Hypothesis Ⅲ

-More than Three Groups-

Contents

1. 일원 분산분석(One-way ANOVA)

2. 이원 분산분석(Two-way ANOVA)

3. 사후분석(Post-hoc analysis)

1. 분산분석(ANOVA)

분산분석(ANOVA)

세 그룹 이상의 평균 비교

< Data Set >

ID	Age	Sex	Stage	Cr
1021	45	1	1	0.7
1078	58	1	2	1.8
1110	72	2	3	3.1
1124	37	2	1	0.5
1145	32	1	1	0.5
1187	66	2	3	2.0
1211	61	2	2	1.7
⋮	⋮	⋮	⋮	⋮

< One-way ANOVA >

	Stage			
	1	2	3	
Stage 별 Cr 값 정리	0.7	1.8	3.1	
	0.5	1.7	2.0	
	0.5			
	⋮	⋮	⋮	
합계	$T_{1\cdot}$	$T_{2\cdot}$	$T_{3\cdot}$	T
평균	$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$	$\bar{y}_{3\cdot}$	$\bar{y}$

분산분석(ANOVA)

분산분석 정의

분산분석(ANalysis Of VAriance; ANOVA)

- **반응치의 산포를 요인 별로 분해**하여, 반응치의 유의한 영향을 주는 요인을 찾아내는 통계적 기법

- H_0: $\mu_1 = \mu_2 = \cdots = \mu_r$
- H_1: 적어도 한 개의 μ_i는 나머지와 같지 않다.

- 세 가지 이상의 수준(level) 또는 범주(category)을 갖는 범주형 자료가 예측변수이고, 연속형변수가 결과변수일 때 사용한다.
- ANOVA test는 모든 범주에서 평균 차이가 발생하는지를 확인하는 것이 주된 목적이며, 어떤 범주에서 차이가 발생하는지를 확인하는 것이 사후분석(post-hoc analysis)이다.

	Stage			
	1	**2**	**3**	
Stage 별 Cr 값 정리	0.7	1.8	3.1	
	0.5	1.7	2.0	
	0.5			
	⋮	⋮	⋮	
합계	$T_{1\cdot}$	$T_{2\cdot}$	$T_{3\cdot}$	T
평균	$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$	$\bar{y}_{3\cdot}$	$\bar{y}$

분산분석(ANOVA)

일원 분산분석(One-way ANOVA)

	수준(level)				
	$\boldsymbol{A_1}$	$\boldsymbol{A_2}$	$\cdots$	$\boldsymbol{A_r}$	
반응치	y_{11}	y_{21}	$\cdots$	y_{r1}	
	y_{12}	y_{22}	$\cdots$	y_{r2}	
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	
	y_{1n_1}	y_{2n_2}	$\cdots$	y_{rn_r}	
합계	$T_{1\cdot}$	$T_{2\cdot}$		$T_{r\cdot}$	T
평균	$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$		$\bar{y}_{r\cdot}$	$\bar{y}$

- 반응치의 산포를 요인 별로 분해하여, 반응치의 유의한 영향을 주는 요인을 찾는다.
- 한 개의 변수 또는 요인(factor)만을 고려하는 경우를 일원 분산분석(one-way ANOVA)이라 한다.
- 변수 또는 요인(factor) A가 r개의 수준을 가질 때, y_{ij}에서 i는 요인의 수준을 나타내며 j는 각 수준에서의 관측된 반응치 개수를 의미한다. 각 수준 별 반응치 개수는 서로 다를 수 있다.

분산분석(ANOVA)

일원 분산분석(One-way ANOVA)

	수준(level)				
	$\boldsymbol{A_1}$	$\boldsymbol{A_2}$	$\cdots$	$\boldsymbol{A_r}$	
반응치	y_{11}	y_{21}	$\cdots$	y_{r1}	
	y_{12}	y_{22}	$\cdots$	y_{r2}	
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	
	y_{1n_1}	y_{2n_2}	$\cdots$	y_{rn_r}	
합계	$T_{1\cdot}$	$T_{2\cdot}$		$T_{r\cdot}$	T
평균	$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$		$\bar{y}_{r\cdot}$	$\bar{y}$

- 전체 수준에서의 모평균을 μ, 수준 A_i에서 반응치의 모평균을 μ_i라 하면, 다음과 같은 구조식으로 나타낼 수 있다.

$$y_{ij} = \mu_i + \varepsilon_{ij} = \mu + a_i + \varepsilon_{ij}, \qquad i = 1,2,\cdots,r, \qquad j = 1,2,\cdots,n_i$$

- 여기서 $a_i \equiv \mu_i - \mu$를 요인 A의 주효과라고 하고, 오차 ε_{ij}들은 서로 독립이고 $N(0,\sigma^2)$ 분포를 가정한다.

