분산분석(ANOVA)

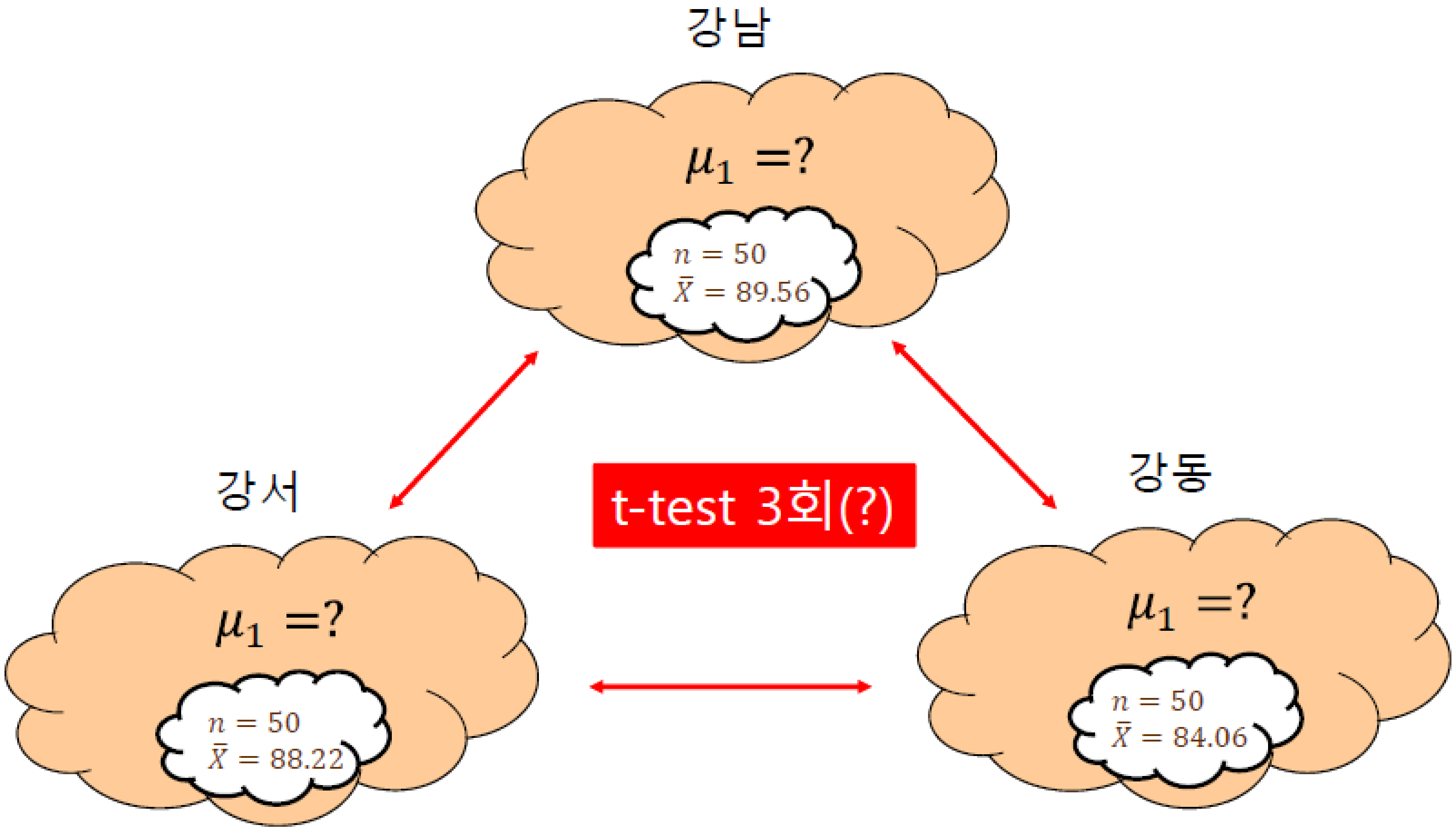
보건빅데이터통계분석

이새봄
삼육대학교 SW융합교육원

분산분석이란?

ANOVA란?

- 모집단에서 표본추출



ANOVA란?

- 다중검정
- t-test를 여러 번 하기 위한 귀무가설 (3개의 귀무가설)

$$H_{0\_1}: \mu_1 = \mu_2, H_{0\_2}: \mu_1 = \mu_3, H_{0\_3}: \mu_2 = \mu_3$$

- 여러 번 t-test 를 해주게 되면 1종 오류(α)를 범할 확률이 증가
- Bonferroni inequality

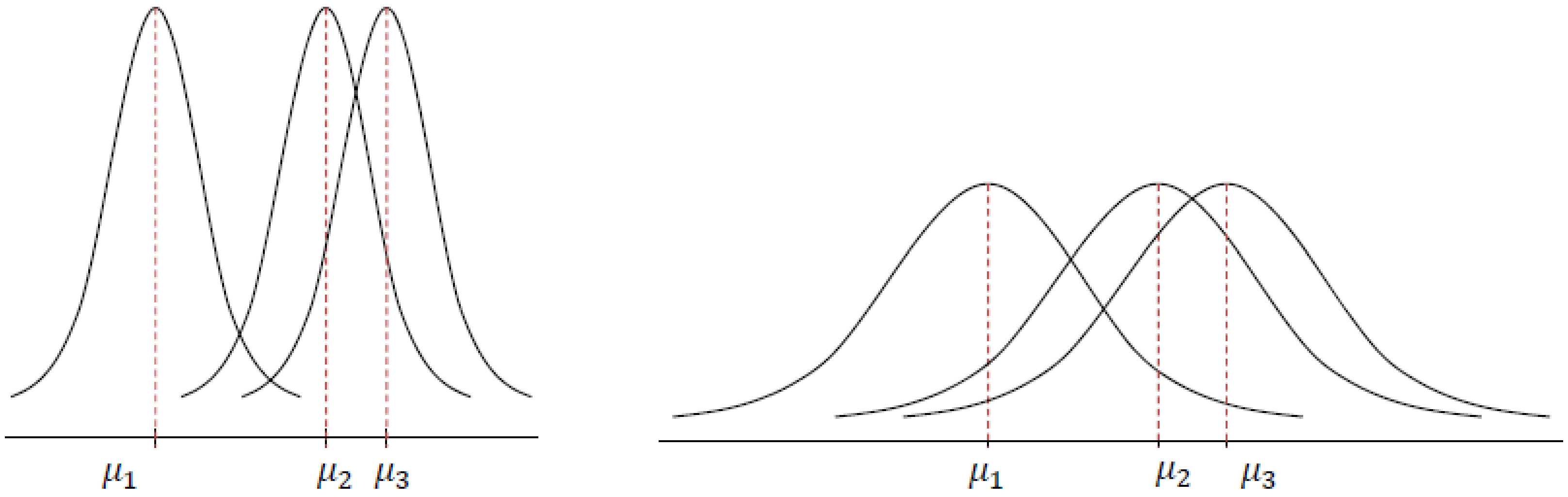
$$\mathrm{P}(H_{1\_1} \cap H_{1\_2} \cap H_{2\_3}) = \max(n\alpha) = \max(3\alpha) = \max(15\%)$$

- 최대 15%까지 가능
- 집단이 3개 이상일 때는 모든 집단의 평균이 같은지를(1=2=3) 분산을 이용하여 분석

* 평균의 차이를 분석하는 것인데 왜 분산을 이용하지요?

ANOVA란?

- ANOVA(Analysis of Variance)
  - 그룹내 분산의 비율을 통해 평균의 차이를 비교할 수 있음
  - 분산분석을 실시한 후에 집단 간에 평균이 차이가 있다고 결과가 나타나면, 각 집단간 평균을 비교
  - Post-hoc 분석: Bonferroni, Duncan, Tukey 등 오류를 보정한 방법 이용



ANOVA란?

