13 강

통계적추론

가설검정 2

한국방송통신대학교 통계·데이터과학과
이긍희 교수

학습내용

1. 가능도비 검정을 이해한다.
2. 유의확률을 이용한 검정을 이해한다.
3. 적합도 검정을 이해한다.
4. 독립성 검정을 이해한다.

01

가능도비 검정

1 최강력 검정

최강력 검정

- 단순가설 $H_0 : \theta = \theta_0 \;\; vs \;\; H_1 : \theta = \theta_1$ 에 대한 최강력 검정
  - 두 결합확률밀도함수의 비에 의하여 결정
- 복합가설 : 최강력검정을 사용 제약

2 가능도비 검정

가능도비 검정(Likelihood Ratio Test)

- 가설 $H_0 : \theta \in \Omega_0 \quad vs \quad H_1 : \theta \in \Omega_1$
- 최대가능도비 : $\lambda(X) = \dfrac{\sup_{\theta\in\Omega_1} L(\theta|X)}{\sup_{\theta\in\Omega_0} L(\theta|X)} = \dfrac{L(\hat{\theta}_1)}{L(\hat{\theta}_0)}$
- 검정법 $\delta : \lambda(X) > C \Rightarrow H_0$ 기각

2 가능도비 검정

가능도비 검정(Likelihood Ratio Test)

- 가설 $H_0 : \theta \in \Omega_0 \quad vs \quad H_1 : \theta \in \Omega_1$
- 최대가능도비 : $\lambda^*(X) = \dfrac{\sup_{\theta \in \Omega} L(\theta|X)}{\sup_{\theta \in \Omega_0} L(\theta|X)}$
- 검정법 $\delta : \lambda^*(X) > C' \Rightarrow H_0$ 기각

2 가능도비 검정

예 $X_1, \cdots, X_n \sim N(\theta, 1)$ 확률표본. 가능도비 검정은?

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

2 가능도비 검정

예 $X_1, \cdots, X_n \sim N(\theta, 1)$ 확률표본. 가능도비 검정은?

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

2 가능도비 검정

예 $X_1, \cdots, X_n \sim N(\theta, 1)$ 확률표본. 가능도비 검정은?

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

2 가능도비 검정

예 $X_1, \cdots, X_n \sim N(\theta, 1)$ 확률표본. 가능도비 검정은?

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

2 가능도비 검정

예 $X_1, \cdots, X_n \sim N(\theta, \sigma^2)$ 확률표본, σ^2미지. 가능도비 검정은?

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

2 가능도비 검정

예 $X_1, \cdots, X_n \sim N(\theta, \sigma^2)$ 확률표본, σ^2미지. 가능도비 검정은?

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

2 가능도비 검정

예 $X_1, \cdots, X_n \sim N(\theta, \sigma^2)$ 확률표본, σ^2미지. 가능도비 검정은?

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

2 가능도비 검정

예 $X_1, \cdots, X_n \sim N(\theta, \sigma^2)$ 확률표본, σ^2미지. 가능도비 검정은?

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

3 가능도비 검정통계량

검정통계량

- 가능도비 검정통계량은 충분통계량에 의존

3 가능도비 검정통계량

검정통계량

- 유의수준 α 가능도비 검정의 정확한 기각역
  - 귀무가설이 참일 때 가능도비 $\frac{f(X|\hat{\theta}_0)}{f(X|\hat{\theta})}$의 확률분포를 알아야함

3 가능도비 검정통계량

검정통계량

- 귀무가설이 참일 때 로그변환 가능도비 근사적으로 카이제곱 (χ^2) 분포 따름

$$-2\log\frac{f(X|\hat{\theta}_0)}{f(X|\hat{\theta})} \sim \chi^2(df)$$

02

유의성 검정

1 네이만-피어슨의 검정

네이만-피어슨의 검정

- 귀무가설, 대립가설에 대하여 제1종 오류를 범할 확률과 제2종 오류를 범할 확률에 기초한 방법
- 주어진 제1종 오류를 범할 확률 α에 대하여 대립가설을 고려하여 최적의 기각역을 구하여 검정

1 네이만-피어슨의 검정

- 정규분포 모평균의 가설검정

2 유의성 검정

유의성 검정

- 주어진 귀무가설에 대한 p-값에 바탕으로 한 검정
- p –값 : 귀무가설 하에서 주어진 관측값보다 더 극단적인 값을 얻을 확률 → 귀무가설에 대한 반대 증거

2 유의성 검정

- 정규분포 모평균의 가설검정

2 유의성 검정

예 $X_1, \cdots, X_n \sim N(\theta, 25)$, 전년 통계학 점수 평균 70점 올해 9명 조사 결과, 평균 75점. 올해 평균이 전년보다 올랐는지 유의수준 5%에서 검정.

