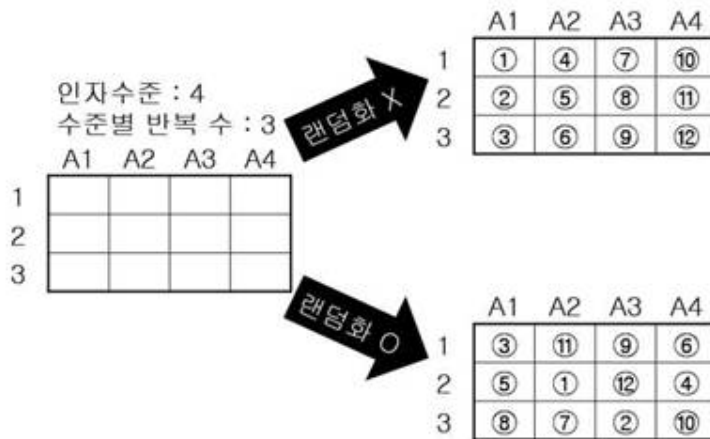


제 2 장 일원배치 분산분석

① 랜덤화 과정

1. 일원배치법의 특징

- 수준수와 각 수준에서 취해지는 측정치의 반복수에는 제한이 없다.
- 반복수가 모든 수준에 대하여 같지 않아도 된다.
- 실험의 측정은 실험의 장 전체를 완전히 랜덤화하여 모든 특성치를 랜덤한 순서에 의해 측정



② 자료 구조와 모형

1. 자료의 구조와 모형

처리					
	1	2	...	l	합계
	Y_{11}	Y_{21}	...	Y_{l1}	$Y_{.1}$
	Y_{12}	Y_{22}	...	Y_{l2}	$Y_{.2}$
	\vdots	\vdots	\vdots	\vdots	\vdots
	Y_{1m}	Y_{2m}	...	Y_{lm}	$Y_{.m}$
합계	$Y_{1.}$	$Y_{2.}$...	$Y_{l.}$	$Y_{..}$
평균	$\overline{Y}_{1.}$	$\overline{Y}_{2.}$...	$\overline{Y}_{l.}$	$\overline{Y}_{..}$

$$① Y_{ij} = \mu_i + \epsilon_{ij} \quad (i = 1, 2, \dots, l, j = 1, 2, \dots, m), \quad \epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$② Y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad (i = 1, 2, \dots, l, j = 1, 2, \dots, m), \quad \epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$\mu = \frac{1}{l} \sum_{i=1}^l \mu_i \quad (\text{총 평균, overall mean})$$

$$\tau_i = \mu_i - \mu (i = 1, 2, \dots, l) \text{ (} i \text{ 번째 처리 효과)}, \sum_{i=1}^l \tau_i = 0$$

$$\Rightarrow \sum_{i=1}^l \tau_i = \sum_{i=1}^l (\mu_i - \mu) = \sum_{i=1}^l \mu_i - l\mu = l \left(\frac{1}{l} \sum_{i=1}^l \mu_i - \mu \right) = 0$$

③ 분산분석의 원리

1. 통계적 가설

$$1) H_0 : \mu_1 = \mu_2 = \dots = \mu_m \text{ vs } H_1 : \text{not } H_0$$

$$2) H_0 : \tau_1 = \tau_2 = \dots = \tau_m = 0 \text{ vs } H_1 : \text{not } H_0$$

2. 모수의 직관적인 추정

1) 모수의 추정

$$\textcircled{1} \hat{\mu} = \sum \frac{Y_{ij}}{N} = \bar{Y}_{..}$$

$$\textcircled{2} \hat{\mu}_i = \sum \frac{Y_{ij}}{l} = \bar{Y}_{i.}$$

$$\textcircled{3} \hat{\tau}_i = \hat{\mu}_i - \hat{\mu} = \bar{Y}_{i.} - \bar{Y}_{..}$$

$$\textcircled{4} \hat{\epsilon}_i = Y_{ij} - \bar{Y}_{i.}$$

2) 최소제곱법(least squares models)에 의한 추정

; 오차들의 제곱합이 최소가 되도록 모수들의 추정량을 구하는 방법으로 최소제곱법으로 구한 추정량들은 일반적으로 매우 바람직한 성질을 가짐이 알려져 있다.

① 최소제곱법의 방법

$$; Q = \sum \sum \epsilon_{ij}^2 = \sum \sum (Y_{ij} - \mu - \tau_i)^2$$

$\Rightarrow Q$ 를 최소화하는 μ 와 τ_i 들은 다음 연립방정식의 해로 결정된다.

$$\frac{\partial Q}{\partial \mu} = -2 \sum \sum (Y_{ij} - \mu - \tau_i) = 0$$

$$\frac{\partial Q}{\partial \tau_i} = -2 \sum_j (Y_{ij} - \mu - \tau_i) = 0 \quad (i = 1, 2, \dots, l)$$

$$\therefore lm\mu + m \sum \tau_i = Y_{..}$$

$$m\mu + m\tau_i = Y_{i.} \quad (i = 1, 2, \dots, l)$$

이 정규방정식의 해를 $\hat{\mu}$, $\hat{\tau}_1$, $\hat{\tau}_2$, ..., $\hat{\tau}_l$ 이라 하고 식을 정리하면 다음과 같다.

$$N\hat{\mu} + r\hat{\tau}_1 + r\hat{\tau}_2 + \dots + r\hat{\tau}_l = N\bar{Y}_{..}$$

$$r\hat{\mu} + r\hat{\tau}_1 = r\bar{Y}_{1.}$$

$$r\hat{\mu} + r\hat{\tau}_2 = r\bar{Y}_{2.}$$

$$r\hat{\mu} + r\hat{\tau}_3 = r\bar{Y}_{3.}$$

$$\vdots \quad \quad \quad \vdots$$

$$r\hat{\mu} + r\hat{\tau}_l = m\bar{Y}_{..}$$

일반적으로 조건 $\sum \tau_i = 0$ 을 삽입하여 해를 구하면

$\hat{\mu} = \bar{Y}_{..}$, $\hat{\tau}_i = \bar{Y}_i - \bar{Y}_{..}$, $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{Y}_i$ 의 값을 구할 수 있다.

