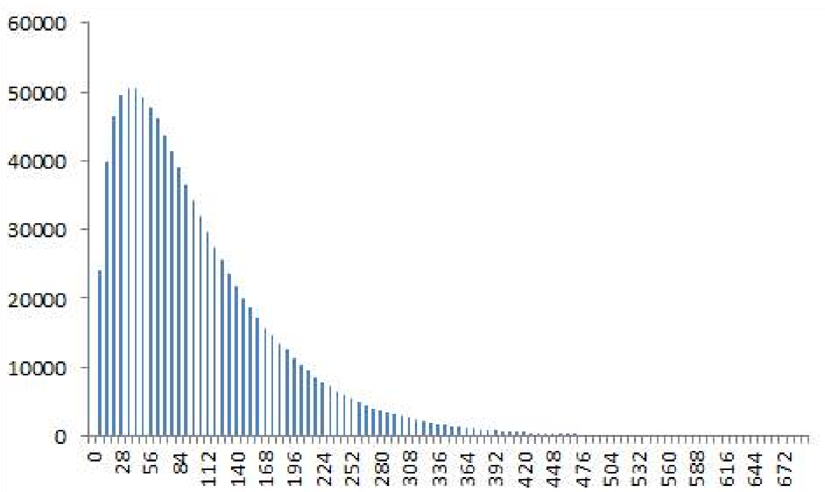
제 11 장 모집단 분산의 추론

실험 1. $x \sim N(100,\ 10^2)$, n=4; 실험횟수 1,000,000

	x_1	x_2	x_3	x_4	$\overline{x}$	S^2
1	130.05	104.20	94.44	116.26	111.24	236.86
2	105.30	94.40	107.87	103.61	102.80	34.37
3	84.79	101.77	96.08	96.33	94.74	50.94
4	96.35	99.32	107.02	117.35	105.01	87.89
5	80.56	107.91	108.23	94.27	97.74	173.54
6	108.84	106.54	105.19	104.51	106.27	3.65
7	105.71	90.55	115.92	93.96	101.54	134.12
8	108.55	119.81	102.84	89.64	105.21	157.49
9	90.45	96.07	100.07	85.96	93.14	38.44
10	103.36	96.42	99.07	106.35	101.30	19.52
…						
1,000,000	89.58	112.11	97.90	108.86	102.11	106.73
평균					99.99964	99.87746

S^2의 도수분포그래프



$\overline{X}$와 S^2의 백분위수

	$\overline{x}$	S^2
0.00%	76.5	0.0
10.0%	93.6	19.5
20.0%	95.8	33.6
30.0%	97.4	47.5
40.0%	98.7	62.4
50.0%	100.0	79.0
60.0%	101.3	98.3
70.0%	102.6	122.3
80.0%	104.2	154.8
90.0%	106.4	208.5
100.0%	124.2	1,027.4

	$\overline{x}$	S^2
최소	72.85	0.01
중앙	100.00	78.88
최대	124.44	1,129.66

연습문제 1.

문제 1. $\overline{x}$와 s^2의 분포의 특징에 대해 논하시오. (40.0% 줄의 첫 번째 값 98.7은 1,000,000개의 $\overline{x}$ 값들 중 하위 40%에 해당하는 값이 98.7, 두 번째 값 62.4는 1,000,000개의 s^2 값들 중 하위 40%에 해당하는 값이 62.4임을 의미한다.)

문제 2. s^2의 (이론적) 최소값은?

제1절 단일 모집단 분산의 추론

1. 카이자승 (chi-square) 분포

가정: $X \sim N(\mu_X, \sigma_X^2)$

확률변수 χ_{n-1}^2 (chi-square)의 정의

$$\chi_{n-1}^2 \equiv \frac{(n-1)S^2}{\sigma^2}, \text{ where } n = \text{sample size}$$

Note 1. 모집단이 정규분포 (또는 정규분포에 근사한 경우)가 아니면, $\frac{(n-1)S^2}{\sigma^2}$는 χ^2 분포를 갖지 못한다.

Note 2. $\chi_{n-1}^2 \equiv \frac{(n-1)S^2}{\sigma^2}$에서 변수는 S^2 하나이며, n과 σ^2은 상수이다. χ^2 분포는

S^2의 분포와 닮은 꼴이며, 단지 scale을 $\dfrac{(n-1)}{\sigma^2}$으로 조정한 것 뿐이다.

실험 2. s^2와 χ^2_{n-1}의 관계: $x_1 \sim N(100, 10^2)$, $x_2 \sim N(200, 20^2)$, $n_1 = n_2 = 4$; 실험횟수 10,000

	모집단 1 $\sigma_1^2 = 10^2$		모집단 2 $\sigma_2^2 = 20^2$		χ^2 이론값
	s_1^2	χ^2	s_2^2	χ^2	
0.0%	0.04	**0.0011**	0.48	**0.0036**	**-**
10.0%	19.54	**0.5861**	78.18	**0.5863**	**0.5844**
20.0%	33.33	**0.9999**	135.65	**1.0173**	**1.0052**
30.0%	47.17	**1.4151**	191.88	**1.4391**	**1.4237**
40.0%	61.78		252.93		**1.8692**
50.0%	78.76	**2.3627**	320.32	**2.4024**	**2.3660**
60.0%	98.34	**2.9503**	397.79	**2.9834**	**2.9462**
70.0%	121.02	**3.6307**	494.47	**3.7085**	**3.6649**
80.0%	152.35	**4.5706**	628.33	**4.7125**	**4.6416**
90.0%	205.84	**6.1751**	842.82	**6.3212**	**6.2514**
95.0%	259.57		1,047.08		**7.8147**
100.0%	691.98	**20.7594**	2,484.51	**18.6339**	**77.3963**

연습문제 2.

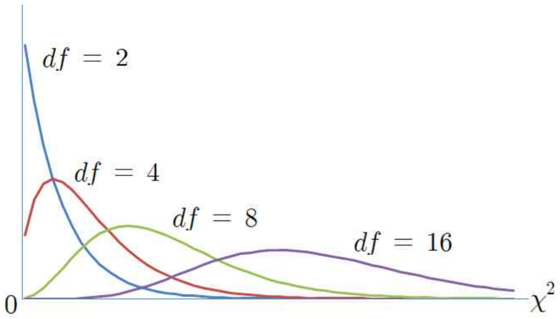
문제 1. 모집단 1에서 하위 40%에 해당하는 χ^2값과, 모집단 2에서 하위 40%에 해당하는 χ^2을 구하시오. (단, $n = 4$이다.)

문제 2. 모집단 1에서 하위 95%에 해당하는 χ^2값과, 모집단 2에서 하위 95%에 해당하는 χ^2을 구하시오. (단, $n = 4$이다.)

