

Lecture 08 - Confidence Interval and Hypothesis Testing

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



서울과학기술대학교 IT정책전문대학원

1 I. Confidence Interval

2 II. Hypothesis Testing

I. Confidence Interval

신뢰구간 (Confidence Interval, CI)

- 신뢰구간은 모수(parameter)가 확률적으로 어느 범위 안에 있는지를 나타낸다.
- 신뢰수준(confidence level)은 신뢰구간이 모수를 포함하고 있을 확률을 의미한다.
- 신뢰수준(confidence level)에 따라서 구간의 길이가 달라진다.
 - 신뢰수준이 높을 수록 구간의 길이가 길어진다.
 - 신뢰수준이 낮을 수록 구간의 길이가 짧아진다.
- **모수의 분포를 알면 신뢰구간을 구할 수 있다.**

Review on CI - 모평균 (μ)

모평균(μ)의 신뢰구간 (σ 를 알 때) (L07,p17)

- 분포: $\frac{\mu - \bar{X}}{\sigma/\sqrt{n}} \sim Z$ (or, $\mu \sim N(\bar{X}, \sigma^2/n)$)
- 예시(L07,p18): $\bar{X} = 20, \sigma = 2, n = 100$
 - 신뢰수준 95% CI: $\mu \in [20 - 1.96 \times \frac{2}{\sqrt{100}}, 20 + 1.96 \times \frac{2}{\sqrt{100}}]$
 - 신뢰수준 99% CI: $\mu \in [20 - 2.58 \times \frac{2}{\sqrt{100}}, 20 + 2.58 \times \frac{2}{\sqrt{100}}]$
 - 신뢰수준 100(1- α)% CI: $\mu \in [20 - z_{\alpha/2} \times \frac{2}{\sqrt{100}}, 20 + z_{\alpha/2} \times \frac{2}{\sqrt{100}}]$
- 신뢰수준 100(1- α)% CI: $\mu \in [\bar{X} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}]$

모평균(μ)의 신뢰구간 (σ 를 모를때) (L07,p19)

- 분포: $\frac{\mu - \bar{X}}{s/\sqrt{n}} \sim t_{n-1}$
- 예시(L07,p22): $\bar{X} = 20, s = 2, n = 100$
 - 신뢰수준 95% CI: $\mu \in [20 - 1.99 \times \frac{2}{\sqrt{100}}, 20 + 1.99 \times \frac{2}{\sqrt{100}}]$
- 신뢰수준 100(1- α)% CI: $\mu \in [\bar{X} - t_{\alpha/2, n-1} \times \frac{\sigma}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \times \frac{\sigma}{\sqrt{n}}]$

t-분포

- 스튜던트 t 분포 (Student's t-distribution)
 - 스튜던트라는 필명으로 발표된 분포
 - 정규분포의 평균을 측정하기 위해서 주로 사용된다.

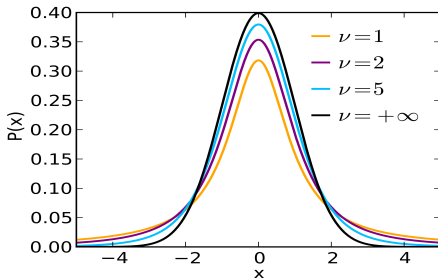


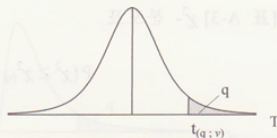
그림 1: 자유도에 따른 t-분포의 pdf

- 자유도는 사용한 샘플의 숫자에 따라서 변화하며,
- 자유도가 높아질수록 (샘플의 숫자가 커질수록) 정규분포에 가까워진다.

t-분포표

[표 A-2] t-분포표

$$P\{T \geq t_{(q; v)}\} = q$$



자유도 v	꼬리확률 q									
	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221

14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.792	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Review on CI - 이항시행의 성공횟수 (X)

- 분포
 - $X \sim \text{bin}(n, p) \rightarrow N(np, npq)$
 - $\frac{X - np}{\sqrt{npq}} \sim Z$
- 예시 (L06, p24): $n = 1000, p = 0.5$
 - 분포: $X \sim \text{bin}(1000, 0.5) \rightarrow N(500, 16^2)$
 - 신뢰수준 95% CI: $X \in [500 - 1.96 \times 16, 500 + 1.96 \times 16]$
- 신뢰수준 $100(1-\alpha)\%$ CI:

New CI - 이항실험 결과를 이용한 모비율의 추정 (p)