분산분석(ANOVA)

일원 분산분석(One-way ANOVA)

$$y_{ij} = \mu_i + \varepsilon_{ij} = \mu + a_i + \varepsilon_{ij}, \qquad i = 1,2,\cdots,r, \qquad j = 1,2,\cdots,n_i$$

- 오차에 대한 네 가지 가정, 1) 독립성, 2) 정규성, 3) 불편성($\mathrm{E}(\varepsilon_{ij}) = 0$), 4) 등분산성이 내포되어 있다. 오차의 가정에 따라 반응변수는 다음과 같은 특성을 갖는다.

① $\mathrm{E}(\varepsilon_{ij}) = 0$이므로 $\mathrm{E}(y_{ij}) = \mu_i = \mu + a_i$

② $\mathrm{Var}(\varepsilon_{ij}) = \sigma^2$이므로 $\mathrm{Var}(y_{ij}) = \mathrm{Var}(\varepsilon_{ij}) = \sigma^2$

③ ε_{ij}가 서로 독립이므로 y_{ij}도 서로 독립이다.

④ $y_{ij} \sim N(\mu + a_i,\ \sigma^2)$

분산분석(ANOVA)

일원 분산분석(One-way ANOVA)

일원 분산분석 제곱합의 분해
● $SS_T = SS_A + SS_E$ ✓ 총제곱합: $SS_T = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N}$ ✓ 처리제곱합: $SS_A = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{y}_{i\cdot}-\bar{y})^2 = \sum_{i=1}^{r}\frac{T_{i\cdot}^2}{n_i} - \frac{T^2}{N}$ ✓ 오차제곱합: $SS_E = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2 = SS_T - SS_A$ 여기서, $N = \sum_{i=1}^{r} n_i$ and $T = \sum_{i=1}^{r}\sum_{j=1}^{n_i} y_{ij}$

- 분산분석을 수행하기 위해 반응치의 전체 변동을 요인별 변동으로 분해할 필요가 있으며, 총편차는 다음과 같이 수준 간 편차와 수준 내 편차(=잔차)로 분해할 수 있다.

$$y_{ij} - \bar{y} = y_{ij} - \bar{y}_{i\cdot} + \bar{y}_{i\cdot} - \bar{y} = (\bar{y}_{i\cdot} - \bar{y}) + (y_{ij} - \bar{y}_{i\cdot})$$

- 양변을 제곱하고 더하면 위의 box 안의 식처럼 정리할 수 있다.

분산분석(ANOVA)

자유도

	수준(level)				
	A_1	A_2	$\cdots$	A_r	
반응치	y_{11}	y_{21}	$\cdots$	y_{r1}	
	y_{12}	y_{22}	$\cdots$	y_{r2}	
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	
	y_{1n_1}	y_{2n_2}	$\cdots$	y_{rn_r}	
평균	$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$		$\bar{y}_{r\cdot}$	$\bar{y}$

- 각 제곱합의 자유도를 구하고자 한다. 자유도는 제곱합에 포함된 독립적인 항의 개수이다.
- 총제곱합에는 N개의 편차가 있으나 그 편차들을 모두 더하면 0이 되는 제약조건이 있으므로 자유도는 $\phi_T = N - 1$이 된다.
- 처리제곱합에는 r개의 편차가 있으나 모두 더하면 $\sum_{i=1}^{r} n_i(\bar{y}_{i\cdot} - \bar{y}) = 0$ 이 되므로 자유도는 $\phi_A = r - 1$이 된다.
- 오차제곱합에는 N개의 편차가 있지만, 각 수준마다 $\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot}) = 0$인 제약이 있으므로 총 r개의 제약으로 인해 자유도는 $\phi_E = N - r$이 된다.

