01 강

통계적 추론

통계적 추론의 개요와 확률의 개념

한국방송통신대학교 통계·데이터과학과
이긍희 교수

학습내용

1. 통계적 추론의 기초를 이해한다.
2. 통계학의 역사를 이해한다.
3. 확률의 개념을 이해한다.
4. 조건부 확률을 이해한다.

01

통계적 추론의 기초 개념

1 통계적 추론의 개요

통계학

- (기술 통계학) 관심대상으로부터 데이터를 수집, 요약
- (추론 통계학) 데이터로부터 일반성을 찾아내고,
  이를 근거로 불확실한 사실에 대한 결론 및 규칙성 도출
  → 통계적 추론

1 통계적 추론의 개요

통계적 추론의 출발점

- **관심 대상은 불확실**

  → 불확실성은 확률로 표현

- **관심대상을 모두 측정할 수 없음**

  → 일부를 측정해서 관심대상 전체를 추론

2 통계적 추론의 용어

모집단과 표본

- **모집단(population) : 관심 대상 전체**
- **표본 (sample) : 모집단의 일부 → 확률표본**

2 통계적 추론의 용어

확률변수와 데이터

- **확률변수(random variable) : 사건을 실수로 바꾸어 주는 함수**
- **데이터 : 관측값**

2 통계적 추론의 용어

확률분포

- **확률변수는 확률분포를 따름**
- **확률분포 : 몇 개의 모수(parameter)를 가진 수학함수로 가정**

2 통계적 추론의 용어

통계량과 표본분포

- **통계량 : 표본의 함수**

  (예) 표본평균, 표본분산

- **표본분포 : 통계량의 확률분포**

3 통계적 추론의 구조

통계적 추론 : 추정과 검정

추정

- 수많은 꽃씨 중 흰색 꽃씨의 비율
- 호수에 사는 A 어종의 수
- 지지율 조사

검정

- 동전은 평평한가?
- 신약은 효과 있는가?

