

제 12장

분산분석



고려대학교 경영대학 박 광태

분산분석(ANOVA)

- ◆ 분산분석(ANOVA)
 - ◆ 셋 이상의 모평균에 대한 검정
- ◆ ANOVA검정의 가정
 - ◆ 모집단들은 등분산을 가짐 (Hartley 검정 이용)
 - ◆ 모집단들은 정규분포를 따름 (카이제곱 검정 이용)
 - ◆ 관찰치는 서로 독립임



일원분류 ANOVA 검정-1)

일원분류 ANOVA 검정

- ◆ 하나의 특정한 독립변수 또는 요인이 종속변수의 값에 영향을 미치는 지에 대한 검정

- ◆ $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

$H_1 : H_0$ 중 적어도 한 처리효과는 0이 아니다.

- ◆ ANOVA 표의 작성

변동요인	제곱합	자유도	평균제곱합	F-ratio
표본간	SSB	$k-1$	$MSB=SSB/(k-1)$	MSB/MSW
표본내	SSW	$n-k$	$MSW=SSW/(n-k)$	
합	SST	$n-1$		

일원분류 ANOVA 검정-2)

◆ 일원분류 ANOVA 검정(계속)

- ◆ 분자의 자유도 ($k-1$)과 분모의 자유도 ($n-k$)를 갖는 F 분포를 사용하여 기각치 $F_{\alpha, k-1, n-k}$ 결정

- ◆ 기각역:

$$\frac{MSB}{MSW} = F > F_{\alpha, k-1, n-k}$$

이면 H_0 기각



예제 12-1

국내 자동차기업인 H사, 해외 자동차기업인 B사와 V사의 자동차 연비(단위: 천km)에 차이가 있는지를 분석하고자 한다. 각 회사별로 다섯 대의 자동차를 임의로 추출하여 연비를 기록한 결과가 표 12-2에 주어 있다. 이 표를 사용하여 자동차 3사의 연비에 차이가 있는지를 유의수준 0.05에서 검정하시오.

표 12-2 자동차 3사의 연비조사 결과

순서	B사	V사	H사
1	18.2	19.8	21.2
2	19.4	21.0	21.8
3	19.6	20.0	22.4
4	19.0	20.8	22.0
5	18.8	20.4	21.6
\bar{x}	19.0	20.4	21.8





등분산성에 대한 하틀리 검정



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

$H_A : H_0$ 은 사실이 아니다.



기각역 :

$H > H_\alpha$ 이면 H_0 기각



검정통계량

$$H = \frac{\max(S_1^2, S_2^2, \dots, S_k^2)}{\min(S_1^2, S_2^2, \dots, S_k^2)}$$