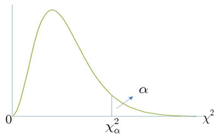
문제 3. 위의 표는 무엇을 방증하는가?

χ^2_{n-1} 분포의 특징

비대칭이고 오른쪽으로 꼬리 → 자유도(df, $n-1$)가 증가할수록 정규분포에 가깝게 된다. 또한 모든 χ^2의 값은 항상 양수이다.



χ^2_{n-1} **분포표**



(여기서 α는 하위가 아닌 상위 개념이다.)

					α				
ν	0.99	0.975	0.95	0.9	0.5	0.1	0.05	0.025	0.01
1	0.000	0.001	0.004	0.016	0.455	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	1.386	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	2.366	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	3.357	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	4.351	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	5.348	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	6.346	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	7.344	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	8.343	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	9.342	15.987	18.307	20.483	23.209
19	7.633	8.907	10.117	11.651	18.338	27.204	30.144	32.852	36.191

연습문제 3.

문제 1. 모집단은 정규분포를 따른다. n = 4일 때, 상위 5%에 해당하는 χ^2값은?

문제 2. 모집단은 정규분포를 따른다. n = 4일 때, 하위 5%에 해당하는 χ^2값은?

문제 3. 모집단은 정규분포를 따르고 σ^2 = 50이다. n = 6인 표본을 구성하여 s^2의 값을 구하고자 한다. 상위 10%에 해당하는 s^2의 값은?

문제 4. 모집단은 정규분포를 따르고 σ^2 = 50이다. n = 6인 표본을 구성하여 s^2의 값을 구하고자 한다. 하위 10%에 해당하는 s^2의 값은?

2. 분산의 추정 및 가설검정

(1) σ^2에 대한 (1-α)100% 신뢰구간

① 가정: $X \sim N(\mu_X, \sigma_X^2)$

② 신뢰구간: $\dfrac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}$, $\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$

(Lower Confidence Limit, Upper Confidence Limit; LCL, UCL)

(1-α)100% σ^2에 대한 신뢰구간 도출

신뢰구간 개념에서 도출하면 $P(\ \chi^2_{1-\frac{\alpha}{2},n-1} < \chi^2 < \chi^2_{\frac{\alpha}{2},n-1}\ ) = 1 - \alpha$

좌변을 정리하면

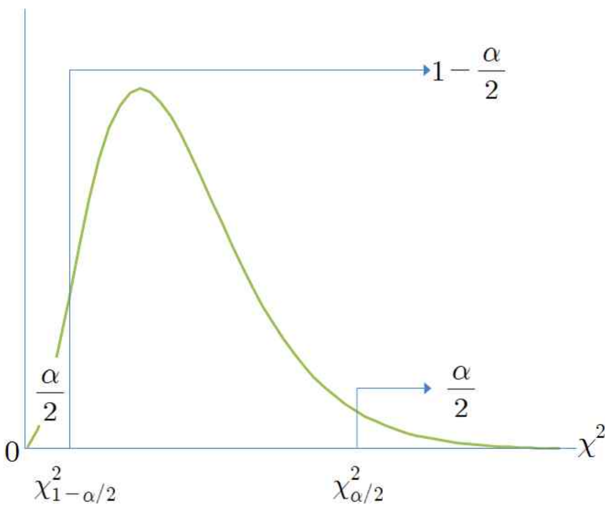
$(\ \chi^2_{1-\frac{\alpha}{2},n-1} < \chi^2 < \chi^2_{\frac{\alpha}{2},n-1}\ )$

$= (\ \chi^2_{1-\frac{\alpha}{2},n-1} < \dfrac{(n-1)s^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2},n-1}\ )$ χ^2을 $\dfrac{(n-1)s^2}{\sigma^2}$로 대체

$= \left(\dfrac{1}{\chi^2_{1-\frac{\alpha}{2},n-1}} > \dfrac{\sigma^2}{(n-1)s^2} > \dfrac{1}{\chi^2_{\frac{\alpha}{2},n-1}}\right)$ 각 항의 역수; 주의 - 역수를 취하면 부등호 방향이 바뀜

$= \left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$ 각 항에 $(n-1)s^2$을 곱하고, 부등호 방향을 <로 정리

③ $\chi^2_{1-\alpha/2,n-1}$와 $\chi^2_{\alpha/2,n-1}$를 그림에서 이해하기



예제. $X \sim N(\mu_X, \sigma_X^2)$, $n = 20$, $S = 50$. σ^2의 90% 수준의 신뢰구간을 구하시오.

LCL: $\dfrac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} = \dfrac{(20-1)50^2}{\chi^2_{0.05,\ 19}} = \dfrac{(20-1)50^2}{30.14} = 1{,}576.0$

UCL: $\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}} = \dfrac{(20-1)50^2}{\chi^2_{0.95,\ 19}} = \dfrac{(20-1)50^2}{10.12} = 4{,}639.7$

연습문제 4.

문제 1. σ^2에 대한 (1-α)100% 신뢰구간을 $\dfrac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}}$, $\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}$으로 설정하였다. 이에 필요한 가정은?

문제 2. $x \sim U[0,1]$인 경우, σ^2에 대한 (1-α)100% 신뢰구간을 $\dfrac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}}$, $\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}$으로 설정할 수 있나?

문제 3. $\chi^2_{n-1,\ \alpha/2}$과 $\chi^2_{n-1,\ 1-\alpha/2}$의 의미를 설명하시오.

문제 4. 모집단이 정규분포를 따르고 표본의 크기는 n이라 하자. 크기가 n인 표본을 10,000개 구성하고 각 표본의 분산값 s^2을 구한 후, $\chi^2_{n-1}(=\dfrac{(n-1)S^2}{\sigma^2})$도 구했다고 하자. 표본마다 s^2과 χ^2_{n-1}을 구했으니, 총 10,000개의 s^2과 χ^2_{n-1}을 구한 것이 된다. 10,000개의 χ^2_{n-1} 중 임의로 하나를 선택한다고 하자. 이때 선택한 χ^2_{n-1}이 $\chi^2_{n-1,\ \alpha/2}$보다 클 확률은 얼마인가?

	$\sigma_1^2 = 10^2$, n=4	
	s_1^2	χ^2
0.0%	0.04	0.0011
5.0%	12.18	0.3654
10.0%	19.54	0.5861
95.0%	259.57	7.7871
100.0%	691.98	20.7594

문제 5. (문제 4의 연속) 10,000개의 χ^2_{n-1} 중 임의로 하나를 선택한다고 하자. 이때 선택한 χ^2_{n-1}이 $\chi^2_{n-1,\ 1-\alpha/2}$보다 작을 확률은 얼마인가?