2 유의성 검정

예 $X_1, \cdots, X_n \sim N(\theta, 25)$, 전년 통계학 점수 평균 70점 올해 9명 조사 결과, 평균 75점. 올해 평균이 전년보다 올랐는지 유의수준 5%에서 검정.

03

카이제곱 검정

1 적합도 검정

분할표

- m개 범주 빈도수 : $N_1, N_2, \dots, N_m$

	범주1	범주2	…	범주m	합계
X	N_1	N_2	…	N_m	n
확률	p_1	p_2	…	p_m	1

- 빈도수의 분포 : 다항분포

$$f(n_1, \cdots, n_m | p_1, \cdots, p_m) = \frac{n!}{n_1! \cdots n_m!} p_1^{n_1} \cdots p_m^{n_m}$$

1 적합도 검정

귀무가설과 대립가설

- $H_0: (p_1, \cdots, p_m) \in \Omega_0 \;\; vs \;\; H_1: (p_1, \cdots, p_m) \in \Omega_1$
  - $H_0: p_i = p_{i0} \;\; vs \;\; H_1: \text{not } H_0$
- 모수 전체 영역 : p_i의 최대가능도추정량 $\hat{p}_i = \frac{N_i}{n}$

1 적합도 검정

최대가능도비

- 최대가능도비 : $\dfrac{f(n_1,\cdots,n_m|\hat{p}_{10},\cdots,\hat{p}_{m0})}{f(n_1,\cdots,n_m|\hat{p}_1,\cdots,\hat{p}_m)} = \left(\dfrac{\hat{p}_{10}}{\hat{p}_1}\right)^{n_1}\cdots\left(\dfrac{\hat{p}_{m0}}{\hat{p}_m}\right)^{n_m}$
- $-2\log(\text{최대가능도}) = 2\sum_{i=1}^{m} n_i \log\left(\dfrac{n_i}{n\hat{p}_{i0}}\right)$

1 적합도 검정

분포와 기각역

- 유의수준 α 가능도비검정의 기각역 :

$$2\sum_{i=1}^{m} N_i \log\left(\frac{N_i}{n\hat{p}_{i0}}\right) > \chi^2_{\alpha}(d.f.)$$

1 적합도 검정

- 테일러 급수를 이용한 근사
  - 테일러 급수를 이용한 근사 : $\sum_{i=1}^{m} \frac{(N_i - n\hat{p}_{io})^2}{n\hat{p}_{io}} > \chi_{\alpha}^2(d.f.)$

    - N_i : i 범주 빈도수

    - $n\hat{p}_{i0}$: 귀무가설이 참일 때 기대빈도수

1 적합도 검정

테일러 급수를 이용한 근사

- 테일러 급수를 이용한 근사 : $\sum_{i=1}^{m} \frac{(N_i - n\hat{p}_{io})^2}{n\hat{p}_{io}} > \chi_{\alpha}^2(d.f.)$

  - N_i: i 범주 빈도수

  - $n\hat{p}_{i0}$: 귀무가설이 참일 때 기대빈도수

1 적합도 검정

피어슨의 적합도 검정

- $H_0 : p_i = p_{i0} \;\; vs \;\; H_1 : \text{not } H_0$
  - N_i : i 범주 빈도수, np_{i0} :기대빈도수
- 검정통계량 $\chi^2 = \sum_{i=1}^{m} \frac{(N_i - np_{i0})^2}{np_{i0}} \sim \chi^2(m-1)$,
  - $np_{i0} \geq 5$ 인 경우 사용 가능
- 검정법 : $\chi^2 > C$ 면 H_0 기각, $\mathrm{P}(\chi^2 > C | H_0) = \alpha$

1 적합도 검정

예 A 상품을 100개를 추출하여 조사한 결과 불량품이 17개. 불량률이 10%인지 여부를 5% 유의수준에서 검정하시오.