- 분산분석(ANOVA)
  - 영국의 통계학자인 피셔에 의해 개발
  - 세 그룹 이상의 분산을 검정하여, 각 집단의 모평균 차이를 비교 검정
  - 변동(분산)에 영향을 주는 변수를 몇 개의 레벨로 나누어서 대상에게 실험
  - 변동(분산)의 분해: 실험요인들에 의한 변동과 오차에 의한 변동으로 구분

- 실험계획법에서 가장 많이 사용되는 분석방법
  - 사례) 통닭의 맛을 결정하는 온도는?
  - 반응값(특성치) : 통닭 맛
  - 요인 : 튀기는 온도
  - 요인의 수준 수 : 4
    - 튀기는 온도가 120℃, 140℃, 160℃, 180℃인 경우
  - 수준별 반복수 : 30
    - 각 수준의 요인에서 30명이 점수를 매긴 경우 : 30

실험설계법

- 일원분산분석
  - 완전확률화설계(completely randomized design)

	요인A		
	강남	강서	강동
반응값	93 91 …	90 89 …	95 91 …

- 이원분산분석
  - 반복 없음
    - 확률화블록설계(Randomized block design)
    - 라틴방격설계(Latin square design)

		요인B		
		강남	강서	강동
요인A	A1	93	91	87
	A2	91	95	86
	A3	…	…	…

- 반복 있음 (상호작용 효과검정)
  - 요인실험(Factorial experiment)

		요인B		
		강남	강서	강동
요인A	A1	93 91 …	90 89 …	95 91 …
	A2	87 83 …	90 91 …	88 89 …
	A3	…	…	…

- 범주형+연속형
  - ANCOVA

출처: SPSS를 이용한 실험설계와 분산분석, 한상태 외, spss 아카데미 참조

실험설계법

- 요인 수준 선택에 따른 ANOVA 종류
  - 고정효과모형(fixed effect model) : 요인을 실험자가 선택
  - 변량효과모형(random effect model) : 요인을 무작위로 선택
  - 혼합모형(mixed effect model): 두 개의 요인 중에서 하나는 실험자가 하나는 무작위로 선택

실험설계법

- 예) G레스토랑은 서울에 10개의 매장을 가지고 있다.

  - 강남, 강서, 강동의 3개의 매장에 대한 만족도 차이를 보고 싶다.
  - 고정효과모형: 3개 매장의 평균에 관심

$$H_0: \mu_1 = \mu_2 = \mu_3$$

  - 10개 매장 중에서 3개의 매장을 임의로 선택해서 K레스토랑 전체 만족도의 차이를 보고 싶다.
  - 변량효과모형: 분산의 동질성에 대해 관심 (전체 평균에 관심)

$$H_0: \sigma_\mu^2 = 0$$

  - 매장 간에는 메뉴가 서로 다양하다. 강남, 강서, 강동 3개의 매장을 설정하고 메뉴는 임의로 3개씩 선택하였다.
  - 혼합모형: 매장과 메뉴의 만족도

일원 분산분석

One-Way ANOVA

- 문제의 정의
    - S커피회사는 강남(1), 강동(2), 강서(3)에 매장을 보유하고 있다. 매장별로 고객만족도가 차이가 있는지를 조사하기 위해 매장별 고객만족도를 조사하였다.
    - 과연 3곳 매장의 고객만족도는 차이가 있는지?
    - 있다면 어느 레스토랑의 서비스 만족도가 가장 안 좋은가 확인해보자
    - 07_1.OWA.xlsx

- 가설
    - 귀무가설($H0$): 3곳 매장의 고객만족도는 차이가 없다.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

    - 연구가설($H1$): 3곳 매장의 고객만족도 중 적어도 한 쌍은 차이가 있다.

$$H_1: not\ H_0, \mu_1 \neq \mu_2\ or\ \mu_1 \neq \mu_3\ or\ \mu_2 \neq \mu_3$$

One-Way ANOVA

- 가설
  - 귀무가설($H0$): 3곳 매장의 고객만족도는 차이가 없다.

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

  - 연구가설($H1$): 3곳 매장의 고객만족도 중 적어도 한 쌍은 차이가 있다.

$$H_1 : not\ H_0, \mu_1 \neq \mu_2\ or\ \mu_1 \neq \mu_3\ or\ \mu_2 \neq \mu_3$$

  - 다중비교 (Post-hoc)

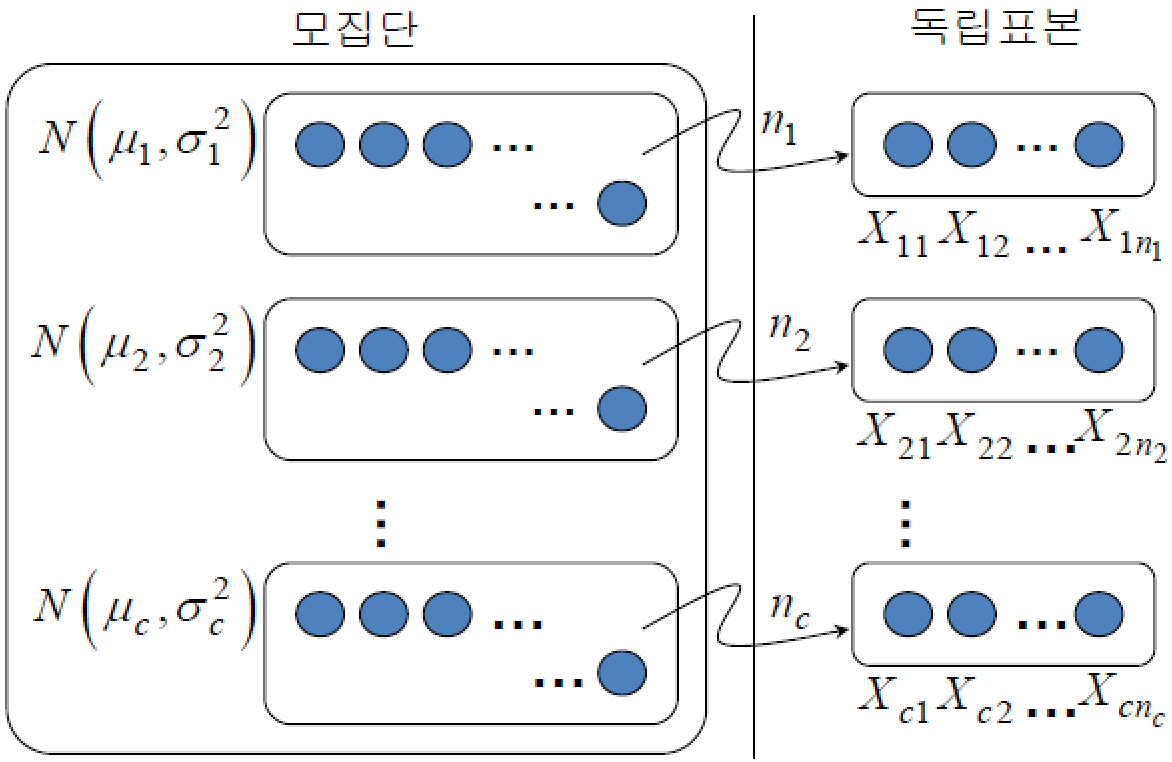
$$H_{10} : \mu_1 = \mu_2 \qquad H_{11} : \mu_1 \neq \mu_2$$

$$H_{20} : \mu_1 = \mu_3 \qquad H_{21} : \mu_1 \neq \mu_3$$

$$H_{30} : \mu_2 = \mu_3 \qquad H_{31} : \mu_2 \neq \mu_3$$

One-Way ANOVA

- **표본추출**



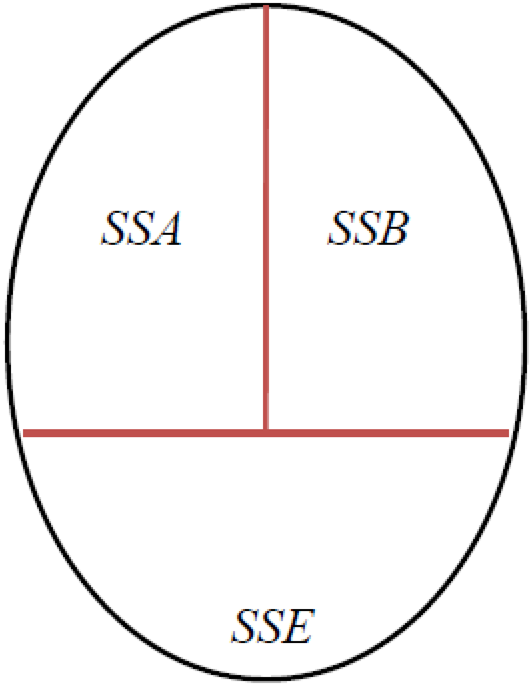
One-Way ANOVA

- 변동의 분해

일원배치 분산분석

$$\boxed{SST} = \boxed{SSA} + \boxed{SSE}$$

이원배치 분산분석



$$SST = SSA + SSB + SSE$$

One-Way ANOVA

- 변동(총제곱합)의 분해
  - 총편차의 분해

$$y_{ij} - \bar{y} = (\bar{y}_j - \bar{y}) + (y_{ij} - \bar{y}_j)$$

총편차= 처리효과 + 측정오차
(실험효과, 요인효과)