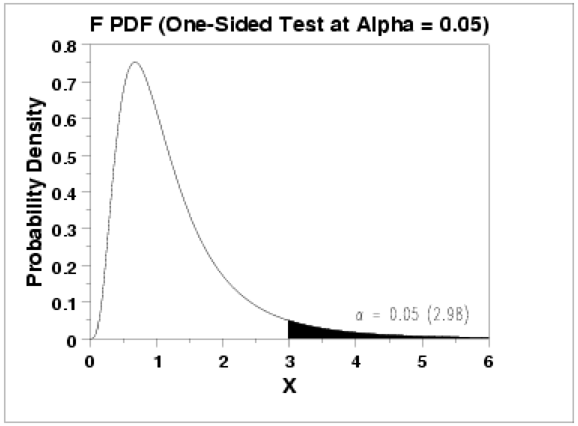
$$\phi_T = N - 1 = (r - 1) + (N - r) = \phi_A + \phi_E$$

분산분석(ANOVA)

가설검정

구분	제곱합	자유도	평균제곱	검정통계량	기각역
처리	SS_A=처리제곱합	$r-1$	$MS_A=\frac{SS_A}{r-1}$=처리평균제곱	$\frac{MS_A}{MS_E}$	$F_{1-\alpha;(r-1,N-r)}$
오차	SS_E=오차제곱합	$N-r$	$MS_E=\frac{SS_E}{N-r}$=오차평균제곱		
계	SS_T=총제곱합	$N-1$			

- $E(MS_E)=\sigma^2$인 성질을 지닌다.
- 요인의 수준변화가 반응치에 영향을 주지 못한다면(귀무가설이 참) $\mu_1=\mu_2=\cdots=\mu_r$ 이므로 수준별 반응치의 평균 $\bar{y}_i.$들은 비슷한 값들이 나올 가능성이 높다. 따라서, 처리제곱합(SS_A)은 상대적으로 작게 나올 것이다.
- 반면에 요인의 수준변화가 주는 영향이 크다면 처리제곱합이 상대적으로 크게 나올 것이며, 검정통계량의 분자가 커지게 된다.



분산분석(ANOVA)

예제

- 다음은 서로 다른 항우울제의 치료 효과를 비교하기 위한 연구이다. 치료제와 방법의 차이를 확인하기 위해서 무작위로 A1에서 A4로 군을 나눠 서로 다른 치료제와 방법의 조합으로 우울증 점수를 측정하였다. 이 네 가지 치료 효과를 유의수준 5% 하에서 비교하라.

	A1	A2	A3	A4	
우울증 점수	79	81	86	76	
	83	89	91	81	
	88	91	93	82	
	78	84	90	79	
	75	86	89		
		82			
합계	**403**	**513**	**449**	**318**	**1683**
평균	**80.6**	**85.5**	**89.8**	**79.5**	**84.15**

분산분석(ANOVA)

예제

- $SS_T = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} y_{ij}^2 \quad - \frac{T^2}{N} = 142171 - \frac{1683^2}{20} = 546.55$

- $SS_A = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{y}_{i\cdot} - \bar{y})^2 = \sum_{i=1}^{r}\frac{T_{i\cdot}^2}{n_i} - \frac{T^2}{N} = \left(\frac{403^2}{5} + \frac{513^2}{6} + \frac{449^2}{5} + \frac{318^2}{4}\right) - \frac{1683^2}{20} = 320.05$

- $SS_E = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2 = SS_T - SS_A = 546.55 - 320.05 = 226.5$

구분	제곱합	자유도	평균제곱	검정통계량(F_0)	기각역($F_{0.95;(3,16)}$)
처리	320.05	$r-1=3$	$MS_A = \frac{SS_A}{r-1}$=106.6833	$\frac{MS_A}{MS_E} = 7.536$	3.239
오차	225.50	$N-r=16$	$MS_E = \frac{SS_E}{N-r}$=14.1563		
계	546.55	$N-1=19$			