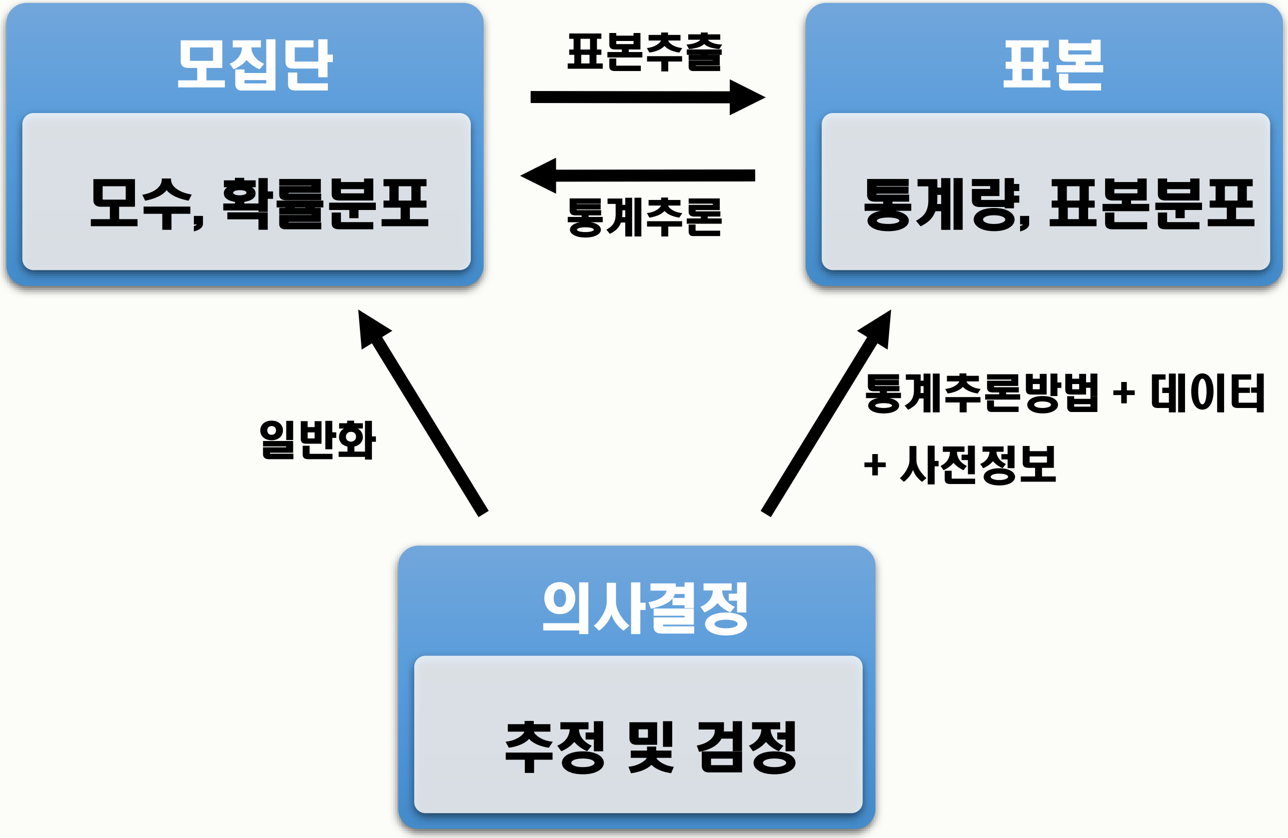
3 통계적 추론의 구조

통계적 추론의 원리

- **가장 가능성 높은 결론 도출 한다.**
- **가능성 낮은 일을 믿지 않는다.**

3 통계적 추론의 구조

통계적 추론의 과정



3 통계적 추론의 구조

통계적 추론의 과정

- $X \sim f(x|\theta)$

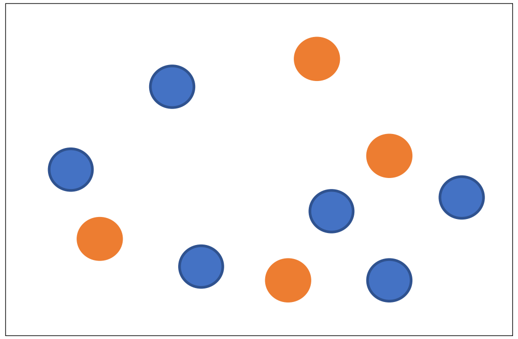
  【예】 $N(\mu, \sigma^2)$
- **확률표본** : $X_1, X_2, \cdots, X_n \sim f(x|\theta)$
- **통계량 (추정량) 도출**

  【예】 $N(\mu, \sigma^2)$: μ를 위한 추정량 $\bar{X}$
- **통계량을 바탕으로 추정 및 검정**

4 통계적 추론의 구분

통계적 추론 이론과 데이터분석

통계적 추론 이론



- 내용을 알고 있는 상자에서 무작위로(randomly) n개의 공을 뽑았을 때, 빨간 공이 x개 나올 확률은?

데이터분석

- 내용을 모르는 상자에서 n개를 뽑았을 때 이 중 x개가 빨간 공이라면 이 상자에는 빨간 공이 몇 %일까?

4 통계적 추론의 구분

- 통계적 추론
  - **이론적 부분 : 연역적 추론**
  - **데이터 분석 : 귀납적 추론**

4 통계적 추론의 구분

통계적 추론의 구성

- **확률 이론 : 객관적 확률, 주관적 확률**
- **추론 이론 : 빈도론적 추론, 베이즈 추론**

02

통계학의 역사

1 통계학의 역사

통계학의 역사

- **확률의 시대 : 17세기 ~ 18세기**

  - B. Pascal(1663 ~ 1662), Jacob Bernoulli(1654 ~ 1705)
  A. de Moivre(1667 ~ 1754), T. Bayes(1701 ~ 1761)

- **오차이론의 시대 : 18세기 중 ~19세기 중**

  -P.–S. Laplace(1749 ~ 1827), C. F. Gauss(1777 ~ 1855)

- **통계의 시대 : 19세기 중, 후**

  -A. Quetelet(1796 ~ 1874), F. Galton(1822 ~ 1911)