  - 총제곱합의 분해

$$\sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_j (\bar{y}_j - \bar{y})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_j)^2$$

$$SST = SSA + SSE$$

총제곱합 = 처리제곱합 + 측정오차제곱합

Total = between subject + within subject

One-Way ANOVA

- 변동(총제곱합)의 분해

	인자의 수준				
	A_1	A_2	…	A_j	
반응값 (반복실험)	y_{11} y_{21} y_{31} … …	y_{12} y_{22} y_{32} … …	… … … … …	y_{1j} y_{2j} y_{3j} … …	
N	n_1	n_2	…	n_j	
합계	T_1	T_2	…	T_j	
평균	$\bar{y}_1$	$\bar{y}_2$	…	$\bar{y}_j$	$\bar{y}$

One-Way ANOVA

- **변동(총제곱합)의 분해**

	인자의 수준			
	강남	강동	강서	
반응값 (반복실험)	93 94 95 86 87 …	85 82 88 83 93 …	86 85 84 93 90 …	
N	50	50	50	150
합계	4,478	4,411	4.203	13,092
평균	89.56	88.22	84.06	87.28

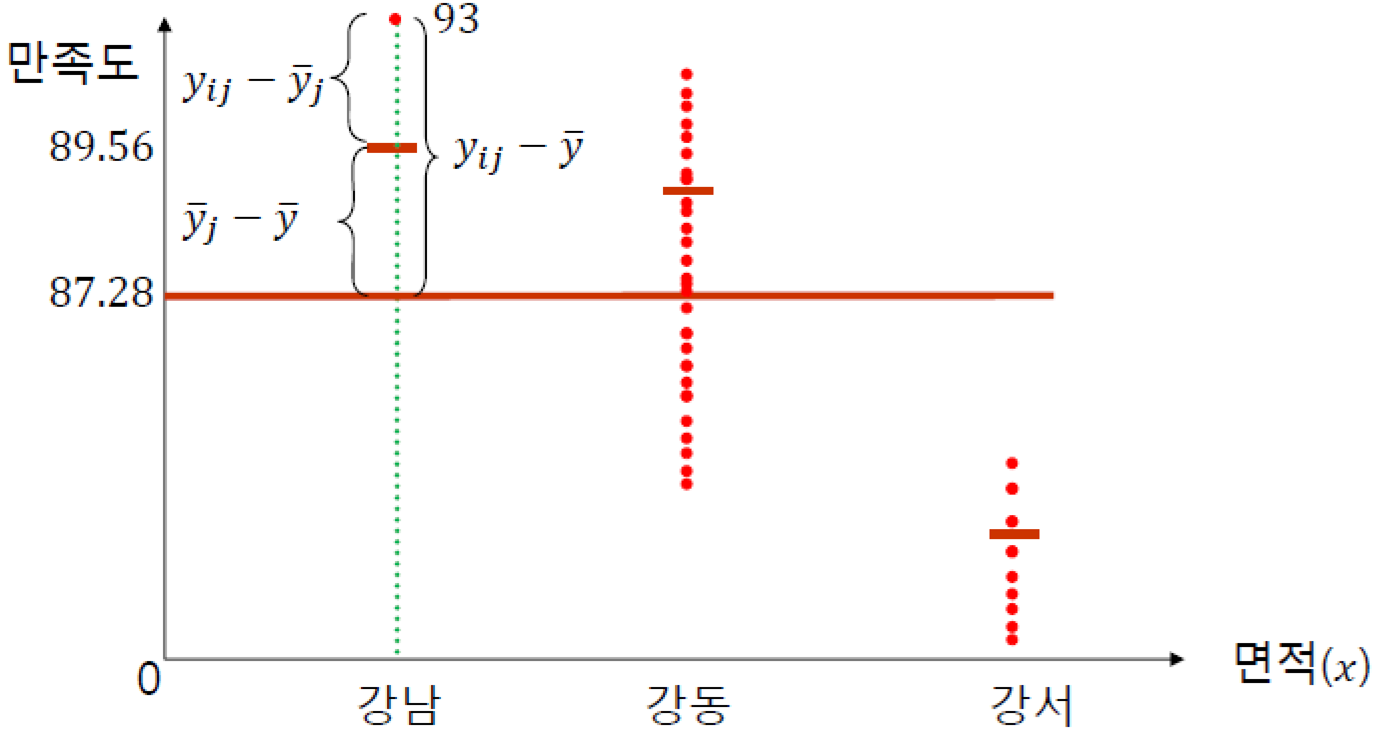
$$y_{ij} - \bar{y} = (\bar{y}_j - \bar{y}) + (y_{ij} - \bar{y}_j)$$

$$5.72(93 - 87.28) = 2.28(89.56 - 87.28) + 3.44(93 - 89.56)$$

총편차 = 처리효과 + 측정오차

One-Way ANOVA

- **변동의 분해(분산분석)**



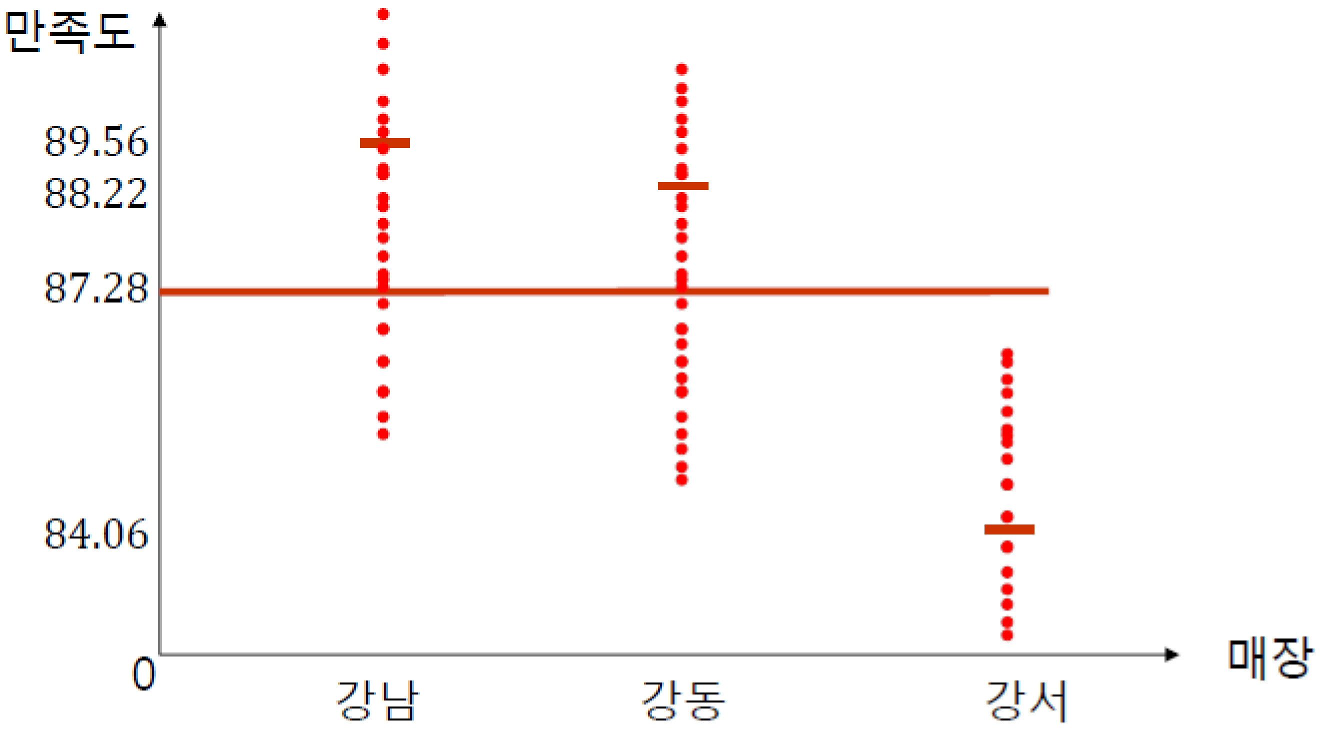
$$y_{ij} - \bar{y} = (\bar{y}_j - \bar{y}) + (y_{ij} - \bar{y}_j)$$

$$5.72(93 - 87.28) = 2.28(89.56 - 87.28) + 3.44(93 - 89.56)$$

총편차 = 처리효과 + 측정오차

One-Way ANOVA

- 변동의 분해(분산분석)



$$\sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_j (\bar{y}_j - \bar{y})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_j)^2$$

