Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

11 CHAPTER

方差分析

第11-2章 方差分析

1. 方差分析的概念
2. 单因素方差分析
3. 双因素方差分析

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

方差分析

- 预测因果关系 → 得出好的决策
- 方差分析
  - 分析某一因素是否影响响应值
- 方差分析的用途
  - 通过方差分析，筛选出有意义的因素，如果事先找到可以得出有价值的响应值的水平范围的话，这将有助于后续的分析。
- 回归分析
  - 分析因素与响应值的函数关系

3/50

1. 方差分析的概念

- 因素(factor)或因子 :在预测会影响响应值的原因中需要分析考虑的原因
- 水平(level)：在试验或者观测中筛选出来的因素的值
- 响应值(response value)：在因素的各水平下所得到的因变量的观测值

[定义 13-1] 方差分析(Analysis of Variance：ANOVA)
：根据因素类别分解响应值的散布，找出对响应值有显著性影响的因素的统计学方法。

- 方差分析: 平方和(sum of squares)来表示响应值的散布，在用各因素的平方和去分解该平方和，筛选出与误差相比具有显著性影响的因素的分析方法。
- 随机化的原理: 为了得出公正的方差分析结果，除了所考虑的因素外，其他可能影响响应值的因素也要维持在一定水平内，并要按序或随机进行试验(观测)。

4/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

2. 单因素方差分析

- 单因素方差分析(one-way ANOVA): 只考虑一个因素的方差分析

2.1 数据的结构

	A_1	A_2	$\cdots$	A_r	
试验的重复	y_{11}	y_{21}	$\cdots$	y_{r1}	
	y_{12}	y_{22}	$\cdots$	y_{r2}	
	$\vdots$	$\vdots$		$\vdots$	
	$\vdots$	y_{2n_2}		$\vdots$	
	y_{1n_1}		$\cdots$	y_{rn_r}	
合计	$T_{1.}$	$T_{2.}$	$\cdots$	$T_{r.}$	T
均值	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\cdots$	$\bar{y}_{r.}$	$\bar{y}$

$\mu_1 \equiv \mu + a_1$ $\mu_2 \equiv \mu + a_2$ $\mu_r \equiv \mu + a_r$

$$\boxed{y_{ij} = \mu_i + \varepsilon_{ij} = \mu + a_i + \varepsilon_{ij}, \quad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, n_i}$$

因素A的主要效果(main effect) $a_i \equiv \mu_i - \mu$

2. 单因素方差分析

- 对于误差的假设 $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ $y_{ij} = \mu + a_i + \varepsilon_{ij}$

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

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

(3) ε_{ij}들이 독립 $\Rightarrow$ y_{ij}들도 독립

2.2 平方和的分解公式

- 和与均值

$$T_{i.} = \sum_{j=1}^{n_i} y_{ij}, \quad i = 1, 2, \ldots, r \qquad \bar{y}_{i.} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} = \frac{T_{i.}}{n_i}, \quad i = 1, 2, \ldots, r$$

$$T = \sum_{i=1}^{r} T_{i.} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} y_{ij} \qquad \bar{y} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{n_i} y_{ij}}{N} = \frac{T}{N} = \frac{\sum_{i=1}^{r} n_i \bar{y}_{i.}}{N}$$

$$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.}) = \sum_{j=1}^{n_i} y_{ij} - n_i \bar{y}_{i.} = \sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} y_{ij} = 0$$

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

2. 单因素方差分析

2.2 平方和的分解公式(接上)

① **총편차** $(y_{ij}-\bar{y})$: 각각의 표본자료(y_{ij})와 전체 평균($\bar{y}$)과의 차이

② **수준간 편차** $(\bar{y}_{i.}-\bar{y})$: 각 수준의 평균($\bar{y}_{i.}$)과 전체 평균($\bar{y}$)과의 차이

③ **수준내 편차** $(y_{ij}-\bar{y}_{i.})$: 각 표본자료(y_{ij})와 그 수준의 평균($\bar{y}_{i.}$)과의 차이

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

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}[(\bar{y}_{i.}-\bar{y})+(y_{ij}-\bar{y}_{i.})]^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\bar{y})^2+\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2+2\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\bar{y})(y_{ij}-\bar{y}_{i.})$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\bar{y})^2+\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2$$

- 正交化的原理: 进行所有偏差都正交的试验设计

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\bar{y})(y_{ij}-\bar{y}_{i.})=\sum_{i=1}^{r}(\bar{y}_{i.}-\bar{y})\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})=0$$