Hartley의 H 통계량에 대한 기각치

$$P(H \geq H_\alpha) = \alpha$$

$$(\alpha = 0.05)$$

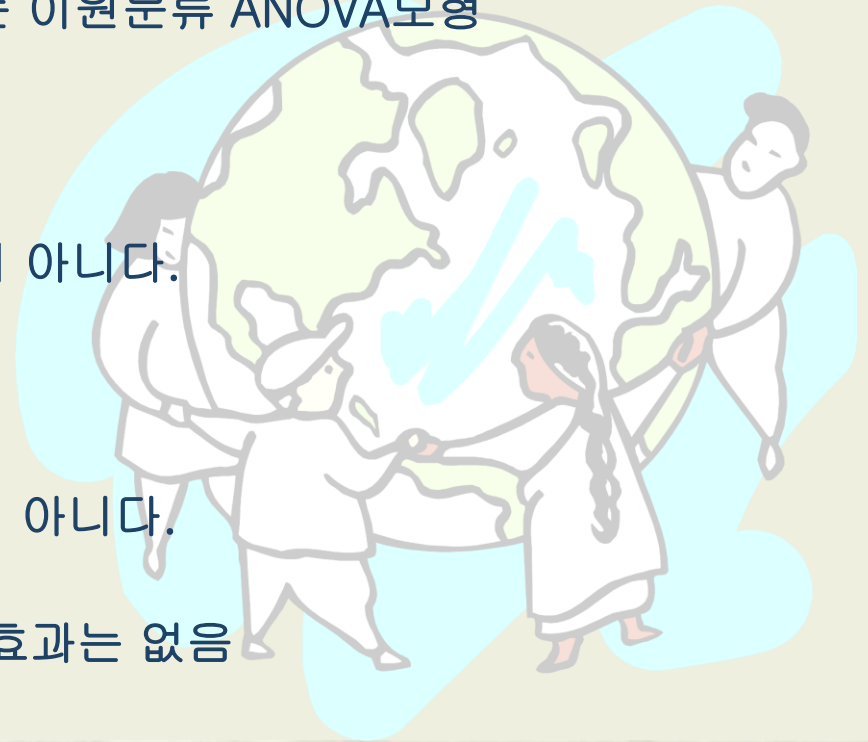
n	k										
	2	3	4	5	6	7	8	9	10	11	12
3	39.0	87.5	142	202	266	333	403	475	550	626	704
4	15.4	27.8	39.2	50.7	62.0	72.9	83.5	93.9	104	114	124
5	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.1	44.6	48.0	51.4
6	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
7	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
8	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
9	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
10	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
11	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
13	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
16	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
21	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
31	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
61	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

예) $n=5$ 이고 $k=3$ 이면 $P(H \geq 15.5) = \alpha$ 이다.

이원분류 ANOVA검정-1)

▶ 이원분류 ANOVA검정

- ◆ 두개의 독립변수 또는 요인이 종속변수의 값에 영향을 미치는지에 대한 검정
- ◆ 각 cell에 하나의 관찰치가 있는 이원분류 ANOVA모형
- ◆ $H_0 : A_1 = A_2 = \dots = A_a = 0$
(요인 A의 효과가 없음)
 $H_1 : 적어도 하나의 A_i 은 0이 아니다.$
- ◆ $H_0 : B_1 = B_2 = \dots = B_b = 0$
(요인 B의 효과가 없음)
 $H_1 : 적어도 하나의 B_j 은 0이 아니다.$
- ◆ 주의 : 요인 A와 B사이의 교호효과는 없음



이원분류 ANOVA검정 -2)

- ANOVA 표의 작성

변동요인	제곱합	자유도	평균제곱합	F -ratio
요인 A	SSA	$a-1$	$MSA=SSA/(a-1)$	$F_A=MSA/ MSE$ $F_B=MSB/ MSE$
요인 B	SSB	$b-1$	$MSB=SSB/(b-1)$	
오차	SSE	$(a-1)(b-1)$	$MSE=SSE/(a-1)(b-1)$	
합	SST	$ab-1$		