3. 제곱합과 자유도

1) 총편차의 분해식

① 관측값 = 총 평균 + 처리효과 편차 + 오차편차

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_i - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_i)$$

② 총편차 = 처리효과 편차 + 오차편차

$$(Y_{ij} - \bar{Y}_{..}) = (\bar{Y}_i - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_i)$$

2) 제곱합의 분해

; $(Y_{ij} - \bar{Y}_{..}) = (\bar{Y}_i - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_i)$ 의 양변을 제곱하고 모든 관측값들에 대하여 합을 취한다.

$$\begin{aligned} \Rightarrow \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_{..})^2 &= \sum_{i=1}^l \sum_{j=1}^m [(\bar{Y}_i - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_i)]^2 \\ &= \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2 + 2 \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})(Y_{ij} - \bar{Y}_i) \\ &= \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2 + 2 \sum_{i=1}^l (\bar{Y}_i - \bar{Y}_{..}) \sum_{j=1}^m (Y_{ij} - \bar{Y}_i) \\ &= \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2 + 2 \sum_{i=1}^l (\bar{Y}_i - \bar{Y}_{..}) \left(\sum_{j=1}^m Y_{ij} - m \bar{Y}_i \right) \\ &= \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2 + 2 \sum_{i=1}^l (\bar{Y}_i - \bar{Y}_{..}) \left(\sum_{j=1}^m Y_{ij} - m \frac{1}{m} \sum_{j=1}^m Y_{ij} \right) \\ &= \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2 + 0 \\ &= \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2 \\ \therefore \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_{..})^2 &= \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2 \end{aligned}$$

3) 제곱합의 간단식

$$\begin{aligned}
 \textcircled{1} \quad SS_T &= \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_{..})^2 \\
 &= \sum_{i=1}^l \sum_{j=1}^m (Y_{ij}^2 - 2Y_{ij}\bar{Y}_{..} + \bar{Y}_{..}^2) \\
 &= \sum_{i=1}^l \sum_{j=1}^m Y_{ij}^2 - 2\bar{Y}_{..} \sum_{i=1}^l \sum_{j=1}^m Y_{ij} + lm\bar{Y}_{..}^2 \\
 &= \sum_{i=1}^l \sum_{j=1}^m Y_{ij}^2 - 2\frac{Y_{..}}{N} Y_{..} + N\left(\frac{Y_{..}}{N}\right)^2 \\
 &= \sum_{i=1}^l \sum_{j=1}^m Y_{ij}^2 - 2\frac{Y_{..}^2}{N} + \frac{Y_{..}^2}{N} \\
 &= \sum_{i=1}^l \sum_{j=1}^m Y_{ij}^2 - \frac{Y_{..}^2}{N} \\
 &= \sum_{i=1}^l \sum_{j=1}^m Y_{ij}^2 - CT
 \end{aligned}$$

$$\begin{aligned}
 \textcircled{2} \quad SS_{tr} &= \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 \\
 &= m \sum_{i=1}^l (\bar{Y}_i - \bar{Y}_{..})^2 \\
 &= m \sum_{i=1}^l (\bar{Y}_i^2 - 2\bar{Y}_{..}\bar{Y}_i + \bar{Y}_{..}^2) \\
 &= m \sum_{i=1}^l \bar{Y}_i^2 - 2m\bar{Y}_{..} \sum_{i=1}^l \bar{Y}_i + lm\bar{Y}_{..}^2 \\
 &\because l \times \frac{\sum_{i=1}^l \bar{Y}_i}{l} = l \times \frac{\bar{Y}_1 + \bar{Y}_2 + \dots + \bar{Y}_l}{l} = l \times \bar{Y}_{..} \\
 &= m \left(\sum_{i=1}^l \bar{Y}_i^2 - 2m\bar{Y}_{..} \times l\bar{Y}_{..} + N\bar{Y}_{..}^2 \right) \\
 &= m \left(\sum_{i=1}^l \bar{Y}_i^2 - N\bar{Y}_{..}^2 \right) \\
 &= m \sum_{i=1}^l \bar{Y}_i^2 - lm \left(\frac{Y_{..}}{N} \right)^2 \\
 &= \sum_{i=1}^m Y_i^2 - N \left(\frac{Y_{..}}{N} \right)^2 \\
 &= \sum_{i=1}^m Y_i^2 - \frac{Y_{..}^2}{N}
 \end{aligned}$$

3) 제곱합의 자유도

; SS_T 의 자유도 = SS_{tr} 의 자유도 + SS_E 의 자유도

$$\Rightarrow (N-1) = (l-1) + (N-l)$$

cf. SS_E 의 자유도가 왜 $l(m-1)$ 일까?

$$\begin{aligned} SS_E &= \sum_{j=1}^l \sum_{i=1}^m (Y_{ij} - \bar{Y}_{i.})^2 \\ &= \sum_{j=1}^m (Y_{1j} - \bar{Y}_{1.})^2 + \sum_{j=1}^m (Y_{2j} - \bar{Y}_{2.})^2 + \cdots + \sum_{j=1}^m (Y_{lj} - \bar{Y}_{l.})^2 \\ \text{따라서 자유도} &: (m-1) + (m-1) + \cdots + (m-1) = l(m-1) = N-l \end{aligned}$$

4. 평균제곱

1) $MS_{tr} = \frac{SS_{tr}}{l-1}$: 처리 평균제곱(treatment MS)

2) $MS_E = \frac{SS_E}{N-l}$: 오차평균제곱(error MS)

MS_E 는 σ^2 의 비편향 추정량이다(즉, $E(MS_E) = \sigma^2$)

5. 추정량

; i 번째 처리의 표본 : $Y_{i1}, Y_{i2}, \dots, Y_{im} \sim N(\mu_i, \sigma^2)$

1) $\hat{\mu}_i = \bar{Y}_{i.} = \sum_{j=1}^m \frac{Y_{ij}}{m}$: μ_i 의 비편향 추정량

2) $\hat{\sigma}^2 = s_i^2 = \frac{\sum_{j=1}^m (Y_{ij} - \bar{Y}_{i.})^2}{m-1}$: σ^2 의 비편향 추정량

3) 공통분산의 추정

$$\begin{aligned} s^2 &= \frac{(m-1)s_1^2 + (m-1)s_2^2 + \cdots + (m-1)s_l^2}{(m-1) + (m-1) + \cdots + (m-1)} \\ &= \frac{\sum_{i=1}^l (m-1)s_i^2}{l(m-1)} \\ &= \frac{\sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_{i.})^2}{N-l} \\ &= \frac{SS_E}{N-l} \\ &= MS_E \end{aligned}$$

4) MS_E 와 MS_{tr} 의 기댓값

$$\begin{aligned} \textcircled{1} \quad \bar{Y}_{i\cdot} &= \frac{1}{m} \sum_{j=1}^m Y_{ij} = \frac{1}{m} \sum_{j=1}^m (\mu + \tau_i + \epsilon_{ij}) = \frac{1}{m} (m\mu + m\tau_i + \sum_{j=1}^m \epsilon_{ij}) \\ &= \mu + \tau_i + \bar{\epsilon}_{i\cdot} \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad \bar{Y}_{..} &= \frac{1}{lm} \sum_{i=1}^l \sum_{j=1}^m Y_{ij} = \frac{1}{lm} \sum_{i=1}^l \sum_{j=1}^m (\mu + \tau_i + \epsilon_{ij}) = \frac{1}{lm} (lm\mu + m \sum_{i=1}^l \tau_i + \sum_{i=1}^l \sum_{j=1}^m \epsilon_{ij}) \\ &= \mu + \bar{\epsilon}_{..} \end{aligned}$$

$$\begin{aligned} \textcircled{3} \quad E(MS_E) &= E \left(\frac{\sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_{i\cdot})^2}{N-l} \right) \\ &= \frac{1}{N-l} E \left(\sum_{i=1}^l \sum_{j=1}^m (\mu + \tau_i + \epsilon_{ij} - \mu - \tau_i - \bar{\epsilon}_{i\cdot})^2 \right) \\ &= E \left(\frac{\sum_{i=1}^l \sum_{j=1}^m (\epsilon_{ij} - \bar{\epsilon}_{i\cdot})^2}{N-l} \right) \\ &= \sigma^2 \end{aligned}$$

