8.7절~8.8절
7장 연습문제풀이

12주1강 가설과 검정4

8.7절

8.7 두 모분산의 동일성에 대한 검정

F-분포

- **F-분포**는 두 개의 독립인 χ^2-분포의 비율에 의하여 정의되는 확률분포이다.
- 두 확률변수 X_1과 X_2가 서로 독립이며 각각 자유도가 v_1 과 v_2 를 갖는 χ^2-분포를 따를 때, 확률변수 F는 다음과 같이 정의된다.

$$X_1 \sim \chi^2_{(v_1)} \ , \ X_2 \sim \chi^2_{(v_2)} \qquad F = \frac{X_1 / v_1}{X_2 / v_2} \sim F_{(v_1, v_2)}$$

두 모분산의 동일성 검정

- 두 모분산의 동일성에 대한 검정은 두 모집단의 분포가 서로 독립인 정규분포를 따른다는 것을 전제로 한다.
- 두 모평균의 동일성에 대한 검정에서 두 집단의 모분산이 동일한 경우와 동일하지 않은 경우의 검정통계량의 형태가 서로 다름을 알 수 있다.
- 즉 두 모분산의 동일성에 대한 검정은 두 모평균의 동일성에 대한 검정에 앞서 실시하여야 한다.

① 가설의 설정

(a) 양측검정 : $H_0 : \sigma_1^2 = \sigma_2^2,\ H_1 : \sigma_1^2 \neq \sigma_2^2$

(b) 단측검정 : $H_0 : \sigma_1^2 = \sigma_2^2,\ H_1 : \sigma_1^2 > \sigma_2^2$(또는 $\sigma_1^2 < \sigma_2^2$)

② 검정통계량과 분포

검정에 필요한 검정통계량과 분포는 다음과 같다.

$$F = \frac{S_1^{\ 2}}{S_2^{\ 2}} \sim F_{(n_1-1,\, n_2-1)}$$

③ 검정

* 단측검정인 경우 유의수준은 α로 하여 기각역을 설정한다.
* 양측검정인 경우에는 유의수준을 2α로 하여 기각역을 설정한다.

검정통계량이 기각역에 속하면 귀무가설을 기각한다.

예8-14.

흰쥐를 이용한 실험에서 쥐의 몸무게가 유사한 것이 바람직하다고 한다. 두 공급원으로부터 공급된 흰쥐의 무게가 동일한 형태의 분포를 갖는가를 보기 위하여 조사한 결과가 다음 표와 같다. 이 통계자료에 의할 때 두 공급원에 의하여 공급되는 흰쥐의 몸무게의 분산이 같다고 볼 수 있는지를 10% 유의수준 하에서 검정하라.

통계 \ 공급원	공급원 1	공급원 2
표본수(n)	13	18
표본평균($\bar{X}$)	4.21g	4.18g
표본분산(S^2)	0.049g	0.019g

(sol)

(1) 가설의 설정 : 공급원 1에 의하여 공급된 흰 쥐의 몸무게의 분산을 σ_1^2, 공급원 2에 의하여 공급된 흰쥐의 몸무게의 분산을 σ_2^2이라고 하면 두 모분산이 동일한가를 검정하고자 하므로 귀무가설과 대립가설은 각각 다음과 같이 설정할 수 있다.

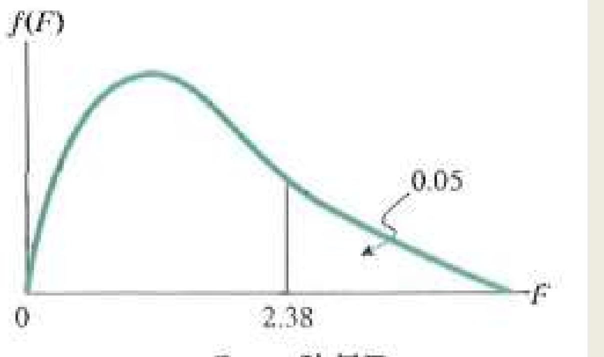
$$H_0 : \sigma_1^2 = \sigma_2^2, \qquad H_1 : \sigma_1^2 \neq \sigma_2^2$$

(2) 검정통계량과 분포 : 표본분산이 각각 $S_1{}^2 = 0.049$ 와 $S_2{}^2 = 0.019$ 이므로 귀무가설 하에서의 검정통계량의 값은 다음과 같이 계산된다.

$$F = \frac{S_1^2}{S_2^2} = \frac{0.049}{0.019} = 2.58$$

F는 자유도가 $(n_1 - 1, n_2 - 1) = (12, 17)$인 F-분포를 따른다.