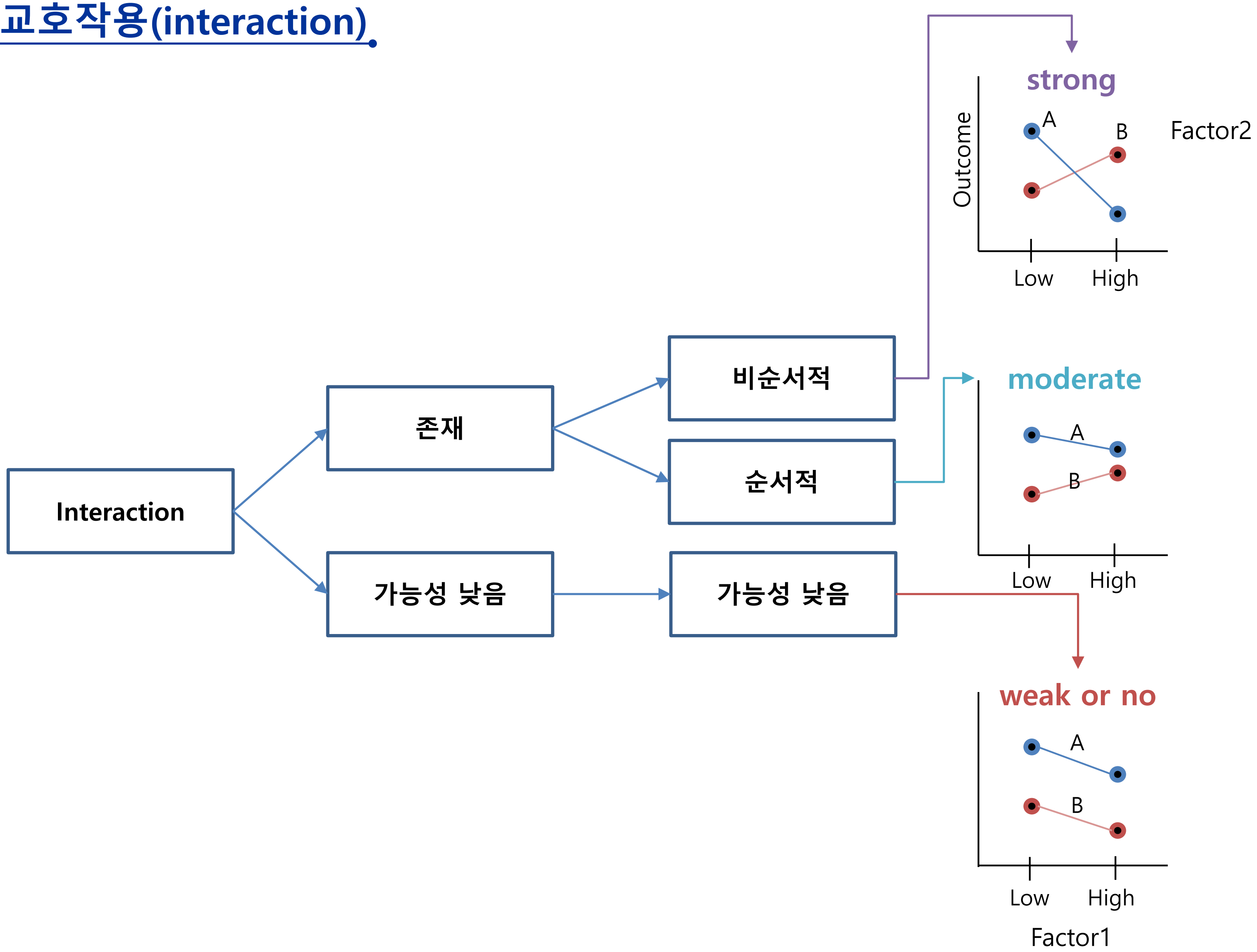
- 검정통계량 $F_0 = 7.536$이므로 기각치 $F_{0.95;(3,16)} = 3.239$보다 크므로 유의수준 5%에서 귀무가설을 기각한다. 즉, 이 4수준의 우울증 점수 변화는 우울증 치료에 영향을 미친다는 증거가 있다.

