김은영

제8장 정규 모집단에서의 소표본 추론

8.1 서론

(1) 대표본 추론
: '**중심극한정리**'가 모집단의 아주 다양한 분포에 대해 잘 적용되기 때문에 모집단의 분포모양에 관한 구체적인 정보가 없이도 추론을 할 수 있다.

(2) 소표본($n \le 30$) 추론
① 많은 통계적 조사, 특히 비용이나 시간이 많이 소요되는 실험의 경우에는 소표본으로부터 통계적 추론을 해야 한다.
② n이 크지 않을 때 표본평균 $\overline{X}$의 표본분포는 무엇인가?
→ n이 작을 때 $\overline{X}$의 분포는 모집단의 분포의 형태에 많이 의존한다.
⇒ <u>**모집단의 분포가 특정조건을 만족하면**</u> 적절한 추론방법을 찾을 수 있다.

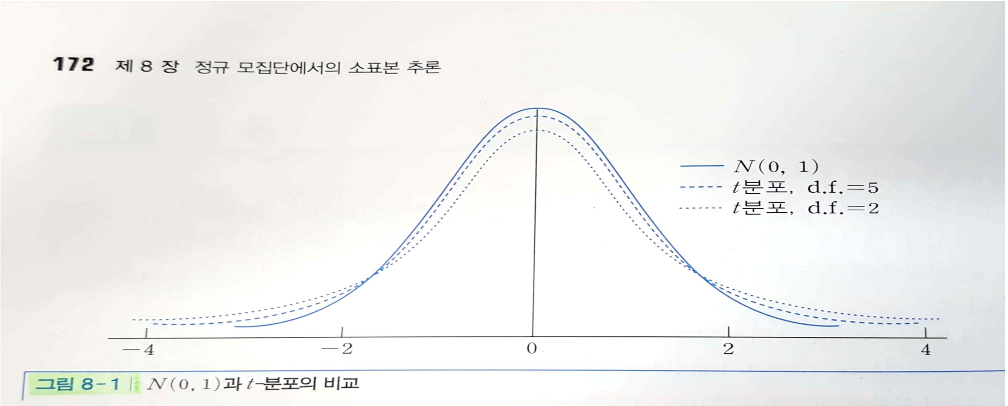
8.2 t-분포 (Student's t-distribution)

① t-분포 (W. S. Gosset, 1908)
❑ $X_1, X_2, \cdots, X_n$: 정규모집단 $N(\mu, \sigma^2)$에서 추출된 확률표본이고

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ , \ s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} \text{ 이면,}$$

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$: 자유도가 $n-1$인 $t-$분포, $T \sim t(n-1)$

② t-분포의 확률밀도곡선의 성질
❑ $t=0$을 중심으로 대칭
❑ $N(0,1)$의 확률밀도곡선에 비해 두꺼운 꼬리를 가진다.
❑ 자유도 $(n-1)$가 커짐에 따라 t-분포의 확률밀도곡선은 $N(0,1)$의 확률밀도곡선에 가까워진다.



[예제 8.1]

부록 [표 3]의 t-분포표를 이용하여 자유도 5인 t-분포의 상위 10%점과 하위 10%점을 구하라.

【풀이】

$t_{0.10}(5) = 1.476,\ -t_{0.10}(5) = -1.476$

[예제 8.2]

확률변수 T가 자유도 9인 t-분포를 따를 때 $P[-b < T < b] = 0.90$을 만족시키는 값 b를 구하라.

【풀이】 $\alpha = 0.05,\ d.f. = 9 \ \rightarrow \ t_{0.05}(9) = 1.833 \ \Rightarrow \ \therefore \ b = 1.833$

8.3 μ의 추론-소표본인 경우

8.3.1 μ의 신뢰구간

(1) μ의 신뢰구간 도출

❑ $T = \dfrac{\overline{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$

❑ $P\left[-t_{\alpha/2} < \dfrac{\overline{X} - \mu}{s/\sqrt{n}} < t_{\alpha/2}\right] = 1 - \alpha$

$P\left[\overline{X} - t_{\alpha/2}\dfrac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\dfrac{s}{\sqrt{n}}\right] = 1 - \alpha$

❑ 모평균 μ의 신뢰구간: $\left(\overline{X} - t_{\alpha/2}\dfrac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2}\dfrac{s}{\sqrt{n}}\right)$

(2) 신뢰구간의 의미

❑ 모평균 μ의 신뢰구간은 표본에 따라 달라질 수 있다.

