9강

추정 II

통계·데이터과학과 이긍희 교수

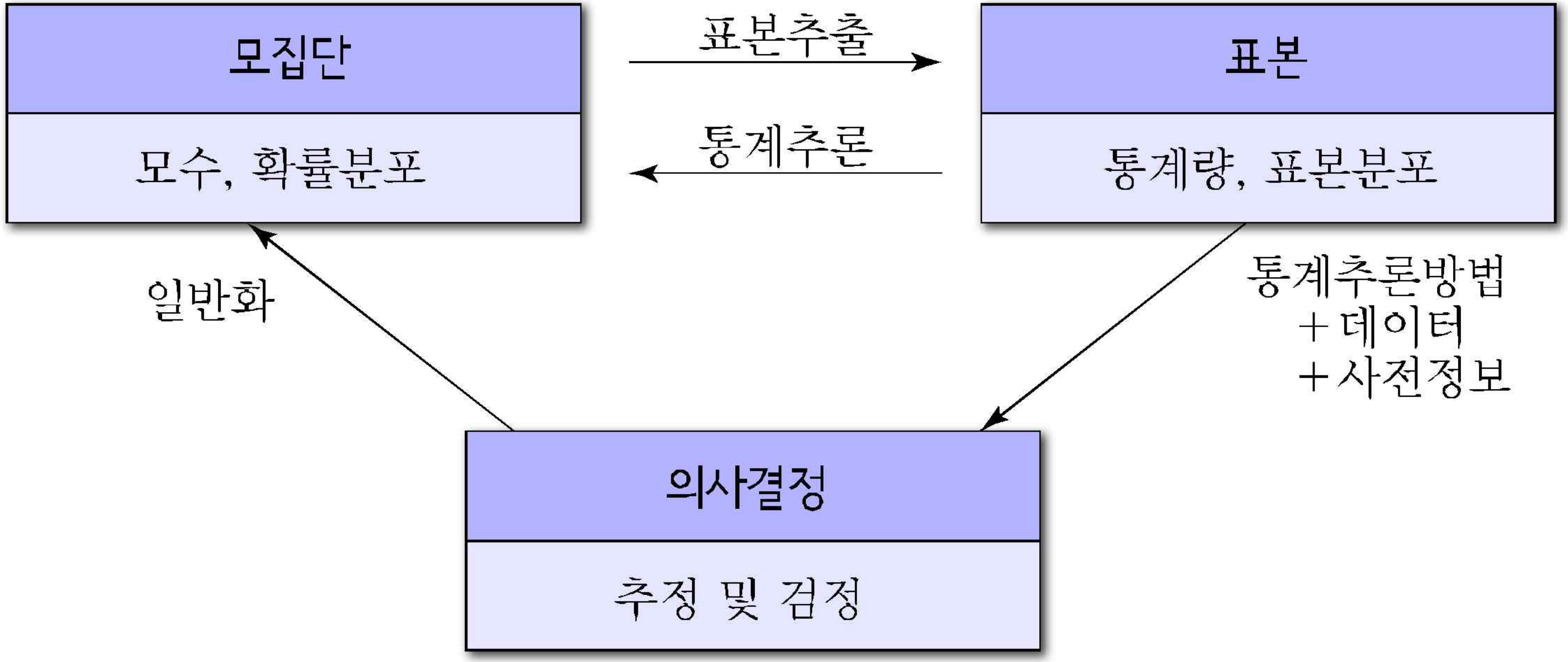
목차

01

추정과 통계량

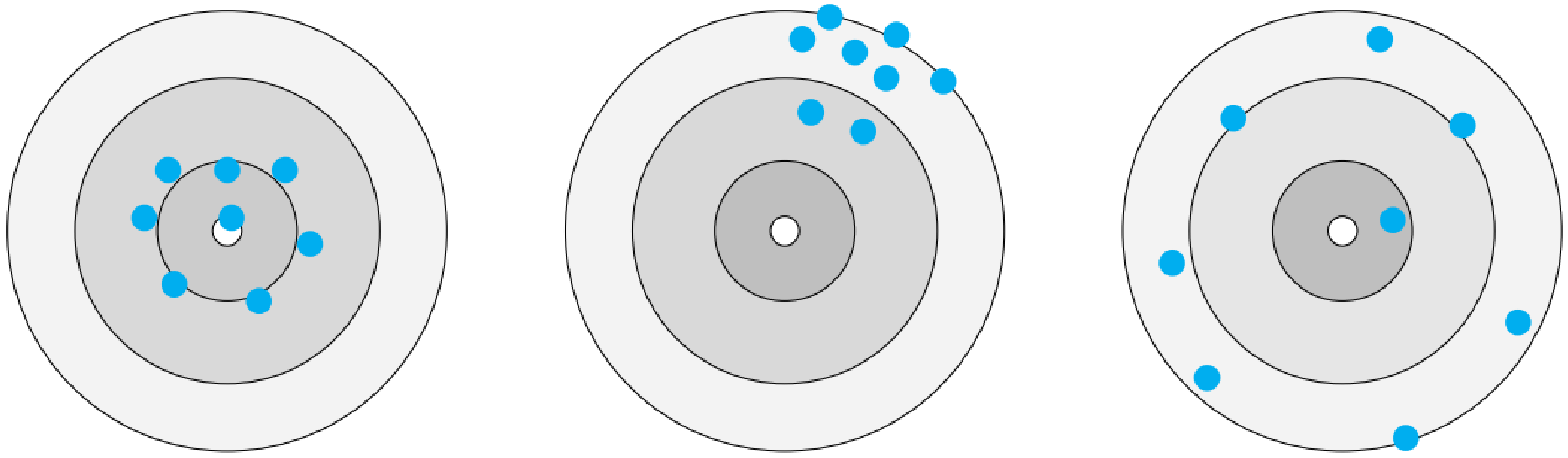
통계적 추정의 개요

통계적 추정



바람직한 통계량

- 바람직한 통계량 : $\bar{X},\ S^2,\ \hat{p}$
  - 불편성, 일치성, 효율성



(a) A 사수 (b) B 사수 (c) C 사수

02

모비율의 추정

모비율과 표본비율

- **비율 추정이 중요 :** 지지율, 불량률, 실업률

- **모집단이 2개 배반사건(찬성, 반대)으로 구성**

$$X_1, X_2, \cdots, X_n \sim B(1, p)$$

$\rightarrow$ n개 표본 중 찬성자 수 $X = \sum_{i=1}^{n} X_i \sim B(n, p)$

모비율과 표본비율

- 표본비율 : 모집단의 찬성비율 추정

$$\hat{p} = \frac{X}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

표본비율의 특성

- 표본비율은 모비율의 불편추정량 : $E(\hat{p}) = p$
- 표본비율의 분산 : $Var(\hat{p}) = \frac{p(1-p)}{n}$

$$\rightarrow \widehat{Var}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}$$

모비율의 특성

- 중심극한정리 :

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, p(1-p))$$

모비율의 신뢰구간

- $$P\left(-z_{\frac{\alpha}{2}} \le \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

모비율의 신뢰구간

- 모비율의 $100(1-\alpha)\%$ 신뢰구간

$$\left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

모비율의 추정

- 부동산정책 : 500명 조사, 200명이 찬성. 찬성비율의 추정값을 구하고 95% 신뢰구간을 구하라.

모비율의 추정

- 부동산정책 : 500명 조사, 200명이 찬성. 찬성비율의 추정값을 구하고 95% 신뢰구간을 구하라.