총제곱합 = 처리제곱합 + 측정오차제곱합

One-Way ANOVA

▪ 분산분석표

요인	제곱합 (SS)	자유도(df)	평균제곱 (MS)	F
처리(SSA)	$SSA = \sum_{j=1}^{k} (\bar{y}_j - \bar{y})^2$	$k-1$	$MSA = \frac{SSA}{k-1}$	$F = \frac{MSA}{MSE}$
잔차(SSE)	$SSE = \sum_{i=1}^{n}\sum_{j=1}^{k} (y_{ij} - \bar{y}_j)^2$	$n-k$	$MSE = \frac{SSE}{n-k}$	
총계(SST)	$SST = \sum_{i=1}^{n}\sum_{j=1}^{k} (y_{ij} - \bar{y})^2$	$\mathrm{n}-1$		

▪ 검정통계량

$$F = \frac{MSA}{MSE} = \frac{\frac{SSA}{k-1}}{\frac{SSE}{n-k}} \sim F_{(\alpha, k-1, n-k)}$$

$$\boxed{F\uparrow = \frac{MSA\uparrow}{MSE\downarrow} \qquad F\downarrow = \frac{MSA\downarrow}{MSE\uparrow}}$$

One-Way ANOVA

- 분산분석표

요인	제곱합 (SS)	자유도(df)	평균제곱 (MS)	F
처리(SSA)	822.52	$2(k-1=3-1)$	$411.26(=\frac{822.52}{2})$	14.804 $(=\frac{411.26}{27.78})$
잔차(SSE)	4,083.72	$147(\mathrm{n}-\mathrm{k}=150-3)$	$27.78(=\frac{2{,}083.72}{147})$	
총계(SST)	4,906.24	$149(\mathrm{n}-1=150-1)$		

- 검정통계량

$$F_{cal}=\frac{411.26}{27.78}=14.804>F_{(0.05,\,2,\,149)}=3.06$$

- P-value 계산

$$p-value=P(F>14.804)=0.000<\alpha=0.05$$

F분포는 분산(σ^2)을 검증하기 때문에 큰지만 검정: $P(F>14.804)$

One-Way ANOVA

▪ 다중비교

종 류	설 명
Fisher's LSD (least significant difference)	• t-검정을 사용하여 집단 평균 간 모든 대응별 비교
Bonferroni's MSD	
Seidak	
Scheffe	• F-검정 사용
R-E-G-W F	• F-검정 사용 Ryan-Einot-Gabriel-Welsch의 F
R-E-R-W Q	• Ryan-Einot-Gabriel-Welsch의 Q
S-N-K	• 스튜던트화 범위 분포를 사용한 평균간 대응비교
Tukey'HSD	• 평균(스튜던트화) 범위 통계량
Duncan	

One-Way ANOVA

- 다중비교

종 류	설 명
Fisher's LSD	$t_{\frac{\alpha}{2}(n-k)}$
Bonferroni's MSD	$t_{(\frac{\alpha}{2k},n-k)}$ $\frac{q_{(\alpha,k,n-k)}}{\sqrt{2}}$
Scheffe	$\sqrt{(k-1)F_{(\alpha,k-1,n-k)}}$
Tukey'HSD	$\frac{q_{(\alpha,k,n-k)}}{\sqrt{2}}$, $q_{(k,n-k)}$ = 표분화범위분포

One-Way ANOVA

- Fisher's LSD(least significant difference)
  - $\boxed{s_p^2\text{(공통분산)를 사용하지 않고}}$, 분산분석후에 사용된 MSE를 이용하여 분석

$$T_{cal} = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{\boxed{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > t_{(\frac{\alpha}{2}, n-k)} \qquad t_{cal} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\boxed{s_p^2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$T_{강남-강서} = \frac{|90.61 - 89.59|}{\sqrt{22.573\left(\frac{1}{38} + \frac{1}{30}\right)}} = \frac{1.02}{1.16} = 0.880 > t_{(0.025,135)} = 1.978$$

- Bonferroni
  - LSD의 경우 가설 검정의 수에 따라 유의수준이 증가 → Bonferroni로 보정

$$t_{(\boxed{\frac{\alpha}{2}}, n-k)} \rightarrow t_{(\boxed{\frac{\alpha}{2k}}, n-k)}$$

$$t_{(\frac{\alpha}{2}, n-k)} = 1.978 \rightarrow t_{(\frac{\alpha}{2k}, 135)} = t_{(\frac{0.05}{\boxed{2*6=12}}, 135)} = 2.915$$

One-Way ANOVA

- Tukey's test(Q)

$$T_{cal} = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > \boxed{T_{(k,n-k)}}$$

$$T_{cal} = \frac{|90.61 - 89.59|}{\sqrt{22.573\left(\frac{1}{38} + \frac{1}{30}\right)}} = \frac{1.02}{1.16} = 0.880 > \boxed{T_{(4,135)} = 3.63}$$

$$* \; t_{(0.025,135)} = 1.978$$

Critical Values for the Tukey Q Test

d.f. = N – K (ANOVA Error or Within d.f.) for ANOVA: Single Factor
d.f. = N – R*C (ANOVA Within d.f.) for ANOVA: Two-Factor With Replication
d.f. = (R – 1)*(C – 1) (ANOVA Error d.f.) for ANOVA: Two-Factor Without Replication

Number of Groups (Treatments) = K

Error df	Number of Groups (Treatments)								
	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

One-Way ANOVA

- ANOVA 모형의 가정(잔차검정)
  - ANOVA는 분산을 이용해 검정하는 기법으로 잔차의 가정이 중요함
  - 잔차의 등분산성: 잔차그림, 등분산검정, Levene test
  - 잔차의 정규성: Shapiro-Wilk, jarque-Bera
  - 잔차의 독립성: 무작위로 표본 추출(ANOVA에서는 특별히 강조하지 않음)
  - 변동의 분해

$$y_{ij} - \bar{y} = (\bar{y}_j - \bar{y}) + (y_{ij} - \bar{y}_j)$$

총편차 = 처리효과 + 측정오차

  - ANOVA 모형식

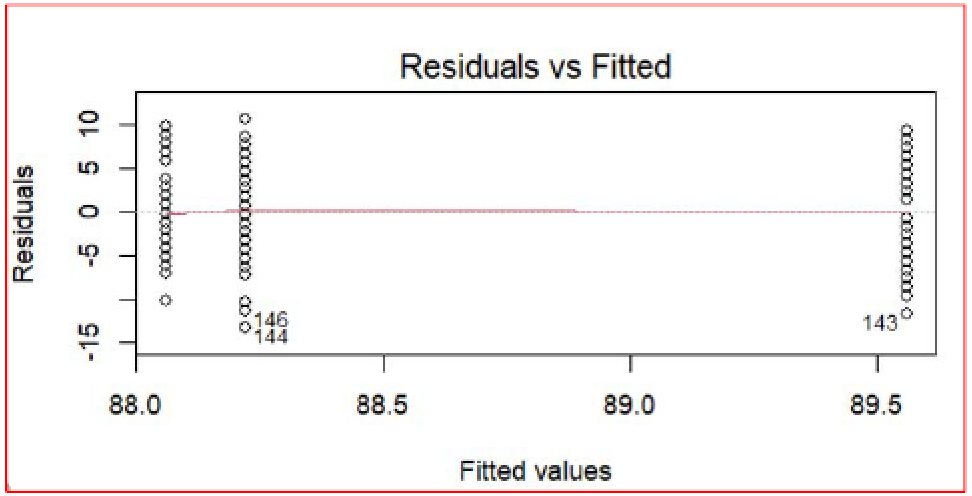
$$Y_{ij} - \mu = (\mu_i - \mu) + \varepsilon_{ij}$$