$$\begin{aligned} \textcircled{4} \quad E(MS_{tr}) &= E \left(\frac{m \sum_{i=1}^l (\bar{Y}_{i\cdot} - \bar{Y}_{..})^2}{l-1} \right) \\ &= \frac{1}{l-1} E \left(m \sum_{i=1}^l (\mu + \tau_i + \bar{\epsilon}_{i\cdot} - (\mu + \bar{\epsilon}_{..}))^2 \right) \\ &= \frac{1}{l-1} E \left(m \sum_{i=1}^l (\tau_i + (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..}))^2 \right) \\ &= \frac{1}{l-1} E \left(m \sum_{i=1}^l (\tau_i^2 + 2\tau_i(\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..}) + (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..})^2) \right) \\ &= \frac{1}{l-1} \left[m E \left(\sum_{i=1}^l \tau_i^2 \right) + 2m E \left(\sum_{i=1}^l \tau_i (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..}) \right) + m E \left(\sum_{i=1}^l (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..})^2 \right) \right] \\ &= \frac{1}{l-1} \left[m \sum_{i=1}^l \tau_i^2 + 2m \sum_{i=1}^l \tau_i E(\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..}) + m E \left(\sum_{i=1}^l (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..})^2 \right) \right] \\ &= \frac{1}{l-1} \left[m \sum_{i=1}^l \tau_i^2 + m E \left(\sum_{i=1}^l (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..})^2 \right) \right] \quad (\because \bar{\epsilon}_{ij} \sim N(0, \sigma^2)) \\ &= m \sum_{i=1}^l \frac{\tau_i^2}{l-1} + m E \left(\frac{\sum_{i=1}^l (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{..})^2}{l-1} \right) \\ &= m\sigma_A^2 + m \times \frac{\sigma^2}{m} \quad (\because \bar{\epsilon}_{i\cdot} \sim N(0, \frac{\sigma^2}{m})) \\ &= m\sigma_A^2 + \sigma^2 \end{aligned}$$

6. 분산분석 (분산비 검정 (variance-ratio test), F-검정)

1) 분산분석

; l 개의 처리효과 사이에 차이가 없다는 귀무가설 $H_0 : \tau_1 = \tau_2 = \dots = \tau_l = 0$ 을 검정하고자 한다. 물론 대립가설은 μ_i 들이 모두 같지는 않다는 것이다. 귀무가설을 검정하기 위한 규칙을 찾기 위해서, 모든 처리의 모평균이 같으면 ($\bar{Y}_i - \bar{Y}_{..}$)가 작을 것으로 기대된다. 따라서 처리의 평균제곱 $m \sum_{i=1}^l (\bar{Y}_i - \bar{Y}_{..})^2 / (m-1)$ 이 작게된다. 반면에 모평균들의 차이가 현저하면 처리의 평균제곱도 크게 될 것이다. 오차의 분산 σ^2 의 추정량으로 잔차의 평균제곱이 이용되며, 이것은 처리의 평균제곱이 유의적 차이를 나타내기 전에 그것이 어느 정도 큰지를 결정하는 척도로 사용될 수 있다.

분산분석

a) 모형 $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$, $\epsilon_{ij} \sim N(0, \sigma^2)$ ($i = 1, \dots, l, j = 1, \dots, m$)

b) 가설 $H_0 : \tau_1 = \tau_2 = \dots = \tau_l = 0$ (처리 효과 차이가 없다)

$H_1 : \text{not } H_0$ (그렇지 않다.)

b) 검정 통계량 $F = \frac{MS_{tr}}{MS_E} \sim F_{l-1, N-l}$

c) 기각역 $F > F_{l-1, N-l, \alpha}$

2) 분산분석표

source	SS	df	MS	F
처리(Treatments)	SS_{tr}	$l-1$	MS_{tr}	MS_{Trt} / MS_E
오차(Error)	SS_E	$N-l$	MS_E	
합계(Total)	SS_T	$N-1$		

3) 처리 평균에 대한 추론

① μ_i 에 대한 $100(1-\alpha)\%$ 신뢰구간

; $\bar{Y}_i : \mu_i$ 의 점추정량, $\bar{Y}_i \sim N(\mu_i, \sigma^2/m)$, $MS_E = s^2 : \sigma^2$ 에 대한 추정량

$$\Rightarrow \bar{Y}_i \pm t_{N-l, \alpha/2} \frac{\sqrt{MS_E}}{\sqrt{m}}$$

② $(\mu_i - \mu_j)$ 에 대한 $100(1-\alpha)\%$ 신뢰구간

$$\Rightarrow (\bar{Y}_i - \bar{Y}_j) \pm t_{N-l, \alpha/2} \frac{\sqrt{2MS_E}}{\sqrt{m}} \quad (i \neq j)$$

예제 2-1. 어느 공장에서 제품을 생산하는데 열처리 온도에 따라서 제품의 강도가 차이를 보이는지 조사하기 위해 열처리 온도를 $A_1 = 125^\circ\text{C}$, $A_2 = 150^\circ\text{C}$, $A_3 = 175^\circ\text{C}$, $A_4 = 200^\circ\text{C}$ 로 변화시키고, 각 열처리 온도를 16개의 제품을 표본으로 추출하여 강도를 측정한 결과 다음의 자료를 얻었다.

125 °C	150 °C	175 °C	200 °C
12	18	15	39
15	22	26	46
16	26	43	49
17	30	20	38

다음 물음에 답하여라.

(a) 주어진 데이터에서 모든 관측값을 분할하여 SS_T , SS_{tr} , SS_E 를 구하여라.

sol1) 제공함

$$\text{관측값}(Y_{ij}) = \begin{bmatrix} 12 & 18 & 15 & 39 \\ 15 & 22 & 26 & 46 \\ 16 & 26 & 43 & 49 \\ 17 & 30 & 20 & 38 \end{bmatrix} \quad \bar{Y}_1 = 15, \bar{Y}_2 = 24, \bar{Y}_3 = 26, \bar{Y}_4 = 43, \bar{Y}_{..} = 27$$

$$\begin{aligned} \text{총편차}(Y_{ij} - \bar{Y}_{..}) &= \text{처리효과}(\bar{Y}_i - \bar{Y}_{..}) + \text{잔차}(Y_{ij} - \bar{Y}_i) \\ \begin{bmatrix} -15 & -9 & -12 & 12 \\ -12 & -5 & -1 & 19 \\ -11 & -1 & 16 & 22 \\ -10 & 3 & -7 & 11 \end{bmatrix} &= \begin{bmatrix} -12 & -3 & -1 & 16 \\ -12 & -3 & -1 & 16 \\ -12 & -3 & -1 & 16 \\ -12 & -3 & -1 & 16 \end{bmatrix} + \begin{bmatrix} -3 & -6 & -11 & -4 \\ 0 & -2 & 0 & 3 \\ 1 & 2 & 17 & 6 \\ 2 & 6 & -6 & -5 \end{bmatrix} \end{aligned}$$