문제 6. 임의로 χ^2_{n-1}를 하나 선택했을 때, 그 값이 $\chi^2_{n-1,\ 1-\alpha/2}$보다는 크고 $\chi^2_{n-1,\ \alpha/2}$보다는 작을 확률은?

문제 7. $P(\chi^2_{n-1,\ 1-\alpha/2} < \chi^2_{n-1} < \chi^2_{n-1,\ \alpha/2}) = 1-\alpha$을 사용하여 σ^2에 대한 (1-α)100% 신뢰구간 공식을 도출하시오.

문제 8. 모집단은 정규분포를 따른다. n = 20, s^2 = 95. σ^2의 90% 수준의 신뢰구간을 구하시오.

문제 9. 모집단은 정규분포를 따른다. n = 20인 표본을 10,000개 추출한 후 각 표본마다 σ^2의 90% 수준의 신뢰구간을 설정한다고 하자. (총 10,000개의 신뢰구간이 설정된다.) 10,000개의 신뢰구간 중 σ^2를 포함하고 있는 신뢰구간은 몇 개쯤 될까?

문제 10. (문제 9 연속) 10,000개 중 σ^2를 포함하고 있지 않은 신뢰구간의 수는?

문제 11. (문제 10 연속) 신뢰구간이 σ^2를 포함하지 않을 확률은?

문제 12. α의 의미는?

(2) σ^2에 가설검정

① 가정: $X \sim N\left(\mu_X, \sigma_X^2\right)$

② H_0: $\sigma^2 = \theta^2$, H_A: $\sigma^2 \neq \theta^2$ (양측검정)

③ Test Statistic: $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$

④ Rejection Region: $\chi^2 > \chi^2_{\alpha/2}$, $\chi^2 < \chi^2_{1-\alpha/2}$

예제. S^2 = 120, n = 30, α = 0.10; σ^2 = 100 여부에 대한 가설검정

① H_0: $\sigma^2 = 100$, H_A: $\sigma^2 \neq 100$

② Test Statistic: $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$

③ For α = 0.10,

Rejection Region: $\chi^2 > \chi^2_{\alpha/2} = \chi^2_{0.05} = 42.56$, $\chi^2 < \chi^2_{1-\alpha/2} = \chi^2_{0.95} = 17.71$

④ Value of the statistic: $\chi^2 = \dfrac{(30-1)120}{100} = 34.8$

⑤ Conclusion: Do not reject H_0.

연습문제 5.

문제 1. "H_0: $\sigma^2 = \theta^2$란 모집단의 분산 σ^2 이 θ^2이라는 의미이다."라는 진술을 평가하시오.

문제 2. 제품 A의 중량의 분산은 10^2이어야 한다. 최근 제품 A의 중량 분산이 10^2이 아니라는 불만이 제기되어 조사를 하려 한다. 제품 20개를 선택하여 표본 조사하려 한다. 제품의 중량은 정규분포를 따른다. 유의수준 α = 0.05로 한다.

Test Statistic으로 $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$을 사용하는 이유를 설명하시오.

문제 3. 모집단은 정규분포를 따른다. n = 20, s^2 = 95. α= 0.1

H_0: σ^2 = 10^2이고, H_A: σ^2 $\neq$ 10^2이다. 가설검정을 실시하시오.

제2절 두 모집단 분산의 추론

1. F 분포

가정: $X_1 \sim N(\mu_{X_1}, \sigma^2_{X_1})$, $X_2 \sim N(\mu_{X_2}, \sigma^2_{X_2})$

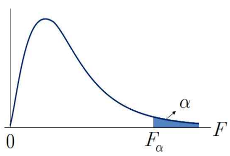
확률변수 F의 정의

$F_{n_1-1,\, n_2-1} = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$, where n_1 = 첫 번째 표본의 크기, n_2 = 두 번째 표본의 크기

***F* 분포의 특징**

비대칭이고 오른쪽으로 꼬리 → (n_1-1)과 (n_2-1)이 증가할수록 정규분포에 가깝게 된다. 또한 모든 F의 값은 항상 양수이다.

$F_{n_1-1,\, n_2-1}$ **분포표**



α = 0.01

df_2 \ df_1	1	2	3	4	5	6	7	8	9
1	4,052.18	4,999.50	5,403.35	5,624.58	5,763.65	5,858.99	5,928.36	5,981.07	6,022.47
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94

연습문제 6.

문제 1. 표준정규분포에서 z_α의 의미는?
문제 2. 표준정규분포에서 하위 α에 해당하는 z값은?
문제 3. 표준정규분포에서 z_α와 $z_{1-\alpha}$의 관계는?
문제 4. 문제 3의 등식이 성립하는 이유는?
문제 5. $F_{\alpha,\, n_1-1,\, n_2-1}$의 의미를 서술하시오.
문제 6. $F_{1-\alpha,\, n_1-1,\, n_2-1}$의 의미를 서술하시오.
문제 7. $F_{1-\alpha,\, n_1-1,\, n_2-1} = -F_{\alpha,\, n_1-1,\, n_2-1}$가 성립하는 가?

$$\boxed{F_{1-\alpha,\, n_1-1,\, n_2-1} = \frac{1}{F_{\alpha,\, n_2-1,\, n_1-1}}}$$

$F_{1-\alpha,\, n_1-1,\, n_2-1}$란 $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ 중 상위 1-α에 해당하는 값이다.

$\dfrac{1}{F_{1-\alpha,\, n_1-1,\, n_2-1}}$는 $\dfrac{1}{\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}}$ 중 상위 얼마에 해당하는 값일까?

역수를 취했으므로 상위 1-α가 아니라 상위 α에 해당하는 값이 된다. 즉,

$\dfrac{1}{F_{1-\alpha,\, n_1-1,\, n_2-1}}$는 $\dfrac{1}{\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}}$ 중 상위 α에 해당하는 값이다.

$\rightarrow$ $\dfrac{1}{F_{1-\alpha,\, n_1-1,\, n_2-1}}$는 $\dfrac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2}$ 중 상위 α에 해당하는 값이다.

$\dfrac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2}$ 중 상위 α에 해당하는 값은 F_{α, n_2-1, n_1-1}이므로,

$F_{\alpha,\, n_2-1,\, n_1-1} = \dfrac{1}{F_{1-\alpha,\, n_1-1,\, n_2-1}}$이 성립한다. 양변을 역수로 하면,

$F_{1-\alpha,\, n_1-1,\, n_2-1} = \dfrac{1}{F_{\alpha,\, n_2-1,\, n_1-1}}$이 성립한다.

연습문제 7. 제품 A와 B의 중량은 정규분포를 따른다. 두 제품 모두 중량의 분산은 10^2라 알려져 있다. 제품 A의 표본 크기는 5, 제품 B의 6이며, 조사한 결과 제품 A의 표본 분산은 12^2, 제품 B의 표본 분산은 8^2이다.