- 유의수준 α 에서 행효과 A_j 가 모두 0이라는 귀무가설 검정

$$F_A > F_{\alpha, a-1, (a-1)(b-1)}$$

기각

- 유의수준 α 에서 행효과 B_j 가 모두 0이라는 귀무가설 검정

$$F_B > F_{\alpha, b-1, (a-1)(b-1)}$$

이면 H_0 기각

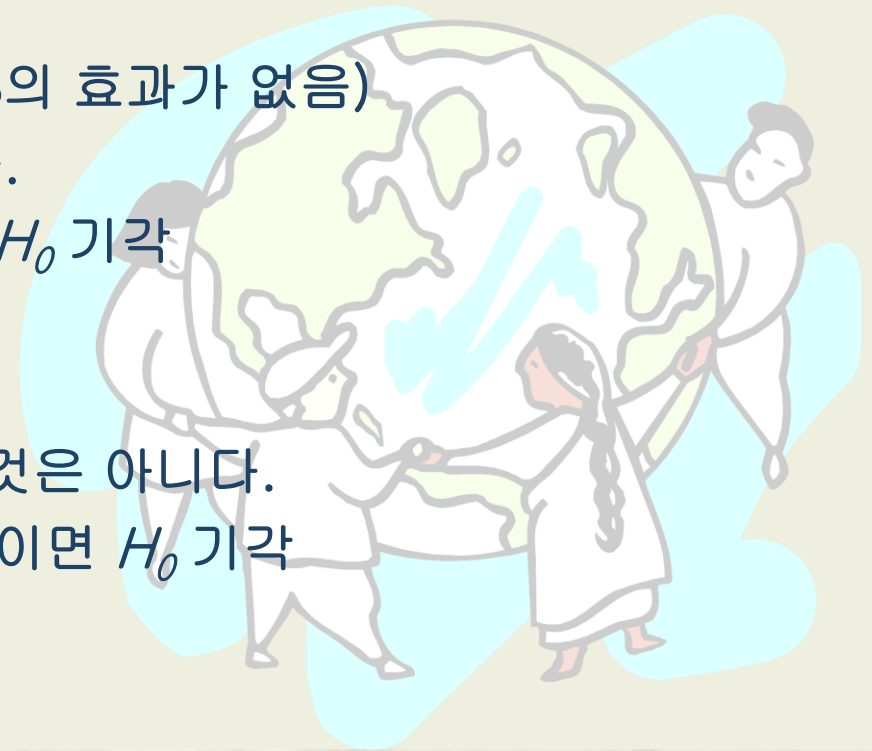
이원분류 ANOVA검정 -3)

- ◆ 각 cell 에 다수의 관찰치가 있는 이원분류 ANOVA모형
- 교호효과를 분석할 수 있음
- ◆ ANOVA 표의 작성

변동요인	제곱합	자유도	평균제곱합	<i>F</i> -ratio
요인 A	<i>SSA</i>	<i>a-1</i>	$MSA=SSA/(a-1)$	$F_A=MSA/ MSE$ $F_B=MSB/ MSE$ $F_I=MSI/ MSE$
요인 B	<i>SSB</i>	<i>b-1</i>	$MSB=SSB/(b-1)$	
교호효과	$SSA \times B$	$(a-1)(b-1)$	$MSI=SSI/(a-1)(b-1)$	
오차	<i>SSE</i>	$ab(c-1)$	$MSE=SSE/ab(c-1)$	
합	<i>SST</i>	$abc-1$		

이원분류 ANOVA검정 -4)

- ◆ $H_0 : \mu_1 = \mu_2 = \dots = \mu_a$ (요인 A의 효과가 없음)
 $H_1 : \text{평균이 모두 같지 않다.}$
 $\Rightarrow F_A > F_{\alpha, a-1, ab(c-1)}$ 이면 H_0 기각
- ◆ $H_0 : \mu_1 = \mu_2 = \dots = \mu_b$ (요인 B의 효과가 없음)
 $H_1 : \text{평균이 모두 같지 않다.}$
 $\Rightarrow F_B > F_{\alpha, b-1, ab(c-1)}$ 이면 H_0 기각
- ◆ $H_0 : l_{ij} = 0 (\forall i, j)$
 $H_1 : \text{교호효과가 모두 0인 것은 아니다.}$
 $\Rightarrow F_1 > F_{\alpha, (a-1)(b-1), ab(c-1)}$ 이면 H_0 기각



예제 12-4

의사는 환자들의 평균입원기간(단위: 일)이 약의 종류와 비타민의 종류와 관련이 있는지를 조사하고자 한다. 세 가지 종류의 약 M_1, M_2, M_3 와 세 가지 종류의 비타민 V_1, V_2, V_3 가 사용되었다고 하자. 9개 조합 $M_i V_j$ 의 각각에 $c=4$ 명의 환자로 이루어진 확률표본이 주어져 있다.