$$Y_{ij} = \mu + (\mu_i - \mu) + \varepsilon_{ij}$$

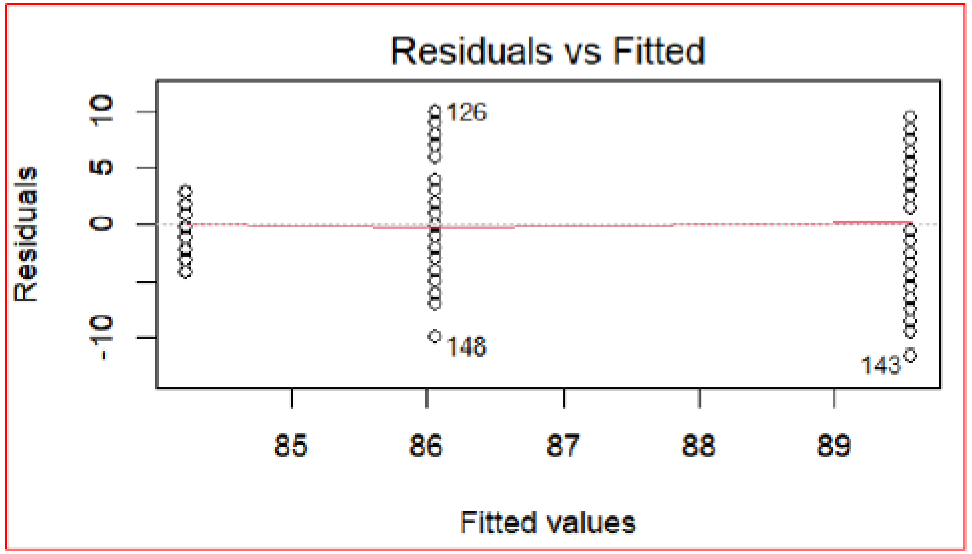
전체평균 + 처리효과 + 측정오차

  - 오차 → 잔차의 가정 $\boxed{\varepsilon_{ij} \to e_{ij} = (y_{ij} - \bar{y}_j) \sim N(0, \sigma^2)}$

One-Way ANOVA

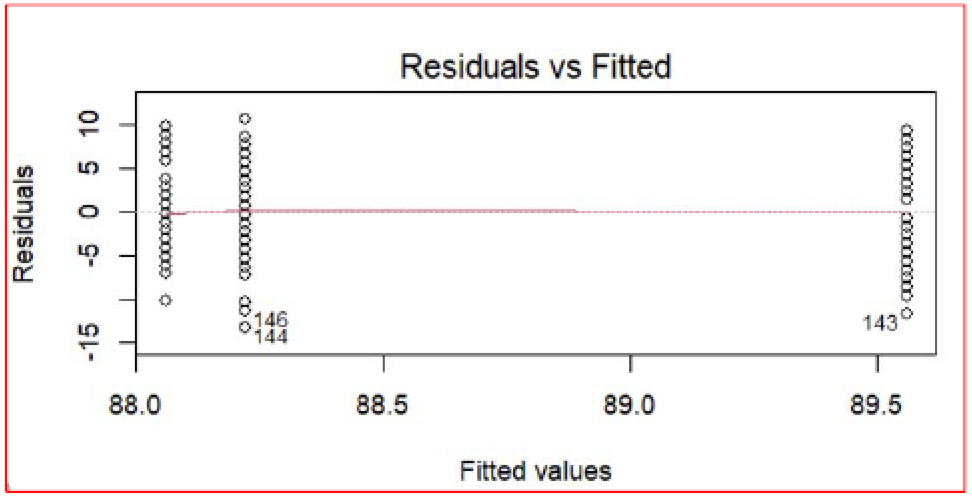


잔차의 등분산성(O)
잔차의 정규성(O)

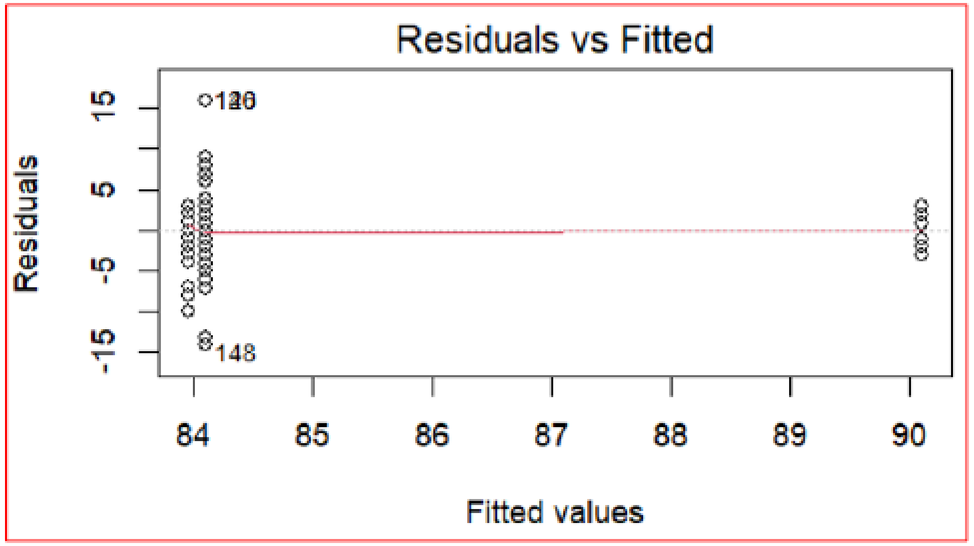


잔차의 등분산성(X)
잔차의 정규성(O)

One-Way ANOVA



잔차의 등분산성(O)
잔차의 정규성(O)



잔차의 등분산성(X)
잔차의 정규성(X)

One-Way ANOVA 분석절차

반복측정 분산분석

Repeated Measures ANOVA

- 문제의 정의
  - S대학에서 운영하는 6개월 어학 프로그램이 있다.
  - 어학 프로그램의 효과를 측정하기 위해 학습 프로그램에 참여한 학생을 대상으로 프로그램 참여전, 중간시험, 최종시험 영어실력을 테스트하였다.
  - 과련 어학프로그램은 효과가 있는지? 있다면 언제부터 효과가 나타났을지를
  - 검증해 보자
  - 08_1.RMA.xlsx

- 가설
  - 귀무가설($H0$): 어학프로그램은 시점에 따라 차이가 없다.

$$H_0: \delta_1 = \delta_2 = \delta_3$$

  - 연구가설($H1$): 어학프로그램은 시점에 따라 차이가 있다.

$$H_1: not\ H_0$$

Repeated Measures ANOVA

- **일원 반복측정 분산분석**

개체	인자의 수준			
	실시전	3주후	6주후	
1	10	16	25	
2	11	17	26	
3	9	15	21	
4	9	14	20	
5	9	20	22	
	…	…	…	
평균	11.08	19.15	29.46	

Repeated Measures ANOVA

- 변량분해

$$TSS \quad = \quad SSB \; + \; SSW \;\; (SST + SSE)$$

총제곱합 = 개체간제곱합+개체내제곱합 (처리제곱합 + 측정오차제곱합)

Total = Between subject + Within subject (Treatment + Error)

- 검정통계량

$$F = \frac{MST}{MSE} = \frac{\frac{SST}{k-1}}{\frac{SSE}{(n-1)(K-1)}} \sim F_{(\alpha,\, k-1,\, (n-1)(k-1))}$$