문제 1. 제품 A를 첫 번째 표본으로, 제품 B를 두 번째 표본으로 간주하여 F값을 계산하시오.

문제 2. (문제 1의 연속) $F_{4,\,5}$ 값들 중 2.25는 상위 %에 해당하는 값인가?

문제 3. n_1 = 5, n_2 = 6일 때, 상위 0.1987에 해당하는 F값이 2.25임을 수식으로 표현하시오.

문제 4. 제품 B를 첫 번째 표본으로, 제품 A를 두 번째 표본으로 간주하여 F값을 계산하시오.

문제 5. (문제 4의 연속) $F_{5,\,4}$ 값들 중 $\frac{1}{2.25}$=0.44는 상위 %에 해당하는 값인가?

문제 6. n_1 = 6, n_2 = 5일 때, 상위 0.8013에 해당하는 F값이 0.44임을 수식으로 표현하시오.

문제 7. $F_{0.1987,\,4,\,5}$과 $F_{0.8013,\,5,\,4}$의 관계를 등식으로 표현하시오.

문제 8. 문제 7을 일반화하시오.

연습문제 8.

문제 1. $F_{0.05,\,7,\,9}$ = 3.2927이다. $F_{0.95,\,9,\,7}$ =

문제 2. $F_{0.10,\,8,\,12}$ = 2.2446이다. $F_{0.90,\,12,\,8}$ =

2. 분산의 추정 및 가설검정

(1) $\frac{\sigma_1^2}{\sigma_2^2}$에 대한 (1-$\alpha$)100% 신뢰구간

① 가정: $X_1 \sim N(\mu_{X_1},\, \sigma_{X_1}^2),\ X_2 \sim N(\mu_{X_2},\, \sigma_{X_2}^2)$

② 신뢰구간: $\text{LCL} = \left(\frac{S_1^2}{S_2^2}\right)\frac{1}{F_{\alpha/2,\, n_1-1, n_2-1}}$

$$\text{UCL} = \left(\frac{S_1^2}{S_2^2}\right)\frac{1}{F_{1-\alpha/2,\, n_1-1, n_2-1}} = \left(\frac{S_1^2}{S_2^2}\right)F_{\alpha/2,\, n_2-1,\, n_1-1}$$

(1-α)100% $\frac{\sigma_1^2}{\sigma_2^2}$에 대한 신뢰구간 도출

신뢰구간 개념에서 도출하면

$$P\left(F_{1-\frac{\alpha}{2}, n_1-1, n_2-1} < F < F_{\frac{\alpha}{2}, n_1-1, n_2-1}\right) = 1 - \alpha$$

좌변을 정리하면

$$\left(F_{1-\frac{\alpha}{2}, n_1-1, n_2-1} < F < F_{\frac{\alpha}{2}, n_1-1, n_2-1}\right)$$

$$= \left(F_{1-\frac{\alpha}{2}, n_1-1, n_2-1} < \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} < F_{\frac{\alpha}{2}, n_1-1, n_2-1} \right)$$

F를 $\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$로 대체

$$= \left(F_{1-\frac{\alpha}{2}, n_1-1, n_2-1} < \frac{s_1^2}{s_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} < F_{\frac{\alpha}{2}, n_1-1, n_2-1} \right)$$

$\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$를 $\dfrac{s_1^2}{s_2^2} \cdot \dfrac{\sigma_2^2}{\sigma_1^2}$으로 대체

$$= \left(\frac{1}{F_{1-\frac{\alpha}{2},\, n_1-1,\, n_2-1}} > \frac{s_2^2}{s_1^2} \cdot \frac{\sigma_1^2}{\sigma_2^2} > \frac{1}{F_{\frac{\alpha}{2},\, n_1-1,\, n_2-1}} \right)$$

각 항의 역수; 주의 - 역수 취하면 부등호 방향이 바뀜

$$= \left[\left(\frac{S_1^2}{S_2^2} \right) \frac{1}{F_{\alpha/2,\, n_1-1, n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{S_1^2}{S_2^2} \right) \frac{1}{F_{1-\alpha/2,\, n_1-1, n_2-1}} \right]$$

각 항에 $\dfrac{s_1^2}{s_2^2}$을 곱하고, 부등호 방향을 < 로 정리

예제. 두 모집단 모두 정규분포, $S_1^2 = 9.4$, $S_2^2 = 6.8$, $n_1 = 41$, $n_2 = 31$; σ_1^2 / σ_2^2의 90% 수준의 신뢰구간을 구하시오.

$$\text{LCL: } \left(\frac{S_1^2}{S_2^2} \right) \frac{1}{F_{\alpha/2,\, n_1-1, n_2-1}} = \left(\frac{9.4}{6.8} \right) \frac{1}{F_{0.05,\, 40,\, 30}} = \left(\frac{9.4}{6.8} \right) \frac{1}{1.79} = 0.77$$

$$\text{UCL: } \left(\frac{S_1^2}{S_2^2} \right) \frac{1}{F_{1-\alpha/2,\, n_1-1, n_2-1}} = \left(\frac{9.4}{6.8} \right) F_{0.05,\, 30,\, 40} = \left(\frac{9.4}{6.8} \right) (1.74) = 2.41$$

연습문제 9.

두 모집단은 정규분포를 이룬다. 첫 번째 표본 크기는 n_1, 두 번째 표본 크기는 n_2이다. $F_{n_1-1,\, n_2-1} = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$이다.

문제 1. 상위 $\dfrac{\alpha}{2}$와 하위 $\dfrac{\alpha}{2}$에 해당하는 F값은?

문제 2. $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$가 $F_{\frac{\alpha}{2},\, n_1-1,\, n_2-1}$와 $F_{1-\frac{\alpha}{2},\, n_1-1,\, n_2-1}$ 사이에 존재할 확률은?

문제 3. 문제 2를 수식으로 표현하시오.

문제 4. 문제 3의 수식으로부터 σ_1^2/σ_2^2에 대한 (1-α)100% 신뢰구간을 도출하시오.

연습문제 10.

문제 1. 두 모집 모두 정규분포, $S_1^2 = 12$, $S_2^2 = 15$, $n_1 = 21$, $n_2 = 16$; σ_1^2 / σ_2^2의 95% 수준의 신뢰구간을 구하시오. 단, 상위 2.5% 및 하위 2.5%에 해당하는 $F_{n_1-1,\, n_2-1}$ 값은 2.7559와 0.3886이다.

(2) $\dfrac{\sigma_1^2}{\sigma_2^2}$에 가설검정

① 가정: $X_1 \sim N(\mu_{X_1}, \sigma_{X_1}^2)$, $X_2 \sim N(\mu_{X_2}, \sigma_{X_2}^2)$

② H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$, H_A: $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$ (양측검정)

③ Test Statistic: $F_{n_1-1, n_2-1} = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \dfrac{S_1^2}{S_2^2}$

(H_0가 $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$이면 $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \dfrac{S_1^2}{S_2^2}$이 성립한다.)

