

제 10장

두 모집단의 모수에 대한 추론



고려대학교 경영대학 박 광태

모평균의 가설검정-1)-a)

- ◆ 모평균의 가설검정(표본이 서로 독립적일 때:모분산이
기지(既知)이거나 표본이 클 경우)

$$H_0 : \mu_1 - \mu_2 = \mu_0$$

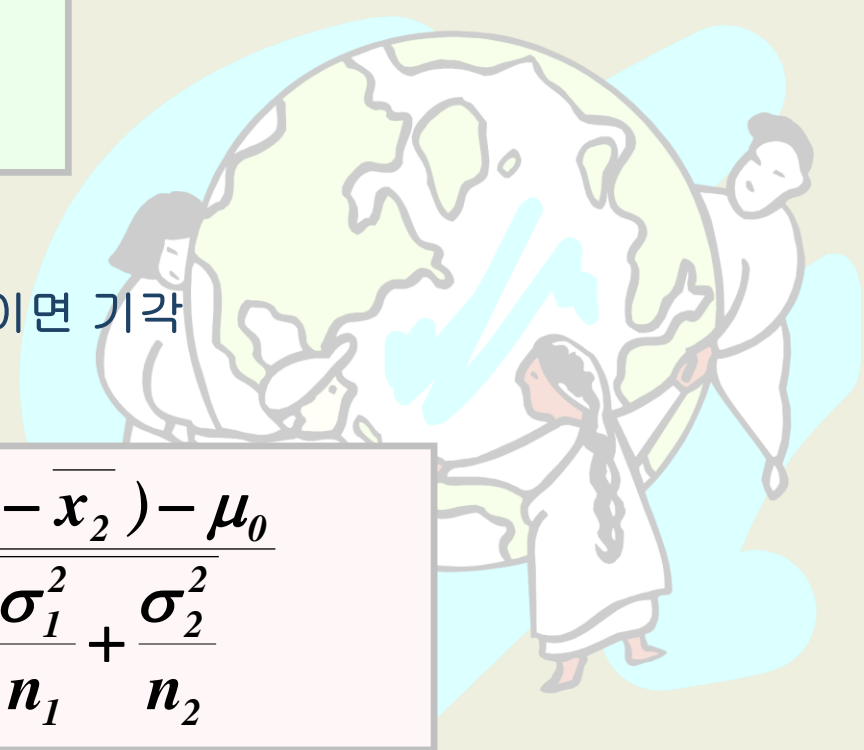
$$H_1 : \mu_1 - \mu_2 \neq \mu_0$$

◆ 기각역:

$$|Z| > z_{\frac{\alpha}{2}} \quad \text{이면 기각}$$

◆ 검정통계량

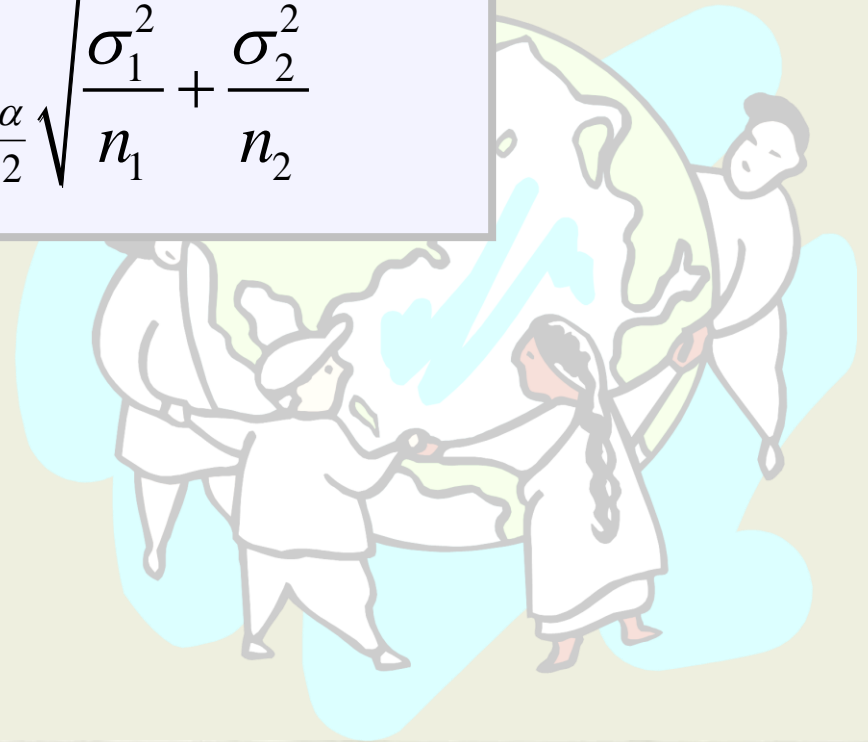
$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$