➧ 표본평균 $\bar{x}$, 신뢰구간의 길이 = $2t_{\alpha/2}\frac{s}{\sqrt{n}}$

❑ μ의 $100(1-\alpha)\%$ 신뢰구간

[예제 8.3]

우주선에 사용될 새로운 합금이 개발되었다. 이 합금 시편 15개를 임의추출하여 인장강도를 측정한 결과 표본평균과 표본표준편차가 각각 39.3, 2.6으로 나타났다. 합금의 인장강도는 정규분포를 따른다고 할 때, ① 이 합금의 평균인장강도 μ의 90% 신뢰구간을 구하라. ② μ가 ①에서 구한 구간에 포함되는가?

【① 풀이】

$n=15,\ d.f.=n-1=15-1=14,\ \bar{x}=39.3,\ s=2.6$

$t_{\alpha/2}(d.f.)=t_{0.1/2}(14)=t_{0.05}(14)=1.761$

< μ의 90% 신뢰구간 >

$$\left(\bar{x}-t_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x}+t_{\alpha/2}\frac{s}{\sqrt{n}}\right)=\left(\bar{x}-t_{0.1/2}\frac{s}{\sqrt{n}},\ \bar{x}+t_{0.1/2}\frac{s}{\sqrt{n}}\right)$$

$$=\left(\bar{x}-t_{0.05}\frac{s}{\sqrt{n}},\ \bar{x}+t_{0.05}\frac{s}{\sqrt{n}}\right)=\left(39.3-1.761\frac{2.6}{\sqrt{15}},\ 39.3+1.761\frac{2.6}{\sqrt{15}}\right)$$

$$=(38.12,\ 40.48)$$

【② 풀이】

$(38.12,\ 40.48)$과 같은 하나의 실현값이 μ를 포함하는지 아닌지는 결코 알 수 없다. 신뢰수준이란 표본들을 반복추출하여 구한 구간들 중에서 μ를 포함하는 구간의 비율을 말하는 것이다.

8.3.1 μ의 가설검정

(1) t-검정(t-test)

① 검정통계량(test statistic): $t=\frac{\overline{X}-\mu_0}{s/\sqrt{n}}$

② H_1의 형태와 이에 상응하는 기각역의 설정 (유의수준 α 하에서)

H_1	기각역
$H_1: \mu > \mu_0$	R: $T \geq t_\alpha$
$H_1: \mu < \mu_0$	R: $T \leq -t_\alpha$
$H_1: \mu \neq \mu_0$	R: $\lvert T \rvert \geq t_{\alpha/2}$

[예제 8.4]

식품의약품안정청에서는 어떤 생수의 단위량당 세균의 평균숫자가 안전수준인 200 이내인지를 조사하고자 한다. 한 조사원이 10개의 표본자료를 검사한 결과 세균의 수는 다음과 같았다.

175, 190, 215, 198, 184

207, 210, 193, 196, 180

이 자료에 의하면 이 생수는 안전하다고 할 수 있는가?

【풀이】

① 가설설정: $H_0: \mu = 200$, $H_1: \mu < 200$

② 검정통계량

$\alpha = 0.01$, $n = 10$, $d.f. = n-1 = 10-1 = 9$

$\overline{x} = 194.8$, $s = 13.14$

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{194.8 - 200}{13.14/\sqrt{10}} = -1.25$$

③ 기각치: $c = -t_\alpha(d.f.) = -t_{0.01}(9) = -2.821$

④ 결론

$c = -2.821$, $t = -1.25$ $\quad\Rightarrow c < t \quad\Rightarrow H_0$을 기각할 수 없다.

⇒ 10개의 관측값에 의하면 모평균이 안전수준 이내에 있다는 강력한 증거가 되지 못한다.