- *R 실습!*

2. 이원 분산분석(Two-way ANOVA)

이원 분산분석(Two-way ANOVA)

교호작용(interaction)



이원 분산분석(Two-way ANOVA)

자료 구조

T_{ijk}: $factor\ A$의 i번째 $level$,

$factor\ B$의 j번째 $level$,

$i\&j$에 해당하는 k번째 대상자

< Data Set >

ID	Age	Sex	Stage	Cr
1021	45	1	1	0.7
1078	58	1	2	1.8
1110	72	2	3	3.1
1124	37	2	1	0.5
1145	32	1	1	0.5
1187	66	2	3	2.0
1211	61	2	2	1.7
⋮	⋮	⋮	⋮	⋮

< Two-way ANOVA >

	Stage			합계	평균
	1	2	3		
Sex =1	0.7	1.8	$T_{31\cdot}$	$T_{\cdot 1\cdot}$	$\bar{y}_{\cdot 1\cdot}$
	0.5	⋮	$\bar{y}_{31\cdot}$		
Sex =2	0.5	1.7	3.1	$T_{\cdot 2\cdot}$	$\bar{y}_{\cdot 2\cdot}$
	⋮	⋮	2.0		
합계	$T_{1\cdot\cdot}$	$T_{2\cdot\cdot}$	$T_{3\cdot\cdot}$	T	
평균	$\bar{y}_{1\cdot\cdot}$	$\bar{y}_{2\cdot\cdot}$	$\bar{y}_{3\cdot\cdot}$		$\bar{y}$

이원 분산분석(Two-way ANOVA)

이원 분산분석(Two-way ANOVA)

- 전체 수준에서의 모평균을 μ, 요인 A의 주효과를 a_i, 요인 B의 주효과를 b_j, 요인 A와 B의 교호작용을 $(ab)_{ij}$이라 하면 아래의 식과 같이 표현할 수 있다.

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk}, \qquad i = 1,2,\cdots,r;\ \ j = 1,2,\cdots,s;\ \ k = 1,2,\cdots,n$$

이원 분산분석(Two-way ANOVA)

이원 분산분석(Two-way ANOVA)

이원 분산분석 제곱합의 분해

- $SS_T = SS_{AB} + SS_E = SS_A + SS_B + SS_{A\times B} + SS_E$

  - 총제곱합: $SS_T = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk} - \bar{y})^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n} y_{ijk}^2 - \frac{T^2}{N}$
  - 요인A 제곱합: $SS_A = sn\sum_{i=1}^{r}(\bar{y}_{i..} - \bar{y})^2 = \sum_{i=1}^{r}\frac{T_{i..}^2}{sn} - \frac{T^2}{N}$
  - 요인B 제곱합: $SS_B = rn\sum_{j=1}^{s}(\bar{y}_{.j.} - \bar{y})^2 = \sum_{j=1}^{s}\frac{T_{.j.}^2}{rn} - \frac{T^2}{N}$
  - 교호작용제곱합: $SS_{A\times B} = SS_{AB} - SS_A - SS_B$
  - 오차제곱합: $SS_E = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2 = SS_T - SS_{AB}$
  - AB총제곱합: $SS_{AB} \equiv n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij.} - \bar{y})^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{T_{ij.}^2}{n} - \frac{T^2}{N}$

  여기서, $N = r\times s\times n$ and $T = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n} y_{ijk}$

이원 분산분석(Two-way ANOVA)

가설검정

가설검정

① 귀무가설: **요인 A**의 수준변화가 반응치에 미치는 영향이 없다.
대립가설: **요인 A**의 수준변화가 반응치에 미치는 영향이 있다.
② 귀무가설: **요인 B**의 수준변화가 반응치에 미치는 영향이 없다.
대립가설: **요인 B**의 수준변화가 반응치에 미치는 영향이 있다.
③ 귀무가설: **요인 A와B**의 교호작용이 없다.
대립가설: **요인 A와B**의 교호작용이 있다.

① $\boldsymbol{H_0: a_1 = a_2 = \cdots = a_r = 0}$

② $\boldsymbol{H_0: b_1 = b_2 = \cdots = b_s = 0}$

③ $\boldsymbol{H_0: (ab)_{11} = (ab)_{12} = \cdots = (ab)_{rs} = 0}$

$\boldsymbol{\text{vs.} \quad H_1: Not\ H_0}$

	요인A			합계	평균
	1	2	3		
요인B =1	0.7	1.8	$T_{31\cdot}$	$T_{\cdot1\cdot}$	$\bar{y}_{\cdot1\cdot}$
	0.5	⋮	$\bar{y}_{31\cdot}$		
요인B =2	0.5	1.7	3.1	$T_{\cdot2\cdot}$	$\bar{y}_{\cdot2\cdot}$
	⋮	⋮	2.0		
합계	$T_{1\cdot\cdot}$	$T_{2\cdot\cdot}$	$T_{3\cdot\cdot}$	T	
평균	$\bar{y}_{1\cdot\cdot}$	$\bar{y}_{2\cdot\cdot}$	$\bar{y}_{3\cdot\cdot}$		$\bar{y}$

