8강

추정 I

통계·데이터과학과 이긍희 교수

한국방송통신대학교

목차

01

추정의 이해

통계적 추론

통계적 추론

- 표본의 관측값(데이터)으로 모집단의 모수를 추측, 판단하는 방법 → 통계적 사고

통계적 추론 : 추정(추측)과 검정(판단)

- 추정 : 점추정, 구간추정
- 추정의 예 : 여론조사
- 검정의 예 : 비타민 C의 감기예방효과

모집단과 표본

- 모집단(population)

  알고 싶은 모든 개체의 집합

- 표본(sample) : 모집단의 일부
  - 임의 추출해서 모집단을 대표

확률변수와 데이터

- **확률변수(random variable)**

  모집단 특성 나타내는 변수 → 불확실(확률분포)

- **확률변수와 데이터(관측값)**

확률분포

- **확률분포** : 몇 개 모수를 가지는 수리적 함수
- **확률변수의 분포 아는 것**

  → 모수(parameter)를 아는 것
- **확률변수의 모수**
  - 이항분포 : 모비율 $B(n, p)$
  - 정규분포 : 모평균과 모분산 $N(\mu, \sigma^2)$

통계적 추정

- 표본(확률변수)으로 모집단의 모수를 추측
  - 모수는 고정된 상수이지만 알 수 없다고 가정
- 통계량(statistic) : 표본 확률변수의 함수
  - 모수를 추정하기 위한 도구 → 추정량
  - 표본분포(sampling distribution) : 통계량의 분포
  - 추정값 : 추정량(수식)에 관측값(데이터)를 대입

통계적 추정

- 모집단에서 표본 : $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$
  - 관측값 : $x_1, x_2, \cdots, x_n$

- 통계량(추정량) : $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$

모수	추정량	추정값	표본분포
모평균	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\bar{x} = 170.1\mathrm{cm}$	정규분포
모분산	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = 2.54$	χ^2분포

통계적 추정

- 점추정(point estimation)