8.4 가설검정과 신뢰구간의 관계

(1) μ의 신뢰구간: $\left(\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right)$

(2) 가설검정: 기각역

$$H_0: \mu = \mu_0 \text{ vs. } H_1: \mu \neq \mu_0 \rightarrow R: \left|\frac{\overline{X} - \mu_0}{s/\sqrt{n}}\right| \geq t_{\alpha/2}$$

cf) 가설검정: 채택역

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0 \rightarrow \overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu_0 < \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

(3) 만일 어떤 귀무가설 μ_0가 $100(1-\alpha)\%$ 신뢰구간에 포함되면 유의수준 α에서 귀무가설이 채택된다. 따라서 μ의 $100(1-\alpha)\%$ 신뢰구간이 구해지면 이 구간에 포함되지 않는 모든 가능한 귀무가설 μ_0들이 유의수준 α에서 기각될 것이고 이 구간에 포함되는 귀무가설 μ_0들은 기각되지 않을 것이다.

[예제 8.5]

[예제 8.3]의 자료에서 구한 신뢰구간과 귀무가설 $H_0 : \mu = 40$과 대립가설 $H_1 : \mu \neq 40$을 유의수준 $\alpha = 0.1$에서 검정하여 비교하여라.

【풀이】

① 가설설정: $H_0 : \mu = 40$, $H_1 : \mu \neq 40$

② 검정통계량

$\alpha = 0.1$, $n = 15$, $d.f. = n-1 = 15-1 = 14$, $\overline{x} = 39.3$, $s = 2.6$

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{39.3-40}{2.6/\sqrt{15}} = -1.04$$

③ 기각치

$c = \pm t_{\alpha/2}(d.f.) = \pm t_{0.1/2}(14) = \pm t_{0.05}(14) = \pm 1.761$

④ 결론

$c_1 = -1.761$, $c_2 = 1.761$, $t = -1.04$ $\Rightarrow$ $c_1 < t < c_2$ $\Rightarrow$ H_0을 기각할 수 없다.

8.5 표준편차 σ에 대한 추론

(1) 카이제곱분포(χ^2-distribution)

$X_1, X_2, \cdots, X_n$: 정규모집단 $N(\mu, \sigma^2)$으로부터의 확률표본
$\chi^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)}{\sigma^2/(n-1)} = \dfrac{(n-1)s^2}{\sigma^2}$: 자유도 $n-1$인 χ^2분포
여기서, $s^2 = \dfrac{\sum(X_i - \overline{X})^2}{n-1}$