- Motivation: 이항실험 100번을 시행하여 40번의 성공을 거두었다. 이 경우에 이항실험의 성공확률 p 의 95% 신뢰구간을 추정하라.
- Terminology: 이항실험의 성공확률 p 를 **모비율(true proportion, population proportion)**이라고 한다.
- Development: 신뢰구간을 추정하기 위해서는 **모수의 분포**를 알아야 한다. 모수 p 의 분포를 아래의 과정을 거쳐서 도출한다.
 1. X 가 성공횟수 일때에 $X \sim \text{bin}(n, p) \rightarrow N(np, npq)$ 이다.
 2. 위의 식의 양변을 n 으로 나누면, $\frac{X}{n} \sim N(p, \frac{pq}{n})$ 이다.
 3. 이를 다시쓰면 $p \sim N(\frac{X}{n}, \frac{pq}{n})$ 이다.
- Issue: 실험결과를 이용해서 마지막 분포식의 우변에 있는 $\frac{X}{n}$ 과 $\frac{pq}{n}$ 의 수치를 알 수 있을까?
 1. 실험결과로서 X 와 n 은 주어져 있지만, p 와 q 는 주어지지 않았다.
 2. 따라서 실험결과를 이용하기 위해 아래의 규칙을 적용한다.
 - p 를 실험내에서의 성공비율로 대체한다.
 - q 를 실험내에서의 실패비율로 대체한다.

● Conclusion

1. 앞의 Development에서 마지막으로 유도된 모수 p 의 분포는 $p \sim N(\frac{X}{n}, \frac{pq}{n})$ 이다.
2. 여기에서 우변의 모수 p 와 q 를 실험 결과를 활용한 추정치 \hat{p} 와 \hat{q} 로 대체한다.
3. 따라서 최종적으로 모수 p 의 분포는 $p \sim N(\frac{X}{n}, \frac{\hat{p}\hat{q}}{n})$ 이다.

● Motivation 문제의 정답

1. $p \sim N(\frac{X}{n}, \frac{\hat{p}\hat{q}}{n})$
2. $n = 100$ and $X = 40$ are given.
3. We have $\hat{p} = X/n = 40/100 = 0.4$ and $\hat{q} = 1 - \hat{p} = 1 - 0.4 = 0.6$
4. Thus, $p \sim N(\frac{40}{100}, \frac{0.4 \cdot 0.6}{100}) = N(0.4, 0.0024) = N(0.4, 0.049^2)$
5. It follows, 95% CI for p is

$$p \in [0.4 - 1.96 \cdot 0.049, 0.4 + 1.96 \cdot 0.049] = [0.304, 0.496]$$

- 이항실험결과
- n
- 과
- X
- 가 주어져 있을때, 모비율
- p
- 의
- $100(1-\alpha)\%$
- 신뢰구간은

$$100(1 - \alpha)\% \text{ CI} : p \in \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right]$$

Exercise 1

남해안에서 잡히는 전복무게의 표준편차는 $4g$ 으로 알려져있다. 200마리의 전복을 채집하여 평균무게를 계산하니 $20g$ 이었다. 전복무게 평균의 90% 신뢰구간을 구하라.

Exercise 2

다이아몬드 0.5캐럿 상품의 가격을 조사하기 위하여 20개의 상점에서 시장조사를 실시하였다. 수집된 데이터의 평균가격은 100만원, 표본표준편차는 5만원이었다. 이를 바탕으로 다이아몬드 0.5캐럿 가격의 평균에 대한 95% 신뢰구간을 구하라.

Exercise 3

정사면체로 되어 있고 각 면에 1,2,3,4가 적혀있는 주사위가 있다. 즉, 각각의 한 면이 나올 확률은 각각 0.25이다. 이 주사위를 100번 던졌을 때에 1의 눈이 나오는 횟수에 대한 99% 신뢰구간을 구하라.

Exercise 4

농구선수의 선발을 위하여 프로테스트를 진행중이다. 한 농구 선수에게 자유투 50개를 던지게 하였더니, 17개를 성공시켰다. 이 농구선수의 자유투 성공확률에 대한 95% 신뢰구간을 구하라.