※ $y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk},\ i = 1,2,\cdots,r;\ \ j = 1,2,\cdots,s;\ \ k = 1,2,\cdots,n$

이원 분산분석(Two-way ANOVA)

가설검정

구분	제곱합	자유도	평균제곱	검정통계량	기각역
A	SS_A	$r-1$	MS_A	$\frac{MS_A}{MS_E}$	$F_{1-\alpha;(r-1,rs(n-1))}$
B	SS_B	$s-1$	MS_B	$\frac{MS_B}{MS_E}$	$F_{1-\alpha;(s-1,rs(n-1))}$
$A\times B$	$SS_{A\times B}$	$(r-1)(s-1)$	$MS_{A\times B}$	$\frac{MS_{A\times B}}{MS_E}$	$F_{1-\alpha;((r-1)(s-1),rs(n-1))}$
E	SS_E	$rs(n-1)$	MS_E		
T	SS_T	$rsn-1$			

이원 분산분석(Two-way ANOVA)

예제

- 한 의료기기 개발 공정에서 온도 100°C, 150°C, 200°C, 250°C 4수준과 압력 1기압, 2기압, 3기압에서 랜덤한 순서로 2회 반복 실험하여 얻은 수율 데이터가 다음의 표와 같다. 이 공정에서 온도와 압력의 변화가 수율에 영향을 미치는지 유의수준 5%에서 검정하시오.

	100°C	150°C	200°C	250°C
1기압	76	79	87	79
	79	81	91	82
2기압	81	84	91	85
	79	86	94	84
3기압	83	89	88	77
	85	88	86	76

이원 분산분석(Two-way ANOVA)

예제

- 한 의료기기 개발 공정에서 온도 100°C, 150°C, 200°C, 250°C 4수준과 압력 1기압, 2기압, 3기압에서 랜덤한 순서로 2회 반복 실험하여 얻은 수율 데이터가 다음의 표와 같다. 이 공정에서 온도와 압력의 변화가 수율에 영향을 미치는지 유의수준 5%에서 검정하시오.

	100°C	150°C	200°C	250°C	합계	평균
1기압	155	160	178	161	654	81.75
2기압	160	170	185	169	684	85.5
3기압	168	177	174	153	672	84.0
합계	483	507	537	483	2010	
평균	80.5	84.5	89.5	80.5		83.75

이원 분산분석(Two-way ANOVA)

예제

구분	제곱합	자유도	평균제곱	F_0	기각역
A	328.5	3	109.5	39.818	3.490
B	57.0	2	28.5	10.364	3.885
$A \times B$	154.0	6	25.667	9.333	2.996
E	33.0	12	2.75		
T	572.5	23			

- 검정통계량 $F_0 = 39.818$이므로 기각치 $F_{0.95;(3,12)} = 3.490$보다 크므로 유의수준 5%에서 귀무가설을 기각한다.

  즉, 이 공정에서 4수준의 온도 변화는 수율에 영향을 미친다는 증거가 있다.

- 검정통계량 $F_0 = 10.364$이므로 기각치 $F_{0.95;(2,12)} = 3.885$보다 크므로 유의수준 5%에서 귀무가설을 기각한다.

  즉, 이 공정에서 3수준의 압력 변화는 수율에 영향을 미친다는 증거가 있다.

- 검정통계량 $F_0 = 9.333$이므로 기각치 $F_{0.95;(6,12)} = 2.996$보다 크므로 유의수준 5%에서 귀무가설을 기각한다.

  즉, 이 공정에서 온도와 압력의 교호작용은 수율에 영향을 미친다는 증거가 있다.

