인공지능 수학: 통계학
7.검정

윤상민

E-mail : smyoon@kookmin.ac.kr
Office : 02-910-4645

통계적가설검정

- 가설 검정
  - 예)
    - 어떤 고등학교의 1학년 학생들의 평균키가 170.5cm으로 알려져 있었다. 올해 새로 들어온 1학년 학생들 중 30명을 랜덤하게 선택하여 키를 잰 후 평균을 계산했더니 171.3cm이었다.
    - 올해 신입생은 평균키가 170.5cm보다 더 크다고 할 수 있는가?
  - 이러한 주장을 검증하는 것이 가설 검정임
  - 표본평균 $\bar{X}$가 μ_0보다 얼마나 커야 모평균 μ가 μ_0보다 크다고 할 수 있을 것인가?
    - 표본평균은 표본의 선택에 의해 달라짐에 주의

통계적가설검정

- 가설 검정
  - 귀무가설 $H_0: \mu = \mu_0$
  - 대립가설 $H_1: \mu > \mu_0$
  - 귀무가설을 기각하기 위해서는 $\bar{X}$가 좀 큰 값이 나와야 함.
    - 귀무가설이 참이라고 가정할 때, 랜덤하게 선택한 표본에서 지금의 $\bar{X}$가 나올 확률을 계산할 필요
    - 이 확률이 낮다면 귀무가설이 참이 아니라고 판단

통계적가설검정

- 가설 검정
  - 확률이 낮다는 기준점 필요
    - 유의수준 α 도입
  - $P(\bar{X} \geq k) \leq \alpha$가 되는 k를 찾아야 함
  - 표준정규확률변수로 변환=> 검정통계량이라고 함
    - $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$
    - $P(Z \geq z_\alpha) = \alpha$
  - 따라서 $\bar{X}$를 Z로 변환한 후 Z값이 z_α보다 큰지를 검토
    - 크다면 귀무가설 기각
    - 그렇지 않다면 귀무가설 채택

통계적가설검정

- 검정의 단계
  - H_0, H_1 설정
  - 유의수준 α 설정
  - 검정통계량 계산
  - 기각역 또는 임계값 계산
  - 주어진 데이터로부터 유의성 판정

모평균의 검정

- 대립가설
  - 문제에서 검정하고자 하는 것이 무엇인지 파악필요
    - 대립가설 H_1채택을 위한 통계적 증거 확보 필요
    - 증거가 없으면 귀무가설 H_0채택
    - $H_1: \mu > \mu_0$
    - $H_1: \mu < \mu_0$
    - $H_1: \mu \neq \mu_0$

모평균의 검정

- 대립가설
  - 어떤 농장에서 생산되는 계란의 평균 무게는 10.5그램으로 알려져 있다. 새로운 사료를 도입한 후에 생산된 계란 30개의 표본 평균을 계산했더니 11.4그램이 나왔다. 새로운 사료가 평균적으로 더 무거운 계란을 생산한다고 할 수 있는가?
    - $H_0: \mu = 10.5$
    - $H_1: \mu > 10.5$

모평균의 검정

- 대립가설
  - 어떤 농장에서 생산되는 계란의 평균 무게는 10.5그램으로 알려져 있다. 새로운 사료를 도입한 후에 생산된 계란 30개의 표본 평균을 계산했더니 9.4그램이 나왔다. 새로운 사료가 평균적으로 더 가벼운 계란을 생산한다고 할 수 있는가?
    - $H_0: \mu = 10.5$
    - $H_1: \mu < 10.5$

모평균의 검정

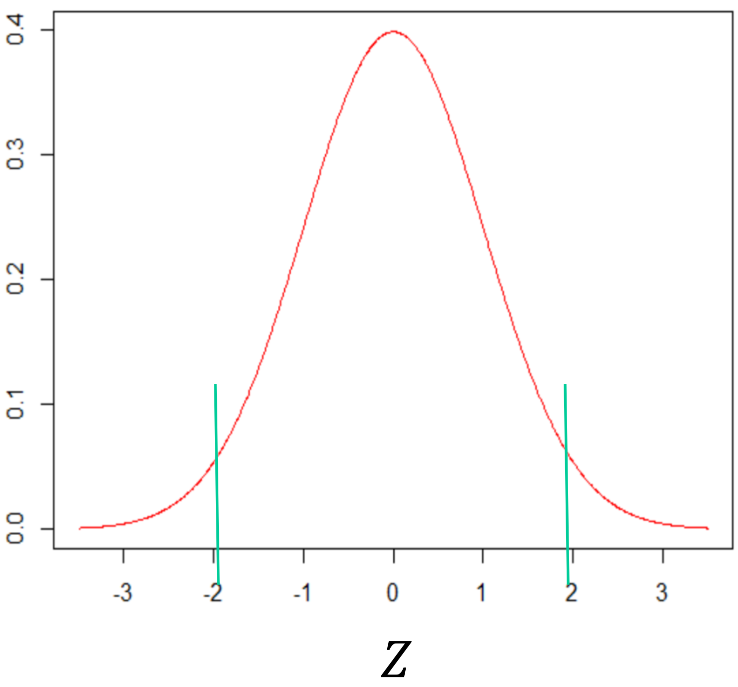
- 대립가설
  - 어떤 농장에서 자신들이 생산하는 계란의 평균 무게가 10.5그램이라고 홍보하고 있다. 이에 생산된 계란 30개의 표본 평균을 계산했더니 9.4그램이 나왔다. 이 농장의 광고가 맞다고 할 수 있나?
    - $H_0: \mu = 10.5$
    - $H_1: \mu \neq 10.5$

