기초통계학II

2. 단일모집단의 기본

(1) 모평균에 대한 추론

(2) 모분산에 대한 추론

(3) 모비율에 대한 추론

□ 모평균 추론에 대한 복습

■ μ의 $100(1-\alpha)\%$ 신뢰구간

$\hat{\mu} = \bar{X}$: 추정량

$\bar{x}$: 추정값(치)

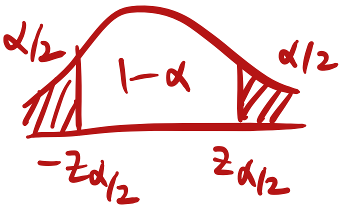
오차에 대한 정보 → 구간추정 (신뢰구간)

$P(L < \mu < U) = 1-\alpha$

기준값 : 확률변수의 확률분포를 알아야함

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$P(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1-\alpha$



∴ 신뢰구간 $\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

표본이 주어지면 실제값 계산.
σ가 주어진 경우 계산가능

σ를 모르는 경우 $\hat{\sigma} = S$ → 오차발생

$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ (양쪽 꼬리가 두꺼움)

* 자유도가 커지면
(표본크기가 커지면)
표준정규분포와 유사 : 중심극한정리

↳ t분포는 같은 α에 대해 자유도에 따라
값이 달라 불편

잘못된 결정을 내리는것 = 오류

제 1종오류 P(H0기각 | H0참)를 max로 잡고 가설결정 : 유의수준 α

◈ 가설검정 절차

(1) 귀무가설과 대립가설을 설정한다. (상반됨, H0이 맞다고 가정)

H0 H1

(2) 주어진 문제의 특성에 따라 유의수준 α를 결정한다.

(3) 표본 자료에서 검정통계량을 계산한다. 표본으로부터 구해짐.

(4) 가설검정결론 : 모든사람이 동일한 결과를 내기위한 기준

(i) 기각역 사용

유의수준에 따라 기각역을 구한다.

검정통계량이 기각역에 속하면 H_0 를 기각한다.

(ii) p 값(유의확률) 사용 : 더 비정상적인 값이 나올확률

- 검정통계량을 이용하여 p값을 구한다.
- p ≤ α이면 H_0를 기각한다.

* 등호는 귀무가설에 붙는다
* 가설은 모수에 대한 주장

→ 모평균을 기준으로 가설이 세워진다

■ μ에 대한 가설검정

① 가설

$\mu = \mu_0$ 일때 기각

→ 모든 $\mu \geq \mu_0$ or $\mu \leq \mu_0$ 에 대해 기각 가능

귀무가설	대립가설	
$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$	양측검정(two-sided test)
	$H_1 : \mu > \mu_0$	단측검정(one-sided test)
	$H_1 : \mu < \mu_0$	단측검정(one-sided test)

② 유의수준 α 정하기 : 0.05, 0.01

모평균의 점추정량 이용 → 표준화

③ 검정통계량 귀무가설이 $H_0 : \mu = \mu_0$ 인 경우 검정통계량

(표준화하지 않으면 모든 표본마다 다시 표준화하는 번거로움)

σ를 알 때	σ를 모를 때, n이 클 때 ($n \geq 30$)	σ를 모를 때, n이 작을 때 ※ 모집단 정규분포 가정
$Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$	$Z = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n}} \sim N(0,1)$	$T = \dfrac{\overline{X} - \mu_0}{s / \sqrt{n}} \sim t(n-1)$

$\bar{x} \approx \mu_0$ (H_0 채택)을 결정하는 기준값 : 임계값

4. 기각역 및 유의확률

			대립가설		
			$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$	$H_1 : \mu \neq \mu_0$
					