$$\therefore \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^l \sum_{j=1}^m (\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^l \sum_{j=1}^m (Y_{ij} - \bar{Y}_i)^2$$

$$SS_T = (-15)^2 + (-9)^2 + \cdots + 11^2 = 2266$$

$$SS_{tr} = (-12)^2 + (-12)^2 + \cdots + 16^2 = 1640$$

$$SS_E = (-3)^2 + 0^2 + \cdots + (-5)^2 = 626$$

sol2) 간단식

	125 °C	150 °C	175 °C	200 °C	
	12	18	15	39	
	15	22	26	46	
	16	26	43	49	
	17	30	20	38	
합계	60	96	104	172	432
평균	15	24	26	43	27

$$CT = \frac{Y_{..}^2}{lm} = \frac{432^2}{16} = 11664$$

$$\begin{aligned}
 SS_T &= \sum_{i=1}^l \sum_{j=1}^m Y_{ij}^2 - CT \\
 &= \sum_{i=1}^4 \sum_{j=1}^4 Y_{ij}^2 - CT \\
 &= 12^2 + 18^2 + \dots + 38^2 - 11664 \\
 &= 13930 - 11664 \\
 &= 2266
 \end{aligned}$$

$$\begin{aligned}
 SS_{tr} &= \sum_{i=1}^l \frac{Y_{i.}^2}{m} - CT \\
 &= \sum_{i=1}^4 \frac{Y_{i.}^2}{4} - CT \\
 &= \frac{60^2 + 96^2 + 104^2 + 172^2}{4} - 11664 \\
 &= 1640
 \end{aligned}$$

$$SS_E = SS_T - SS_{tr} = 2266 - 1640 = 626$$

(b) 분산분석표를 작성하라

source	SS	df	MS	F
처리(Treatments)	1640	3	546.67	10.479
오차(Error)	626	12	52.17	
합계(Total)	2266	15		

(c) 위 data에 의하면 열처리 온도에 따라서 제품의 강도 사이에 유의적 차이가 존재 하는 지를 $\alpha = 0.05$ 에서 검정하여라.

① $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ vs $H_1 : \text{not } H_0$

② [엑셀코드]

$$F_{3, 12, 0.05} = \text{F.INV}(0.95, 3, 12) = 3.49$$

$$F \geq 3.49$$

③ $F = 10.479$

④ $\alpha = 0.05$ 에서 H_0 reject

⑤ 집단간 평균차이가 있다.

⑥ 사후 검정 필요.

(d) 각 수준의 모평균의 95% 신뢰구간을 구하여라.

$$\bar{Y}_i \pm t_{N-l, \alpha/2} \frac{\sqrt{MS_E}}{\sqrt{m}}, \quad \bar{Y}_i \pm t_{12, 0.025} \frac{\sqrt{52.17}}{\sqrt{4}}$$

$$15 \pm 2.179 \times \frac{\sqrt{52.17}}{\sqrt{4}} = 15 \pm 7.869 = (7.131, 22.869)$$

$$24 \pm 2.179 \times \frac{\sqrt{52.17}}{\sqrt{4}} = 24 \pm 7.869 = (16.131, 31.869)$$

$$26 \pm 2.179 \times \frac{\sqrt{52.17}}{\sqrt{4}} = 26 \pm 7.869 = (18.131, 33.869)$$

$$43 \pm 2.179 \times \frac{\sqrt{52.17}}{\sqrt{4}} = 43 \pm 7.869 = (35.131, 50.869)$$

(e) 모든 두 수준간의 모평균 차의 95% 신뢰구간을 구하여라.

$$(\bar{Y}_i - \bar{Y}_j) \pm t_{N-t, \alpha/2} \frac{\sqrt{2MS_E}}{\sqrt{l}}, \quad (\bar{Y}_i - \bar{Y}_j) \pm t_{12, 0.025} \frac{\sqrt{2 \times 52.17}}{\sqrt{4}}$$

$$(15 - 24) \pm 2.179 \times \frac{\sqrt{2 \times 52.17}}{\sqrt{4}} = (15 - 24) \pm 11.129 = (-20.129, 2.129)$$

$$(15 - 26) \pm 2.179 \times \frac{\sqrt{2 \times 52.17}}{\sqrt{4}} = (15 - 26) \pm 11.129 = (-22.129, 0.129)$$

$$(15 - 43) \pm 2.179 \times \frac{\sqrt{2 \times 52.17}}{\sqrt{4}} = (15 - 43) \pm 11.129 = (-39.129, -16.871)$$

$$(24 - 26) \pm 2.179 \times \frac{\sqrt{2 \times 52.17}}{\sqrt{4}} = (24 - 26) \pm 11.129 = (-13.129, 9.129)$$

$$(24 - 43) \pm 2.179 \times \frac{\sqrt{2 \times 52.17}}{\sqrt{4}} = (24 - 43) \pm 11.129 = (-30.129, -7.871)$$

$$(26 - 43) \pm 2.179 \times \frac{\sqrt{2 \times 52.17}}{\sqrt{4}} = (26 - 43) \pm 11.129 = (-28.129, -5.871)$$

(f) SPSS 분석결과 : 일원배치 분산분석

① 분산분석

기술통계

강도

	N	평균	표준편차	표준오차	평균에 대한 95% 신뢰구간		최소값	최대값
					하한값	상한값		
125℃	4	15.00	2.160	1.080	11.56	18.44	12	17
150℃	4	24.00	5.164	2.582	15.78	32.22	18	30
175℃	4	26.00	12.193	6.096	6.60	45.40	15	43
200℃	4	43.00	5.354	2.677	34.48	51.52	38	49
합계	16	27.00	12.291	3.073	20.45	33.55	12	49

분산의 등질성 검정

강도

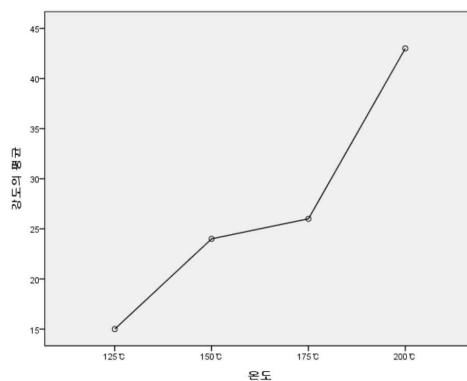
Levene 통계량	df1	df2	유의확률
2.202	3	12	.141

분산분석

강도

	제곱합	df	평균 제곱	거짓	유의확률
집단-간	1640.000	3	546.667	10.479	.001
집단-내	626.000	12	52.167		
합계	2266.000	15			

평균 도표



- i) $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ vs $H_1 : \text{not } H_0$
- ii) $\alpha = 0.05 > p\text{-value} = 0.001$
- iii) $\alpha = 0.05$ 에서 H_0 reject
- iv) 집단간 평균차이가 있다.
- v) 사후 검정 필요.

