Lecture 09 - Hypothesis Testing 1

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr

국립 서울과학기술대학교 IT정책전문대학원

1. I. 유의수준, 1종/2종 오류, 양측/단측 검정
2. II. 단일모집단의 모평균에 대한 검정
3. III. 단일모집단의 모비율에 대한 검정
4. IV. 두개의 모집단의 모평균에 대한 검정

I. 유의수준, 1종/2종 오류, 양측/단측 검정

유의수준 (α, significance level)

- 귀무가설과 대립가설이 $[H_0 : \mu = 30 \text{ vs } H_A : \mu \neq 30]$와 같은 가설검정을 생각해보자.
- 실제로 $\mu = 30$이 진실인데 $\mu \neq 30$로 결론내리기도 하고, $\mu \neq 30$이 진실인데 $\mu = 30$로 결론내리기도 한다. 이를 각각 1종오류와 2종오류라고 한다.

- 1종 오류
  - H_0가 참인데 H_0를 기각하는 오류
  - H_0가 참인데도 불구하고, 우연히 H_0을 지지하지 않는 실험결과가 나와서 H_0를 기각하는 오류
  - $\mu = 30$인데, $\mu \neq 30$이라고 결론내리는 오류
  - 무죄인데도 우연히 일어난 정황 증거들의 결합으로 인해 유죄를 선고받는 오류
- 2종 오류
  - H_0가 거짓인데 H_0를 채택하는 오류
  - H_0가 거짓인데도 불구하고, H_0를 기각할 수 있는 증거가 부족하여 H_0를 채택하는 오류
  - $\mu \neq 30$인데, $\mu = 30$이라고 결론 내리는 오류
  - 유죄인데도 증거가 부족해서 무죄를 선고받는 오류

- 유의수준의 의미
  - 유의수준 α는 1종 오류를 일으킬 확률을 α 이하로 제한함을 의미한다.
  - 유의수준 $\alpha = 0.05$이면, 가설검정에서 귀무가설을 잘못 기각할 가능성이 5% 임을 인정하는 것이다.
  - 유의수준 $\alpha = 0.05$라면 신뢰수준 95% 에 대응된다.
  - 가설검정에 있어서, 귀무가설이 옳을 확률(이를 **유의확률**이라고 한다)이 α보다 클 경우에는 H_0를 채택하고 α보다 작을 경우에는 H_0를 기각하고 H_A를 채택한다.

	H_0 is True	H_0 is False
Accept H_0	Correct Decision	Type II Error
Reject H_0	Type I Error	Correct Decision

- 1종 오류의 확률(α)와 2종 오류의 확률(β)은 역의 관계에 있다.
  - 1종 오류를 범할 확률이 낮게 설정하면, 2종 오류를 범할 확률이 높아진다.
  - 1종 오류를 범할 확률을 높게 설정하면, 2종 오류를 범할 확률이 낮아진다.
  - α와 β는 역의 관계에 있지만, 그렇다고 $\alpha + \beta = 1$인 것은 아니다.

유의수준 설정의 예시 - 암진단

상황

- 암의 진단에 있어서도 1종오류(정상인데 암으로 판단)과 2종오류(암인데 정상으로 판단) 두 가지의 오류가능성이 있다.
  - 1종오류는 불필요한 수술을 하게되는 불편을 야기한다.
  - 2종오류는 치료기회를 놓쳐서 환자의 생명을 위협할 수 있다.
- 진단을 해야하는 의사는 어느 수준의 α를 설정해야 할 것인가?

접근

- 만약에 1종오류로 인한 환자의 피해를 x, 2종오류로 인한 환자의 피해를 y라는 수치로 측정할 수 있다면⋯
- 진단을 하는 의사는 $\alpha x + \beta y$를 최소화하는 수준에서 α를 설정해야 할 것이다.
- 이 상황에서는 2종오류의 피해수치가 매우크기 때문에 β를 낮추어야 하고, α가 높은것이 자연스럽다.
- (환자의 생명을 다루는 의사라면 최악의 상황을 항상 생각해야 한다.)

유의수준 설정의 예시 - 형사판결

상황

- 형사판결에서도 1종오류(무죄를 유죄로 판단)과 2종오류(유죄를 무죄로 판단) 두 가지의 오류가능성이 있다.
  - 1종오류는 억울한 사람이 옥에 갇히게 된다.
  - 2종오류는 범죄자를 처벌하지 못하게 된다.
- 판결을 내리는 판사는 어느 수준의 α를 설정해야 할 것인가?

접근

- 일반적으로 "무죄추정의 원칙''은 α의 수치를 매우 낮게 하는 것을 목표로 하는 형법의 원칙이다.
- 1종오류로 인한 억울한 수감자의 피해를 x, 2종오류로 인한 사회적 피해를 y라고 한다면 오류로 인한 피해의 확률가중 합은 $\alpha x + \beta y$이다.
- x가 y에 비해서 매우 크므로 α를 낮게 설정한다.