검정통계량	$Z=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 또는 $Z=\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$	기각역	$R=\{z \geq z_\alpha\}$	$R=\{z \leq -z_\alpha\}$	$R=\left\{\begin{matrix} z \leq -z_{\alpha/2}, \\ z \geq z_{\alpha/2} \end{matrix}\right\}$ $\lvert z\rvert \geq z_{\alpha/2}$
		유의확률 (p-값)	$P(Z \geq z)$	$P(Z \leq z)$	$2 \times P(Z \geq \lvert z\rvert)$ α/2가 아닌 α가 기준이므로
	$T=\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$	기각역	$R=\{t \geq t_{\alpha,n-1}\}$	$R=\{t \leq -t_{\alpha,n-1}\}$	$R=\{\lvert t\rvert \geq t_{\alpha/2,n-1}\}$
		유의확률 (p-값)	$P(T \geq t)$	$P(T \leq t)$	$2 \times P(T \geq \lvert t\rvert)$

내가구한검정통계량값보다
더 비정상적인 표본이 나올확률

□ 분산(표준편차)

- 모집단 가정: $N(\mu, \sigma^2)$ ⇦ 정규성에 대한 가정확인 필요
- 확률표본 $X_1, X_2, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2)$

○ 점추정

자유도로 나누기 : 불편성 만족하는 이유

- 모수: σ^2 ⇦ 표본분산: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
- 모수: σ ⇦ 표본표준편차: $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$

↓

선택된 기준 : 불편성, 효율성, 일치성

★ 불편성 : 추정량의 기댓값이 모수와 같을때

○ 점추정량의 통계적 성질

○ $X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$이면

- $E(S^2) = \sigma^2$ (⇒수리통계학) : 불편성 만족 → 오차를 알려면 S^2에 대한 분포를 알아야한다.
- $Var(S^2) = 2\sigma^4/(n-1)$ (⇒수리통계학)

확률변수 >0

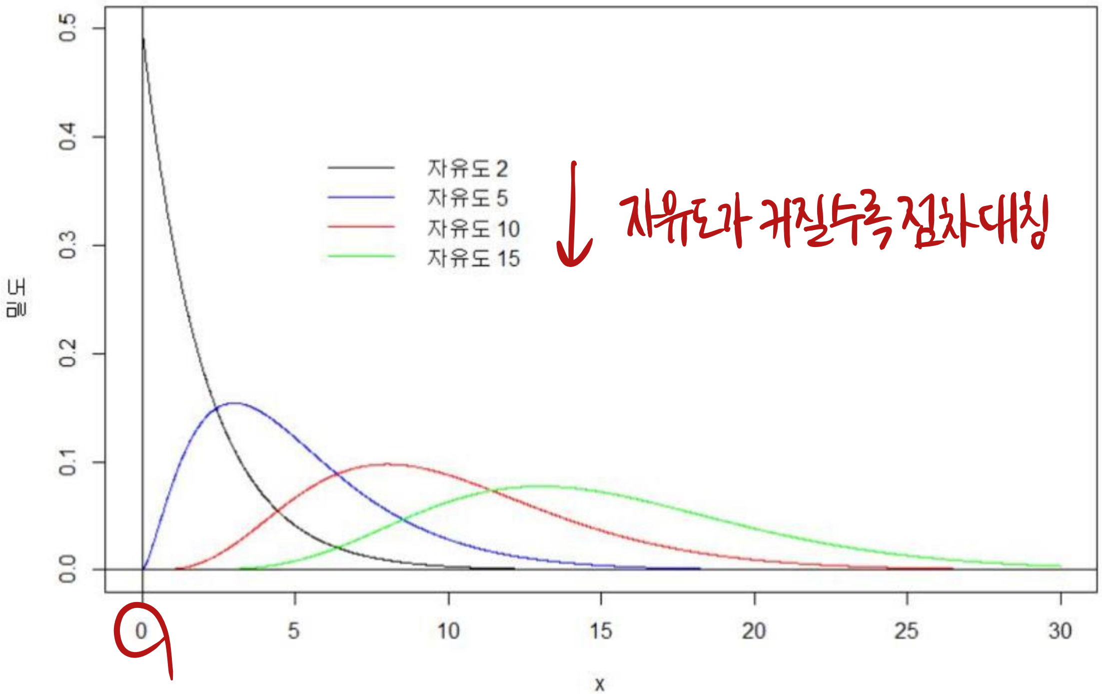
$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \boxed{\chi^2_{n-1}}$$

- χ^2_{n-1} : 자유도가 $n-1$ 인 카이제곱(chi-square)분포

○ **중심축량**: $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

자유도에 따라 모양이 바뀜.