모평균의 가설검정-1)-a)

$\mu_1 - \mu_2$ 에 대한 $100(1-\alpha)\%$ 신뢰구간

$$CI = (\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$



모평균의 가설검정-1)-a)

예제 10-1

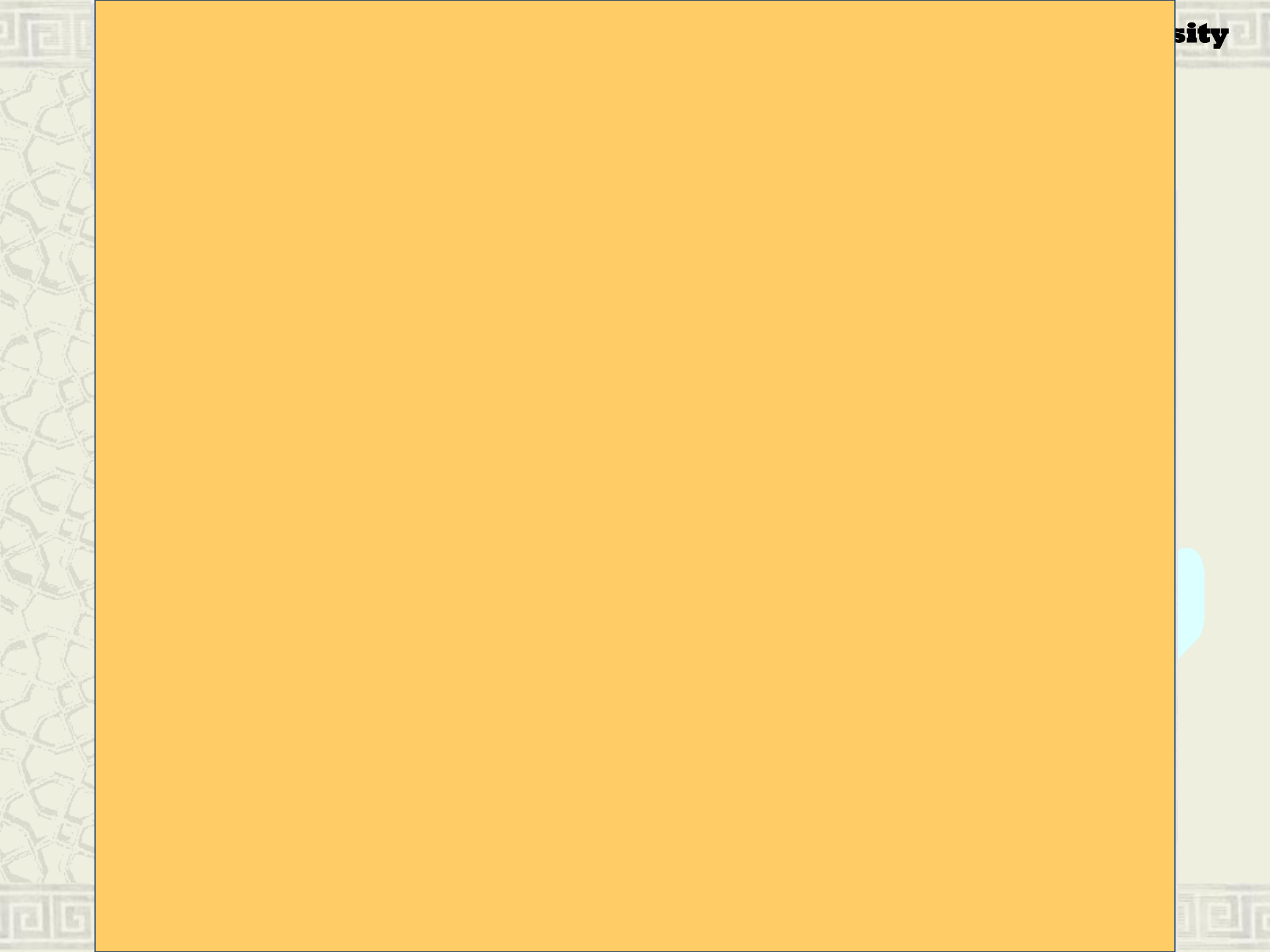
해동치약은 자사의 죽염치약이 어린이 충치예방에 효과가 있다고 주장하고 있다. 죽염치약을 사용하지 않는 어린이 100명(n_1)을 무작위추출하여 충치 수를 관찰한 결과, $\bar{X}_1 = 2.7$ 이고 $S_1^2 = 1.1$ 로 나타났다. 또 죽염치약을 사용하는 어린이 120명(n_2)에 대한 \bar{X}_2 와 S_2^2 은 각각 2.4와 0.9였다. 일반치약과 죽염치약의 평균충치 수의 차이 ($\mu_1 - \mu_2$)에 대한 90% 신뢰구간을 구하시오.