II. Hypothesis Testing

귀무가설과 대립가설

- 통계적 가설 검정에서는 보수적이고 기존의 관념에 해당하는 **귀무가설**을 가정한 상태에서, 이를 반박할 수 있는 강력한 통계적 근거에 해당하는 **대립가설**의 채택을 검토한다.
- 귀무가설 (Null hypothesis)
 - H_0 로 흔히 표기한다.
 - 기존에 알려진 지식, 관념, 편견, status quo에 해당한다.
 - 반증의 대상이 되는 가설이며, 흔히 실험결과가 유의미 하려면 귀무가설을 기각할 수 있어야 한다.
- 대립가설 (Alternative hypothesis)
 - H_1 이나 H_A 로 흔히 표기한다.
 - 대부분의 실험에서 입증하려고 노력하는 가설이다.
 - 귀무가설에 대한 반증이 성공하였을 때에 채택되는 가설이다.

● 귀무가설 예시

1. 다르지 않다.
2. 효과가 없다.
3. μ 가 20과 다르지 않다.
($H_0 : \mu = 20$)
4. 동전의 p 가 0.5와 다르지 않다.
($H_0 : p = 0.5$)
5. 의약품의 p 가 0.7보다 크지 않다.
($H_0 : p \leq 0.7$)
6. 유죄를 입증할 근거가 부족하다.
(무죄추정의 원칙)

● 대립가설 예시

1. 다르다고 말할 수 있는 근거가 충분하다.
2. 효과가 있다고 말할 근거가 충분하다.
3. μ 가 20과 다르다. ($H_1 : \mu \neq 20$)
4. 동전의 p 가 0.5와 다르다 말할 근거가 충분하다. ($H_1 : p \neq 0.5$)
5. 의약품의 p 가 0.7보다 크다고 말할 통계적 근거가 충분하다.
($H_1 : p > 0.7$)
6. 유죄를 입증할 근거가 충분하다.

가설 검정의 방법

1. 신뢰구간을 구성한다.
2. 검정통계량(test-statistics)을 사용한다.
3. 유의확률(P-value)을 사용한다.

가설 검정 방법 (L07, p18 예시 이용)

- 모집단의 표준편차가 2cm로 알려진 100명의 남성을 추출하여 손의 크기를 재보니 표본평균이 20cm였다. 남성 손 크기의 평균이 20.4cm가 아니더라도 할 수 있는가? 신뢰수준 95%에서 검증하라.

0. 가설을 명시한다.

- $H_0 : \mu = 20.4$
- $H_1 : \mu \neq 20.4$

1. 신뢰구간을 이용하는 방법

- CLT에 의해서 $\mu \sim N(\bar{X}, \sigma^2/n)$ 이며, 표본평균 $\bar{X} = 20$, 모표준편차 $\sigma = 2$, 표본의 수 $n = 100$ 을 이용하면, 이는 $\mu \sim N(20, 2^2/100) = N(20, 0.2^2)$ 이다.
- 따라서 95% CI는 $[20 - 1.96 \cdot 0.2, 20 + 1.96 \cdot 0.2] = [19.608, 20.392]$ 이다.
- 가설에서 사용된 수치(이를 μ_0 라고 표기한다)인 20.4가 신뢰구간에 포함되지 않는다.
- 따라서 귀무가설을 기각한다. 따라서 남성 손 크기의 평균이 20.4cm가 아니더라도 말할 충분한 통계적 근거가 있다.

2. 검정통계량(test-statistics)을 사용하는 방법

● Motivation

- 앞의 신뢰구간을 이용한 가설검정에서는 95% 신뢰수준에 해당하는 수치 1.96 ($=z_{0.025}$)을 사용하였다.
- 만약에 문제에서 99% 신뢰수준을 제시했다면, 그에 해당하는 수치 2.58 ($=z_{0.005}$)을 사용했을 것이며, 이로 인해 가설검정의 결론이 달라졌을 것이다.
- 1.96이나 2.58과 비교할 수 있는 어떤 수치를 먼저 계산한다면, 분포표에서 수치($z_{\alpha/2}$)를 찾자마자 가설검정의 결론을 바로 내릴수 있지 않을까?

● 검정통계량

- 이 문제에서 검정통계량의 수치는 아래와 같다.

$$\frac{20.4 - 20}{2/\sqrt{100}} \quad \left(= \frac{\mu_0 - \bar{X}}{\sigma/\sqrt{n}} \right)$$

- 즉, 검정통계량의 수치는 2이며,
 - 신뢰수준 95%에서는 $z_{0.025} = 1.96$ 보다 크기 때문에 H_0 를 기각한다.
 - 신뢰수준 99%에서는 $z_{0.005} = 2.58$ 보다 작기 때문에 H_0 를 채택한다.
 - 신뢰수준 $100(1-\alpha)\%$ 에서는 $z_{\alpha/2}$ 와 비교한다.
- 참고: H_0 를 기각할 수 있는 검정통계량의 범위를 기각역이라고 한다.

