제6장 모집단에 대한 추론

추리통계학

- 추리통계학이란?
  - 표본의 특성을 분석하여 모집단의 특성을 파악하는 학문
  - 통계량을 통하여 모수(parameter)를 추측하는 추론(inference)의 과정

- 모수 추정의 방법
  - 추정(Estimation)
    - 점추정 (point estimation): 모수의 값을 하나의 값으로 추정
    - 구간추정 (interval estimation) : 모수의 값은 속하는 범위를 추론

- 가설 검정 (Hypothesis test)
  - 가설: 관심있는 모수에 대한 진술 또는 주장
    - 예) 모집단의 평균은 5이다.
  - 가설 검정: 가설에 대한 통계적으로 채택이나 기각하는 판단

모집단 평균의 추정

- 통계적 추정의 의의
  - 전수 검사의 비용 절감
    - 대통령 선거결과 예측을 위해 일부 유권자를 대상으로 하는 지지후보를 조사하는 경우
  - 전수 검사가 불가능한 경우
    - 검사로 인해 대상물의 속성이 변경되는 경우
      - 예) 품질검사: 형광등, 타이어의 수명조사
    - 무한 모집단
  - 시간적 제약

- 추정치와 추정량
  - 추정치 (estimation): 추정하여 나온 결과 값
  - 추정량 (estimator) : 추정하기 위해 쓰인 변수

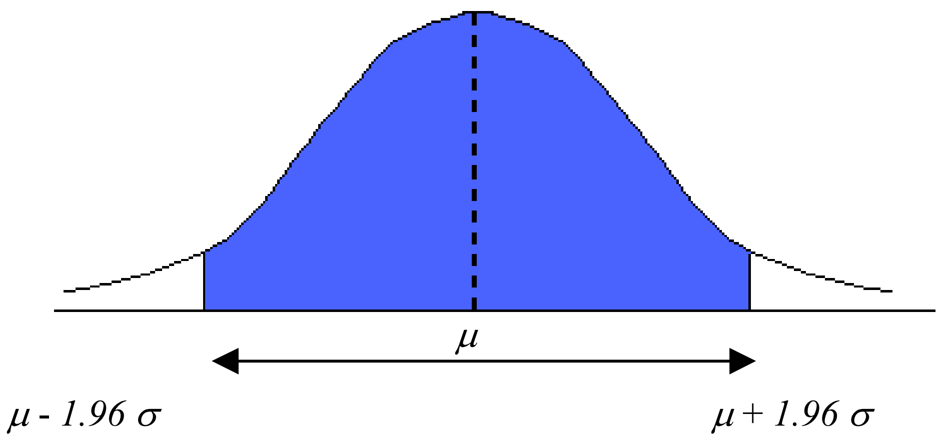
모집단 평균의 추정

- 모집단의 모수 추정방법
  - 점추정: 모수의 진정한 값으로 추측되는 단일 값을 예측
  - 구간 추정: 모수의 진정한 값이 포함되는 것을 기대하는 측정치의 범위를 예측

- 점추정
- 구간추정
  - 신뢰구간 (Confidence Interval) : 일정 구간내에 모수가 포함될 가능성이 있는 구간
  - 신뢰구간에 모수가 있을 가능성과 그 구간을 함께 제시