(3) 검정 : 양측검정이므로 α=0.10일 때의 기각역은 부록V의 [표 6]에서 $\alpha = 0.05$인 표에서 구할 수 있다. $v_1 = 12$, $v_2 = 17$ 일 때, $F_{0.05} = 2.38$ 이므로 한쪽 기각역은 아래 그림과 같다.



$F_{(12,\ 17)}$의 분포

검정통계량의 값 $F = 2.58$은 기각역에 속하므로 귀무가설은 기각된다.
즉 10% 유의수준 하에서 두 집단의 모분산이 동일하다고 볼 수 없다.

예8-15.

주식투자자가 투자종목을 설정하는 데 있어서 두 종목의 주식을 고려한다. 주식 1은 주식 2에 비하여 가격이 높은데, 일일 가격의 변화량이 주식 2보다 크기 때문에 위험한 투자라고 판단된다. 과연 주식 1의 일일가격 변화량이 주식 2의 변화량 보다 큰가를 분석하기 위하여 두 주식의 일일가격 변화량을 25일 동안 관찰한 결과 다음과 같은 통계를 구하였다. 이 통계를 이용하여 주식 1의 분산이 주식 2의 분산보다 큰가를 유의수준 5%하에서 검정하라.

통계 \ 주식	주식 1	주식 2
표본의 수(n)	25	25
표본평균($\overline{X}$)	250	125
표본분산(S^2)	76	46

(sol)

(1) 가설의 설정

$\sigma_1{}^2$ =주식 1의 가격의 분산, $\sigma_2{}^2$ =주식 2의 가격의 분산

이라고 할 때 $\sigma_1{}^2 > \sigma_2{}^2$ 인가를 검정하고자 하므로 귀무가설과 대립가설은 각각 다음과 같이 설정할 수 있다.

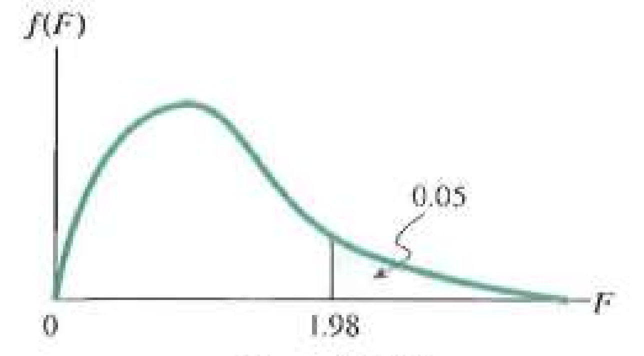
$$H_0 : \sigma_1^2 = \sigma_2^2, \qquad H_1 : \sigma_1^2 > \sigma_2^2$$

(2) 검정통계량과 분포 : $S_1^2 = 76, S_2^2 = 46$이므로 귀무가설 하에서 검정통계량 값은 다음과 같이 계산된다.

$$F = \frac{S_1^2}{S_2^2} = \frac{76}{46} = 1.652$$

F는 자유도가 $v_1 = 24, v_2 = 24$ 인 F-분포를 따른다.

(3) 검정 : 부록 V의 [표 6]에서 $v_1 = v_2 = 24$ 일 때 $F_{0.05} = 1.98$ 이므로 단측검정에서 $\alpha = 0.05$ 일 때의 기각역은 아래 그림의 빗금 친 부분과 같다.



$F_{(24,\ 24)}$의 분포

검정통계량의 값 $F = 1.652$는 기각역에 속하지 않으므로 귀무가설은 기각되지 않는다. 즉 5% 유의수준 하에서 두 주식의 분산은 동일하다고 볼 수 있다.