모평균의 가설검정-1)-a)

예제 10-3

예제 1의 문제에서 대립가설 $H_A: \mu_1 - \mu_2 > 0$ 에 대하여 유의수준 0.05에서 검정하시오.





모평균의 가설검정-1)-b)

- ◆ 모평균의 가설검정 (표본이 서로 독립적일 때: 모분산이 미지(未知)이나 서로 같고 표본이 작을 때)

- ◆ 검정통계량

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- ◆ $\mu_1 - \mu_2$ 에 대한 $100(1-\alpha)\%$ 신뢰구간

$$CI = (\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2, \frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

예제 10-5

K사는 이동통신업체의 새로운 강자로 부상하고 있는 신흥업체이다. K사는 최근 새로운 스마트폰 제품을 통하여 모바일사업 분야로 그 영역을 넓혀가려고 한다. 이러한 새로운 사업진입을 위한 마케팅전략을 수립하기 위해 L과장이 영입되었다.

L과장은 고려전자의 K스마트폰과 안압전자의 A스마트폰 중 어느 제품을 통해 모바일사업을 확장해야 할지 고민에 빠지게 되었다. 따라서 L과장은 M리서치 회사에 두 제품에 대한 소비자 만족도 조사를 의뢰하였다.

M리서치 회사는 두 제품의 소비자 만족도를 측정하기 위해 무작위로 소비자들을 추출하여 각 제품에 대한 만족도를 100점 만점으로 측정하였다.

그 결과는 표 10-1과 같다. 만족도는 대체로 정규분포를 따르며, 두 스마트폰에 대한 만족도의 분산은 동일하다고 가정한다.

표 10-1 두 표본의 소비자 만족도

표 본	소비자 만족도					
	K스마트폰	65	76	63	68	71
A스마트폰	79	73	67	80	77	

- (a) 두 표본의 소비자 만족도 차이 $\mu_1 - \mu_2$ 에 대한 95% 신뢰구간을 구하시오.
- (b) K스마트폰의 소비자 만족도가 A스마트폰의 소비자 만족도보다 높다고 할 수 있는지를 5%의 유의수준에서 검정하시오.

표 10-1의 자료를 그림 10-2와 같이 입력한다. 그리고 나서 데이터 메뉴에서 데이터 분석을 선택하여 두 집단 간 등분산인지 아닌지를 그림 10-2와 같이 *F*검정을 이용하여 분석하면 그림 10-3의 결과가 나타난다.

그림 10-2 *F*검정 선택 및 대화상자

The figure consists of two screenshots of Microsoft Excel. The left screenshot shows the 'Data Analysis' dialog box with 'F-검정: 분산에 대한 두 집단' selected. The right screenshot shows the 'F-검정: 분산에 대한 두 집단' dialog box with input fields for '변수 1 입력 범위', '변수 2 입력 범위', '유의 수준', and '출력 옵션'.