④ Rejection Region: $F > F_{\alpha/2}$ 또는 $F < F_{1-\alpha/2}$

② H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$, H_A: $\dfrac{\sigma_1^2}{\sigma_2^2} < 1$ → R.R: $F < F_{1-\alpha}$ (왼쪽 꼬리검정)

② H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$, H_A: $\dfrac{\sigma_1^2}{\sigma_2^2} > 1$ → R.R: $F > F_{\alpha}$ (오른쪽 꼬리검정)

예제. 앞의 예제에서 $\alpha = 0.10$로 $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ 가설검증

① H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$, H_A: $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$

② Test Statistic: $F_{n_1-1,\, n_2-1} = \dfrac{S_1^2}{S_2^2}$ (H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$이므로 $F = \dfrac{S_1^2}{S_2^2}$이다.)

③ For $\alpha = 0.10$,
Rejection Region: $F > F_{n_1-1,\, n_2-1,\, \alpha/2} = 1.79$ 또는

$F < F_{n_1-1,\, n_2-1,\, 1-\alpha/2} = \dfrac{1}{F_{n_2-1,\, n_1-1,\, \alpha/2}} = \dfrac{1}{1.74} = 0.57$

④ Value of the statistic: $F = \dfrac{9.4}{6.8} = 1.38$

⑤ Conclusion: Do not reject H_0.

연습문제 정답

1. (1) $\overline{x}$의 분포는 평균을 중심으로 좌우 대칭임을 알 수 있다.
   ① 평균과 하위 50%의 값이 거의 일치한다.
   ② 하위 50%에서 하위 40% 값을 제한 값(100.0-98.7 = 1.3)과 하위 60%에서 하위 50% 값을 제한 값(101.3-100.0=1.3)이 거의 일치한다. 하위 10%에서 하위 0% 값을 제한 값(18.3)과 하위 100%에서 하위 90% 값을 제한 값(18.4)이 거의 일치한다.

   s^2의 분포는 비대칭이며, 오른쪽으로 늘어진 형태이다.
   ① 하위 50%의 값보다 평균이 더 크다.
   ② 하위 10%에서 하위 0% 값을 제한 값(19.5)보다 하위 100%에서 하위 90% 값을 제한 값(818.9)이 월등히 크다.

학습 포인트: x가 정규분포를 따르면 $\overline{x}$의 분포는 평균을 중심으로 좌우대칭인 정규분포를 따른다. s^2의 분포는 오른쪽으로 늘어진 비대칭이다. (하위 50%값이 평균보다 작다.)

   (2) $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$이다. 제곱의 합을 일정한 수로 나눈 값이므로 0보다 작을 수 없다.
   $s^2 = 0$이 되려면, 표본의 데이터값들이 모두 동일해야만 한다. 연속확률 변수의 속성상 동일한 값이 반복될 확률은 0이다. 결론적으로 s^2은 음수는 불가능하며, 0에 가까운 양수는 가능하다.

2. (1) 모집단 1의 하위 40% $\chi^2 = \dfrac{(n-1)\cdot \text{하위 } 40\%\ s_1^2}{\sigma_1^2} = \dfrac{3(61.78)}{10^2} = 1.8534$

   모집단 2의 하위 40% $\chi^2 = \dfrac{(n-1)\cdot \text{하위 } 40\%\ s_2^2}{\sigma_2^2} = \dfrac{3(252.93)}{20^2} = 1.8970$

   모집단 1의 하위 40% s^2값(61.78)과 모집단 2의 하위 40% s^2값(252.93)은 다르다.
   그렇지만 각각의 값들을 (n-1)로 곱하고 자신의 모집단 분산으로 나눈 값, 즉
   $\dfrac{(n-1)\cdot \text{하위 } 40\%\ s_1^2}{\sigma_1^2}$과 $\dfrac{(n-1)\cdot \text{하위 } 40\%\ s_2^2}{\sigma_2^2}$은 동일하다.

   위와 같이 s^2을 (n-1)로 곱하고 자신의 모집단 분산으로 나눈 값, $\dfrac{(n-1)s^2}{\sigma^2}$을 χ^2_{n-1}라 한다.

학습 포인트: $\chi^2_{n-1} \equiv \dfrac{(n-1)s^2}{\sigma^2}$, where n = sample size

(2) 모집단 1의 하위 95% $\chi^2 = \dfrac{(n-1)\cdot \text{하위 } 95\%\ s_1^2}{\sigma_1^2} = \dfrac{3(259.57)}{10^2} = 7.7871$

모집단 2의 하위 95% $\chi^2 = \dfrac{(n-1)\cdot \text{하위 } 95\%\ s_2^2}{\sigma_2^2} = \dfrac{3(1{,}047.08)}{20^2} = 7.8531$

모집단 1의 하위 95% s^2값(259.57)과 모집단 2의 하위 40% s^2값(1,047.08)은 다르다. 그렇지만 각각의 값들을 (n-1)로 곱하고 자신의 모집단 분산으로 나눈 값, 즉 $\dfrac{(n-1)\cdot \text{하위 } 95\%\ s_1^2}{\sigma_1^2}$과 $\dfrac{(n-1)\cdot \text{하위 } 95\%\ s_2^2}{\sigma_2^2}$은 동일하다.

위와 같이 s^2을 (n-1)로 곱하고 자신의 모집단 분산으로 나눈 값, $\dfrac{(n-1)s^2}{\sigma^2}$을 χ^2_{n-1}라 한다.

학습 포인트: $\chi^2_{n-1} \equiv \dfrac{(n-1)S^2}{\sigma^2}$, where n = sample size

(3) 모집단이 정규분포이고 표본의 크기가 동일하면, χ^2 분포는 동일하다.

3. (1) 위 표에서 ν(nu라고 읽는다.)는 자유도를 의미한다. 자유도는 n-1이다. 표 상단에 있는 그래프를 보면 α는 큰 쪽을 향하고 있다. (상위라는 이야기이다.)

n = 4이므로 ν = 3이고, 상위 5%이므로 α = 0.05이다. 이에 해당하는 χ^2 = 7.815

학습 포인트: $\chi^2_{n-1,\alpha}$란 상위 α에 속하는 χ^2값이다. (표본의 크기는 n이다.)

(2) 위 표에서 ν(nu라고 읽는다.)는 자유도를 의미한다. 자유도는 n-1이다. 표 상단에 있는 그래프를 보면 α는 큰 쪽을 향하고 있다. 상위라는 이야기이다.
표는 상위를 기준으로 작성되어 있으니 하위를 상위로 변환한다. 하위 5%는 상위 95%에 해당한다.

n = 4이므로 ν = 3이고, 상위 95%이므로 α = 0.95이다. 이에 해당하는 χ^2 = .352이다.