8.8절

8.8 유의수준과 P-값

유의수준과 P-값

① 유의수준

가설검정이란 귀무가설하에서 구한 검정통계량 값의 희귀성에 의하여 귀무가설의 채택여부를 결정하는 데 이 희귀성이 어느 정도일 때 귀무가설을 기각할 것인가를 판단하는 기준이 **유의수준 α** 이다.

② P-값

P-값이란 귀무가설하에서 검정통계량의 값이 나타날 가능성을 측정하는 확률로 귀무가설의 채택여부를 결정하는 기준인 유의수준 α와는 직접적인 관계가 있다.

P-값이 α보다 작으면 귀무가설은 기각되며, P-값이 α보다 크면 귀무가설은 기각되지 않는다.

③ 유의수준과 P-값의 관계

가설검정에서 P-값과 유의수준 α와는 다음과 같은 관계가 있다.

P-값 $\geq \alpha$: 귀무가설 채택

P-값 $< \alpha$: 귀무가설 기각

예 8-16 다음의 주어진 가설의 검정에서 검정통계량의 값이 $T(X) = 2.05$ 이며 $T(X) \sim N(0,1)$이라고 할 때 귀무가설은 기각될 수 있는가?

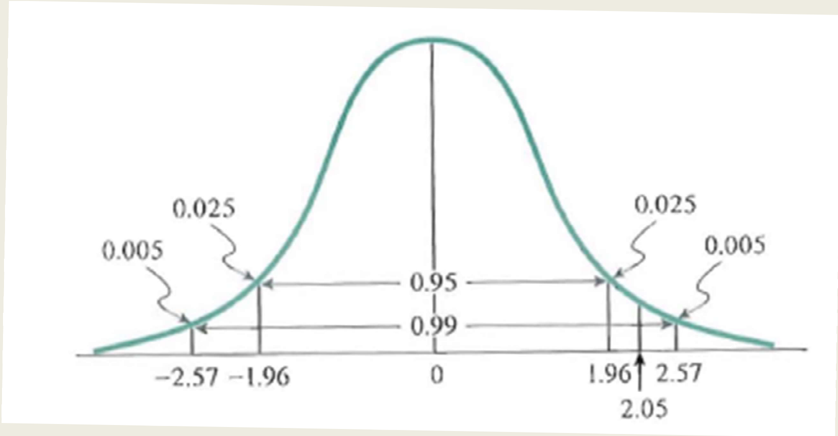
$$H_0 : \mu_1 = \mu_2, \qquad H_1 : \mu_1 \neq \mu_2$$

(sol) 정답은 “예”와 “아니오” 모두 맞는다. 이 검정은 양측검정이므로 표준정규분포표에 의하여 유의수준 α와 기각역은 다음과 같은 관계가 있다.

$\alpha = 0.05 \iff$ 기각역 : $|T(X)| > 1.96$

$\alpha = 0.01 \iff$ 기각역 : $|T(X)| > 2.57$

이것을 그림으로 표현하면 다음과 같다.



따라서 $\alpha = 0.05$인 경우에는 귀무가설은 기각되지만 $\alpha = 0.01$인 경우에는 귀무가설은 기각되지 않는다.

7장 연습문제풀이

01. 다섯 개의 값 $\{1, 2, 3, 4, 5\}$에서 세 개의 원소로 이루어진 표본을 추출하는 경우

(a) 표본을 비복원으로 추출할 때 가능한 표본의 수는 몇 개인가?

풀이) $\binom{5}{3} = 10$가지

(b) 모집단 {1, 2, 3, 4, 5}의 평균 μ, 분산 σ^2, 그리고 표준편차 σ를 구하라.

풀이) $\mu = \frac{1}{5}(1 + 2 + 3 + 4 + 5) = 3$

$$\sigma^2 = \frac{1}{5}[(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2)]$$

$$= \frac{1}{5}(4 + 1 + 1 + 4) = 2$$

$$\sigma = \sqrt{2} = 1.414$$

(c) (a)에서 구한 모든 가능한 표본들을 모두 나열하고, 각 표본에서 평균 $\overline{X}$, 중위수 $\tilde{X}$, 그리고 최솟값 $X_{(1)}$을 구하라.