표 본	K 스마트폰	A 스마트폰
1	65	79
2	75	73
3	63	67
4	68	80
5	71	77
6	77	

두 집단의 등분산에 대한 가설검정결과와 $H_0: \sigma_K^2 = \sigma_A^2$, $H_A: \sigma_K^2 \neq \sigma_A^2$ 인 가설에서 p 값을 이용한 가설검정결과 그리고 F 검정통계량과 F 기각치를 비교한 결과, 대립가설을 기각하고 귀무가설을 채택하므로 두 집단 간 분산이 같다.

그림 10-3 F 검정의 결과

	A	B	C	D	E	F
1	표	K 스마트폰	A 스마트폰	F-검정: 분산에 대한 두 집단		
2	1	65	79			
3	2	76	73		K 스마트폰	A 스마트폰
4	3	63	67	평균	70	75.2
5	4	68	80	분산	32.8	28.2
6	5	71	77	관측수	6	5
7	6	77		자유도	5	4
8				F 비	1.163120567	
9				P(F<=f) 단측 검정	0.454512608	
10				F 기각치: 단측 검정	6.256056502	

두 집단의 분산에 대한 검정결과, 등분산이라는 귀무가설이 채택되었으므로 엑셀의 '데이터-데이터 분석'에서 그림 10-4와 같이 『 t -검정: 등분산가정 두 집단』을 선택한다. 만약 문제에서 두 집단의 분산이 동일하지 않은 것으로 주어지면 『 t -검정: 이분산가정 두 집단』을 선택하면 된다. 그림 10-4에서 확인을 누르면 그림 10-5의 대화상자가 나타나게 된다.

그림 10-4 t검정: 등분산가정 두 집단 선택

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E
1	표 본	K 스마트폰	A 스마트폰		
2	1	65	79		
3	2	76	73		
4	3	63	67		
5	4	69	80		
6	5	71	77		
7	6	77			

The 't-Test: Two-Sample Assuming Equal Variances' dialog box is open, showing the following settings:

- 변수 1의 범위 (Variable 1 Range): \$B\$1:\$B\$7
- 변수 2의 범위 (Variable 2 Range): \$C\$1:\$C\$7
- 표본 평균을 같음 (Assume Equal Variances) is checked.
- 결과를 출력할 위치 (Output Range): \$D\$1

변수 1의 입력범위는 B1부터 B7까지 지정하고, 변수 2의 입력범위에는 C1에서 C7까지를 지정한다. 또한 가설평균차는 두 집단 간에 차이가 있는지 없는지에 관한 것이기 때문에 0을 입력한다. 또한 변수 입력 시 K스마트폰, A스마트폰이 포함되어 있으므로 첫 번째 행은 이름표로 사용하도록 선택한다.