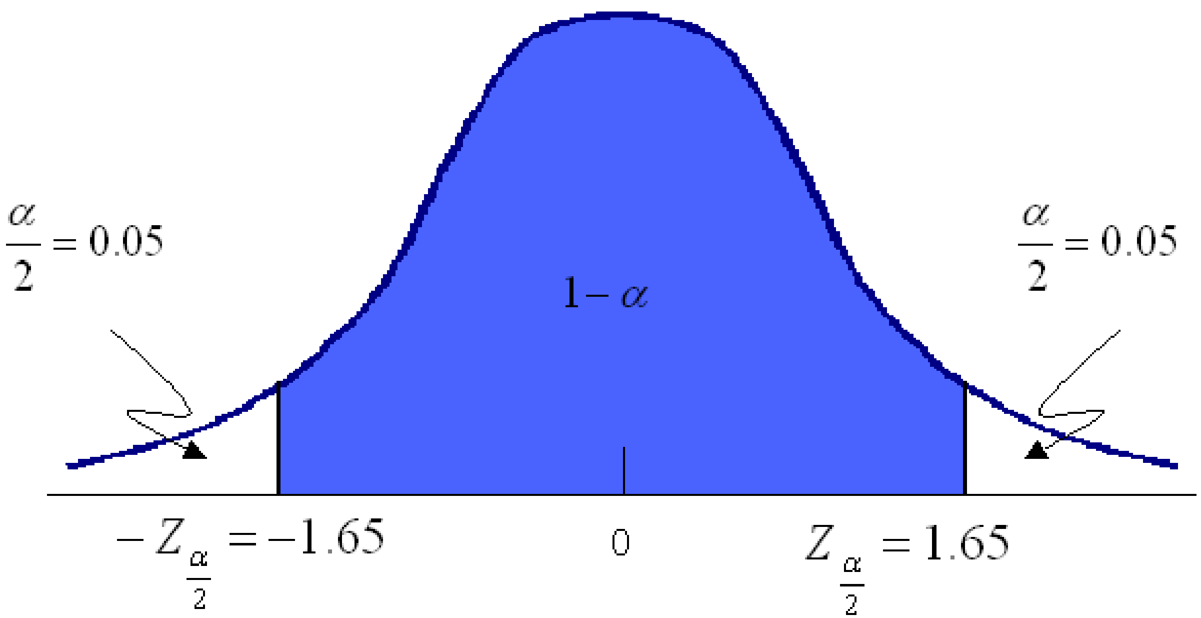
모집단 평균의 추정

- **모평균의 신뢰구간**
  - **95% 신뢰수준이 널리 사용됨**
  - μ를 중심으로 한 95% 신뢰구간



모집단 평균의 추정

- **모집단분산 σ^2를 알고 있는 경우**
  - **표본분포 $X \sim N(\mu, \sigma^2/n)$**
  - $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$



모집단 평균의 추정

- **모집단분산 σ^2 를 알고 있는 경우 – 대표본(大標本)**
  - **σ^2 대신 S^2 를 사용**

- **모집단분산 σ^2 를 알고 있는 경우 – 소표본(小標本)**
  - **표본크기가 n인 경우는 다음의 분포를 따름**
  - $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
  - **t 분포의 모양과 특성**
    - **오른쪽 꼬리 분포**
    - **n이 커지면 정규분포에 근사, n>=30**

모비율의 신뢰구간

- **비율의 표본분포의 평균과 표준오차**
  - $\boldsymbol{X_p = p}$
  - $\boldsymbol{S_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}}, \quad \boldsymbol{p + q = 1}$

- **표본크기가 모집단 크기의 5%이상이고 비복원추출인 경우**
  - $\boldsymbol{S = \sqrt{\frac{pq}{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$, $\boldsymbol{N}$은 모집단의 크기, $\boldsymbol{n}$은 표본의 크기

가설검정

- **가설: 실증적인 증명 이전에 잠정적으로 세우는 모집단 특성에 대한 진술**
- **귀무가설 (Null hypothesis): 표본추출 오차여부에 대한 검정대상이 되는 가설, H_0으로 표시**
- **연구가설 (Research hypothesis): 논리적 대안으로서 귀무가설이 기각될 때 채택되는 가설, H_1 으로 표시**
- **통조림 무게에 대한 가설 (통상적인 무게가 400g)**
  - **H_0: 통조림의 평균 무게는 400g이다.**
  - **H_1: 통조림의 평균 무게는 400g이 아니다.**

가설검정

- **가설검정의 오류**

실제상태 / 의사결정	진실한 H_0	거짓된 H_0
H_0 채택	올바른 결정	제2종 오류: β오류
H_0 기각	제1종 오류: α오류	올바른 결정 (Power)

  - **P(제1종 오류) = P(H_0 기각 | H_0 진실) = α**
  - **P(제2종 오류) = P(H_0 채택 | H_0 거짓) = β**

  - **검정력(Power): H_0 가 거짓일때 이것을 기각하는 확률**
    - **Power = 1- β**

가설검정

- **가설검정의 종류**

	양측검정	단측검정	
		왼쪽꼬리검정	오른쪽꼬리검정
일반적인 경우	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$
통조림 무게에 대한 예제	$H_0 : \mu = 400$ $H_1 : \mu \neq 400$	$H_0 : \mu \geq 400$ $H_1 : \mu < 400$	$H_0 : \mu \leq 400$ $H_1 : \mu > 400$

가설검정

- **가설검정의 순서**
  1. **귀무가설과 연구가설의 설정**
  2. **유의수준과 임계치 설정**
  3. $\boldsymbol{H_0}$ **의 채택영역과 기각영역의 결정**
  4. **통계량의 계산**
  5. **통계량과 임계치의 비교 및 결론**

모평균 가설검정 분류

- 모분산을 알고 있는 경우 VS 모분산을 모르는 경우
  - 모분산을 아는 경우
    - 표본분산 대신 모분산을 사용해야 함
  - 모분산을 모르는 경우
    - 표본의 분산을 사용할 수 밖에 없음

- 검정의 종류
  - 양측 검정 : 기각역이 중심값에서 양쪽으로 (아주 크거나 또는 아주 작게) 있는 경우의 검정
    - 중심에서 일정한 범위(신뢰구간)내에 있으면 귀무가설 기각할 수 없음
  - 단측검정: 기각역이 중심값에서 한쪽으로 있는 경우의 검정

모비율의 가설검정

- **표본비율 p의 표본 분포와 평균과 표준오차**
    - $\boldsymbol{\mu_p} = \pi$
    - $\boldsymbol{\sigma_p} = \sqrt{\frac{\boldsymbol{\pi(1-\pi)}}{\boldsymbol{n}}}$
    - $\boldsymbol{Z} = \frac{\boldsymbol{p-\pi}}{\boldsymbol{\sigma_p}} = \frac{\boldsymbol{p-\pi}}{\sqrt{\frac{\boldsymbol{\pi(1-\pi)}}{\boldsymbol{n}}}}$