따라서 각각의 cell에 4개의 관찰치가 있고 전체 9개의 cell에 총 36개의 관찰치가 주어져 있다. 네 명의 환자에게 $M_1 V_1$ 조합이 주어지고, 또 다른 네 명의 환자에게는 $M_1 V_2$ 조합 등등이 주어진다. 그 자료가 표 12-8에 나타나 있다. 5% 유의수준을 사용하여 약의 효과, 비타민의 효과 그리고 교호효과가 0이라는 세 가지 귀무가설을 검정하시오.

표 12-8 평균 입원기간의 자료

요인 A \ 요인 B	M_1	M_2	M_3	
V_1	10	9	12	
	6	3	10	
	8	5	12	
	4	7	10	
	$\bar{x}_{11.} = 7$	$\bar{x}_{12.} = 6$	$\bar{x}_{13.} = 11$	$\bar{x}_{1..} = 8$
V_2	9	8	6	
	9	6	8	
	6	14	10	
	12	12	8	
	$\bar{x}_{21.} = 9$	$\bar{x}_{22.} = 10$	$\bar{x}_{23.} = 8$	$\bar{x}_{2..} = 9$
V_3	6	12	8	
	4	13	12	
	4	15	11	
	6	16	13	
	$\bar{x}_{31.} = 5$	$\bar{x}_{32.} = 14$	$\bar{x}_{33.} = 11$	$\bar{x}_{3..} = 10$
$\bar{x}_{.1.} = 7$	$\bar{x}_{.2.} = 10$	$\bar{x}_{.3.} = 10$	$\bar{x} = 9$	

그림 12-11과 같이 예제 4의 수치를 입력한다.

이제 예제 4의 가설검정을 위해 데이터 메뉴에서 데이터 분석을 선택하면 그림 12-12의 화면이 나타난다. 여기서 『분산분석: 반복 있는 이원배치법』을

선택하고 확인을 누르면 그림 12-13이 나타나게 된다.

그림 12-11 자료입력

	A	B	C	D	E	F
1	요인 A					
2	요인 A	M1	M2	M3		
3		10	9	12		
4		6	3	10		
5	Y1	8	5	12		
6		4	7	10		
7		7	6	11		
8		9	8	6		
9		9	6	8		
10	Y2	6	14	10		
11		12	12	6		
12		9	10	6		
13		6	12	6		
14		4	13	12		
15	Y3	4	15	11		
16		6	16	13		
17		5	14	11		
18						
19						