② 다중비교

다중 비교

종속 변수: 강도

	(I) 온도	(J) 온도	평균차(I-J)	표준오차	유의확률	95% 신뢰구간	
						하한값	상한값
Tukey HSD	125℃	150℃	-9.00	5.107	.337	-24.16	6.16
		175℃	-11.00	5.107	.192	-26.16	4.16
		200℃	-28.00*	5.107	.001	-43.16	-12.84
	150℃	125℃	9.00	5.107	.337	-6.16	24.16
		175℃	-2.00	5.107	.979	-17.16	13.16
		200℃	-19.00*	5.107	.013	-34.16	-3.84
	175℃	125℃	11.00	5.107	.192	-4.16	26.16
		150℃	2.00	5.107	.979	-13.16	17.16
		200℃	-17.00*	5.107	.027	-32.16	-1.84
	200℃	125℃	28.00*	5.107	.001	12.84	43.16
		150℃	19.00*	5.107	.013	3.84	34.16
		175℃	17.00*	5.107	.027	1.84	32.16
LSD	125℃	150℃	-9.00	5.107	.103	-20.13	2.13
		175℃	-11.00	5.107	.052	-22.13	.13
		200℃	-28.00*	5.107	.000	-39.13	-16.87
	150℃	125℃	9.00	5.107	.103	-2.13	20.13
		175℃	-2.00	5.107	.702	-13.13	9.13
		200℃	-19.00*	5.107	.003	-30.13	-7.87
	175℃	125℃	11.00	5.107	.052	-.13	22.13
		150℃	2.00	5.107	.702	-9.13	13.13
		200℃	-17.00*	5.107	.006	-28.13	-5.87
	200℃	125℃	28.00*	5.107	.000	16.87	39.13
		150℃	19.00*	5.107	.003	7.87	30.13
		175℃	17.00*	5.107	.006	5.87	28.13

관측평균을 기준으로 합니다.

오류 조건은 평균 제곱(오류) = 52.167입니다.

*. 평균차는 0.05 수준에서 유의합니다.

강도

온도	N	집단군	
		1	2
Student-Newman-Keuls ^a b	125℃	4	15.00
	150℃	4	24.00
	175℃	4	26.00
	200℃	4	43.00
	유의확률	.120	1.000
Tukey HSD ^{a, b}	125℃	4	15.00
	150℃	4	24.00
	175℃	4	26.00
	200℃	4	43.00
	유의확률	.192	1.000

동일 집단군에 있는 집단에 대한 평균이 표시됩니다.
관측평균을 기준으로 합니다.

오류 조건은 평균 제곱(오류) = 52.167입니다.

a. 조화평균 표본 크기 4.000을(를) 사용합니다.

b. 유의수준 = 0.05.

SNK 방법에 의한 사후 검정 결과 125℃, 150℃, 175℃ 사이에는 유의한 차이가 없었고 이들 각각과 200℃ 사이에는 유의한 차이가 있었다. HSD 방법에 의한 사후 검정 결과 125℃, 150℃, 175℃ 사이에는 유의한 차이가 없었고 이들 각각과 200℃ 사이에는 유의한 차이가 있었다.

(g) SPSS 분석결과 : 일변량분석

SPSS 일변량 분석 옵션

① 분산분석

기술통계량

종속 변수: 온도

온도	평균	표준편차	N
125°C	15.00	2.160	4
150°C	24.00	5.164	4
175°C	26.00	12.193	4
200°C	43.00	5.354	4
합계	27.00	12.291	16

오차 분산의 동일성에 대한 Levene의 검정^a

종속 변수: 온도

F	df1	df2	유의확률
2.202	3	12	.141

여러 집단에서 종속변수의 오차 분산이 동일한 영가설을 검정합니다.

a. Design: 절편 + 온도

개체-간 효과 검정

종속 변수: 온도

소스	제 III 유형 제곱합	자유도	평균 제곱	F	유의확률
수정 모형	1640.000 ^a	3	546.667	10.479	.001
절편	11664.000	1	11664.000	223.591	.000
온도	1640.000	3	546.667	10.479	.001
오차	626.000	12	52.167		
합계	13930.000	16			
수정 합계	2266.000	15			

a. R 제곱 = .724 (수정된 R 제곱 = .655)

[등분산성]

- i) H_0 : 분산이 같다. vs H_1 : 분산이 다르다.
- ii) $\alpha = 0.05 < p\text{-value} = 0.141$
- iii) $\alpha = 0.05$ 에서 H_0 not reject
- iv) 분산이 같다.

[분산분석]

- i) $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ vs $H_1 : \text{not } H_0$
- ii) $\alpha = 0.05 > p\text{-value} = 0.001$
- iii) $\alpha = 0.05$ 에서 H_0 reject
- iv) 집단간 평균차이가 있다.
- v) 사후 검정 필요.

온도

종속 변수: 강도

온도	평균	표준오차	95% 신뢰구간	
			하한값	상한값
125℃	15.000	3.611	7.132	22.868
150℃	24.000	3.611	16.132	31.868
175℃	26.000	3.611	18.132	33.868
200℃	43.000	3.611	35.132	50.868

각 온도의 모평균의 95% 신뢰구간은 위 신뢰구간과 같다.

② 다중비교

다중 비교

종속 변수: 강도

(I) 온도		(J) 온도	평균차(I-J)	표준오차	유의확률	95% 신뢰구간	
						하한값	상한값
Tukey HSD	125℃	150℃	-9.00	5.107	.337	-24.16	6.16
		175℃	-11.00	5.107	.192	-26.16	4.16
		200℃	-28.00*	5.107	.001	-43.16	-12.84
	150℃	125℃	9.00	5.107	.337	-6.16	24.16
		175℃	-2.00	5.107	.979	-17.16	13.16
		200℃	-19.00*	5.107	.013	-34.16	-3.84
	175℃	125℃	11.00	5.107	.192	-4.16	26.16
		150℃	2.00	5.107	.979	-13.16	17.16
		200℃	-17.00*	5.107	.027	-32.16	-1.84
	200℃	125℃	28.00*	5.107	.001	12.84	43.16
		150℃	19.00*	5.107	.013	3.84	34.16
		175℃	17.00*	5.107	.027	1.84	32.16
LSD	125℃	150℃	-9.00	5.107	.103	-20.13	2.13
		175℃	-11.00	5.107	.052	-22.13	.13
		200℃	-28.00*	5.107	.000	-39.13	-16.87
	150℃	125℃	9.00	5.107	.103	-2.13	20.13
		175℃	-2.00	5.107	.702	-13.13	9.13
		200℃	-19.00*	5.107	.003	-30.13	-7.87
	175℃	125℃	11.00	5.107	.052	-.13	22.13
		150℃	2.00	5.107	.702	-9.13	13.13
		200℃	-17.00*	5.107	.006	-28.13	-5.87
	200℃	125℃	28.00*	5.107	.000	16.87	39.13
		150℃	19.00*	5.107	.003	7.87	30.13
		175℃	17.00*	5.107	.006	5.87	28.13

관측평균을 기준으로 합니다.

오류 조건은 평균 제곱(오류) = 52.167입니다.

*. 평균차는 0.05 수준에서 유의합니다.