학습 포인트: $\chi^2_{n-1,\alpha}$란 하위 1-α에 속하는 χ^2값이다. (표본의 크기는 n이다.)

(3) n = 6 (ν = 5)일 때 상위 10%에 해당하는 $\chi^2 = \chi^2_{5,0.1}$ = 9.236이다.

$\chi^2_{n-1} = \dfrac{(n-1)S^2}{\sigma^2}$이므로 $9.236 = \dfrac{(6-1)s^2}{50} \rightarrow s^2 = \dfrac{9.236(50)}{5} = 92.36$

학습 포인트: 상위 α에 해당하는 s^2의 값은 $\dfrac{\chi^2_{n-1,\alpha}\cdot\sigma^2}{n-1}$이 된다.

(4) n = 6 (ν = 5)일 때 하위 10%에 해당하는 χ^2
= 상위 90%에 해당하는 $\chi^2 = \chi^2_{5,\ 0.9}$ = 1.610

$\chi^2_{n-1} = \dfrac{(n-1)S^2}{\sigma^2}$이므로 $1.610 = \dfrac{(6-1)s^2}{50} \rightarrow s^2 = \dfrac{1.610(50)}{5} = 16.10$

학습 포인트: 하위 α에 해당하는 s^2의 값은 $\dfrac{\chi^2_{n-1,\ 1-\alpha}\cdot\sigma^2}{n-1}$이 된다.

4. (1) 모집단이 정규분포를 이룰 때만 $\frac{(n-1)s^2}{\sigma^2}$이 χ^2분포를 따른다. 실험 2를 참조한다.

(2) 없다.

모집단이 최소값이 0, 최대값이 1인 일양분포를 따른다고 하자. 이 경우 σ^2 = $\frac{(\text{최대값}-\text{최소값})^2}{12}$ = $\frac{(1-0)^2}{12}$ = $\frac{1}{12}$이다.

n=4 (ν=3)인 표본 10,000개를 구성하여 s^2을 추출한 결과를 정리하였다.

	$x \sim U[0,1]$		χ^2_3 이론값
	s^2	$\frac{(n-1)s^2}{\sigma^2}$	
0%	0.0002	0.01	-
5%	0.0130	0.47	0.35
10%	0.0214	0.77	0.58
15%	0.0287	1.03	0.80
50%	0.0767	2.76	2.37
90%	0.1545	5.56	6.25
95%	0.1745	6.28	7.81
99%	0.2094	7.54	11.34

본 실험은 정규분포가 아닌 일양분포를 따르는 데이터 4개씩을 추출하여 각 표본의 분산을 백분위로 표현한 것이다. 즉, 하위 5%에 속하는 s^2는 0.0130이고, 하위 5%에 속하는 $\frac{(n-1)s^2}{\sigma^2}$은 $\frac{(4-1) \cdot 0.0130}{1/12}$ = 0.47이다. 그런데, n=4(ν=3) 일 때 하위 5%에 속하는 χ^2은 $\chi^2_{3,\,0.95}$ = 0.352(χ^2_{n-1} 분포표에서 또는 컴퓨터로 구한 값)이다. 결론적으로 모집단이 정규분포를 따르지 않으면, $\frac{(n-1)s^2}{\sigma^2}$값은 χ^2값과 같지 않으며, σ^2에 대한 (1-α)100% 신뢰구간을 $\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}$, $\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$으로 설정할 수 없다.

학습 포인트: $\frac{(n-1)s^2}{\sigma^2}$ = χ^2이 성립할 조건은 x가 정규분포를 따르는 것이다.

모집단이 정규분포 (또는 정규분포에 근사한 경우)가 아니면,

$\frac{(n-1)S^2}{\sigma^2}$는 χ^2 분포를 갖지 못한다.

(3) $\chi^2_{n-1,\,\alpha/2}$란 표본의 크기가 n인 χ^2_{n-1} ($\frac{(n-1)S^2}{\sigma^2}$) 중 상위 $\frac{\alpha}{2}$에 해당하는 χ^2_{n-1}를, $\chi^2_{n-1,\,1-\alpha/2}$란 표본의 크기가 n인 χ^2_{n-1} ($\frac{(n-1)S^2}{\sigma^2}$) 중 하위 $\frac{\alpha}{2}$에 해당하는 χ^2_{n-1}를 의미한다. 하위 $\frac{\alpha}{2}$와 상위 1-$\frac{\alpha}{2}$는 같은 의미이다.

학습 포인트: $\chi^2_{n-1,\ \alpha/2}$는 상위 $\dfrac{\alpha}{2}$에 해당하는 χ^2_{n-1}

$\chi^2_{n-1,\ 1-\alpha/2}$는 하위 $\dfrac{\alpha}{2}$(또는 상위 1-$\dfrac{\alpha}{2}$)에 해당하는 χ^2_{n-1}

(4) α = 0.10이라 하자. 그러면 $\dfrac{\alpha}{2}$ = 0.05이다. 앞에서 실험한 결과를 보면, 상위 5%(위의 표에서는 상위 95%)의 $\chi^2_{3,\ 0.05}$의 값은 7.7871이다. 즉, 10,000개의 χ^2_3 중 95,000개는 7.7871보다 작고 5,000개는 7.7871보다 크다는 것이다. 이를 식으로 표현하면,

$P(\chi^2_3 < \chi^2_{3,\ 0.05}) = P(\chi^2_3 < 7.7871) = 0.95$ ....

10,000개의 χ^2_3 중 95,000개는 $\chi^2_{3,\ 0.05}$ (=7.7871)보다 작다.

$P(\chi^2_3 > \chi^2_{3,\ 0.05}) = P(\chi^2_3 > 7.7871) = 0.05$ ....

10,000개의 χ^2_3 중 5,000개는 $\chi^2_{3,\ 0.05}$ (=7.7871)보다 크다.

이를 일반화하면, χ^2_{n-1}이 $\chi^2_{n-1,\ \alpha/2}$보다 클 확률 = $P(\chi^2_{n-1} > \chi^2_{n-1,\ \alpha/2}) = \dfrac{\alpha}{2}$

학습 포인트: χ^2_{n-1}이 $\chi^2_{n-1,\ \alpha/2}$보다 클 확률 = $P(\chi^2_{n-1} > \chi^2_{n-1,\ \alpha/2}) = \dfrac{\alpha}{2}$

(5) α = 0.10이라 하자. 그러면 $\dfrac{\alpha}{2}$ = 0.05이다. 앞에서 실험한 결과를 보면, 하위 5%의 χ^2_3 (표기는 상위 95% χ^2_3, 즉 $\chi^2_{3,\ 0.95}$로 한다.)의 값은 0.3654이다. 즉, 10,000개의 χ^2_3 중 5,000개는 0.3654보다 작고 95,000개는 0.3654보다 크다는 것이다. 이를 식으로 표현하면,

$P(\chi^2_3 < \chi^2_{3,\ 0.95}) = P(\chi^2_3 < 0.3654) = 0.05$ ....