모평균의 검정

- 검정통계량
  - $n \geq 30$인 경우
    - 중심극한 정리 사용
    - $Z = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim N(0,1)$
  - 모집단이 정규 모집단이고, 모표준편차 σ가 주어진 경우
    - $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
  - 기타의 경우는 본 강의의 범위를 벗어남

모평균의 검정

- 기각역
  - $H_0: \mu = 10.5$
  - 유의수준: α
  - 기각역
    - $H_1: \mu > 10.5 \Rightarrow Z > z_\alpha$
    - $H_1: \mu < 10.5 \Rightarrow Z < -z_\alpha$
    - $H_1: \mu \neq 10.5 \Rightarrow |Z| > z_{\frac{\alpha}{2}}$



모평균의 검정

- 검정의 예
  - 어떤 농장에서 자신들이 생산하는 계란의 평균 무게가 10.5그램이라고 홍보하고 있다.
  - 이에 생산된 계란 30개의 표본을 뽑았더니 그 무게가 아래와 같다.
    - w=[10.7, 11.7, 9.8, 11.4, 10.8, 9.9, 10.1, 8.8, 12.2, 11.0, 11.3, 11.1, 10.3, 10.0, 9.9, 11.1, 11.7, 11.5, 9.1, 10.3, 8.6, 12.1, 10.0, 13.0, 9.2, 9.8, 9.3, 9.4, 9.6, 9.2]
  - 이 농장의 홍보가 맞는지 유의수준 5%로 검정하시오.

모평균의 검정

- 검정의 예
  - $H_0: \mu = 10.5$
  - $H_1: \mu \neq 10.5$
  - $\alpha = 0.05$
  - $\bar{X} = 10.43$
  - $s = 1.11$
  - $Z = \frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{10.43-10.5}{1.11/\sqrt{30}} = -0.351$
  - $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96 > |-0.351| = |Z|$
  - 따라서 귀무가설을 기각할 수 없다.

모평균의 검정

- 검정의 예
  - import numpy as np
  - w=[10.7, 11.7, 9.8, 11.4, 10.8, 9.9, 10.1, 8.8, 12.2, 11.0, 11.3, 11.1, 10.3, 10.0, 9.9, 11.1, 11.7, 11.5, 9.1, 10.3, 8.6, 12.1, 10.0, 13.0, 9.2, 9.8, 9.3, 9.4, 9.6, 9.2]
  - mu=10.5
  - xbar=np.mean(w)
  - sd=np.std(w, ddof=1)
  - print("평균 %.2f, 표준편차: %.2f" %(xbar, sd))
  - z=(xbar-mu)/(sd/np.sqrt(len(w)))
  - print("검정통계량: ", z)
  - alpha=0.05
  - import scipy.stats
  - cri = scipy.stats.norm.ppf(1-alpha/2)
  - print("임계값: ", cri)

모평균의 검정

- 검정의 예
  - import numpy as np
  - w=[10.2, 11.2, 9.3, 10.9, 10.3, 9.4, 9.6, 8.3, 11.7, 10.5, 10.8, 10.6, 9.8, 9.5, 9.4, 10.6, 11.2, 11.0, 8.6, 9.8, 8.1, 11.6, 9.5, 12.5, 8.7, 9.3, 8.8, 8.9, 9.1, 8.7]
  - mu=10.5
  - xbar=np.mean(w)
  - sd=np.std(w, ddof=1)
  - print("평균 %.2f, 표준편차: %.2f" %(xbar, sd))
  - z=(xbar-mu)/(sd/np.sqrt(len(w)))
  - print("검정통계량: ", z)
  - alpha=0.05
  - import scipy.stats
  - cri = scipy.stats.norm.ppf(1-alpha/2)
  - print("임계값: ", cri)

W의 모든 값에서 0.5를 뺌

모평균의 검정

- 검정의 예
  - 평균 9.93, 표준편차: 1.11
  - 검정통계량: -2.8129403330804132
  - 임계값: 1.959963984540054

  - 귀무가설 기각