1 통계학의 역사

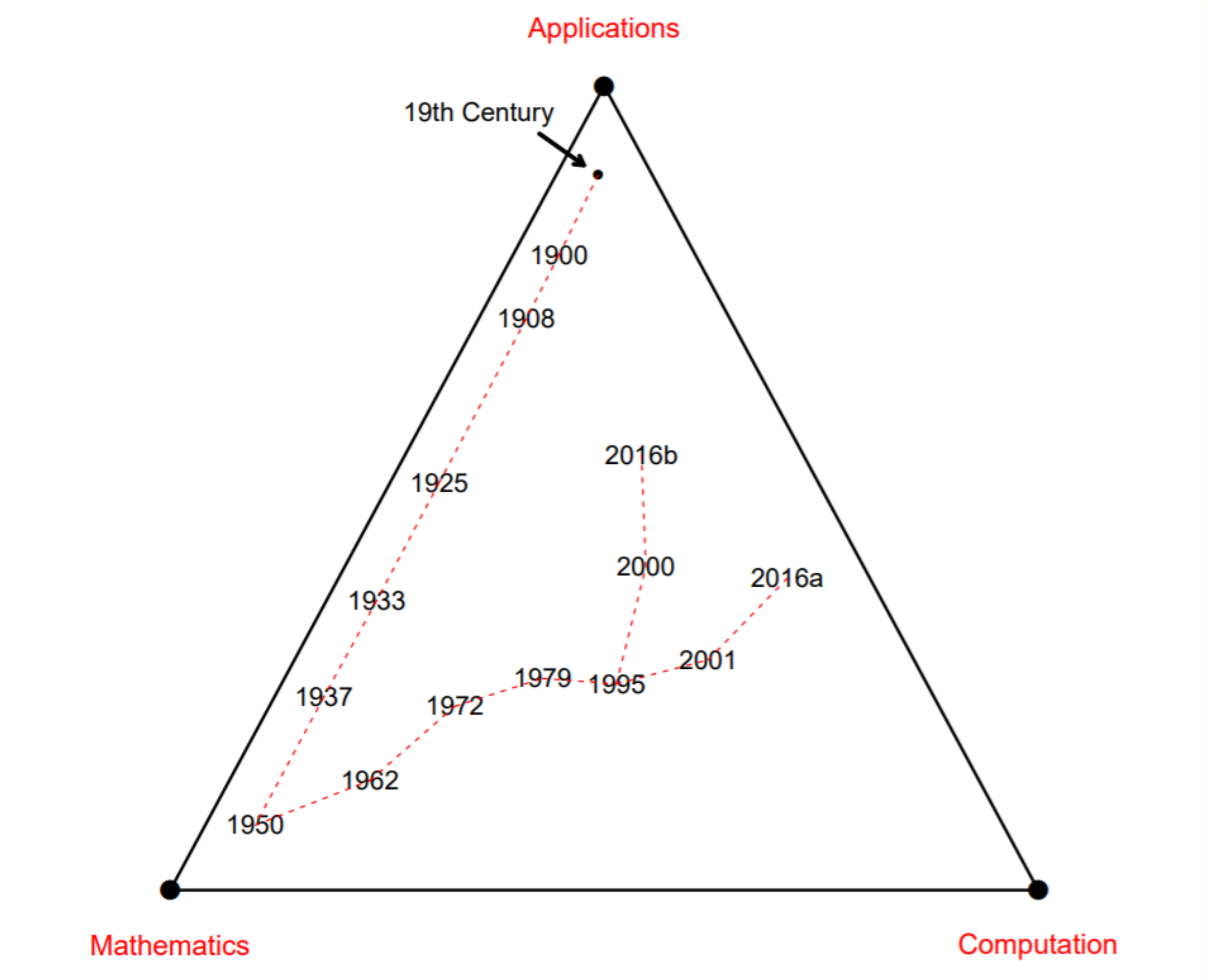
통계학의 역사

- **통계적 추론의 시대 : 20세기 초 중**
  - Karl Pearson(1857 ~ 1936), R.A. Fisher(1890 ~ 1962)
  - W.S. Gosset(1876 ~ 1937), J. Neyman(1894 ~ 1981)

- **데이터과학의 시대 : 21세기**

1 통계학의 역사

통계학의 역사



Efron and Hastie (2016)

03

확률의 개념

1 확률의 정의

확률적 실험

- 어떤 실험이 반복될 때 개개의 실험 결과는 미리 알 수 없으나 반복과정에서 "규칙성"을 지니는 실험

표본공간과 사건

- 표본공간(sample space) : 확률적 실험을 통해 일어날 수 있는 모든 가능한 결과의 집합
- 사건(event) : 표본공간의 부분집합