풀이) $\{(1,2,3),(1,2,4),(1,2,5),(1,3,4),(1,3,5),(1,4,5),(2,3,4),(2,3,5),(2,4,5),(3,4,5)\}$

표본	관측값	$\overline{X}$	$\tilde{X}$	$X_{(1)}$
1	$(1,2,3)$	2	2	1
2	$(1,2,4)$	$7/3$	2	1
3	$(1,2,5)$	$8/3$	2	1
4	$(1,3,4)$	$8/3$	3	1
5	$(1,3,5)$	3	3	1
6	$(1,4,5)$	$10/3$	4	1
7	$(2,3,4)$	3	3	2
8	$(2,3,5)$	$10/3$	3	2
9	$(2,4,5)$	$11/3$	4	2
10	$(3,4,5)$	4	4	3

(d) (c)에서 구한 통계량 $\bar{X}, \tilde{X}, X_{(1)}$의 값들을 이용해 각각의 통계량에 대한 평균, 분산, 표준오차를 구하라.

풀이) (i) $\bar{X}$의 평균$= 3$ 분산≈ 0.37 표준오차≈ 0.1825

(ii) $\tilde{X}$의 평균$= 3$ 분산$\approx 2/3$ 표준오차≈ 0.2582

(iii) $X_{(1)}$의 평균$= 1.5$ 분산≈ 0.45 표준오차≈ 0.2236

(e) 세 통계량 $\bar{X}, \tilde{X}, X_{(1)}$ 중에서 μ의 불편추정량은 어느 것인가?

풀이) $\bar{X}$와 $\tilde{X}$ 모두 불편추정량이다.

(f) 세 통계량 $\bar{X}, \tilde{X}, X_{(1)}$ 중에서 최소분산을 갖는 통계량은 어느 것인가?

풀이) 해답. $\bar{X}$가 최소분산을 갖는다.

04. $\overline{X}$가 $\mu = 50$, $\sigma = 10$인 확률분포로부터 추출된 100개 관측값의 표본평균이라 할 때 다음 값을 구하라.

$X \sim N\left(50, \left(\frac{10}{\sqrt{100}}\right)^2\right),$ 표준오차= $\frac{10}{\sqrt{100}} = 1$

(a) $\boldsymbol{P_r(\overline{X} < 52)}$

풀이) $P_r(\bar{X} < 52) = P_r\left(Z < \frac{52-5}{1}\right) = P_r(Z < 2) = 0.5 + 0.4772 = 0.9772$

(b) $\boldsymbol{P_r(|\overline{X}| > 51)}$

풀이) $P_r(|\bar{X}| > 51) = P_r\left(Z < \frac{-51-5}{1}, Z > \frac{51-50}{1}\right) = P_r(Z < -101, Z > 1)$

(c) $\boldsymbol{P_r(|\overline{X} - 50| < 1)}$

풀이) $P_r(|\bar{X} - 50| < 1) = P_r(|Z| < 1) = 2 \times 0.3413 = 0.6826$

04. $\overline{X}$가 $\mu = 50$, $\sigma = 10$인 확률분포로부터 추출된 100개 관측값의 표본평균이라 할 때 다음 값을 구하라.

$$X \sim N\left(50, \left(\frac{10}{\sqrt{100}}\right)^2\right), \qquad \text{표준오차} = \frac{10}{\sqrt{100}} = 1$$

(d) $\boldsymbol{P_r(\overline{X} > K) = 0.975}$**인** $\boldsymbol{K}$**값**

풀이) $P_r(\overline{X} > K) = P_r\left(Z > \frac{K-50}{1}\right) = 0.975$

$K - 50 = -1.96 \Rightarrow K = 50 - 1.96 = 48.04$

(e) $\boldsymbol{P_r(|\overline{X} - 50| > K) = 0.10}$**인** $\boldsymbol{K}$**값**

풀이) $P_r(|\overline{X} - 50| > K) = P_r\left(|Z| > \frac{K}{1}\right) = 0.10, K = 1.645$

05. 한 농구선수는 자유투 성공률이 60%라고 한다. 이 선수가 자유투를 100회 던졌을 때 성공한 횟수의 비율을 $\hat{P}$라 할 때 다음 확률을 구하라.