양측검정과 단측검정 (Side of tails)

1. 양측검정
   - Two-sided test, both-sided test
   - Ex) $H_0 : \mu = 30$ vs $H_A : \mu \neq 30$. μ가 30과 같을 확률이 α보다 작다면 H_0를 기각하고, 그렇지 않다면 H_0를 채택한다. 가설검정에서는 "30과 다르다는 충분한 통계적 근거가 있는지 여부''에 대해서 다룬다.
2. 단측검정
   - One-sided test
   - Ex) $H_0 : \mu \geq 30$ vs $H_A : \mu < 30$. μ가 30보다 크거나 같을 확률이 α보다 작다면 H_0를 기각하고, 그렇지 않다면 H_0를 채택한다. 가설검정에서는 "30보다 작다는 충분한 통계적 근거가 있는지 여부''에 대해서 다룬다.
   - Ex) $H_0 : \mu \leq 30$ vs $H_A : \mu > 30$. μ가 30보다 작거나 같을 확률이 α보다 작다면 H_0를 기각하고, 그렇지 않다면 H_0를 채택한다. "30보다 크다는 충분한 통계적 근거가 있는지 여부''에 대해서 다룬다.

가설검정의 분류 - 가설검정의 목적과 가정에 따라서 분류할 수 있다.

1. 모집단의 수
   - 단일 모집단
   - 2개의 모집단
   - 3개의 모집단 - ANOVA
   - …
2. 모수의 종류
   - 모평균 (μ)
   - 모비율 (p)
   - 모분산 (σ)
   - 모상관계수 (ρ)
   - …
3. Sideness
   - 양측검증
   - 단측검증
4. 다른 조건
   - Ex) 모평균에 대한 검증인데 모분산(σ)가 알려져 있는지 여부
   - Ex) 두 개의 모집단이 서로 독립이거나 서로 대응이거나 한 관계
   - Ex) 샘플의 갯수가 충분한지에 따라 사용하는 분포가 다름
   - …

	단일모집단	두개의 모집단
모평균에 대한 검정		
모비율에 대한 검정		
모상관계수에 대한 검정		
모분산에 대한 검정		

두개의 모집단에 대한 비교

- 많은 경우에 실험은 실험군과 대조군을 나누어 수행한다. 두 집단의 모수를 통계적으로 비교하여 실험군에 가해진 행위의 유의성을 입증한다.

- 예시
  - 혈압약 A를 복용한 후에 혈압이 개선되는가?
  - 왼손과 오른손의 악력에 차이가 있는가?
  - 공법 A와 B는 불량률이 서로 다른가?
  - 공법 A는 공법 B에 비해서 불량률이 낮은가?
  - Bulls' one shot을 넣기 전후로 연비가 좋아지는가?
  - Federer가 Clay 구장과 Grass 구장에서 serve ace의 비율이 같은가?

II. 단일모집단의 모평균에 대한 검정

단일모집단의 모평균에 대한 검정 - σ known

- CLT에 기반함 (표본평균이 정규분포에 수렴)
- 분포

$$\frac{\mu - \overline{X}}{\sigma/\sqrt{n}} \sim Z, \quad \text{or,} \quad \mu \sim N(\overline{X}, \sigma^2/n)$$

- 검정 통계량

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

- $100(1-\alpha)\%$ 신뢰구간

$$\mu \in [\overline{X} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}]$$

- Case 1. 양측검정
  - 가설: $H_0 : \mu = \mu_0$ vs $H_A : \mu \neq \mu_0$
  - 기각역: $|Z| \geq z_{\alpha/2}$
- Case 2. 단측검정
  - 가설: $H_0 : \mu \leq \mu_0$ vs $H_A : \mu > \mu_0$
  - 기각역: $Z \geq z_{\alpha}$
- Case 3. 단측 검정
  - 가설: $H_0 : \mu \geq \mu_0$ vs $H_A : \mu < \mu_0$
  - 기각역: $Z \leq -z_{\alpha}$

단일모집단의 모평균에 대한 검정 - σ unknown

- CLT(표본평균이 정규분포에 수렴)에 기반하며 σ를 s로 대체하기에 t분포를 사용
- 분포

$$\frac{\mu - \overline{X}}{s/\sqrt{n}} \sim t_{n-1}$$

- 검정 통계량

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

- $100(1-\alpha)\%$ 신뢰구간

$$\mu \in [\overline{X} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n}]$$

- Case 1. 양측검정
  - 가설: $H_0 : \mu = \mu_0$ vs $H_A : \mu \neq \mu_0$
  - 기각역: $|T| \geq t_{\alpha/2,n-1}$
- Case 2. 단측검정
  - 가설: $H_0 : \mu \leq \mu_0$ vs $H_A : \mu > \mu_0$
  - 기각역: $T \geq t_{\alpha,n-1}$
- Case 3. 단측검정
  - 가설: $H_0 : \mu \geq \mu_0$ vs $H_A : \mu < \mu_0$
  - 기각역: $T \leq -t_{\alpha,n-1}$