3. 유의확률 (P-value)를 사용하는 방법

- Motivation

- 이 문제에서 검정통계량의 수치는 2였고, 검정통계량을 이용한 방법에서는 이를 $z_{\alpha/2}$ 와 비교하였다.
- 검정통계량인 2를 α 의 도메인으로 변환할 수 있다면, 신뢰수준 몇 %에서 귀무가설이 채택되고 기각되는지를 한번에 알 수 있지 않을까?

- 유의확률의 계산

- 아래의 R code 블록 및 결과치를 확인하라. 검정통계량을 구한 이후에 아래의 마지막 수치인 0.0455는 신뢰수준에 따른 H_0 의 기각여부를 바로 판단할 수 있게 해준다.
- 즉, 0.0455라는 수치로 보아서 아래의 결론이 바로 도출된다.
 - 신뢰수준이 90%라면, H_0 는 기각된다.
 - 신뢰수준이 95%라면, H_0 는 기각된다.
 - 신뢰수준이 99%라면, H_0 는 채택된다.

```
pnorm(2)
```

```
## [1] 0.9772499
```

```
1-pnorm(2)
```

```
## [1] 0.02275013
```

```
2*(1-pnorm(2))
```

```
## [1] 0.04550026
```

Exercise 5

남해안에서 잡히는 전복무게의 표준편차는 4g으로 알려져있다. 전복 판매업자는 전복의 평균무게가 22g이라고 주장하고 있다. 전복 판매업자의 주장을 검증하려고 200마리의 전복을 채집하여 평균무게를 계산하니 20g이었다.

1. 귀무가설과 대립가설을 명시하라
2. 검정통계량을 계산하고, 90%, 95%, 99%의 신뢰수준에서 각각 H_0 를 기각할 수 있는지 여부를 판단하라.
3. 유의확률을 계산하라.

논문 작성시의 표기 방법

- 유의확률을 계산하여 독자가 가진 신뢰수준에 따라 실험결과를 해석할 수 있는 여지를 주는 것이 일반적이다.

6  M. K. SIM AND D. G. CHOI

Table 4. Test of path dependency on different surfaces.

Current state	Last point won by	Grass			Hard			Clay		
		Server's winning prob. in the next point (%)	Number of obs.	Z -statistics	Server's winning prob. in the next point (%)	Number of obs.	Z -statistics	Server's winning prob. in the next point (%)	Number of obs.	Z -statistics
15:15	Server	66.59	7,079	-0.19	64.40	34,302	0.76	62.06	17,513	-0.53
	Receiver	66.74	7,445		64.12	35,576		62.34	17,627	
30:15	Server	66.45	9,683	1.16	63.67	44,903	1.44	62.13	21,857	-1.67*
	Receiver	65.45	4,504		63.09	21,400		63.10	10,132	
15:30	Server	63.85	2,606	-0.41	61.83	13,130	1.11	60.77	6,829	0.44
	Receiver	64.33	4,841		61.25	24,975		60.45	13,283	
40:15	Server	69.20	9,382	1.77*	67.12	42,091	2.93***	65.06	19,972	3.60***
	Receiver	67.48	3,023		65.77	13,933		62.58	6,390	
30:30	Server	66.76	4,778	0.82	63.60	23,414	-0.12	60.58	12,180	-1.91**
	Receiver	65.97	4,805		63.65	24,212		61.78	12,017	
15:40	Server	63.66	853	0.18	61.50	4,893	0.66	60.28	2,694	1.1
	Receiver	63.32	2,669		60.96	14,691		59.08	7,932	
40:30	Server	65.74	6,360	0.06	63.52	30,304	-1.16	61.98	14,803	1.25
	Receiver	65.69	3,873		64.04	18,608		61.18	9,369	
30:40	Server	64.49	2,233	1.20	61.55	11,965	-0.04	60.06	6,310	0.43
	Receiver	62.89	3,223		61.57	17,322		59.72	9,394	
Duce	Server	64.37	6,289	-0.42	62.79	32,997	1.09	60.82	17,458	0.08
	Receiver	64.72	6,327		62.38	33,012		60.78	17,220	

*p < 0.1; **p < 0.05; ***p < .01.

- 독자의 신뢰수준에 따라서 실험결과의 유의미성에 대한 해석이 달라질 수 있다.
 - 신뢰수준 90%를 가진 독자라면,
 - 신뢰수준 95%를 가진 독자라면,
 - 신뢰수준 99%를 가진 독자라면,