그림 10-5 t검정: 등분산가정 두 집단 대화상자

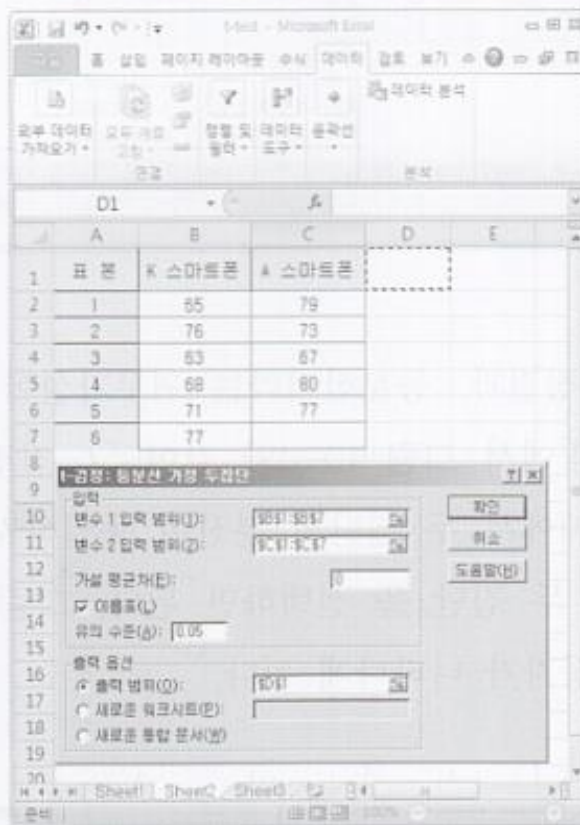


그림 10-5와 같이 입력을 하고 확인을 누르면 그림 10-6과 같은 결과가 출력된다. t 통계량과 기각치를 비교한 분석결과, 가설검정에 양측검정을 이용하게 되므로 귀무가설을 채택하고 대립가설을 기각한다고 해석할 수 있다. 또한 p 값에 의한 가설검정결과, 귀무가설을 채택하고 대립가설을 기각하는 것으로 나타났다.

그림 10-6 t검정: 등분산가정 두 집단의 결과

표본	K 스마트폰	A 스마트폰	t-검정: 등분산 가정 두 집단		
1					
2	1	65	78		
3	2	76	73		
4	3	63	67	평균	70
5	4	68	80	분산	32.8
6	5	71	77	관측수	6
7	6	77		공통(Pooled) 분산	30.75555556
8				가설 평균차	0
9				자유도	9
10				t 통계량	-1.540480855
11				P(T<=t) 단측 검정	0.07795715
12				t 기각치 단측 검정	1.833112933
13				P(T<=t) 양측 검정	0.1559143
14				t 기각치 양측 검정	2.262157163

그림 10-6에서 보는 바와 같이 양측검정과 단측검정 모두 p 값($=0.07795715$)이 유의수준 0.05보다 크므로 귀무가설을 채택하고 대립가설을 기각하게 된다. 따라서 K스마트폰이 A스마트폰보다 소비자 만족도가 크다고 할 수 없다.

모비율에 대한 가설검정

◆ 모비율의 가설검정

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 \neq 0$$

◆ 검정통계량

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

◆ ※ 참고 : 이 경우 두 모집단이 동일한 모비율을 갖고 있다는 의미로

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

을 이용한다.

예제 10-2

두 종류의 냉장고 A와 B에 대한 품질보증기간은 각각 1년이다. A상표의 냉장고 중 50개를 무작위추출하여 관찰하였더니 그중 12개가 품질보증기간 중에 고장이 발생하였다. 또 B상표에 대해서도 60개를 임의로 추출하여 조사한 결과, 12개가 품질보증기간 중에 고장이 발생하였다. 품질보증기간 중 이 두 상표의 고장률의 차이 ($p_1 - p_2$)에 대한 98% 신뢰구간을 구하시오.



예제 10-4

예제 2에서 귀무가설과 대립가설이 각각 $H_0: p_1 = p_2$, $H_A: p_1 \neq p_2$ 로 주어졌을 때, 유의수준 0.05에서 가설검정을 하시오.



모분산에 대한 가설검정

- ▶ 두 정규모집단의 분산 비교
 - ◆ t분포를 이용 모평균을 비교하려면 등분산의 가정을 전제로 하여야 하는데 이때 가정이 적합한지에 대한 분석이 선행되어야 한다.
- ▶ 모분산에 대한 가설검정

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

 \Leftrightarrow

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

◆ 표본분산 s_1^2, s_2^2 계산

◆ 기각역

$$F < F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}$$

또는

$$F > F_{n_1-1, n_2-1, \frac{\alpha}{2}}$$

이면 H_0 기각 (양측검정)

모분산에 대한 가설검정

◆ 검정통계량

$$F = \frac{S_1^2}{S_2^2}$$

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

$$F_{n_2-1, n_1-1, 1-\frac{\alpha}{2}} \frac{S_1^2}{S_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq F_{n_2-1, n_1-1, \frac{\alpha}{2}} \frac{S_1^2}{S_2^2}$$