10,000개의 χ^2_3 중 5,000개는 $\chi^2_{3,\ 0.95}$ (=0.3654)보다 작다.

$P(\chi^2_3 > \chi^2_{3,\ 0.95}) = P(\chi^2_3 > 0.3654) = 0.95$ ....

10,000개의 χ^2_3 중 95,000개는 $\chi^2_{3,\ 0.95}$ (=0.3654)보다 크다.

이를 일반화하면, χ^2_{n-1}이 $\chi^2_{n-1,\ 1-\alpha/2}$보다 작을 확률 = $P(\chi^2_{n-1} > \chi^2_{n-1,\ 1-\alpha/2}) = \dfrac{\alpha}{2}$

포인트: χ^2_{n-1}이 $\chi^2_{n-1,\ 1-\alpha/2}$보다 작을 확률 = $P(\chi^2_{n-1} < \chi^2_{n-1,\ 1-\alpha/2}) = \dfrac{\alpha}{2}$

(6) χ^2_{n-1}이 $\chi^2_{n-1,\ 1-\alpha/2}$보다는 크고 $\chi^2_{n-1,\ \alpha/2}$보다는 작을 확률

$= P(\chi^2_{n-1,\ 1-\alpha/2} < \chi^2_{n-1} < \chi^2_{n-1,\ \alpha/2}) = 1-\alpha$

학습 포인트: $P(\chi^2_{n-1,\ 1-\alpha/2} < \chi^2_{n-1} < \chi^2_{n-1,\ \alpha/2}) = 1-\alpha$

(7) $P(\chi^2_{n-1,\ 1-\alpha/2} < \chi^2_{n-1} < \chi^2_{n-1,\ \alpha/2})$

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

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

..... 역수로 하면 부등호 방향이 바뀐다.

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

$= P\left(\frac{(n-1)s^2}{\chi^2_{n-1,\ \alpha/2}} < \sigma^2 < \right)$ ..... 각 항에 $(n\text{-}1)s^2$을 곱한다.

= 1-α

학습 포인트: σ^2에 대한 (1-α)100% 신뢰구간은 $\frac{(n-1)s^2}{\chi^2_{n-1,\ \alpha/2}}$, $\frac{(n-1)s^2}{\chi^2_{n-1,\ 1-\alpha/2}}$이다.

(8) 90%의 신뢰구간을 구하는 것이므로, α = 0.1 (=10%)이다.

n = 20이므로, 자유도는 19이다.

90%의 신뢰구간을 구하기 위해서는 상위 5%의 χ^2_{19}와 하위 5%의 χ^2_{19}의 값이 필요하다.

χ^2_{n-1} 분포표에서 상위 5%의 χ^2_{19}, 즉 $\chi^2_{19,\,0.05}$ = 30.14이고, χ^2_{n-1} 분포표에서 하위 5%의 χ^2_{19}는 상위 95%의 χ^2_{19}, 즉 $\chi^2_{19,\,0.95}$ = 10.117이다.

LCL: $\frac{(n-1)s^2}{\chi^2_{\alpha/2,\ n-1}} = \frac{(20-1)95}{\chi^2_{0.05,\ 19}} = \frac{(20-1)95}{30.14} = 59.89$

UCL: $\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\ n-1}} = \frac{(20-1)95}{\chi^2_{0.95,\ 19}} = \frac{(20-1)95}{10.12} = 178.36$

(9) 9,000개

(10) 1,000개

(11) $\frac{1,000}{10,000}$ = 0.1

(12) 신뢰구간이 σ^2를 포함하지 않을 확률 (문제 11의 경우 α=0.1)

5. (1) σ^2에 대한 가설검정 은 σ^2값을 모르기 때문에 실시한다. 만약 σ^2값을 안다면, 굳이 검정을 실시할 필요가 없다.

가설검정에서 H_0란 사실(Fact)가 아니라 단지 가설(Hypothesis)에 불과하다. 즉, H_0: σ^2 = θ^2란 모집단의 분산 σ^2 이 θ^2이라는 가정하는 것뿐이다.

σ^2 을 알려면 모집단 전체를 조사해야만 한다. 그렇지 않다면 표본을 구성하여 표본의 분산 s^2을 구한다. σ^2 값이 θ^2일 때 너무도 예외적인 s^2값이 표본으로부터 나온다면

① σ^2 값은 정말 θ^2인데, 너무도 예외적인 값이 나왔다라고 결론내릴 수도 있고,

② σ^2 값이 θ^2라고 주장하기에는 너무 예외적인 값이 나왔으므로 σ^2 값은 θ^2이 아니라고 결론 내릴 수도 있다. (H_0를 기각한다.)

가설검정에서는 미리 설정한 s^2 값보다 더 예외적인 s^2가 나오면 ②와 같이 결론 내린다. (물론, 미리 설정한 s^2 값보다 덜 예외적인 s^2가 나오면 H_0의 주장을 기각하지 않는다.)

(2) 지문을 보면 H_0: σ^2 = 10^2이고, H_A: σ^2 $\neq$ 10^2이다. 양측검정이다.

σ^2를 직접 조사할 수 없다면, 표본을 통해 σ^2값을 추정한다. 추정에 사용되는 도구는 s^2이다.

양측검정인데, 유의수준이 0.05이다. 그러므로 ① 표본에서 구한 s^2값이, H_0가 사실일 때 상위 2.5%에 해당하는 s^2값보다 더 크거나 또는 ② 표본에서 구한 s^2값이, H_0가 사실일 때 하위 2.5%에 해당하는 s^2값보다 더 작은 지를 조사한다.

만약 ①과 ② 중 하나라도 해당하면 H_0를 기각하고, 두 조건 중 하나라도 해당되지 않으면 H_0를 기각하지 못한다.

문제는 σ^2 = 10^2일 때 상위 2.5%에 해당하는 s^2과 하위 2.5%에 해당하는 s^2을 모른다는 것이다. 그런데, $\chi^2_{20-1,0.025}$와 $\chi^2_{20-1,0.925}$은 χ^2분포표를 통해 알려져 있고,

$\dfrac{(n-1)\cdot \text{상위 2.5\%에 해당하는 } s^2}{\sigma^2}$ 은 $\chi^2_{20-1,0.025}$이고,

$\dfrac{(n-1)\cdot \text{하위 2.5\%에 해당하는 } s^2}{\sigma^2}$ 은 $\chi^2_{20-1,0.925}$와 동일하므로, 기각하는 규칙을 다음과 같이 변경할 수 있다.