(a) $\boldsymbol{P_r(\hat{P} \geq 0.61)}$

풀이) $P_r(\hat{P} \geq 0.61) = P_r\left(\frac{\hat{P}-P}{\sqrt{\frac{P(1-P)}{n}}} \geq \frac{0.61-0.6}{0.049}\right)$

$= P_r(Z \geq 0.20) = 0.5 - 0.0793 = 0.4207$

(b) $\boldsymbol{P_r(0.58 \leq \hat{P} \leq 0.61)}$

풀이) $P_r(0.58 \leq \hat{P} \leq 0.61) = P_r\left(\frac{0.58-0.6}{0.049} \leq Z \leq \frac{0.61-0.6}{0.049}\right)$

$= P_r(-0.41 \leq Z \leq 0.20)$

$= 0.1591 + 0.0793 = 0.2384$

05. 한 농구선수는 자유투 성공률이 60%라고 한다. 이 선수가 자유투를 100회 던졌을 때 성공한 횟수의 비율을 $\hat{P}$라 할 때 다음 확률을 구하라.

(c) $\boldsymbol{P_r\left(\hat{P} \leq 0.60\right)}$

풀이) $P_r\left(\hat{P} \leq 0.60\right) = P_r\left(Z \leq \frac{0.6-0.6}{0.049}\right) = P_r(Z \leq 0) = 0.5$

(d) $\boldsymbol{P_r(\left|\hat{P} - 0.60\right| \geq 0.01)}$

풀이) $P_r\left(\left|\hat{P} - 0.60\right| \geq 0.01\right) = P_r\left(|Z| \geq \frac{0.01}{0.049}\right) = P_r(|Z| \geq 0.20)$

$= 2 \times (0.5 - 0.0793) = 2 \times 0.4207 = 0.8414$

08. 확률변수 X가 다음 각각의 자유도($d.f.$)를 갖는 t분포를 따를 때, <부록 V>의 [표 4]를 참고해 다음 각각의 경우에 대한 값을 구하라.

(a) $\mathbf{df = 10}, \boldsymbol{P_r(X \leq 1.812)}$

풀이) $P_r(X \leq 1.812) = 0.95$

(d) $\mathbf{df = 20}, \boldsymbol{P_r(-2.086 \leq X \leq 2.086)}$

풀이) $P_r(-2.086 \leq X \leq 2.086) = 0.95$

(c) $\mathbf{df = 10}, \boldsymbol{P_r(X \leq -2.764)}$

풀이) $P_r(X \leq -2.764) = 0.10$

08. 확률변수 X가 다음 각각의 자유도($d.f.$)를 갖는 t분포를 따를 때, <부록 V>의 [표 4]를 참고해 다음 각각의 경우에 대한 값을 구하라.

(d) $\mathbf{df = 19, P_r(X \geq 2.093)}$

풀이) $P_r(X \geq 2.093) = 0.025$

(e) $\mathbf{df = 15, P_r(X \geq K) = 0.025}$ **인** $\boldsymbol{K}$**값**

풀이) $P_r(X \geq K) = 0.025$에서 $K = 2.131$

(f) $\mathbf{df = 10, P_r(X \geq -K) = 0.100}$ **인** $\boldsymbol{K}$**값**

풀이) $P_r(X \geq -K) = 0.100$에서 $K = 1.372$

11. 모평균과 모분산이 각각 $\mu = 50, \sigma^2 = 25$인 정규모집단으로부터 10개의 표본을 관측한 결과 표본평균과 표본분산이 각각 $\bar{X} = 52, S^2 = 38$과 같이 구해졌다.

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(9)}$$

(a) 표본분산 S^2의 분포에 대한 X^2의 값을 구하라.

풀이) $X^2 = \frac{(10-1)\times 38}{25} = \frac{9\times 38}{25} = 13.68$

(b) <부록 V>의 [표 5]를 이용해 이 X^2값 이상이 나타날 확률의 근사값을 구하라.

풀이) $X^2 \sim \chi^2_{(9)}$이므로 <부록 V>의 [표 5]에서 $P_r(X^2 \geq 13.68) \approx 0.10$이다.

12. 한 정규모집단으로부터 관측한 10개 표본값의 평균이 $\bar{X} = 50$이었다. 이 모집단의 모분산이 $\sigma^2 = 25$임을 알고 있을 때 모평균 μ에 대한 95% 신뢰구간을 구하라.