F 분포

$$F_{n_1-1, n_2-1} = \frac{\chi_{n_1-1}^2 / (n_1 - 1)}{\chi_{n_2-1}^2 / (n_2 - 1)}$$

$$F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}} = \frac{1}{F_{n_2-1, n_1-1, \frac{\alpha}{2}}}$$

- ◆ F분포표 보는 방법 pp.504-513 표->다음 슬라이드

0. F分布表

$$P(F \geq F_{\alpha}) = \alpha$$

$$(\alpha = 0.10)$$



$v_2(d.f.)$	$v_1(d.f.)$																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

예제 10-9

증권시장에 상장되어 있는 주식 중 제조업종과 금융업종 주식에 대한 수익률의 분산을 비교하려고 한다. 31(n_1)개의 제조업종과 21(n_2)개의 금융업종 주식으로 이루어진 독립표본을 얻어 각각에 대한 표본 분산 $S_1^2=9.4$ 와 $S_2^2=7.0$ 을 얻었다. σ_1^2/σ_2^2 에 대한 95% 신뢰구간을 구하시오.

예제 10-10

예제 9에서 제조업종 주식의 분산이 금융업종 주식의 분산보다 크다고 할 수 있는지 유의수준 0.05에서 검정하시오.

모평균의 가설검정-2)

▶ 모평균의 가설검정(표본이 서로 종속적일때 (대응 표본))

$$H_0 : \mu_d = \mu_0$$

$$H_1 : \mu_d \neq \mu_0$$

검정통계량

$$t = \frac{\bar{d} - \mu_0}{s_d / \sqrt{n}}$$

여기서

$$d_i = x_{i_1} - x_{i_2}, \bar{d} = \frac{\sum d_i}{n}$$

$$S_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$$

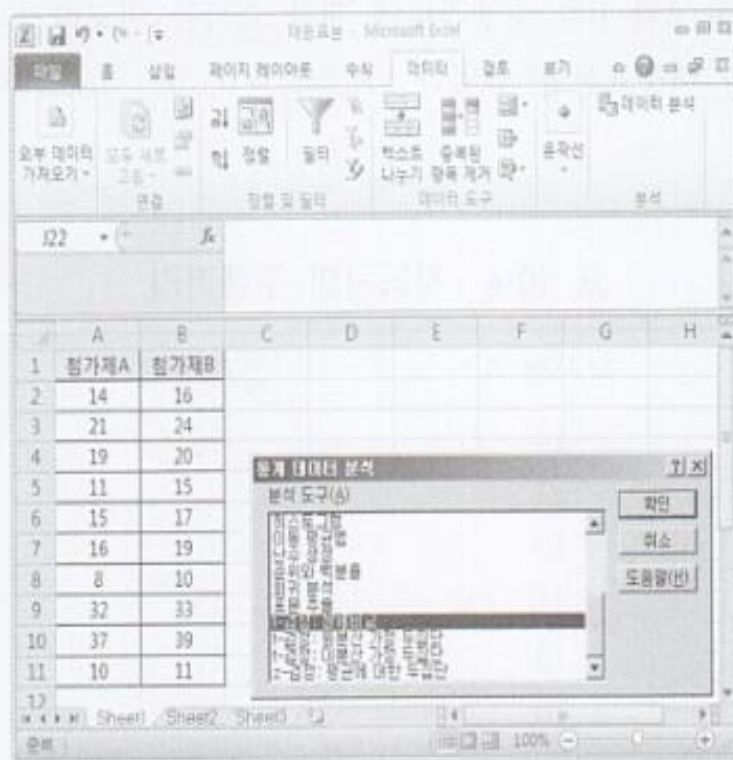
두 가지 종류의 윤활유 첨가제 A와 B의 주행거리에 대한 효과를 비교하기 위해 10대의 차량을 무작위로 추출하여 표 10-4와 같은 대응표본을 얻었다. $H_0: \mu_d = 0$, $H_A: \mu_d \neq 0$ 에 대해 유의수준 0.05에서 검정하시오.