Repeated Measures ANOVA

- Mauchly's sphericity test
  - Mauchly’ W = 1
  - 1940년 John Mauchly 에 의해 개발

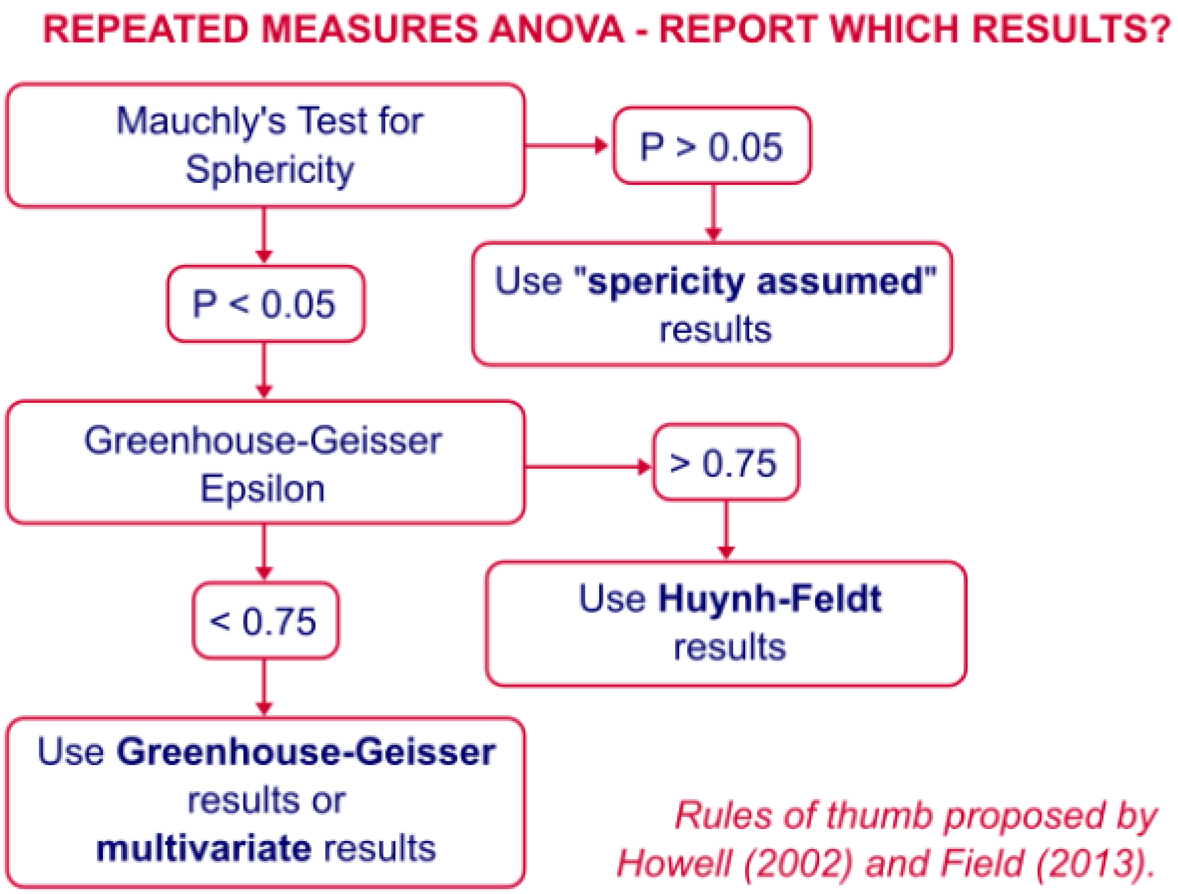
- 공분산행렬 ∑의 복합대칭성 $\sigma_{jk} = Cov(y_j, y_k) = \begin{cases} \sigma_s^2 + \sigma^2, & if\, y_j = y_k \\ \sigma_s^2, & if\, y_j \neq y_k \end{cases}$

- 구형성 검정 가설 $H_0: \sigma^2_{TXA-TXB} = \sigma^2_{TXA-TXC} = \sigma^2_{TXB-TXC}$

Patient	Tx A	Tx B	Tx C	Tx A − Tx B	Tx A − Tx C	Tx B − Tx C
1	30	27	20	3	10	7
2	35	30	28	5	7	2
3	25	30	20	−5	5	10
4	15	15	12	0	3	3
5	9	12	7	−3	2	5
			Variance:	17 >	10.3 ≈	10.3

Repeated Measures ANOVA

- 구형성 조건에 따른 해석 방법 및 절차



Repeated Measures ANOVA

- 구형성 검정

No	사전	3개월	6개월	사전-3개월	사전-6개월	3개월-6개월
1	63	63	68	0	-5	-5
2	60	60	67	0	-7	-7
3	61	61	70	0	-9	-9
4	57	57	68	0	-11	-11
5	58	58	68	0	-10	-10
6	58					10
7	53					14
8	58					8
9	58					9
43	56	59	67	-3	-11	-8
44	59	57	66	2	-7	-9
45	60	62	67	-2	-7	-5
분산				6.32 ≈	5.22 ≈	7.57

Assumptions

Tests of Sphericity

	Mauchly's W	p	Greenhouse-Geisser ε	Huynh-Feldt ε
시점	0.908	0.126	0.916	0.954

Repeated Measures ANOVA

- 분산분석표

총제곱합 = 개체간제곱합+개체내제곱합 (처리제곱합 + 측정오차제곱합)

$$TSS \quad = \quad SSB \quad + \quad SSW \quad (SST + SSE)$$

- 검정통계량

$$F = \frac{MST}{MSE} = \frac{\frac{12.978}{3-1}}{\frac{229.022}{(45-1)(3-1)}} = 2.493 < 3.848 = F_{(\alpha,\, k-1,\, (n-1)(k-1))}$$

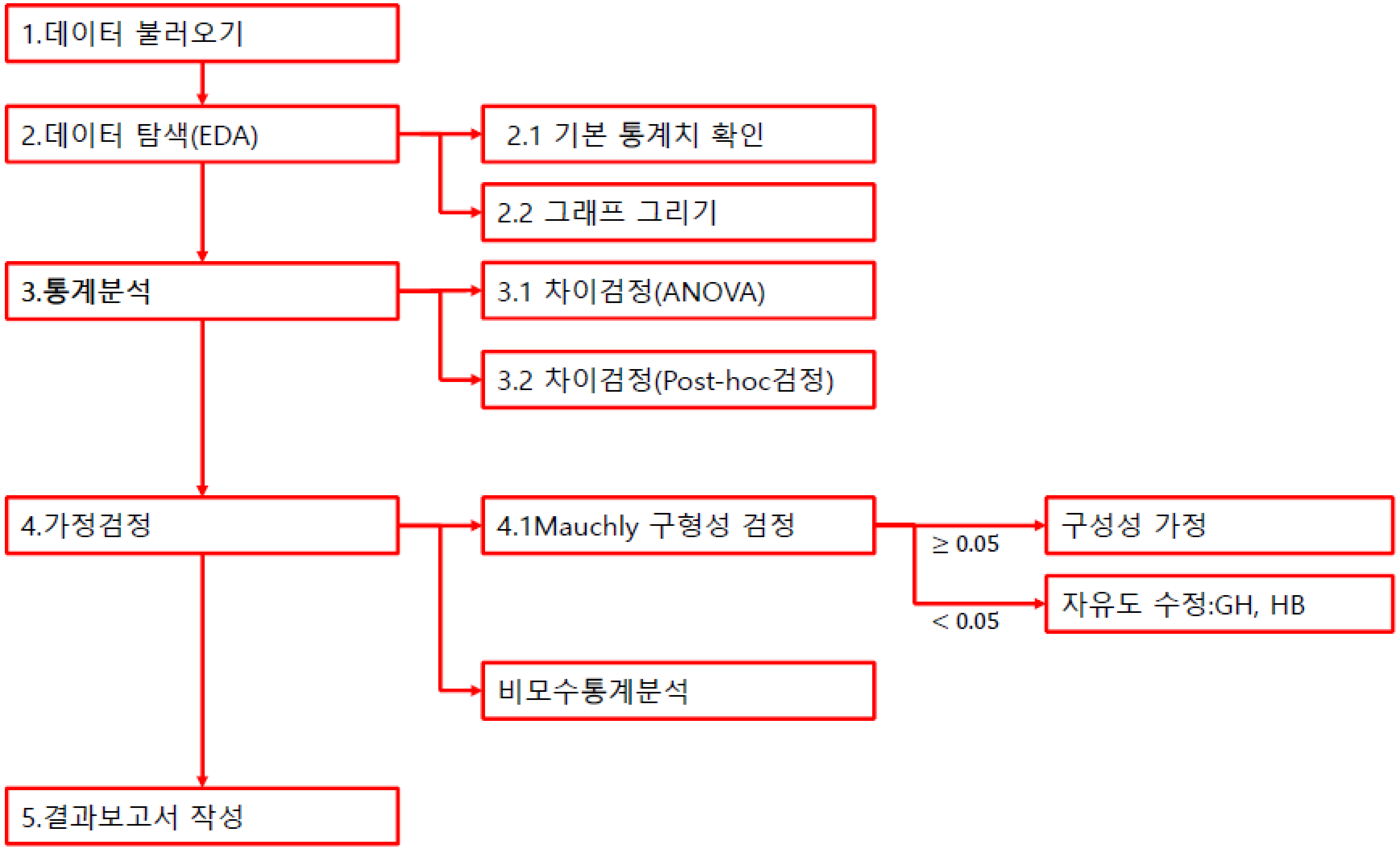
$$p - value = P(F > 2.493) = 0.088 > \alpha = 0.05$$

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p
시점	None	12.978	2	6.489	2.493	0.088
	Greenhouse-Geisser	12.978	1.832	7.084	2.493	0.094
	Huynh-Feldt	12.978	1.908	6.803	2.493	0.091
Residual	None	229.022	88	2.603		
	Greenhouse-Geisser	229.022	80.607	2.841		
	Huynh-Feldt	229.022	83.934	2.729		

Note. Type 3 Sums of Squares

Repeated Measures ANOVA 분석절차



Repeated Measures ANOVA

- 어학 프로그램의 효과를 측정하기 위해 참여전, 3개월 후, 6개월 후에 영어실력을 테스트한 결과, 참여전(M = 60.2), 3개월 후 (M = 61.6), 6개월 후(M = 68.1) 등 시간이 지남에 따라 영어점수가 변화가 있는 것으로 나타났다(F = 248.614, p < .000).
- 시점 간을 비교 분석한 결과, 교육 직후보다 3개월 후, 3개월 후보다 6개월 후 점수가 통계적으로 상승한 것으로 나타났다.