표본에서 구한 $\dfrac{(n-1)s^2}{\sigma^2}$이 $\chi^2_{20-1,0.025}$보다 크거나 또는 표본에서 구한 $\dfrac{(n-1)s^2}{\sigma^2}$이 $\chi^2_{20-1,0.925}$보다 작으면 H_0를 기각하고, 그렇지 않으면 H_0를 기각하지 못한다.

정리하면, 각 σ^2마다 상위 $\dfrac{\alpha}{2}$에 해당하는 s^2값과 하위 $\dfrac{\alpha}{2}$에 해당하는 s^2값을 알고 있다면, 굳이 Test Statistic으로 χ^2을 사용할 필요가 없다. 그 값들을 모르기 때문에 값들이 표로 정리되어 있는 χ^2을 Test Statistic으로 사용한다.

(3) ① H_0: σ^2 = 100, H_A: σ^2 $\neq$ 100

② Test Statistic: $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$

③ For α = 0.10,

Rejection Region: $\chi^2 > \chi^2_{n-1,\ \alpha/2} = \chi^2_{19,\ 0.05} = 30.14$,

$\chi^2 < \chi^2_{n-1,\ 1-\alpha/2} = \chi^2_{19,\ 0.95} = 10.12$

④ Value of the statistic: $\chi^2 = \dfrac{(20-1)95}{100} = 18.05$

⑤ Conclusion: Do not reject H_0.

6. (1) 상위 α에 해당하는 z값 (2) $z_{1-\alpha}$ (3) $z_{1-\alpha}$ = -z_α

(4) z분포는 0을 중심으로 좌우대칭

(5) 첫 번째 자유도가 n_1-1이고, 두 번째 자유도가 n_2-1인 F값들 중 상위 α에 해당하는 값

(6) 첫 번째 자유도가 n_1-1이고, 두 번째 자유도가 n_2-1인 F값들 중 상위 1-α에 해당하는 값; 상위 1-α는 하위 α라는 의미이기도 함

(7) 성립하지 않는다. z분포의 경우, 평균이 0이고 좌우대칭이므로 $z_{1-\alpha}$ = -z_α이

성립하지만, F분포는 모두 양수이고 오른쪽으로 꼬리가 있는 형태로 비대칭이다.

7. (1) $F_{4,5} = \dfrac{12^2/10^2}{8^2/10^2} = 2.25$

(2) 0.1987 (19.87%)

Excel에서 F.DIST(2.25, 4, 5)를 입력하면 첫 번째 자유도 4, 두번째 자유도 5일 때 F_value = 2.25가 하위 %에 해당하는 값인지를 알려준다. F.DIST(2.25, 4, 5, cumulative=1) = 0.801267 (80.1267%). 상위로 변환하려면 1에서 0.801267을 빼면 된다.

(3) $F_{0.1987,\,4,\,5} = 2.25$

(4) $F_{5,4} = \dfrac{8^2/10^2}{12^2/10^2} = \dfrac{1}{2.25} = 0.44$

(5) 0.801267 (80.1267%)

(6) $F_{0.8013,5,4} = 0.44$ $(= \dfrac{1}{2.25})$

(7) $F_{0.1987,\,4,\,5} = \dfrac{1}{F_{0.8013,5,4}}$

(8) $F_{\alpha,n_1-1,n_2-1} = \dfrac{1}{F_{1-\alpha,\,n_2-1,n_1-1}}$

8. (1) $F_{0.95,9,7} = \dfrac{1}{F_{0.05,7,9}} = 1/3.2927 = 0.3037$

(2) $F_{0.90,12,8} = \dfrac{1}{F_{0.10,8,12}} = 1/2.2446 = 0.4455$

9. (1) 상위 $\dfrac{\alpha}{2}$에 해당하는 $F = F_{\frac{\alpha}{2},\,n_1-1,n_2-1}$, 하위 $\dfrac{\alpha}{2}$에 해당하는 $F = F_{1-\frac{\alpha}{2},\,n_1-1,n_2-1}$

(2) 1-α

(3) $P\left(F_{1-\frac{\alpha}{2},\,n_1-1,\,n_2-1} < \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}\right) = 1\text{-}\alpha$

(4) $P\left(F_{1-\frac{\alpha}{2},\,n_1-1,\,n_2-1} < \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}\right)$

$= P\left(F_{1-\frac{\alpha}{2},\,n_1-1,\,n_2-1} < \dfrac{S_1^2}{S_2^2}\cdot\dfrac{\sigma_2^2}{\sigma_1^2} < F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}\right)$

$= P\left(\dfrac{S_2^2}{S_1^2}\cdot F_{1-\frac{\alpha}{2},\,n_1-1,\,n_2-1} < \dfrac{\sigma_2^2}{\sigma_1^2} < \dfrac{S_2^2}{S_1^2}\cdot F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}\right)$

$= P\left(\dfrac{S_2^2}{S_1^2}\cdot F_{1-\frac{\alpha}{2},\,n_1-1,\,n_2-1} < \dfrac{\sigma_2^2}{\sigma_1^2} < \dfrac{S_2^2}{S_1^2}\cdot F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}\right)$

$= P\left(\dfrac{1}{\dfrac{S_2^2}{S_1^2}\cdot F_{1-\frac{\alpha}{2},\,n_1-1,\,n_2-1}} > \dfrac{1}{\sigma_2^2/\sigma_1^2} > \dfrac{1}{\dfrac{S_2^2}{S_1^2}\cdot F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}}\right)$

$= P\left(\dfrac{S_1^2}{S_2^2}\cdot\dfrac{1}{F_{\frac{\alpha}{2},\,n_1-1,\,n_2-1}} < \dfrac{\sigma_1^2}{\sigma_2^2} < \dfrac{S_1^2}{S_2^2}\cdot\dfrac{1}{F_{1-\frac{\alpha}{2},\,n_1-1,\,n_2-1}}\right)$

10. (1) LCL: $\left(\dfrac{S_1^2}{S_2^2}\right)\dfrac{1}{F_{\alpha/2,\, n_1-1, n_2-1}} = \left(\dfrac{12}{15}\right)\dfrac{1}{F_{0.025,\, 20,\, 15}} = \left(\dfrac{12}{15}\right)\dfrac{1}{2.7559} = 0.2903$

UCL: $\left(\dfrac{S_1^2}{S_2^2}\right)\dfrac{1}{F_{1-\alpha/2,\, n_1-1, n_2-1}} = \left(\dfrac{12}{15}\right)\dfrac{1}{F_{0.975,\, 20,\, 15}} = \left(\dfrac{12}{15}\right)\dfrac{1}{0.3886} = 2.0585$