표 10-4 자동차의 주행거리

차 량	첨가제 A	첨가제 B	d_i
1	14	16	-2
2	21	24	-3
3	19	20	-1
4	11	15	-4
5	15	17	-2
6	16	19	-3
7	8	10	-2
8	32	33	-1
9	37	39	-2
10	10	11	-1

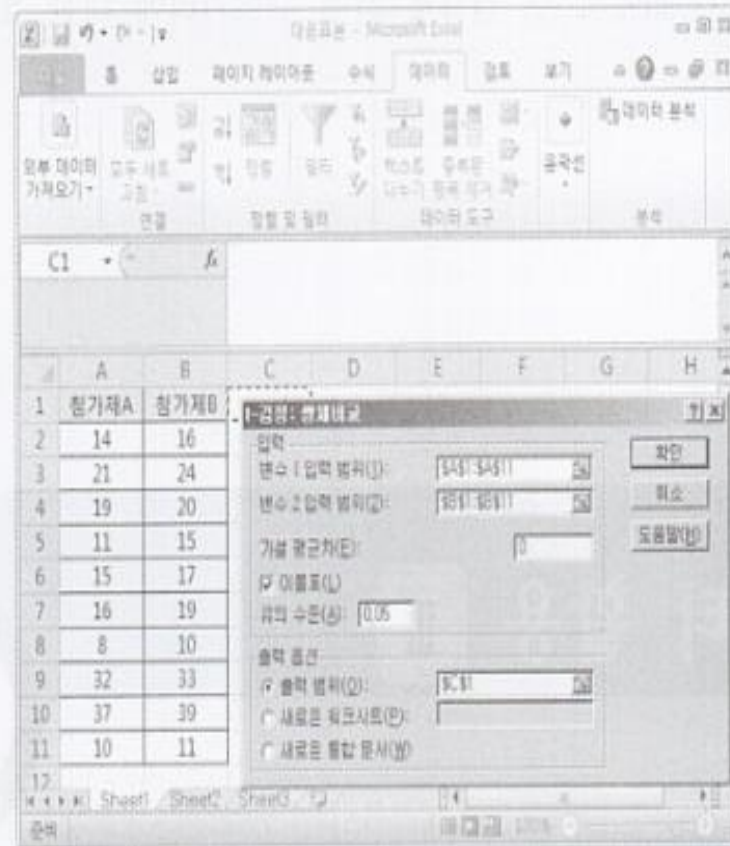
표 10-4의 자료를 그림 10-14와 같이 입력한다.

그림 10-14 t검정: 쌍체비교 선택



대응표본의 경우 데이터 메뉴에서 데이터 분석을 선택하면 그림 10-14의 화면이 나타난다. 여기서 『t-검정: 쌍체비교』를 선택한다. 그림 10-14에서 확인을 누르면 그림 10-15의 대화상자가 나타난다.

그림 10-15 t검정: 쌍체비교 대화상자



여기서 각 입력칸에 셀범위를 지정한다. 현재 귀무가설이 유효유 첨가 전
 후의 주행거리에 차이가 없다는 것이므로 가설평균차에는 0을 입력한다. 유의
 수준 0.05를 입력하고 출력범위를 지정한 후 확인을 누르면 그림 10-16과 같
 이 결과가 주어진다.

그림 10-16 t검정: 쌍체비교의 결과

	A	B		D	E	F
1	첨가제A	첨가제B	t-검정: 쌍체 비교			
2	14	16				
3	21	24		첨가제A	첨가제B	
4	19	20	평균	18.30000	20.40000	
5	11	15	분산	89.78889	86.26667	
6	15	17	관측수	10.00000	10.00000	
7	16	19	피어슨 상관 계수	0.99458		
8	8	10	가설 평균차	0.00000		
9	32	33	자유도	9.00000		
10	37	39	t 통계량	-6.67799		
11	10	11	P(T<=t) 단측 검정	0.00005		
12			t 기각치 단측 검정	1.83311		
13			P(T<=t) 양측 검정	0.00009		
14			t 기각치 양측 검정	2.3		

그림 10-16에서 보는 바와 같이 양측검정의 경우 p값(0.00009)이 유의수
 준 0.05보다 크지 않으므로 귀무가설을 기각하게 된다. 따라서 첨가제의 사용
 이 주행거리에 영향을 미친다고 결론짓는다.