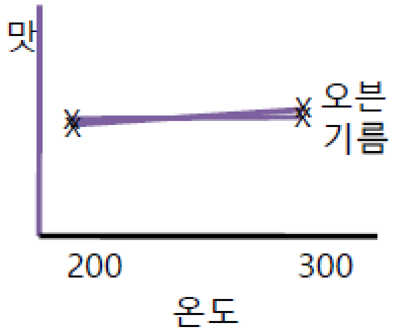
	참여전	3개월후	6개월후	F	p
점수	60.2	61.6	68.1	248.614	.000
	a	b	c		

이원 분산분석

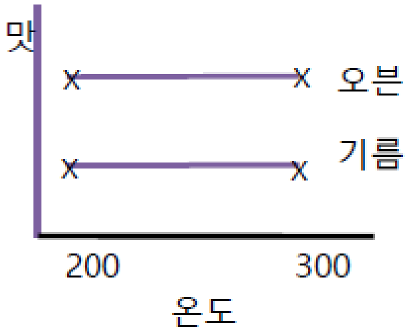
Two-Way ANOVA

- 문제의 정의
  - 치킨의 맛을 결정하는 두 가지 요인은 튀길 때의 기름 온도와 튀기는 방법이다.
  - 튀길 때의 온도를 200도(1)와 300도(2)로 하고, 튀기는 방법을 오븐(1)과 기름(2)으로 하여 치킨을 튀긴 후에 사람들에게 맛을 평가하도록 하였다.
  - 과연 온도와 방법이 맛을 결정하는데 중요한 요인인가? 이 두 가지 요인들간의 상호작용효과는 없었는가?
  - 09_1.TWA.xlsx

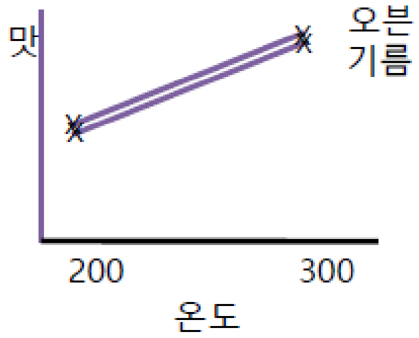
Two-Way ANOVA



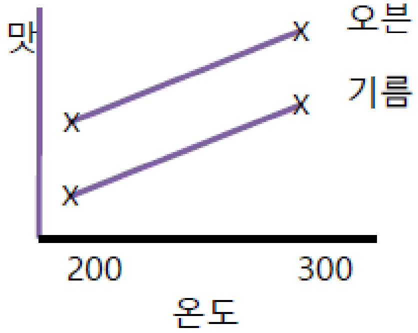
- 방법: X
- 온도: X
- 상호작용: X



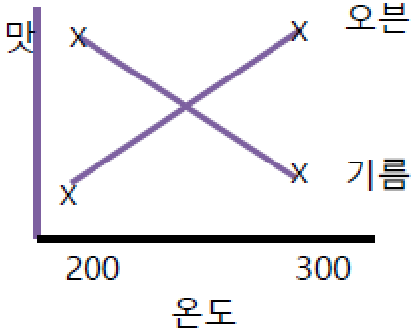
- 방법: O
- 온도: X
- 상호작용: X



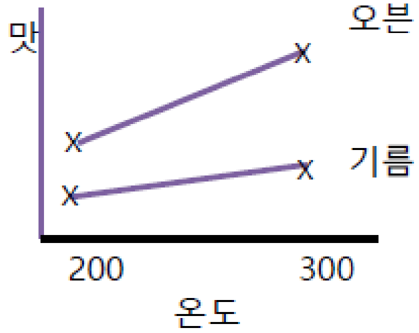
- 방법: X
- 온도: O
- 상호작용: X



- 방법: O
- 온도: O
- 상호작용: X



- 방법: X
- 온도: X
- 상호작용: O



- 방법: O
- 온도: X
- 상호작용: O

Two-Way ANOVA

요인 B \ 요인A		온도 200도	온도 300도	평균
방법	오븐	90 88 88 91 86 .... **(87.5)**	91 88 87 88 90 ... **(94.19)**	91.07
방법	기름	90 91 88 87 91 ... **(94.0)**	86 88 88 92 91 ... **(86.0)**	90.00
평균		90.86	90.23	

Two-Way ANOVA

- 변동(총제곱합)의 분해
  - 총편차의 분해
    - 총편차= A처리효과 + B처리효과 + AB상호작용효과 + 측정오차
  - 검정통계량

$$F_{A0} = \frac{SSA}{a-1} = \frac{MSA}{MSE} \sim F_\alpha(a-1, ab(r-1))$$

$$F_{B0} = \frac{SSA}{b-1} = \frac{MSB}{MSE} \sim F_\alpha(b-1, ab(r-1))$$

$$F_{AB0} = \frac{SSAB}{(a-1)(b-1)} = \frac{MSAB}{MSE} \sim F_\alpha((a-1)(b-1), ab(r-1))$$

Two-Way ANOVA

요인	제곱합 (SS)	자유도(df)	평균제곱 (MS)	F
A 주효과	$SSA = ar\sum_{i}^{b}(\overline{Y}_i - \overline{Y})^2$	$a-1$	$MSA = \frac{SSA}{a-1}$	$\frac{MSA}{MSE}$
B 주효과	$SSb = br\sum_{j}^{a}(\overline{Y}_j - \overline{Y})^2$	$b-1$	$MSB = \frac{SSB}{b-1}$	$\frac{MSB}{MSE}$
상호 작용	$SSAB = r\sum_{i}^{b}\sum_{j}^{a}(\overline{Y}_{ij} - \overline{Y}_i - \overline{Y}_j + \overline{Y})^2$	$(a-1)(b-1)$	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$\frac{MSAB}{MSE}$
잔차	$SSE = \sum_{i}^{b}\sum_{j}^{a}\sum_{k}^{r}(\overline{Y}_{ijk} - \overline{Y}_{ij})^2$	$ab(r-1)$	$MSE = \frac{SSE}{ab(r-1)}$	
총계	$SST = \sum_{j}\sum_{i}(\overline{Y}_{ij} - \overline{Y})^2$	$abr-1$		