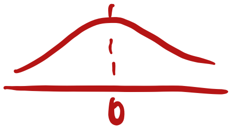
확률변수가 양수만을 가질 때 따르는 여러 분포가 있음.

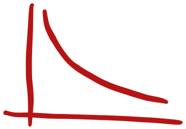

0.5
0.4
0.3
0.2
0.1
0.0
밀도
자유도 2
자유도 5
자유도 10
자유도 15
자유도가 커질수록 점차대칭
0
5
10
15
20
25
30
x

■ 카이제곱분포의 상위 α의 확률을 주는 값

자유도

$Z \sim N(0,1) \rightarrow Z^2 \sim \chi^2_1$





d.f. \ α	.99	.975	.95	.90	.50	.10	.05	.025	.01
1	,0002	,001	,004	,02	,45	2,71	3,84	5,02	6,63
2	,02	,05	,10	,21	1,39	4,61	5,99	7,38	9,21
3	,11	,22	,35	,58	2,37	6,25	7,81	9,35	11,34
4	,30	,48	,71	1,06	3,36	7,78	9,49	11,14	13,28
5	,55	,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09
6	,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81
7	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48
8	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09
9	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67
10	2,56	3,24	3,94	4,87	9,34	15,99	18,31	20,48	23,21
11	3,05	3,81	4,57	5,58	10,34	17,28	19,68	21,92	24,72
12	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22
13	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69
14	4,66	5,62	6,57	7,79	13,34	21,06	23,68	26,12	29,14
15	5,23	6,26	7,26	8,55	14,34	22,31	25,00	27,49	30,58
16	5,81	6,90	7,96	9,31	15,34	23,54	26,30	28,85	32,00
17	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41
18	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81
19	7,63	8,90	10,12	11,65	18,34	27,20	30,14	32,85	36,19
20	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57
21	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93
22	8,54	10,98	12,34	10,04	21,34	30,81	33,92	36,78	40,29
23	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64
24	70,86	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98
25	11,52	13,11	14,61	16,47	24,34	34,38	37,65	40,65	44,31
26	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64
27	12,88	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96
28	13,56	15,30	16,93	18,94	27,34	37,92	41,34	44,46	48,28
29	14,26	16,04	17,71	19,77	28,34	39,09	42,56	45,72	49,59
30	14,95	16,78	18,49	20,60	29,34	40,26	43,77	46,98	50,89
40	22,16	24,42	26,51	29,05	39,34	51,81	55,76	59,34	63,69
50	29,71	32,35	34,76	37,69	49,33	63,17	67,50	71,42	76,15
60	37,48	40,47	43,19	46,46	59,33	74,40	79,08	83,30	88,38
70	45,44	48,75	51,74	55,33	69,33	85,53	90,53	95,02	100,43
80	53,54	57,15	60,39	64,28	79,33	96,58	101,88	106,63	112,33
90	61,75	65,64	69,13	73,29	89,33	107,57	113,15	118,14	124,12
100	70,06	74,22	77,93	82,36	99,33	118,50	124,34	129,56	135,81