그림 12-12 분산분석: 반복 있는 이원배치법 선택

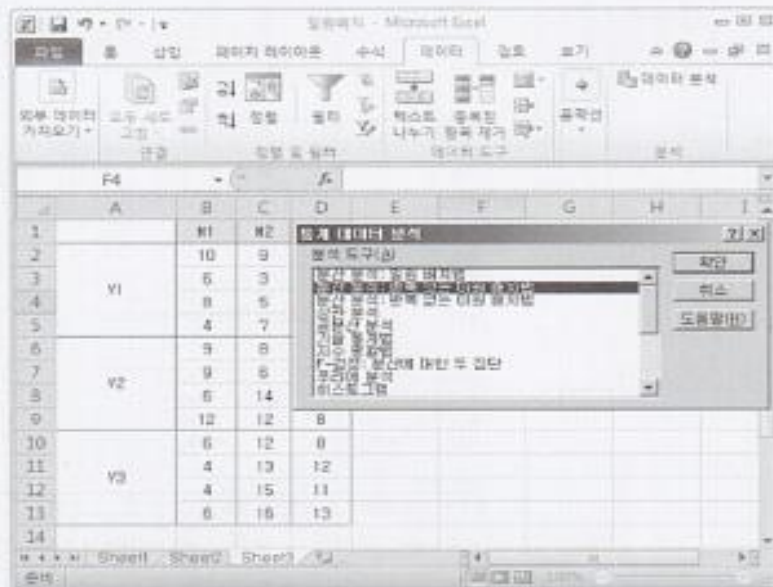


그림 12-13 분산분석: 반복 있는 이원배치법 대화상자

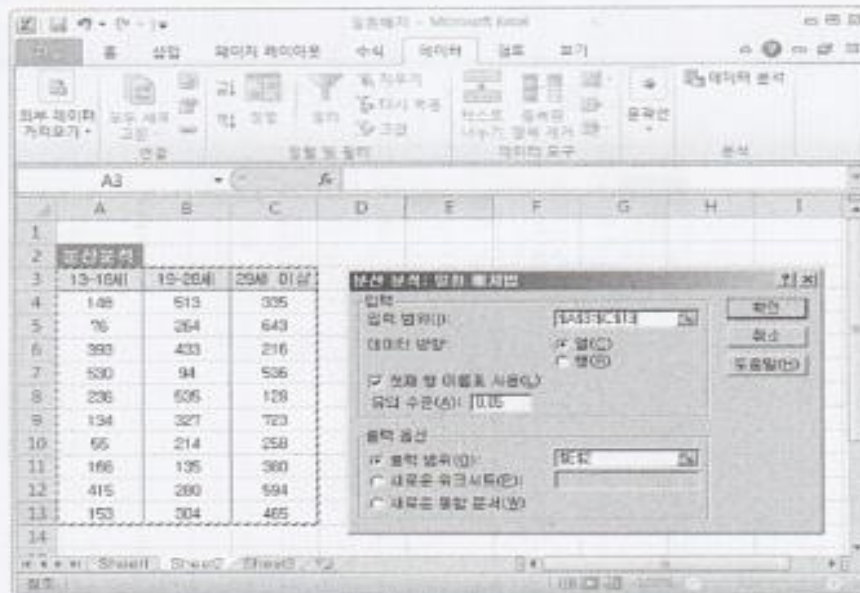


그림 12-14 분산분석: 반복 있는 이원배치법의 결과

	A	B	C	D		F	G	H	I	J	K	
1		M1	M2	M3		분산 분석: 반복 있는 이원 배치법						
2		10	9	12		요약표	M1	M2	M3	계		
3	V1	6	3	10		V1						
4		8	5	12		관측수	4	4	4	12		
5		4	3	10		합	20	24	44	96		
6		5	5	8		평균	7	6	11	8		
7	V2	9	8	8		분산	6.666667	6.666667	1.333333	8.000000		
8		6	14	10		V2						
9		12	12	8		관측수	4	4	4	12		
10		8	10	8		합	36	40	32	108		
11	V3	4	10	12		평균	9	10	8	9		
12		4	15	11		분산	6	13.33333	2.666667	6.727273		
13		6	16	13		V3						
14						관측수	4	4	4	12		
15						합	20	36	44	120		
16						평균	5	14	11	10		
17						분산	1.333333	3.333333	4.666667	17.81818		
18						계						
19						관측수	12	12	12			
20						합	84	120	120			
21						평균	7	10	10			
22						분산	6.727273	18	4.545455			
23						분산 분석						
24						변동 요인	자유도	제곱 평균	F 비	P-값	F 기각치	
25						인자 A(행)	24	2	12	2.347826	0.114793	3.354131
26						인자 B(열)	12	2	36	7.043478	0.003455	3.354131
27						교호작용	160	4	40	7.826087	0.000253	2.727765
28						잔차	138	27	5.111111			
29						계	394	35				

그림 12-13에서 대화상자에 필요한 정보를 입력한다. 유의수준은 0.05로, 표본당 행수는 4개이므로 4로 지정한 후 확인을 누르면 그림 12-14와 같이 결과가 주어진다. 그림 12-14의 결과는 예제 4에 자세히 설명되어 있으므로 예제 4를 참조하기 바란다.