- 구간추정(interval estimation)
  - $100(1-\alpha)\%$ 신뢰구간

$$P(L < \theta < U) = 1 - \alpha$$

02

모평균의 추정

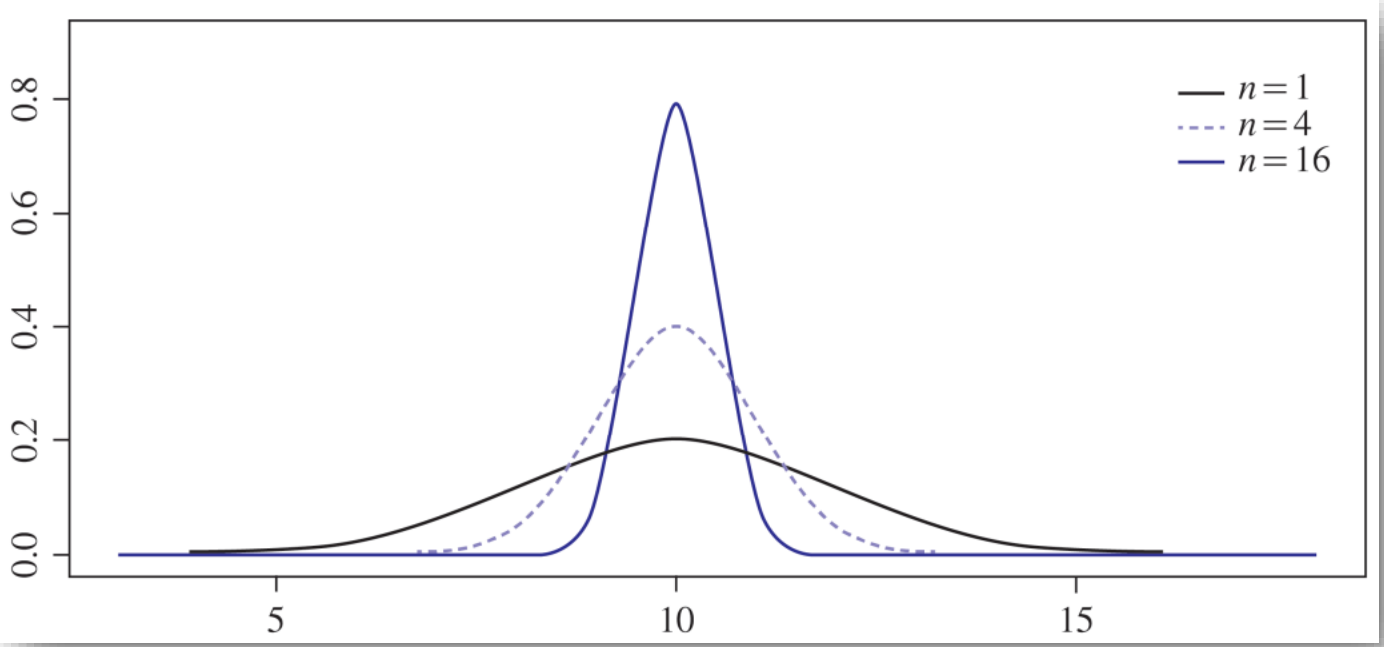
표본평균

- 모집단에서 표본 : $X_1, X_2, \cdots, X_n$
  - $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ → 표본분포

- 표본의 관측값 : $x_1, x_2, \cdots, x_n$
  - $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ → 조사할 때마다 결과가 다름

표본평균의 특성

- 표본평균은 모평균의 불편추정량 : $E(\bar{X}) = \mu$
- 표본평균의 분산은 표본수가 커지면서 작아짐
  - $Var(\bar{X}) = \frac{\sigma^2}{n}$



표본평균의 특성

- **중심극한정리 :** 표본평균은 표본 수가 커지면서 모집단 분포와 관계없이 정규분포를 따름
  - $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$

모평균의 추정 예제

- 어느 대학교 학생 16명을 임의 추출해서 시험 실시, 시험점수가 정규분포 따를 때 어느 대학교 전체학생의 평균(모평균)을 추정하라.