강도

온도		N	집단군	
			1	2
Student-Newman-Keuls ^a , ^b	125℃	4	15.00	
	150℃	4	24.00	
	175℃	4	26.00	
	200℃	4		43.00
	유의확률		.120	1.000
Tukey HSD ^{a, b}	125℃	4	15.00	
	150℃	4	24.00	
	175℃	4	26.00	
	200℃	4		43.00
	유의확률		.192	1.000

동일 집단군에 있는 집단에 대한 평균이 표시됩니다.

오류 조건은 평균 제곱(오류) = 52.167입니다.

a. 조화평균 표본 크기 4.000를(를) 사용합니다.

b. 유의수준 = 0.05.

SNK 방법에 의한 사후 검정 결과 125℃, 150℃, 175℃사이에는 유의한 차이가 없었고 이들 각각과 200℃사이에는 유의한 차이가 있었다. HSD 방법에 의한 사후 검정 결과 125℃, 150℃, 175℃사이에는 유의한 차이가 없었고 이들 각각과 200℃사이에는 유의한 차이가 있었다.

예제 2-2. 국내 4개 회사에서 생산되는 무가당 오렌지 주스(A, B, C, D)의 신 맛을 48명의 평가 요원이 주스마다 12명씩 랜덤 배치되어 9점 척도를 사용하여 평가한 결과이다.

A	B	C	D
8 8	6 5	8 8	9 4
7 9	7 7	8 5	6 5
8 7	6 7	6 7	4 4
9 8	6 8	6 7	5 6
6 8	7 6	7 6	7 5
7 6	8 7	7 8	8 4

(a) 분산분석표를 작성하라

source	SS	df	MS	F
처리(Treatments)	24.896	3	8.299	5.945
오차(Error)	61.417	44	1.396	
합계(Total)	86.313	47		

(b) 위 data에 의하면 오렌지 주스에 따라서 신맛에 유의적 차이가 존재하는 지를 유의수준 5%에서 검정하여라.

① $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ vs $H_1 : \text{not } H_0$

② **[엑셀코드]**

$$F_{3, 44, 0.05} = \text{F.INV}(0.95, 3, 44) = 2.816$$

$$F \geq 2.816$$

③ $F = 5.945$

④ $\alpha = 0.05$ 에서 H_0 reject

⑤ 집단간 평균차이가 있다.

⑥ 사후 검정 필요.

(c) 각 수준의 모평균의 95% 신뢰구간을 구하여라.

[엑셀코드]

$$t_{44, 0.025} = T.INV(0.05, 44) = 2.015$$

$$7.583 \pm 2.015 \times \frac{\sqrt{1.963}}{\sqrt{12}} = (6.896, 8.271)$$

$$6.667 \pm 2.015 \times \frac{\sqrt{1.963}}{\sqrt{12}} = (6.896, 8.271)$$

$$6.917 \pm 2.015 \times \frac{\sqrt{1.963}}{\sqrt{12}} = (6.896, 8.271)$$

$$5.583 \pm 2.015 \times \frac{\sqrt{1.963}}{\sqrt{12}} = (6.896, 8.271)$$

(d) 모든 두 수준간의 모평균 차의 90% 신뢰구간을 구하여라.

[엑셀코드]

$$t_{44, 0.025} = T.INV(0.1, 44) = 1.680$$

$$(7.583 - 6.667) \pm 1.608 \times \frac{\sqrt{2 \times 1.963}}{\sqrt{12}} = (6.896, 8.271)$$

$$(7.583 - 6.917) \pm 1.608 \times \frac{\sqrt{2 \times 1.963}}{\sqrt{12}} = (-0.144, 1.476)$$

$$(7.583 - 5.583) \pm 1.608 \times \frac{\sqrt{2 \times 1.963}}{\sqrt{12}} = (1.190, 2.810)$$

$$(6.667 - 6.917) \pm 1.608 \times \frac{\sqrt{2 \times 1.963}}{\sqrt{12}} = (-1.060, 0.560)$$

$$(6.667 - 5.583) \pm 1.608 \times \frac{\sqrt{2 \times 1.963}}{\sqrt{12}} = (0.274, 1.894)$$

$$(6.917 - 5.583) \pm 1.608 \times \frac{\sqrt{2 \times 1.963}}{\sqrt{12}} = (0.524, 2.144)$$

예제 2-3. 주어진 자료는 돼지 체중증가에 가장 효율적인 사료 내 단백질 첨가 퍼센티지를 결정하려는 비교실험에서 얻어진 측정값들이다. 이 실험에서는 단백질 %가 처리이다. 동일 품종으로 나이와 무게가 비슷한 돼지 20마리를 동원하여, 각 처리별로 5마리씩 랜덤 배치한 다음 일정 기간 후에 돼지의 체중 증가분을 측정하였다.

	10%	20%	30%	40%
	60.8	87.9	102.6	68.7
	57.0	84.2	102.1	67.7
	65.0	83.1	100.2	74.0
	58.6	85.7	96.5	66.3
	61.7	90.3	99.9	69.8

(a) 다음은 SPSS 분석 결과이다. 단백질 %에 따른 체중 증가량에 차이가 있는 지를 유의수준 1%에서 검정하여라.

기술통계량

종속 변수: 체중증가분

처리	평균	표준편차	N
10%	60.620	3.0646	5
20%	86.240	2.8962	5
30%	100.260	2.4048	5
40%	69.300	2.9266	5
합계	79.105	15.9108	20

오차 분산의 동질성에 대한 Levene의 검정^a

종속 변수: 체중증가분

F	df1	df2	유의확률
.155	3	16	.925

여러 집단에서 종속변수의 오차 분산이 동일한 평가설을 검정합니다.

a. Design: 절편 + 처리

개체-간 효과 검정

종속 변수: 체중증가분

소스	제 III 유형 제곱합	자유도	평균 제곱	F	유의확률
수정 모형	4681.378 ^a	3	1560.459	194.280	.000
절편	125152.020	1	125152.020	15581.676	.000
처리	4681.377	3	1560.459	194.280	.000
오차	128.512	16	8.032		
합계	129961.910	20			
수정 합계	4809.889	19			

a. R 제곱 = .973 (수정된 R 제곱 = .968)

[등분산성]

- i) H_0 : 분산이 같다. vs H_1 : 분산이 다르다.
- ii) $\alpha = 0.01 < p\text{-value} = 0.925$
- iii) $\alpha = 0.01$ 에서 H_0 not reject
- iv) 분산이 같다.

[분산분석]

- i) $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ vs $H_1 : \text{not } H_0$
- ii) $\alpha = 0.01 > p\text{-value} = 0.000$
- iii) $\alpha = 0.01$ 에서 H_0 reject
- iv) 집단간 평균차이가 있다.
- v) 사후 검정 필요.

(b) 각 수준의 모평균의 95% 신뢰구간을 구하고 아래 SPSS 결과와 비교하여라.

처리

종속 변수: 체중증가분

처리	평균	표준오차	95% 신뢰구간	
			하한값	상한값
10%	60.620	1.267	57.933	63.307
20%	86.240	1.267	83.553	88.927
30%	100.260	1.267	97.573	102.947
40%	69.300	1.267	66.613	71.987

각 체중증가분에 대한 모평균의 95% 신뢰구간은 위 신뢰구간과 같다.