(2) 카이제곱분포의 확률밀도곡선의 성질

$P\left(\chi^2 \geq \chi^2_{\alpha}(r)\right) = \alpha$ 여기서, $r = n - 1$: 자유도

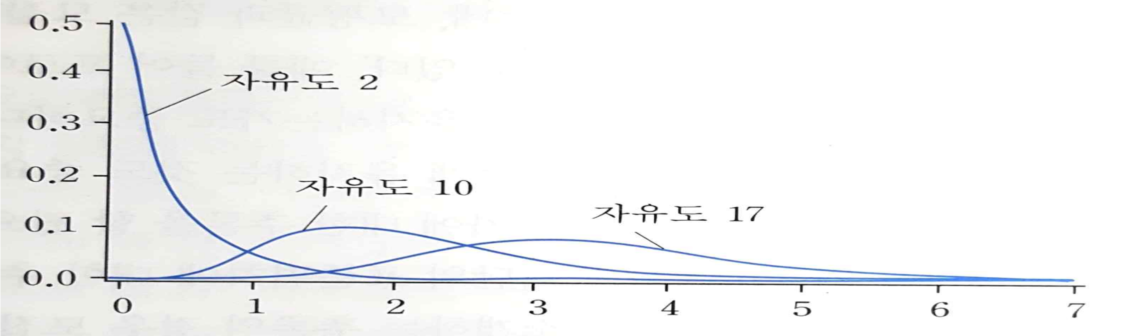


그림 11-5 자유도 2, 10, 17인 χ^2 분포

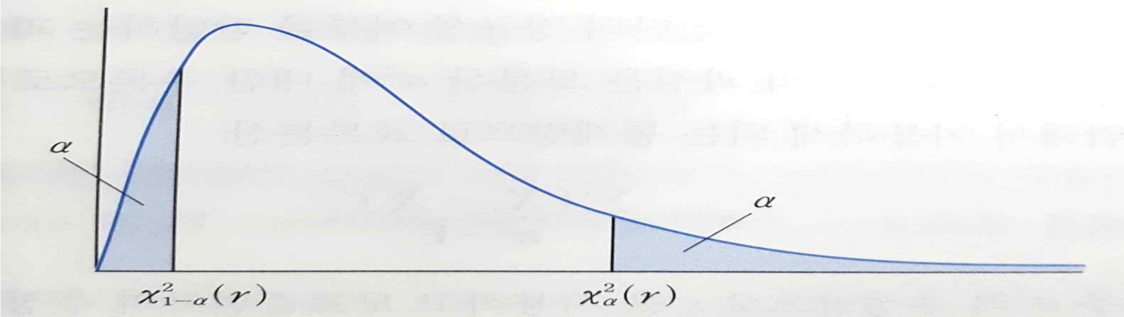


그림 11-6 상, 하위 α의 확률을 주는 값

① 양의 값만 가진다.

② 비대칭이고 오른쪽으로 긴 꼬리를 가진다.

③ 확률밀도곡선의 모양은 자유도에 따라 달라진다.

예제 6 χ^2 분포표에서 자유도가 17인 χ^2 분포의 상, 하위 5%의 확률을 주는 값을 찾아라.

풀이 ▸ 아래와 같이 발췌된 χ^2 분포표로부터 상위 5%의 확률을 주는 값은 $\chi^2_{0.05}(17) = 27.59$ 이고, 하위 5%의 확률을 주는 값은 $\chi^2_{0.95}(17) = 8.67$이다.

자유도 \ α	⋯	0.95	⋯	0.05
⋮		⋮		⋮
17	⋯	8.67	⋯	27.59
⋮	⋯	⋮	⋯	⋮

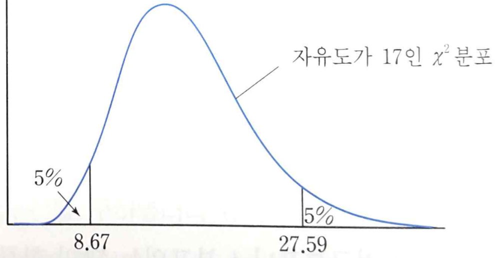


그림 11-7 상, 하위 5%의 확률을 주는 값

[예제 8.6]

자유도가 9인 카이제곱분포의 상위 5백분위수와 하위 5백분위수를 구하라.

【풀이】

$$\chi^2_{\alpha}(d.f.) = \chi^2_{0.05}(9) = 16.92$$

$$\chi^2_{\alpha}(d.f.) = \chi^2_{0.95}(9) = 3.33$$

(3) σ^2에 대한 $100(1-\alpha)\%$ 신뢰구간

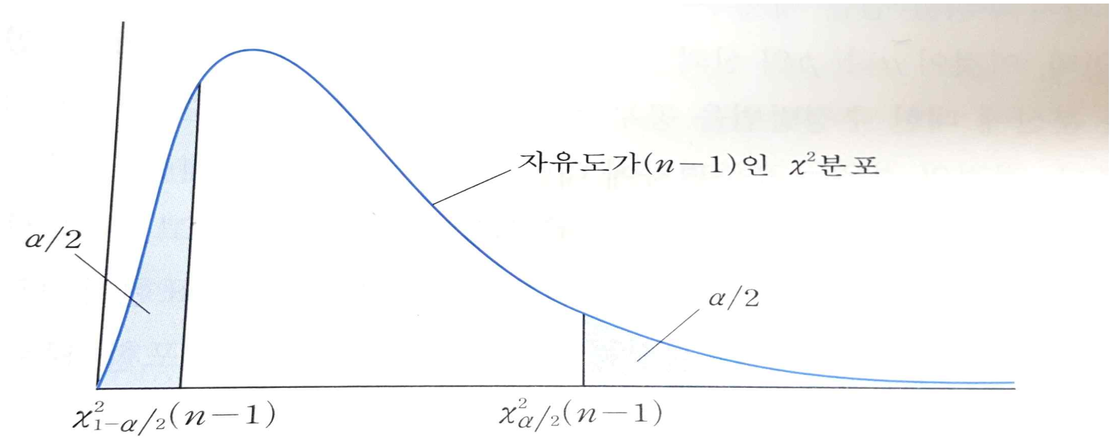


그림 11-8 $\chi^2(n-1)$ 분포에서 상, 하위 $\alpha/2$의 확률을 주는 값

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

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

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

□ σ^2에 대한 $100(1-\alpha)\%$ 신뢰구간: $\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{1-(\alpha/2)}(n-1)}\right]$