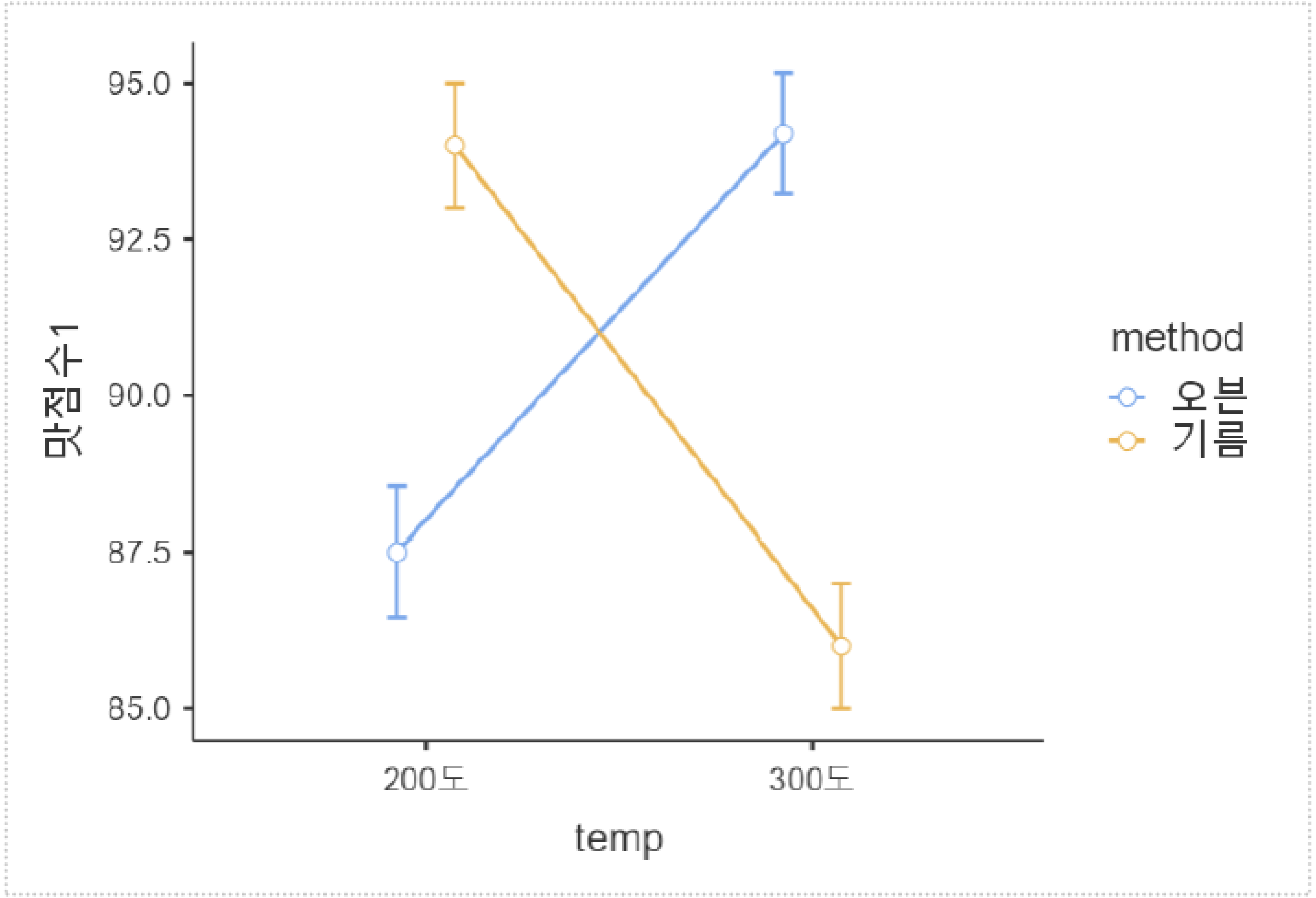
Two-Way ANOVA

- 분산분석표

요인	제곱합 (SS)	자유도(df)	평균제곱 (MS)	F	P-value
방법주효과	10.65	1	10.65	2.84	0.097
온도주효과	6.45	1	6.45	1.72	0.195
상호작용	807.16	1	807.16	215.31	0.001
잔차	209.94	56	3.75		

  - 상호작용효과 유의: 상호작용효과 해석
  - 상호작용효과 유의하지 않은 경우: 상호작용효과를 오차항에 포함해서 다시 검정

Two-Way ANOVA



Two-Way ANOVA

- 통닭을 튀기는 온도와 방법에 따라 맛이 차이가 있는지를 분석한 결과, 튀기는 온도와 방법에 따는 상호작용효과가 있었다(F = 215.31, p = 0.001). 기름일 경우에는 온도가 200도 일 경우에 맛(94.19)이 제일 좋았으며, 오븐일 경우에는 300도로 튀겼을 경우에 맛(94.00)이 좋았다.

		온도			F	P
		200도	300도			
방법	오븐	87.50	94.00	온도	2.84	0.097
				방법	1.72	0.195
	기름	94.19	86.00	방법 * 온도	215.32	0.001

이원 반복측정 분산분석

Two-Way Repeated Measures ANOVA

- 문제의 정의
  - S병원에서는 이번에 새롭게 아로마테라피 치료를 개발하였다.
  - 이 치료가 통증에 효과가 있는지를 검증하기 위해
  - 새롭게 개발한 치료제로 향기요법을 처치 받는 실험군(2)과
  - 일반 향기치료제로 가짜 향기요법을 처치 받는 대조군(1)을 나누고,
  - 치료전과 후에 통증이 차이가 있는지를 검증하였다.
  - 10_1.TWRMA.xlsx

Two-Way Repeated Measures ANOVA

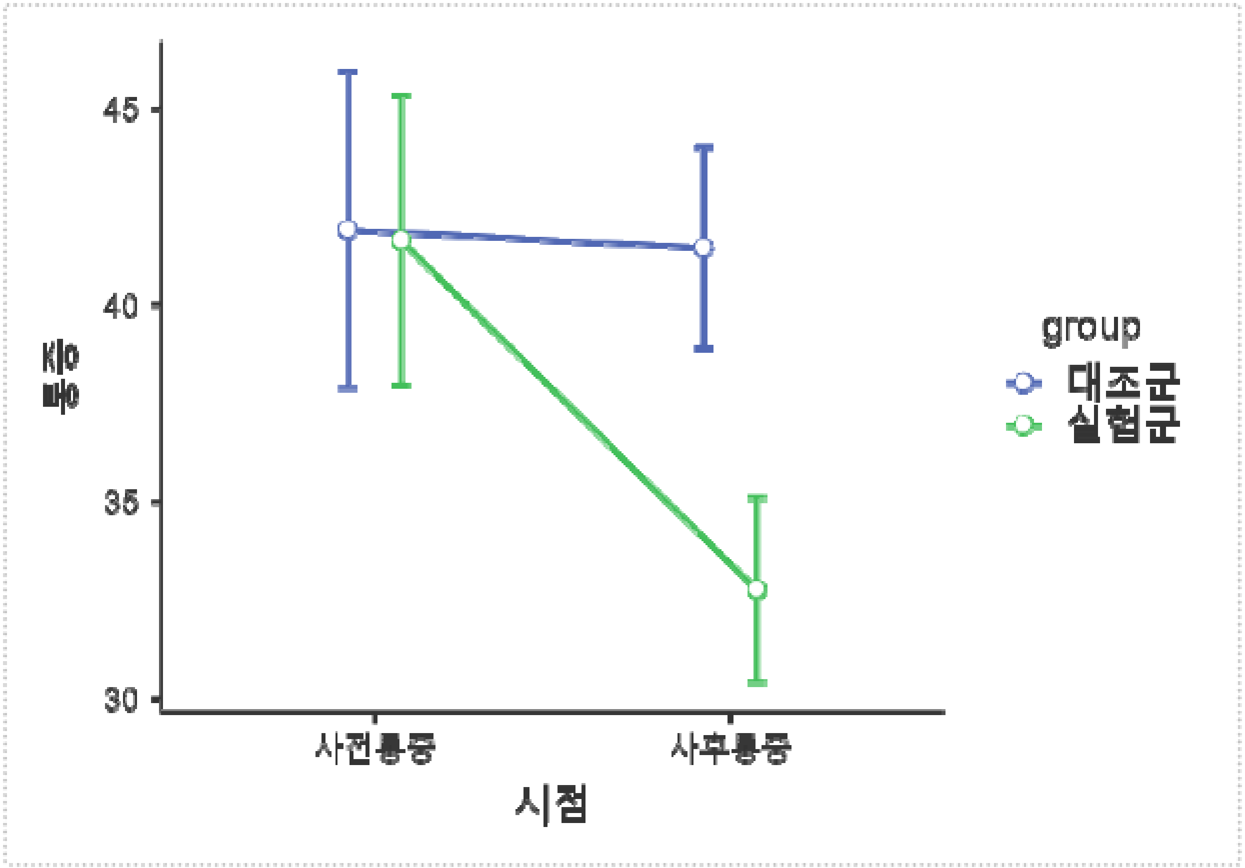
- 가설1
  - 귀무가설(H_0): 실험전에 그룹간 통증은 차이가 없다.
    $$H_0: \mu_{befor\_treatment} = \mu_{before\_control}$$
  - 연구가설(H_1): 실험전에 그룹간 통증은 차이가 있다.
    $$H_0: \mu_{befor\_treatment} \neq \mu_{before\_control}$$
- 가설2
  - 귀무가설(H_0): 그룹과 시점간에는 상호작용이 없다.
    $$H_0: \alpha\beta = 0$$
  - 연구가설(H_1): 그룹과 시점간에는 상호작용이 있다.
    $$H_1: \alpha\beta \neq 0$$
- 가설3
  - 귀무가설(H_0): 실험후에 그룹간 통증은 차이가 없다.
    $$H_0: \mu_{after\_treatment} = \mu_{after\_control}$$
  - 연구가설(H_1): 실험전에 그룹간 통증은 차이가 있다.
    $$H_0: \mu_{after\_treatment} \neq \mu_{after\_control}$$

Two-Way Repeated Measures ANOVA

- 이원 반복측정 분산분석

		요인A (시점)		평균
요인 B	개체	치료전	치료후	
대조군	1 2 3 4 …	57.16 50.88 43.42 45.34 …	55.15 49.58 45.72 46.24 …	41.69
실험군	1 2 3 4 …	40.55 33.70 35.42 33.12 …	30.56 27.54 26.21 24.45 …	37.21
평균		41.77	36.74	

Two-Way Repeated Measures ANOVA



Q&A