예제 2-4. 다음 분석결과는 1990~1991 시즌의 잉글랜드, 이탈리아, 스페인, 독일, 프랑스의 경기당 평균 득점의 평균을 비교한 결과이다. 다음 물음에 답하여라.

경기당득점

	N	평균	표준편차	표준오차	평균에 대한 95% 신뢰 구간		최소값	최대값
					하한값	상한값		
잉글랜드	18	1.3545	.24796	.05844	1.2312	1.4778	.93	1.93
이탈리아	18	1.1356	.39255	.09252	.9404	1.3308	.71	2.18
스페인	18	1.2456	.41833	.09860	1.0376	1.4536	.79	2.29
독일	18	1.3304	.30745	.07247	1.1775	1.4833	.84	2.00
프랑스	18	1.0497	.28241	.06656	.9093	1.1901	.66	1.76
합계	90	1.2232	.34849	.03673	1.1502	1.2962	.66	2.29

경기당득점

Levene 통계량	자유도1	자유도2	유의확률
1.107	4	85	.359

경기당득점

	제곱합	자유도	평균제곱	F	유의확률
집단-간	1.206	4	.302	2.669	.038
집단-내	9.603	85	.113		
합계	10.809	89			

(a) 국가에 경기당 평균득점에 차이가 있는 지를 유의수준 5%에서 검정하여라.

[등분산성]

- i) H_0 : 분산이 같다. vs H_1 : 분산이 다르다.
- ii) $\alpha = 0.05 < p\text{-value} = 0.359$
- iii) $\alpha = 0.05$ 에서 H_0 not reject
- iv) 분산이 같다.

[분산분석]

- i) $H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$ vs $H_1 : \text{not } H_0$
- ii) $\alpha = 0.05 > p\text{-value} = 0.038$
- iii) $\alpha = 0.05$ 에서 H_0 reject
- iv) 집단간 평균차이가 있다.
- v) 사후 검정 필요.

(b) 결정계수를 구하고 설명하여라.

$$R^2 = \frac{SS_{tr}}{SS_T} = \frac{9.603}{10.809} = 0.888$$

총 변동중에서 모형으로 설명되는 변동의 비율이 88.8%다.

(b) 각 수준의 모평균의 90% 신뢰구간을 구하여라.

종속 변수:경기당득점

국가	평균	표준오차	90% 신뢰구간	
			하한값	상한값
잉글랜드	1.354	.079	1.223	1.486
이탈리아	1.136	.079	1.004	1.267
스페인	1.246	.079	1.114	1.377
독일	1.330	.079	1.199	1.462
프랑스	1.050	.079	.918	1.181

종속 변수:경기당득점

국가	평균	표준오차	95% 신뢰구간	
			하한값	상한값
잉글랜드	1.354	.079	1.197	1.512
이탈리아	1.136	.079	.978	1.293
스페인	1.246	.079	1.088	1.403
독일	1.330	.079	1.173	1.488
프랑스	1.050	.079	.892	1.207

(c) 모든 두 수준간의 모평균 차의 95% 신뢰구간을 구하여라.

경기당득점
LSD

(I) 국가	(J) 국가	평균차(I-J)	표준오차	유의확률	95% 신뢰구간	
					하한값	상한값
잉글랜드	이탈리아	.2189	.11204	.054	-.0039	.4416
	스페인	.1089	.11204	.334	-.1139	.3316
	독일	.0241	.11204	.830	-.1987	.2468
	프랑스	.3048*	.11204	.008	.0820	.5276
이탈리아	잉글랜드	-.2189	.11204	.054	-.4416	.0039
	스페인	-.1100	.11204	.329	-.3328	.1128
	독일	-.1948	.11204	.086	-.4175	.0280
	프랑스	.0859	.11204	.445	-.1368	.3087
스페인	잉글랜드	-.1089	.11204	.334	-.3316	.1139
	이탈리아	.1100	.11204	.329	-.1128	.3328
	독일	-.0848	.11204	.451	-.3076	.1380
	프랑스	.1959	.11204	.084	-.0269	.4187
독일	잉글랜드	-.0241	.11204	.830	-.2468	.1987
	이탈리아	.1948	.11204	.086	-.0280	.4175
	스페인	.0848	.11204	.451	-.1380	.3076
	프랑스	.2807*	.11204	.014	.0579	.5035
프랑스	잉글랜드	-.3048*	.11204	.008	-.5276	-.0820
	이탈리아	-.0859	.11204	.445	-.3087	.1368
	스페인	-.1959	.11204	.084	-.4187	.0269
	독일	-.2807*	.11204	.014	-.5035	-.0579

예제 2-5. 어떤 식물의 가공시 처리액의 농도 A 를 인자로 하여 $A_1 = 3.0\%$, $A_2 = 3.3\%$, $A_3 = 3.6\%$, $A_4 = 3.9\%$, $A_5 = 4.2\%$ 에서 반복 각 4회, 전체 20회를 랜덤하게 하여 처리한 후의 인장강도를 측정하여 다음 데이터를 얻었다.

A_1	A_2	A_3	A_4	A_5
46.8	51.2	50.2	40.8	30.2
58.0	62.4	39.8	41.8	25.8
51.4	58.5	45.2	45.5	32.4
56.5	61.9	48.8	35.9	29.2

(a) 처리액의 농도에 따라 인장강도에 차이가 있는가를 유의수준 1%에서 검정하여라.

source	SS	df	MS	F
처리(Treatments)	2030.198	4	507.549	26.044
오차(Error)	292.328	15	19.489	
합계(Total)	2322.525	19		

① $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ vs $H_1 : \text{not } H_0$

② **[엑셀코드]**

$$F_{4, 15, 0.01} = F.INV(0.99, 4, 15) = 4.893$$

$$F \geq 4.893$$

③ $F = 26.044$

④ $\alpha = 0.01$ 에서 H_0 reject

⑤ 집단간 평균차이가 있다.

⑥ 사후 검정 필요.

(b) 결정계수를 구하고 설명하여라.

$$R^2 = \frac{SS_{tr}}{SS_T} = \frac{2030.198}{2322.525} = 0.874$$

총 변동중에서 모형으로 설명되는 변동의 비율이 87.4%다.

(c) A의 각 수준에서 인장강도의 모평균의 95% 신뢰구간을 구하여라.

처리액농도

종속 변수:인장강도

처리액농도	평균	표준오차	95% 신뢰구간	
			하한값	상한값
1	53.175	2.207	48.470	57.880
2	58.500	2.207	53.795	63.205
3	46.000	2.207	41.295	50.705
4	41.000	2.207	36.295	45.705
5	29.400	2.207	24.695	34.105

예제 2-6. 보일러 부식의 문제에 있어서는 연소생성가스 중의 $\text{SO}_3\%$ 가 문제가 된다. 유황분의 함유량에 차이가 있다고 믿어지는 4종류의 기름연료를 사용하였을 때에 연소생성가스 중의 $\text{SO}_3\%$ 를 측정된 결과 다음과 같은 데이터를 얻었다.