7/50

2. 单因素方差分析

[定理 13-1] 单因素方差分析平方和的分解公式

$$y_{ij}-\bar{y}=(\bar{y}_{i.}-\bar{y})+(y_{ij}-\bar{y}_{i.})\Rightarrow SS_T=SS_A+SS_E$$

보정항

총제곱합: $SS_T=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}y_{ij}^2-\frac{T^2}{N}$ $\qquad \phi_T=N-1$

처리제곱합: $SS_A=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\bar{y})^2=\sum_{i=1}^{r}\frac{T_{i.}^2}{n_i}-\frac{T^2}{N}$ $\qquad \phi_A=r-1$

오차제곱합: $SS_E=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2=SS_T-SS_A$ $\qquad \phi_E=N-r$

$$SS_T=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}^2-2\bar{y}y_{ij}+\bar{y}^2)$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}y_{ij}^2-2\bar{y}\sum_{i=1}^{r}\sum_{j=1}^{n_i}y_{ij}+N\bar{y}^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}y_{ij}^2-N\bar{y}^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}y_{ij}^2-T^2/N$$

$$SS_A=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\bar{y})^2=\sum_{i=1}^{r}n_i\bar{y}_{i.}^2+N\bar{y}^2-2\bar{y}\sum_{i=1}^{r}n_i\bar{y}_{i.}=\sum_{i=1}^{r}\frac{T_{i.}^2}{n_i}-\frac{T^2}{N}$$

8/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

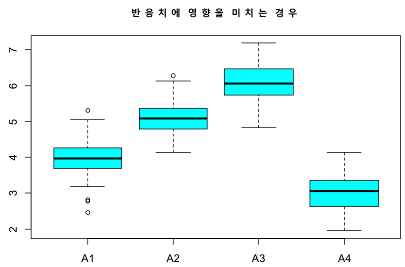
2. 单因素方差分析

2.3 假设检验

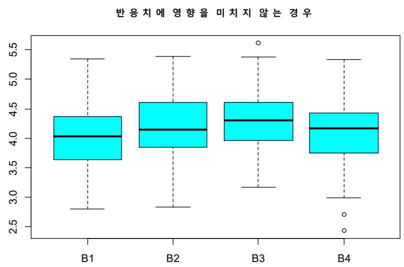
H_0: $\mu_1 = \mu_2 = \cdots = \mu_r$

H_1: 적어도 한 개의 μ_i가 나머지와 같지 않다.

$$\sigma_A^2 \equiv \frac{\sum_{i=1}^{r} a_i^2}{r} \Rightarrow \begin{cases} H_0: \sigma_A^2 = 0 \\ H_1: \sigma_A^2 > 0 \end{cases}$$



$$\frac{SS_A}{SS_E} = \frac{\text{수준간 변동}}{\text{수준내 변동}} \nearrow$$



$$\frac{SS_A}{SS_E} = \frac{\text{수준간 변동}}{\text{수준내 변동}} \searrow$$

2. 单因素方差分析

[表 13-2] 单因素方差分析表

来源	平方和	自由度	平方均值	检验统计量	否定值
处理	SS_A	$r-1$	$MS_A = \dfrac{SS_A}{r-1}$	$\dfrac{MS_A}{MS_E}$	$F_{1-\alpha;(r-1,N-r)}$
误差	SS_E	$N-r$	$MS_E = \dfrac{SS_E}{N-r}$		
合计	SS_T	$N-1$			

$$E(MS_E) = E\left(\frac{SS_E}{\phi_E}\right) = \sigma^2, \ \frac{SS_E}{\sigma^2} \sim \chi^2(\phi_E)$$

$$H_0 \Rightarrow E(MS_A) = E\left(\frac{SS_A}{\phi_A}\right) = \sigma^2, \ \frac{SS_A}{\sigma^2} \sim \chi^2(\phi_A) \Big| H_0$$

$$\Rightarrow F_0 = \frac{MS_A}{MS_E} = \frac{SS_A/\phi_A}{SS_E/\phi_E} = \frac{(SS_A/\sigma^2)/\phi_A}{(SS_E/\sigma^2)/\phi_E} \overset{H_0}{\sim} F(\phi_A, \phi_E)$$

$$\Rightarrow \text{Reject } H_0, \text{ if } F_0 > F_{1-\alpha;(\phi_A,\phi_E)}$$

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

2. 单因素方差分析

[例 13-1] 某一化学工序，温度设为100℃, 150℃, 200℃, 250℃四个水平，按随机随机顺序进行反复试验 → 收率。计算平方和&绘制方差分析表，并检验在显著性水平为5%的情况下，温度变化是否影响收率。