88, 83, 83, 85, 94, 88, 91, 96,
89, 83, 81, 80, 84, 89, 83, 79

03

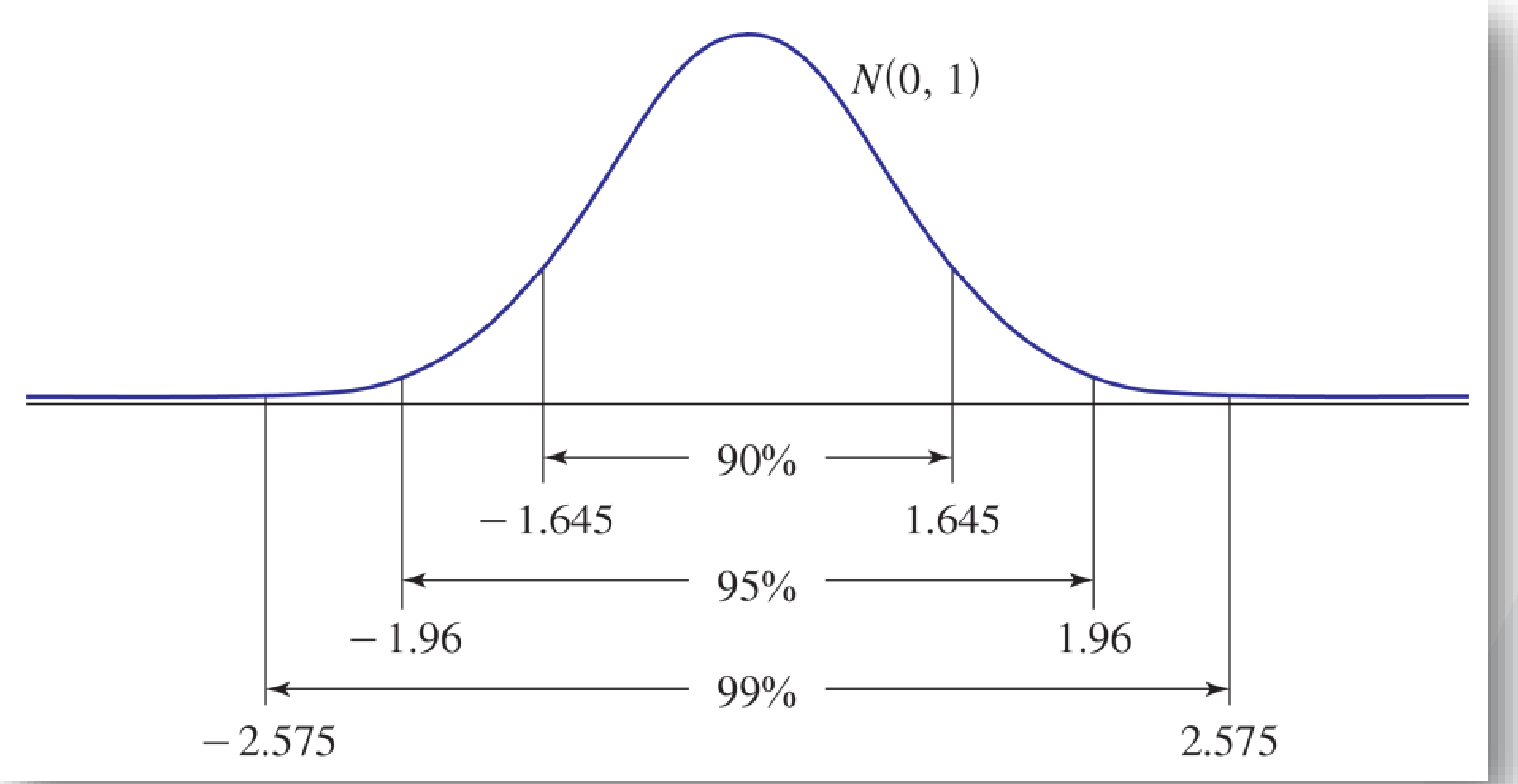
모평균의 구간추정

표본평균의 분포

- $X_1, X_2, \cdots, X_n \sim (\mu, \sigma^2)$
  - 중심극한정리 : $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$
  - $\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{d} N(0, 1)$
  - $P\left(-1.96 < \dfrac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$

모평균의 신뢰구간

- $P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$



모평균의 신뢰구간

- $P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$
- μ의 95% 신뢰구간

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \qquad \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

모평균의 신뢰구간

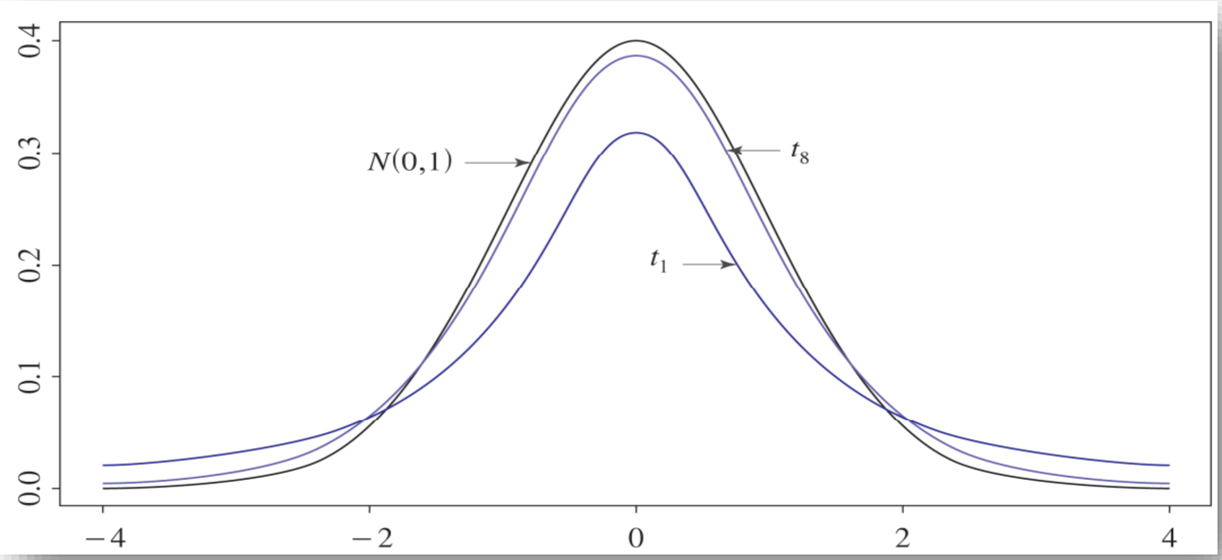
- μ의 $100(1-\alpha)\%$ 신뢰구간

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

모평균의 신뢰구간

- σ를 모를 때 표본표준편차 S를 이용
- $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$$