A_1	A_2	A_3	A_4
6.1	4.4	3.1	2.8
5.8	4.2	3.3	2.6
6.4	3.9	3.1	2.6
		3.2	3.1
		3.6	2.7
		3.5	2.5

(a) 연료 기름의 종류가 다름에 따라 $\text{SO}_3\%$ 의 함량에 차이가 있다고 할 수 있는지 유의수준 1%에서 검정하여라.

source	SS	df	MS	F
처리(Treatments)	24.69	3	8.230	152.609
오차(Error)	0.755	14	0.054	
합계(Total)	25.445	17		

$$CT = \frac{66.9^2}{18} = 248.645$$

$$SS_T = 6.1^2 + 5.8^2 + \dots + 2.5^2 - CT$$

$$= 25.445$$

$$SS_{tr} = \frac{18.3^2}{3} + \frac{12.5^2}{3} + \frac{19.8^2}{6} + \frac{16.3^2}{6} - CT$$

$$= 24.69$$

$$SS_E = SS_T - SS_{tr}$$

$$= 0.755$$

(b) 각 연료기름의 종류에 따른 $\text{SO}_3\%$ 의 95% 신뢰구간을 구하여라.

$$6.1 \pm 2.145 \times \frac{\sqrt{0.054}}{\sqrt{3}} = (5.812, 6.388)$$

$$4.167 \pm 2.145 \times \frac{\sqrt{0.054}}{\sqrt{3}} = (3.879, 4.454)$$

$$3.3 \pm 2.145 \times \frac{\sqrt{0.054}}{\sqrt{6}} = (5.812, 6.388)$$

$$2.717 \pm 2.145 \times \frac{\sqrt{0.054}}{\sqrt{6}} = (2.513, 2.920)$$

기름연료

종속 변수: S03

기름연료	평균	표준오차	95% 신뢰구간	
			하한값	상한값
연료1	6.100	.134	5.812	6.388
연료2	4.167	.134	3.879	4.454
연료3	3.300	.095	3.097	3.503
연료4	2.717	.095	2.513	2.920

(c) 모든 두 수준간의 모평균 차의 95% 신뢰구간을 구하여라.

$$(6.1 - 4.167) \pm 2.145 \times \sqrt{\frac{0.054}{3} + \frac{0.054}{3}} = (1.526, 2.340)$$

$$(6.1 - 3.3) \pm 2.145 \times \sqrt{\frac{0.054}{3} + \frac{0.054}{6}} = (2.448, 3.152)$$

$$(6.1 - 2.717) \pm 2.145 \times \sqrt{\frac{0.054}{3} + \frac{0.054}{6}} = (3.301, 3.736)$$

$$(4.167 - 3.3) \pm 2.145 \times \sqrt{\frac{0.054}{3} + \frac{0.054}{6}} = (0.514, 1.219)$$

$$(4.167 - 2.717) \pm 2.145 \times \sqrt{\frac{0.054}{3} + \frac{0.054}{6}} = (1.098, 1.802)$$

$$(3.3 - 2.717) \pm 2.145 \times \sqrt{\frac{0.054}{6} + \frac{0.054}{6}} = (0.296, 0.871)$$

S03
LSD

(I) 기름연료	(J) 기름연료	평균차(I-J)	표준오차	유의확률	95% 신뢰구간	
					하한값	상한값
연료1	연료2	1.933 [*]	.1896	.000	1.527	2.340
	연료3	2.800 [*]	.1642	.000	2.448	3.152
	연료4	3.383 [*]	.1642	.000	3.031	3.736
연료2	연료1	-1.933 [*]	.1896	.000	-2.340	-1.527
	연료3	.867 [*]	.1642	.000	.514	1.219
	연료4	1.450 [*]	.1642	.000	1.098	1.802
연료3	연료1	-2.800 [*]	.1642	.000	-3.152	-2.448
	연료2	-.867 [*]	.1642	.000	-1.219	-.514
	연료4	.583 [*]	.1341	.001	.296	.871
연료4	연료1	-3.383 [*]	.1642	.000	-3.736	-3.031
	연료2	-1.450 [*]	.1642	.000	-1.802	-1.098
	연료3	-.583 [*]	.1341	.001	-.871	-.296

예제 2-7. 사이프러스에서 발굴된 비잔틴시대에 주조된 동전의 은 함량(%)을 조사하여 주조시기별로 표와 같이 정리하였다. .

1기	2기	3기	4기
5.9	6.9	4.9	5.3
6.8	9.0	5.5	5.6
6.4	6.6	4.6	5.5
7.0	8.1	4.5	5.1
6.6	9.3		6.2
7.7	9.2		5.8
7.2	8.6		5.8
6.9			
6.2			

(a) 다음은 SPSS 일원배치 분산분석 결과이다. 주조시기에 따른 은 함량(%)가 차이가 있는지를 유의수준 1%에서 검정하여라.

기술통계량

종속 변수:은함량

조사시기	평균	표준편차	N
1기	6.744	.5434	9
2기	8.243	1.0998	7
3기	4.875	.4500	4
4기	5.614	.3625	7
합계	6.563	1.3695	27

오차 분산의 동일성에 대한 Levene의 검정^a

종속 변수:은함량

F	df1	df2	유의확률
4.286	3	23	.015

여러 집단에서 종속변수의 오차 분산이 동일한 영가설을 검정합니다.

a. Design: 절편 + 조사시기

개체-간 효과 검정

종속 변수:은함량

소스	제 III 유형 제곱합	자유도	평균 제곱	F	유의확률
수정 모형	37.748 ^a	3	12.583	26.272	.000
절편	1003.449	1	1003.449	2095.181	.000
조사시기	37.748	3	12.583	26.272	.000
오차	11.015	23	.479		
합계	1211.720	27			
수정 합계	48.763	26			

[등분산성]

- i) H_0 : 분산이 같다. vs H_1 : 분산이 다르다.
- ii) $\alpha = 0.01 < p\text{-value} = 0.015$
- iii) $\alpha = 0.01$ 에서 H_0 not reject
- iv) 분산이 같다.

[분산분석]

- i) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs $H_1 : \text{not } H_0$
- ii) $\alpha = 0.01 > p\text{-value} = 0.000$
- iii) $\alpha = 0.01$ 에서 H_0 reject
- iv) 집단간 평균차이가 있다.
- v) 사후 검정 필요.

(b) 결정계수를 구하고 설명하여라.

$$R^2 = \frac{SS_{tr}}{SS_T} = \frac{37.748}{48.763} = 0.774$$

총 변동중에서 모형으로 설명되는 변동의 비율이 77.4%다.

(c) 각 수준에서 은 함량의 모평균의 95% 신뢰구간을 구하고 아래 SPSS 결과와 비교하여라.

조사시기

종속 변수: 은함량

조사시기	평균	표준오차	95% 신뢰구간	
			하한값	상한값
1기	6.744	.231	6.267	7.222
2기	8.243	.262	7.702	8.784
3기	4.875	.346	4.159	5.591
4기	5.614	.262	5.073	6.155