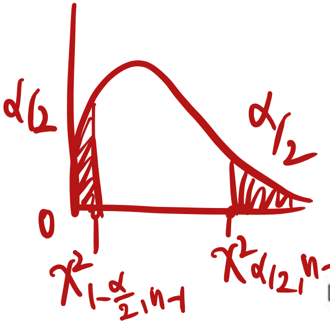
$P(L < \sigma^2 < U) = 1-\alpha$

$\hat{\sigma}^2 = S^2$

○ 신뢰구간

○ σ^2의 $100(1-\alpha)\%$ 신뢰구간

S^2의 직접적인 분포는 모르지만 이용 : 점추정량, 모수 같이 포함



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

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

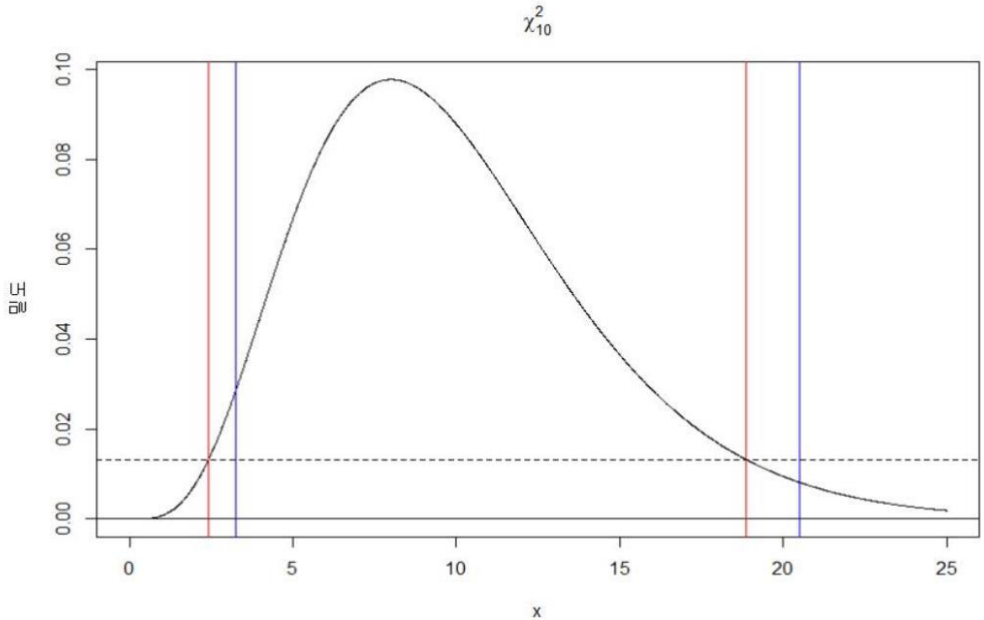
역수취하고 부등호 바꾸기

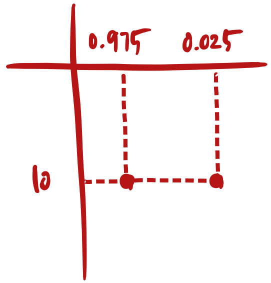
$$\Rightarrow \left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

○ σ의 $100(1-\alpha)\%$ 신뢰구간

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}}\right)$$

높이가 같을때 신뢰구간의 폭이 가장 좁다.





- $\boxed{P(\chi^2_{10} < 3.247) = 0.025,\ P(\chi^2_{10} > 20.483) = 0.025}$ (파란선) ⇦
- $P(\chi^2_{10} < 2.413) = 0.008,\ P(\chi^2_{10} > 18.866) = 0.042$ (빨간선)

빨간선에 비해 신뢰구간의 폭이 넓지만
비대칭 그래프의 특성상 편의를 위해 파란선 사용

◉ 생산된 제품의 평균 강도가 어느 수준에서 안정적으로 생산되는지 알아보기 위해 임의로 8개를 선택하여 제품강도를 측정함

- ○ 안정성은 분산으로 평가

24.3	28.6	30.2	26.5	25.7	27.8	26.9	29.0

  - $s^2 = 3.65$, $\chi^2_{0.975,7} = 1.690$, $\chi^2_{0.025,7} = 16.013$

- ○ σ^2 의 95% 신뢰구간

$$\left(\frac{7\times 3.65}{16.013},\ \frac{7\times 3.65}{1.690}\right) = (1.596,\ 15.122)$$

- ○ σ 의 95% 신뢰구간

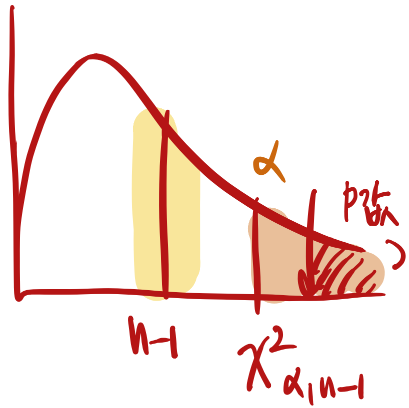
$$(\sqrt{1.596},\ \sqrt{15.122}) = (1.263,\ 3.889)$$