풀이) μ에 대한 95% 신뢰구간은 σ^2을 알고 있으므로 $\left(\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}\right)$이다.

$$\therefore \left(50 \pm 1.96\frac{5}{\sqrt{10}}\right) = (50 \pm 1.96 \times 1.581) = (50 \pm 3.099) = (46.901, 53.099)$$

15. 한강물에 있는 대장균의 분포를 파악하기 위해 한강의 50개 지역으로부터 물을 임의로 취수해 대장균의 수를 측정한 결과 물 1ml당 대장균의 수의 평균과 분산이 각각 $\bar{X} = 150, S^2 = 100$이었다. 한강물 1ml당 대장균 수의 평균에 대한 95% 신뢰구간을 구하라. 또한 99% 신뢰구간을 구하라.

풀이) $\bar{X} = 150, S^2 = 100, n = 50$이므로 정규분포를 이용한다.

(i) 95% 신뢰구간 :

$$\left(\bar{X} \pm 1.96 \frac{S}{\sqrt{n}}\right) = \left(150 \pm 1.96 \frac{10}{\sqrt{50}}\right)$$
$$= (150 \pm 1.96 \times 1.414)$$
$$= (150 \pm 2.77) = (147.23, 152.77)$$

(ii) 99%의 신뢰구간 : 위의 신뢰구간에서 신뢰계수를 2.57로 바꾸면 된다.

$$(150 \pm 2.57 \times 1.414) = (150 \pm 3.63) = (146.37, 153.63)$$

17. 지난 여름에 16일 동안 서울지역의 최고온도를 측정한 결과가 다음과 같다.

30	28	36	34	32	35	29	30
33	37	30	28	33	34	28	36

서울지역 여름의 일별 최고온도가 정규분포를 따른다고 할 때 최고온도의 평균에 대한 95% 신뢰구간을 구하라.

풀이) 평균 $\bar{X} = 32.06$, 표준편차 $S = 3.108$이다.

(i) $n = 16$이므로 자유도 15인 t분포에서 $t_{(0.025;\, d.f.=1\ \ )} = 2.131$이다.

∴ 95% 신뢰구간은

$$\left(32.06 \pm 2.131 \times \frac{3.108}{\sqrt{16}}\right) = (32.06 \pm 2.131 \times 0.777)$$

$$= (32.06 \pm 1.656)$$

$$= (30.404, 33.716)$$

20. 한 국회의원 후보자가 자신에 대한 선거구민들의 지지도를 조사하기 위해 선거구민 100명을 임의로 추출해 조사한 결과 49명이 지지한다고 응답했다.

(a) 이 후보에 대한 지지율의 95% 신뢰구간을 구하라.

풀이) $\hat{P} = 0.49$이므로 95% 신뢰구간은

$$\left(\hat{P} \pm 1.96\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right) = \left(0.49 \pm 1.96\sqrt{\frac{0.49\times0.51}{100}}\right) = (0.49 \pm 1.96 \times 0.05)$$

$$= (0.49 \pm 0.098) = (0.392, 0.588)$$

(b) 그 지역에서는 후보가 두 사람만 출마한다고 할 때, (a)의 결과에 의해 위의 후보가 당선될 가능성이 있는가를 판단하라.

풀이) 그 후보에 대한 실제 지지율의 95% 신뢰구간은 $(0.392, 0.588)$로 50%를 포함하고 있으나 신뢰구간 안에 50% 미만도 포함되어 있으므로 단정적으로 그 후보가 당선가능성이 있다고 말할 수 없다.

23. [연습문제 17]에 주어진 자료를 이용해 모분산 σ^2에 대한 95% 신뢰구간을 구하라.

풀이) [연습문제 17]에서 $S^2 = (3.108)^2 = 9.6597$이다. 자유도 15인 χ^2분포에서

$X_{0.975}^2 = 6.26214, X_{0.025}^2 = 27.4884$이다.

$\therefore$ σ^2에 대한 95% 신뢰구간은

$$\left(\frac{15 \times 9.6597}{27.4884}, \frac{15 \times 9.6597}{6.26214}\right) = (5.2711, 23.1383)$$

끝~~ ♥♥