1 확률의 정의

확률의 정의

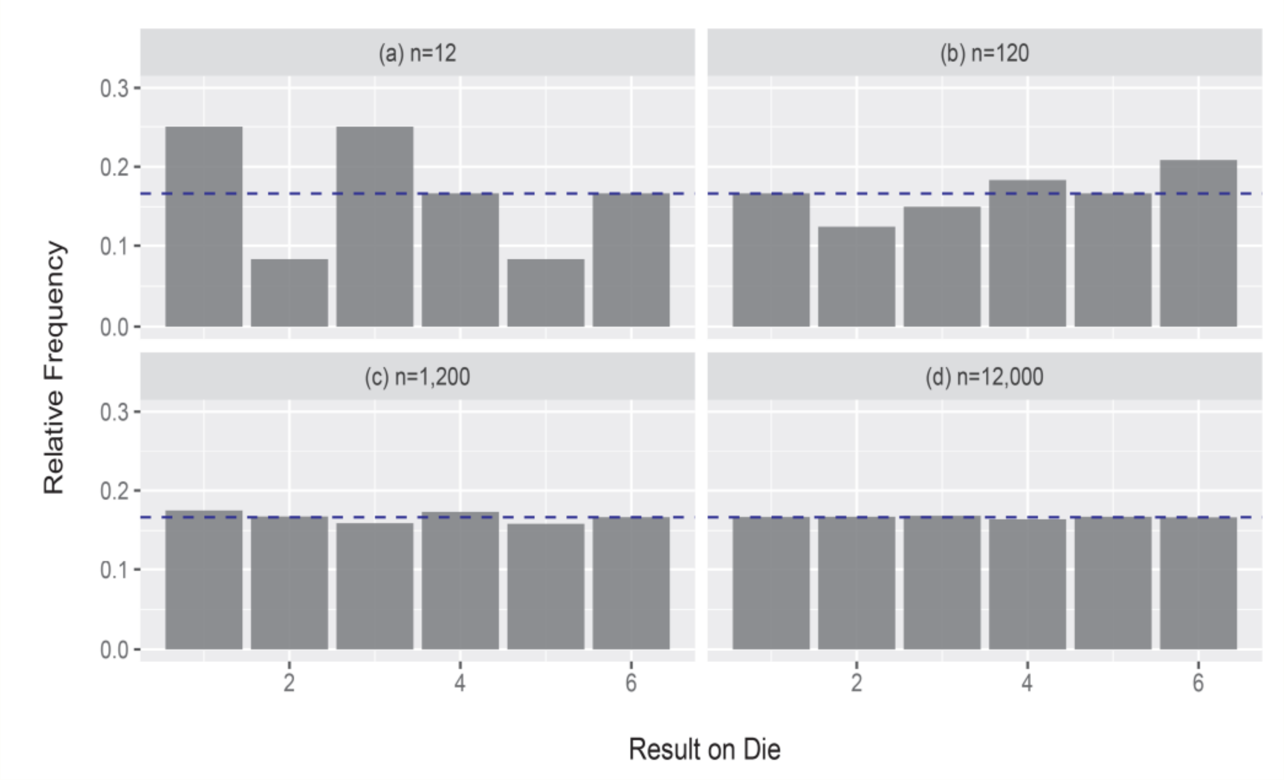
- **어떤 사건이 일어날 가능성을 0과 1 사이의 실수로 표현**

빈도론적 확률

- $P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$

1 확률의 정의

예 주사위를 12회, 120회, 1,200회, 12,000회 던진 결과를 R 프로그램을 이용하여 시뮬레이션하시오.



1 확률의 정의

고전적 확률

- $P(A) = \dfrac{n(A)}{n(S)}$

1 확률의 정의

예 **주사위 눈이 3 이하의 숫자가 나올 확률은?**

1 확률의 정의

- 공리적 확률 : 다음의 공리를 만족하는 P
  - $0 \leq P(A) \leq 1$
  - $P(S) = 1$
  - $A_1, A_2, \ldots$ 서로 배반 사건, $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