모평균의 신뢰구간

- μ의 $100(1-\alpha)\%$ 신뢰구간

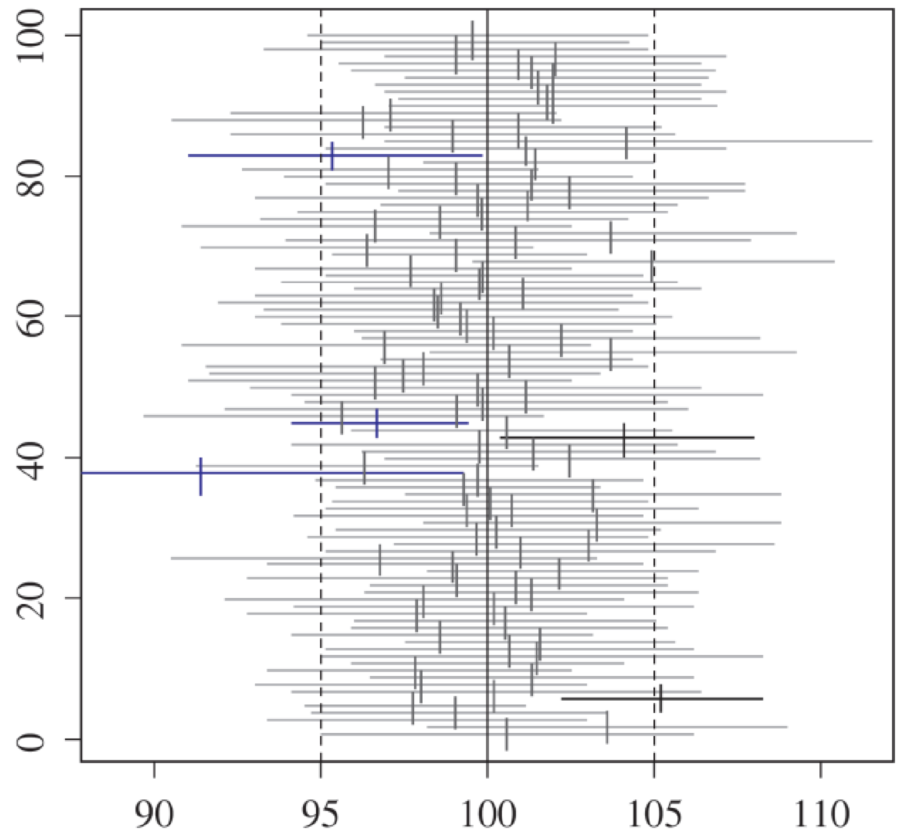
$$\left[\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \qquad \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]$$

신뢰수준과 신뢰구간

- $100(1-\alpha)\%$: 신뢰수준
  - 신뢰수준이 클수록 신뢰구간에 모평균 포함될 가능성 증가
  - 표본수 증가하면 $t_{n-1,\alpha/2}$가 $z_{\alpha/2}$에 근접

신뢰구간의 의미

- **신뢰수준 :** 무한히 반복해서 구한 신뢰구간중 모평균이 포함된 비율



구간추정의 예 1

- 전체학생의 통계학 점수의 평균에 대한 95% 신뢰구간을 구하라.

구간추정의 예 1

- 전체학생의 통계학 점수의 평균에 대한 99% 신뢰구간을 구하고, 95% 신뢰구간과 비교하라.

구간추정의 예 1

- 표본수가 25명으로 증가되었을 때 전체학생의 통계학 점수 평균에 대한 95% 신뢰구간을 구하라.

구간추정의 예 2

- 어느 학교 학생 11명 대상으로 읽은 책 수를 조사. 모집단은 정규분포 가정. 읽은 평균 책 수의 95% 신뢰구간은?

8, 1, 10, 15, 15, 10, 5, 19, 20, 9, 10

구간추정의 예 2

- 어느 학교 학생 11명 대상으로 읽은 책 수를 조사. 모집단은 정규분포 가정. 읽은 평균 책 수의 95% 신뢰구간은?

정리하기

- 통계적 추론은 모집단에서 표본을 추출한 후 구한 통계량으로 모집단과 관련된 모수를 추측하는 과정이다
- 점추정은 모수에 가장 근사한 값으로 추측하는 것인데 모평균은 표본평균으로 추정한다.
- 구간추정은 모수를 포함할 것으로 기대되는 구간을 제시하여 모수를 추정하는 것이다.
- $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$ 일 때 μ의 $100(1-\alpha)\%$ 신뢰구간

$$\left[\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \quad \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]$$

다음시간 안내

9강

추정 II

한국방송통신대학교