값을 비교

유의수준 > P값인 경우	유의수준 < P값인 경우
H_0를 기각	H_0를 기각할 수 없음

가설검정종류

이항 검정

모평균
가설 검정

이항 검정

이항 분포를 이용하여 베르누이 확률 변수의 모수 p 에 대한 가설 조사

베르누이 값을 가지는 확률 변수의 분포를 판단

예) 어떤 동전을 던질 때, 앞면이 나올 확률 $p = 0.5$ 인
공정한 동전인지 알아보는 검정
귀무 가설 : $p = 0.5$ vs 대립 가설 : $p \neq 0.5$

이항 검정

```
scipy.stats.binom_test(x, n, p, alternative=‘’)

# 이항 검정의 유의 확률을 구해주는 함수
```

- x = 검정 통계량, 1이 나온 횟수
- n = 총 시도 횟수
- p = 모수 p 값
- 양측 검정: alternative = 'two-sided'
- 단측 검정: alternative = 'one-sided'

모평균 가설 검정

표본의 크기가 클 때, 모평균 μ 이 정규 분포를 따른다는 가정하에
중심 극한 정리에 의해 정규 분포에 근사함

가설 검정을 하기 위한 검정 통계량 $\bar{X}$ 를 표준화시키면

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

모평균 가설 검정

단측 검정

가설

$\mu = \mu_0$ VS $\mu < \mu_0$

기각역 $R : Z \leq -z_\alpha$

가설

$\mu = \mu_0$ VS $\mu > \mu_0$

기각역 $R : Z \geq -z_\alpha$

양측 검정

가설

$\mu = \mu_0$ VS $\mu \neq \mu_0$

기각역 $R : |Z| \geq z_{\frac{\alpha}{2}}$

모평균 가설 검정

```python
def ztest(stat, mu, sigma):
    z = (stat.mean() - mu) / (sigma*np.sqrt(len(stat)))
    return (2 * (1-sp.stats.norm.cdf(z)))
# 모평균 가설 검정 함수. 유의 확률 출력
```

- stat : 검정 통계량
- mu : 모평균
- sigma : 모표준 편차