○ 가설검정

$S^2 \approx \sigma_0^2 \rightarrow H_0$채택

- $H_0 : \ \sigma^2 = \sigma_0^2$ vs $\quad H_1 : \begin{cases} ① \ \sigma^2 > \sigma_0^2 \\ ② \ \sigma^2 < \sigma_0^2 \\ ③ \ \sigma^2 \neq \sigma_0^2 \end{cases}$

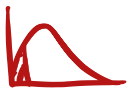
- **검정통계량** : $X^2 = \dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$



- 유의수준을 α라고 하면, 기각역은

① $[\chi^2_{\alpha,n-1}, \infty) \Leftrightarrow \chi^2 \geq \chi^2_{\alpha,n-1}$

② $[0, \chi^2_{1-\alpha,n-1}] \Leftrightarrow \chi^2 \leq \chi^2_{1-\alpha,n-1}$ α



③ $[0, \chi^2_{1-\alpha/2,n-1}], [\chi^2_{\alpha/2,n-1}, \infty) \Leftrightarrow \chi^2 \leq \chi^2_{1-\alpha/2,n-1}, \ \chi^2 \geq \chi^2_{\alpha/2,n-1}$

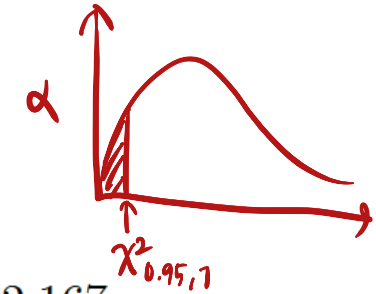
부족표를 이용해 정확한 유의확률 구하기 쉽지 않음. (∵ 자유도)

→ 컴퓨터로 값을 구함.

◉ 표준편차가 2미만일 때 안정적인 품질관리가 유지된다고 할 때 품질관리가 유지되는지 검정

○ $H_0 : \sigma = 2$ vs $H_1 : \sigma < 2$ ⇨ $H_0 : \sigma^2 = 4$ vs $H_1 : \sigma^2 < 4$

○ 검정통계량: $X^2 = \dfrac{(n-1)S^2}{2^2} \sim \chi^2_{n-1}$



○ 5% 유의수준에서 $\chi^2 \le 2.167 = \chi^2_{0.95,7}$

$$\chi^2 = \frac{(8-1)3.65}{4} = 6.389 > 2.167$$

⇨ 안정적인 품질관리가 유지되고 있다고 할 수 없음

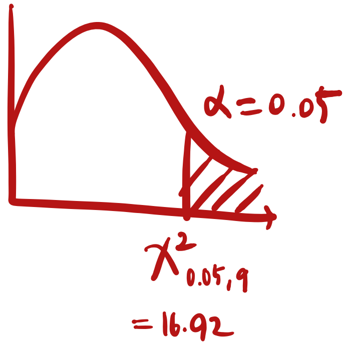
(H0 채택, H1 기각) → 대립가설의 입장에서 결론쓰기.

예제 볼트와 너트를 생산하는 한 공장에서는 제품의 품질이 얼마나 균일하게 유지되는지를 검사하려고 10개의 볼트를 추출하여 지름을 측정하고 그 표준편차를 구하였더니 0.4였다. σ가 0.2보다 크다고 할 수 있는지 유의수준 0.05로 검정하라.

$\sigma^2 = 0.2^2$ $\quad$ $\sigma^2 > 0.2^2$

① 가설 $H_0 : \sigma = 0.2$ vs $H_1 : \sigma > 0.2$

② 검정통계량 : $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{9 \times 0.4^2}{0.2^2} = 36$ → 검정통계량의 값 (검정통계치)



③ 기각역 $R : \chi^2 \geq \chi^2_{0.05}(9) = 16.92$

④ 결론 : 검정통계량의 값이 기각역에 포함되므로 주어진 자료로부터 σ가 0.2보다 크다고 할 수 있다.

(H_0 기각, H_1 채택)