III. 단일모집단의 모비율에 대한 검정

- 이항분포의 정규분포 근사 성질을 이용
- 분포

$$p \sim N(\hat{p}, \frac{\hat{p}\hat{q}}{n})$$

- 검정 통계량

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

- $100(1-\alpha)\%$ 신뢰구간

$$p \in [\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}]$$

- Case 1. 양측검정
  - 가설: $H_0 : p = p_0$ vs $H_A : p \neq p_0$
  - 기각역: $|Z| \geq z_{\alpha/2}$
- Case 2. 단측검정
  - 가설: $H_0 : p \leq p_0$ vs $H_A : p > p_0$
  - 기각역: $Z \geq z_{\alpha}$
- Case 3. 단측검정
  - 가설: $H_0 : p \geq p_0$ vs $H_A : p < p_0$
  - 기각역: $Z \leq -z_{\alpha}$

IV. 두개의 모집단의 모평균에 대한 검정

독립표본 - Setting

- X_1 와 X_2 는 독립된 두 집단에서 각각 얻어진 확률 변수이다.
- Ex) 국산차 100대의 연비와 수입차 50대의 연비를 비교하는 실험
- X_1 과 X_2 의 모평균의 차이를 조사하기 위해 $X_1 - X_2$ 의 분포를 만들고 이 수치를 0과 같은 상수와 비교한다.
- $X_1 \sim some\_dist(\mu_1, \sigma_1^2)$, $X_2 \sim some\_dist(\mu_2, \sigma_2^2)$ 이며, 이를 이용해 $\overline{X_1} - \overline{X_2}$ 의 분포를 먼저 도출한다.

$\overline{X_1} - \overline{X_2}$의 분포 도출

두개의 모집단의 모평균에 대한 검정 - 독립표본 (σ_1, σ_2 unknown)

- CLT(표본평균이 정규분포에 수렴), 두 독립 정규분포의 차이, σ를 s로 대체하기에 t 분포로 사용
- 분포

$$\frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim t_{df}$$

- 검정 통계량

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

- $100(1-\alpha)\%$ 신뢰구간

$$\mu_1 - \mu_2 \in [(\overline{X_1} - \overline{X_2}) \pm t_{\alpha/2} \cdot \sqrt{s_1^2/n_1 + s_2^2/n_2}]$$

- Case 1. 양측검정
  - 가설: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_A : \mu_1 - \mu_2 \neq 0$
  - 기각역: $|T| \geq t_{\alpha/2,df}$
- Case 2. 단측검정
  - 가설: $H_0 : \mu_1 - \mu_2 \leq 0$ vs $H_A : \mu_1 - \mu_2 > 0$
  - 기각역: $T \geq t_{\alpha,df}$
- Case 3. 단측검정
  - 가설: $H_0 : \mu_1 - \mu_2 \geq 0$ vs $H_A : \mu_1 - \mu_2 < 0$
  - 기각역: $T \leq -t_{\alpha,df}$

- 위의 자유도(df)는 $n-1$을 대체할 수 있는 $n_1 - 1$과 $n_2 - 1$을 결합한 수치:

$$df = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

쌍대표본 - Setting

쌍대표본(paired samples)

- 쌍대표본(paired samples)이란 두 개의 모집단이 서로 대응되는 상황을 의미한다.
- 두 모집단에 대한 확률 변수가 X_1과 X_2라고 한다면, 우리의 관심은 개별 확률 변수가 아니라 두 모집단의 차이가 된다.
- 예시
  - 한 사람의 왼손과 오른손의 악력의 차이가 있는가?
  - 첨가제를 넣기 전과 후에 자동차의 연비에 차이가 있는가?
  - 특정질환의 수치를 낮춘다고 주장하는 신약의 효과를 확인하기 위해 복용하게 하였다. 복용전과 후에 수치의 차이가 있는가?
- X_1에 대한 관측치의 갯수 n_1과 X_2에 대한 관측치의 갯수 n_2는 일반적으로 같다. (같지 않다면 어떻게 쌍(pair)을 이루어 조사하겠는가?)

가설 검정

- $X_1 - X_2$에 대한 관측치를 이용한다.
- 즉, 단일 표본의 문제로 변환하여 가설검정을 진행한다.

예 제

- 특정 질환의 수치를 낮추는 신약의 효과를 알아보기 위해 9명의 환자를 대상으로 복용전과 후의 수치를 비교하였다. 실험 전을 A그룹으로, 실험 후를 B그룹으로 할때에 아래의 수치를 얻었다.
- 99%의 신뢰수준에서 신약이 수치를 낮추어주는지 검정하라.

```
A <- c(250, 220, 204, 222, 206,
       259, 204, 205, 198)
B <- c(242, 217, 200, 211, 189,
       240, 199, 200, 195)
X <- A-B
mean(X) # sample mean (X_bar)
```

```
## [1] 8.333333
```

```
sd(X) # sample standard deviation (s)
```

```
## [1] 6.062178
```

```
"This note is made with Rmarkdown"
```

```
## [1] "This note is made with Rmarkdown"
```