□ σ에 대한 $100(1-\alpha)\%$ 신뢰구간: $\left[s\sqrt{\frac{n-1}{\chi^2_{\alpha/2}(n-1)}}, s\sqrt{\frac{n-1}{\chi^2_{1-(\alpha/2)}(n-1)}}\right]$

[예제 8.7] 볼트와 너트를 생산하는 한 공장에서는 제품의 품질이 얼마나 균일하게 유지되는지를 검사하려고 10개의 볼트를 추출하여 지름을 측정하고 그 표준편차를 구하였더니 0.4였다. 그 공장에서 생산되는 볼트의 지름이 정규분포를 따른다는 가정 하에 σ의 90% 신뢰구간을 구하라.

【풀이】

$\alpha=0.1,\ \alpha/2=0.1/2=0.05,\ n=10,\ d.f.=n-1=10-1=9,\ s=0.4$

σ^2의 90% 신뢰구간:

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)},\ \frac{(n-1)s^2}{\chi^2_{1-(\alpha/2)}(n-1)}\right]=\left[\frac{(n-1)s^2}{\chi^2_{0.05}(9)},\ \frac{(n-1)s^2}{\chi^2_{0.95}(9)}\right]=\left[\frac{9\times0.4^2}{16.92},\ \frac{9\times0.4^2}{3.33}\right]$$

$$=(0.085,\ 0.432)$$

σ의 90% 신뢰구간: $\left(\sqrt{0.085},\ \sqrt{0.432}\right)=(0.29,\ 0.66)$

(4) σ^2에 대한 가설검정

모표준편차 σ에 대한 추론 (정규모집단일 때)

자료: 표준편차가 σ인 정규모집단으로부터 추출한 $X_1,\ \cdots,\ X_n$

σ의 추정량: $s=\sqrt{\dfrac{\sum(X_i-\overline{X})^2}{n-1}}$

σ에 대한 $100(1-\alpha)\%$ 신뢰구간: $\left(s\sqrt{\dfrac{n-1}{\chi^2_{\alpha/2}(n-1)}},\ s\sqrt{\dfrac{n-1}{\chi^2_{1-\alpha/2}(n-1)}}\right)$

가설 $H_0:\sigma=\sigma_0$에 대한 검정:

검정통계량 : $\chi^2=\dfrac{(n-1)s^2}{\sigma_0^2}$

각 대립가설에 대한 기각역 :

$H_1:\sigma>\sigma_0$일 때 $\quad R:\chi^2\geq\chi^2_{\alpha}(n-1)$

$H_1:\sigma<\sigma_0$일 때 $\quad R:\chi^2\leq\chi^2_{1-\alpha}(n-1)$

$H_1:\sigma\neq\sigma_0$일 때 $\quad R:\chi^2\geq\chi^2_{\alpha/2}(n-1)$ 혹은 $\chi^2\leq\chi^2_{1-\alpha/2}(n-1)$

[예제 8.8] 앞의 예제 8.7에서 σ가 0.2보다 크다고 할 수 있는지 유의수준 0.05로 검정하라.

【풀이】

① 가설설정: $H_0 : \sigma = 0.2$, $H_1 : \sigma > 0.2$ (단측검정)

② 검정통계량: $\alpha = 0.05$, $n = 10$, $d.f. = n - 1 = 10 - 1 = 9$, $s = 0.4$

$$\chi^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2/(n-1)}{\sigma^2/(n-1)} = \frac{(n-1)s^2}{\sigma^2} = \frac{9 \times 0.16}{0.04} = 36$$

③ 기각치: $c = \chi^2_{\alpha}(n-1) = \chi^2_{0.05}(9) = 16.92$

④ 결론

검정통계량 $\chi^2 = 36$은 기각역에 포함되므로, 귀무가설을 기각하고 대립가설이 맞다고 말할 수 있다. 따라서 주어진 자료로부터 σ가 0.2보다 크다고 할 수 있다.