분산분석 구분

종류

	단변량 변수					다변량 분산분석 (MAN(C)OVA)
	일원배치 분산분석 (One-way ANOVA)	이원배치 분산분석 (Two-way ANOVA)	다원 분산분석 (Multi-way ANOVA)	공분산분석 (ANCOVA)	반복측정 분산분석 (Repeated measures (RM) ANOVA)	
독립변수	**1개**	**2개**	**3개 이상**	**1 + α 개**	**1개 이상**	**1개 이상 (1 + α 개)**
종속변수	**1개**				**1개 (단, 각 대상자마다 반복측정)**	**2개 이상**

< 보건과학통계 SPSS 이야기 >

3. 사후분석(post-hoc analysis)

사후분석(post-hoc analysis)

사후분석

- 두 집단 간 모수 검정(t-test 등)은 중심 값의 차이를 유의확률과 평균을 통해 유의성을 확인하였다.
- 하지만 분산분석은 평균의 직접 비교이기 보다는 분산의 차이를 통해 통계량을 산출하므로 그룹 간 차이를 확인하기 위해선 사후분석(post-hoc analysis)가 필요하다.
- 따라서 세 그룹 이상의 모평균을 비교하는 분산분석의 귀무가설(H_0: $\mu_1 = \mu_2 = \cdots = \mu_r$)은 특정 그룹에서만 유의한 차이가 발생하면 기각시킬 수 있으나, “H_1: 적어도 한 개의 μ_i는 나머지와 같지 않다.”로 대립가설을 정의하기 때문에 사후분석 없이는 어떤 그룹이 차이를 발생하는지 확인하기 어렵기에 추가분석을 필요로 한다.

사후분석(post-hoc analysis)

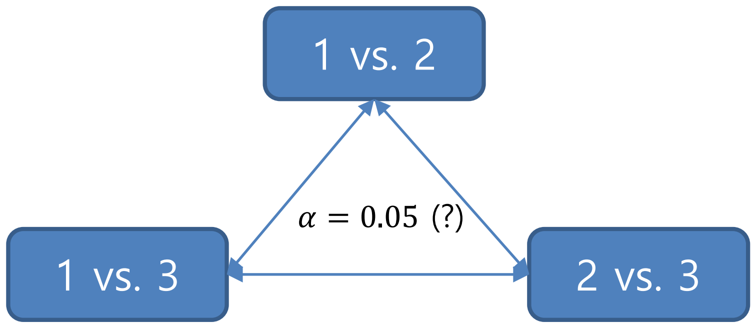
사후분석

- 예를 들어 한 번의 실험으로 얻은 결과에서 3가지 가설을 세우고 각각 5% 유의수준으로 동시에 검정을 시도한다면 5% 유의수준이 3번 사용되므로, 각각이 독립인 경우 $1-(1-0.05)^3 \approx 0.143$의 유의수준을 적용한 것과 같다.
- 20개 또는 100개 이상의 다중 비교를 진행할 시에는 위의 유의수준은 기존에 설정하고자 했던 0.05보다 훨씬 큰 값을 가지게 된다.
- 유전체 자료 분석 등 단일 검정을 여러 번 수행해야 하는 연구에 있어서는 다중 비교 문제를 반드시 해결해야만 한다.

사후분석(post-hoc analysis)

사후분석

- 분산분석의 목표는 모든 그룹에서의 모평균이 같은지를 확인하는 것이다.



개별 검정 시 유의수준 α를 엄격하게 검정(낮춤)함으로 전체 오류 %를 유의수준 $\alpha = 0.05$로 맞춘다.

(예시) Bonferroni correction (3그룹 간 비교) – n개 그룹 간 비교 시 ${}_nC_2$를 계산하여 유의확률에 곱해주면 된다.

1) 유의수준 $\alpha = 0.05/3$ 으로 정의

or

2) 유의확률 p-value에 $\times 3$을 한 후 유의수준 $\alpha = 0.05$ 기준으로 검정

사후분석(post-hoc analysis)

사후분석

- 분산분석에서 등분산이 가정되어 있을 때는

1. Tukey types

> SPSS 사용 시 주로 사용하게 되며, 각 집단의 표본 수가 같을 경우에만 의미가 있는 제약점이 있음

2. Duncan

> 등분산성과 반복수가 동일하다는 제약조건이 있음, 다른 검정에 비해 덜 보수적이다.

3. Scheffe

> 표본 수가 다를 경우에도 사용가능하다.

- 등분산 여부와 관계 없이는

4. Bonferroni correction

> 보수적(p-value를 강하게 제약함)인 방법으로, 표본 수가 다르거나 pairwise 개수가 적을 시 주로 사용한다.

5. False discovery rate (FDR)

> 다중 비교 개수가 많을 시에는 Bonferroni correction 기법이 강한 제약을 완화하고자 함.

Thank you!