03

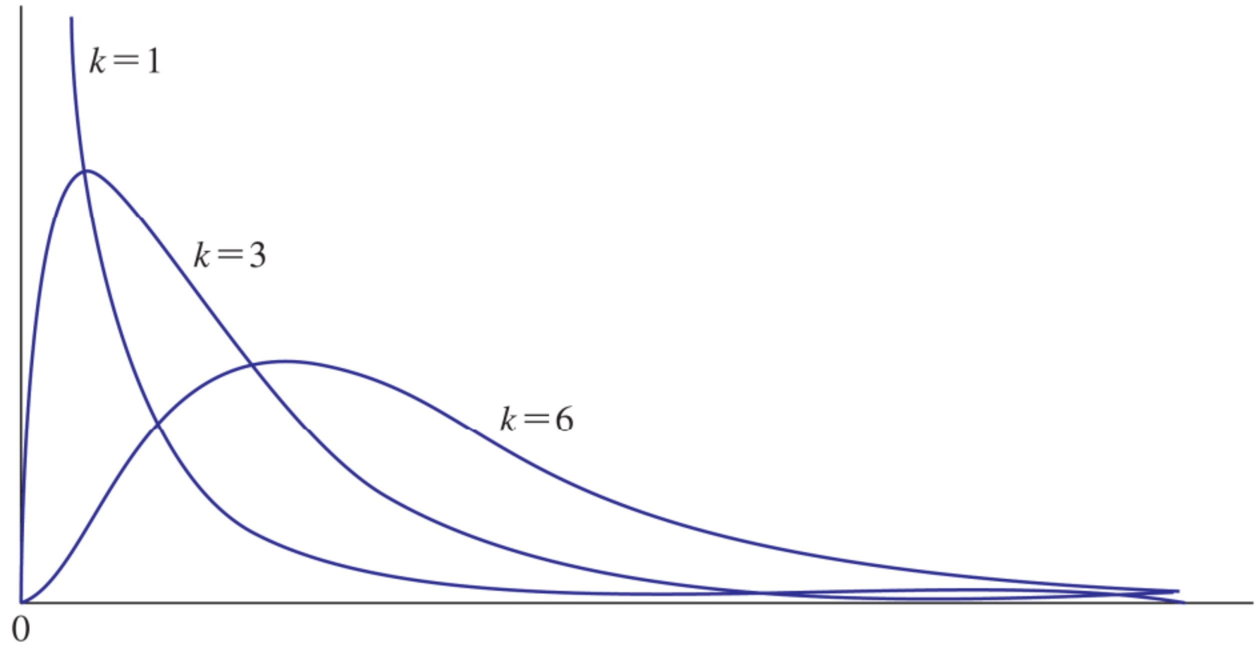
모분산의 추정

표본분산

- 모분산의 분산 추정 : 제품의 품질, 금융시장 변동성 등 파악에 이용
- 표본분산 : $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - $E(S^2) = \sigma^2$

표본분산의 분포

- $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$
- χ^2_k 분포



표본분산의 신뢰구간

- $$P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1,\frac{\alpha}{2}}\right) = 1-\alpha$$

표본분산의 신뢰구간

- σ^2의 $100(1-\alpha)\%$ 신뢰구간

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}},\quad \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right]$$

표본분산의 신뢰구간

- 40명 단순복원 추출, 표본표준편차 4점. 모분산을 점추정하고, 모분산의 95% 신뢰구간을 구하라.

표본분산의 신뢰구간

- 40명 단순복원 추출, 표본표준편차 4점. 모분산을 점추정하고, 모분산의 95% 신뢰구간을 구하라.

04

R을 이용한 실습

모평균의 신뢰구간 추정


```
> # 데이터 입력
> score = c(88, 83, 83, 85, 94, 88, 91, 96, 89, 83, 81, 80, 84, 89, 83, 79)
>
> # 표본평균과 표본표준편차
> bar_x = mean(score); s = sd(score); n = length(score)
>
> # 모평균의 95% 신뢰구간
> qt(0.975,15)
[1] 2.13145
> qt(0.025,15, lower.tail = FALSE)
[1] 2.13145
>
> qt_95 = qt(0.975,15)
> c(bar_x - qt_95*s/sqrt(n), bar_x + qt_95*s/sqrt(n))
[1] 83.36785 88.63215
>
> t.test(score)$conf.int
[1] 83.36785 88.63215
attr(,"conf.level")
[1] 0.95
```

모비율의 신뢰구간 추정


```
> # 모비율의 신뢰구간
> n = 500; X = 200
> p_hat = X/n
> alpha = 0.05
> z_1 = qnorm(1-alpha/2)
> c(p_hat-z_1*sqrt(p_hat*(1-p_hat)/n), p_hat+z_1*sqrt(p_hat*(1-p_hat)/n))
[1] 0.3570593 0.4429407
> prop.test(X, n)$conf.int
[1] 0.3570044 0.4445558
attr(,"conf.level")
[1] 0.95
```

모분산의 신뢰구간 추정


```
> # 모분산의 신뢰구간
> s2 = 4^2
> n = 40
> alpha = 0.05
> q_1 = qchisq(1-alpha/2, n-1)
> q_2 = qchisq(  alpha/2, n-1)
> c((n-1)*s2/q_1, (n-1)*s2/q_2)
[1] 10.73640 26.37995
```

정리하기

- 불편성, 일치성, 효율성을 가지는 통계량을 이용하는 것이 바람직하다.
- 모비율의 $100(1-\alpha)\%$ 신뢰구간 : $[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$
- 모분산의 $100(1-\alpha)\%$ 신뢰구간 : $[\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}]$

다음시간 안내

10강

가설검정 I