1 적합도 검정

예 A 상품을 100개를 추출하여 조사한 결과 불량품이 17개. 불량률이 10%인지 여부를 5% 유의수준에서 검정하시오.

2 독립성 검정

분할표

		Y				합계
		범주 1	범주 2	…	범주 c	
X	범주 1	N_{11}	N_{12}	…	N_{1c}	N_{1+}
	범주 2	N_{21}	N_{22}	…	N_{2c}	N_{2+}
	⋮	⋮	⋮	⋮	⋮	⋮
	범주 r	N_{r1}	N_{r2}	…	N_{rc}	N_{r+}
합계		N_{+1}	N_{+2}	…	N_{+c}	n

2 독립성 검정

분할표

- $r \times c$ 분할표, 각 셀의 확률 p_{ij}
  - $p_{i+} = \sum_j p_{ij}$, $p_{+j} = \sum_i p_{ij}$,
  - $N_{i+} = \sum_j N_{ij}$, $N_{+j} = \sum_i N_{ij}$, $n = \sum_i \sum_j N_{ij}$

2 독립성 검정

독립성 검정

- 귀무가설 : 행변수와 열변수가 서로 독립

  $\Leftrightarrow H_0 : p_{ij} = p_{i+} \cdot p_{+j},\ H_1 : \sim H_0$

- $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(N_{ij} - \hat{B}_{ij})^2}{\hat{B}_{ij}} \sim \chi^2\big((r-1)(c-1)\big)$,

  $\hat{B}_{ij} = n \cdot \hat{p}_{i+} \cdot \hat{p}_{+j} = \frac{N_{i+} \cdot N_{+j}}{n}$

- $\chi^2 > \chi^2_\alpha\big((r-1)(c-1)\big)$이면 H_0 기각

정리하기

- 가능도비 검정은 최대가능도 추정량을 이용한 검정법이다.
  - 가설 $H_0 : \theta \in \Omega_0 \;\; vs \;\; H_1 : \theta \in \Omega_1$
  - 최대가능도비 : $\lambda(X) = \dfrac{\sup_{\theta \in \Omega} L(\theta|X)}{\sup_{\theta \in \Omega_0} L(\theta|X)}$
  - 검정법 $\delta : \lambda(X) > C \;\; \Rightarrow \;\; H_0$ 기각
- 유의성 검정은 유의확률로 주어진 관측값이 이 귀무가설에 얼마나 부합하는지 알아보는 검정이다.

정리하기

- 적합성 검정은 다음과 같다.

  - $H_0 : p_i = p_{i0} \quad vs \quad H_1 : not\, H_0$

  - 검정통계량 $\chi^2 = \sum_{i=1}^{m} \frac{(N_i - np_{i0})^2}{np_{i0}} \sim \chi^2(m-1)$

    p_i : i 번째 유형확률, N_i: i 범주 빈도수, $np_{i0} \geq 5$

  - 검정법 : $\chi^2 > \chi^2_\alpha(m-1)$이면 H_0 기각

정리하기

- 독립성 검정은 다음과 같다.

  - $H_0 : p_{ij} = p_{i+} \cdot p_{+j} \quad vs \quad H_1 : not\, H_0$

    $p_{i+} = \sum_{j=1}^{c} p_{ij}, \quad p_{+j} = \sum_{i=1}^{r} p_{ij}$

  - $\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(N_{ij} - \grave{E}_{ij}\right)^2}{\grave{E}_{ij}} \sim \chi^2\left((r-1)(c-1)\right)$

    $\grave{E}_{ij} = n \cdot \hat{p}_{i+} \cdot \hat{p}_{+j} = \frac{N_{i+} \cdot N_{+j}}{N}$, $N_{ij} : (i,j)$ 범주 빈도수

  - 검정법 : $\chi^2 > \chi_{\alpha}^2\left((\mathrm{r}-1)(\mathrm{c}-1)\right)$이면 H_0 기각

14 강

구간추정

다음시간 안내