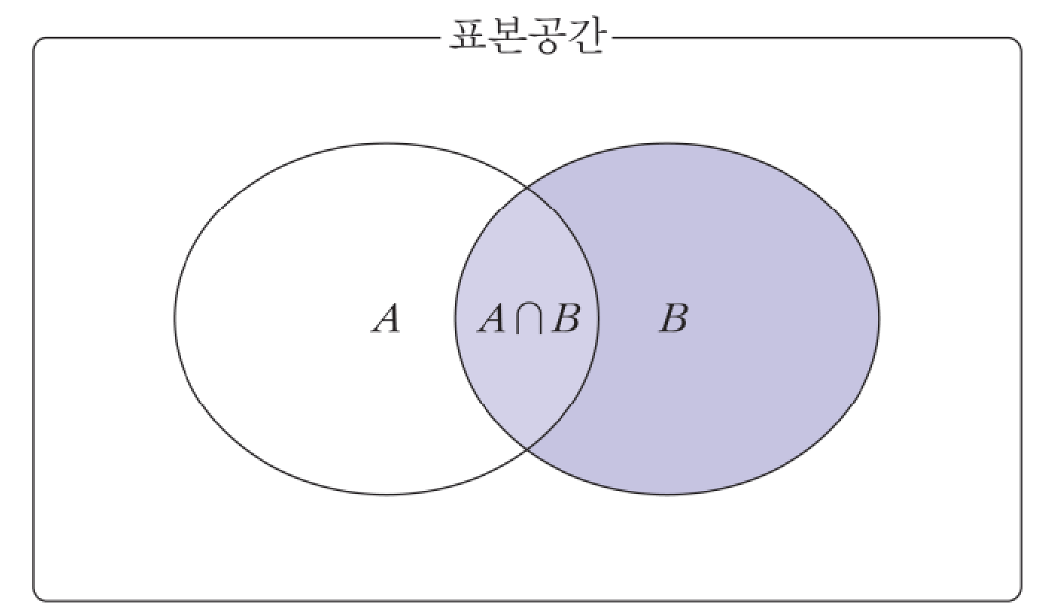
1 확률의 정의

확률의 계산

- $P(A^C) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

2 조건부확률

- 조건부 확률 : 사건 B 조건 하에 사건 A 발생 확률
  - $\mathrm{P(A|B)} = \frac{P(A\cap B)}{P(B)} \Leftrightarrow P(A \cap B) = P(A|B)P(B)$



2 조건부확률

예 주사위 눈이 짝수라는 조건 하에 주사위 눈이 3 이하의 숫자가 나올 확률은?

2 조건부확률

조건부 확률의 성질

- $P(\mathrm{B}|\mathrm{A}) \geq 0$
- $P(S|A) = 1$
- $S = \cup_{i=1}^{\infty} B_i$, B_i 배반 $\Rightarrow P(\cup_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A)$

2 조건부확률

역확률

- $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

2 조건부확률

베이즈 정리

- $S = B \cup B^C$
- $P(A) = P(A|B)P(B) + P(A|B^C)P(B^C)$

2 조건부확률

베이즈 정리

- $P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^C)P(B^C)}$

2 조건부확률

베이즈 정리

- $S = \cup_{i=1}^{k} B_i$, B_i 는 배반사건
- $P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$
- $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}$

2 조건부확률

예 전체 인구의 10%가 어떤 질병을 앓고 있다. 이 질병의 진단시약을 조사한 결과, 질병에 걸린 사람 중 90%는 양성 반응, 질병에 걸리지 않은 사람 중 80%는 음성 반응. 어떤 사람의 진단 시약 검사 결과가 양성 반응일 때 이 사람이 질병에 걸렸을 확률은?

2 조건부확률

예 어떤 사람의 진단 시약 검사 결과가 양성 반응일 때 이 사람이 질병에 걸렸을 확률은?

2 조건부확률

독립

- **두 사건 A, B 간 독립**

$$P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$$

2 조건부확률

예 $P(A) = P(B) = 0.5$일 때 $P(A \cup B)$? (A.B독립)

정리하기

- 통계추론은 데이터를 기반으로 불확실한 사실에 대한 결론이나 예측을 하는 데 필요한 이론과 방법에 관한 학문 분야로 통계학의 중심이다.
- 확률변수는 확률적 실험에서 실험결과를 관심의 대상이 되는 수 값으로 나타낸 것이다.
- 확률은 어떤 사건이 일어날 가능성을 0과 1사이의 실수로 표현한 것이다. 확률은 빈도론적, 고전적, 공리적으로 정의된다.

정리하기

- □ **조건부 확률 :** $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- □ **베이즈정리 :** $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}$, $S = \cup_{i=1}^{k} B_i$, B_i **배반사건**
- □ **독립성 :** $P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$

02 강

다음시간 안내

확률분포 1