反复 \ 水平	A1	A2	A3	A4
1	79	81	86	76
2	83	89	91	81
3	88	91	93	82
4	78	84	90	79
5	75	86	89	
6		82		

$$\sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}^2 = 142{,}171$$

水平	A1	A2	A3	A4	总体
合计	403	513	449	318	1683
均值	80.60	85.50	89.80	79.50	84.15

$$SS_T = \sum_{i=1}^{r}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{N} = 142{,}171 - \frac{1{,}683^2}{20} = 546.55$$

$$SS_A = \sum_{i=1}^{r}\frac{T_{i.}^2}{n_i} - \frac{T^2}{N} = \left(\frac{403^2}{5}+\frac{513^2}{6}+\frac{449^2}{5}+\frac{318^2}{4}\right) - \frac{1{,}683^2}{20} = 320.05$$

$$SS_E = SS_T - SS_A = 546.55 - 320.05 = 226.5$$

11/50

2. 单因素方差分析

来源	平方和	自由度	平方均值	F_0	$F_{0.95;(3,16)}$
处理	320.05	3	106.6833	7.536	3.239
误差	226.50	16	14.1563		
合计	546.55	19			

$$\phi_T = N-1 = 20-1 = 19,$$
$$\phi_A = r-1 = 4-1 = 3,$$
$$\phi_E = N-r = 20-4 = 16$$

$$F_0 = \frac{MS_A}{MS_E} \doteq \frac{106.6833}{14.1563} \doteq 7.536$$

검정통계량 $F_0 \doteq 7.536$으로서 $F_{0.95;(3,16)} \doteq 3.239$보다 크므로

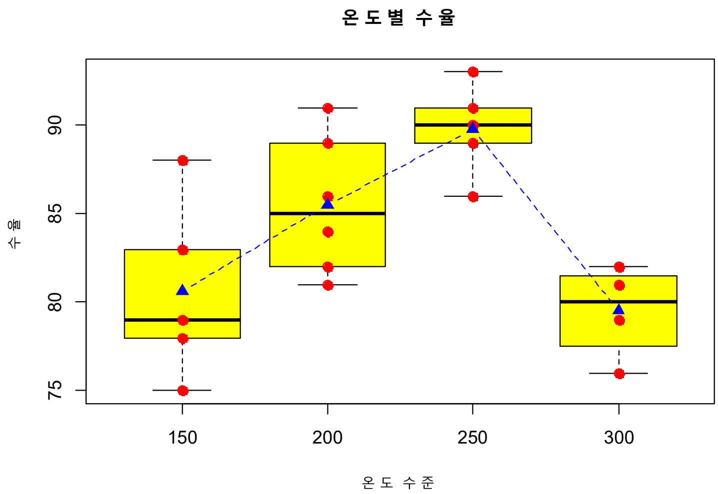
→ 在显著性水平为5%的情况下，原假设被拒绝

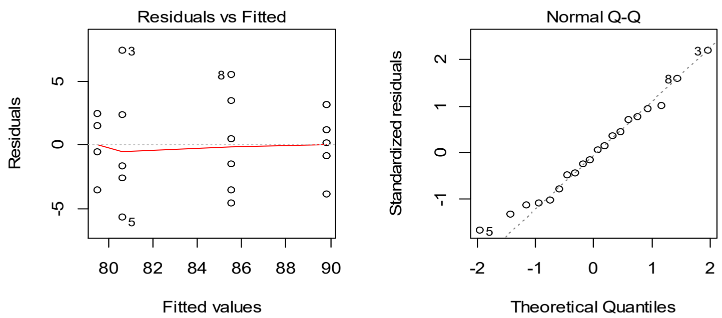
→ 有证据表明此工艺中四个水平的温度变化对收率有影响

12/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

☞ 收率到温度水平3为止一直增加，在水平4处减小
☞ 散布在温度水平1处最大，随之温度水平上，逐渐减小





☞ 残差按照不同水平类别比较均匀地分散，由于未发现特别异常的残差，因此同方差假设没有大问题。

☞ 残差之间没有特别显眼的规则或习性，由此可以看出独立性假设中没有大问题。

☞ 正态概率图没有较大脱离直线，因此可知正态性假设中没有大问题。

☞ 但是，3号、5号、8号点有些显眼，因此需对这些数据进行探讨。

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

2. 单因素方差分析

2.4 方差分析后的估计

(1) 不同水平下总体均值的估计 $E(\bar{y}_{i.}) = \mu_i \equiv \mu + a_i$

$$Var(\bar{y}_{i.}) = Var(\bar{\epsilon}_{i.}) = Var\left[\frac{1}{n_i}(\epsilon_{i1} + \epsilon_{i2} + \cdots + \epsilon_{in_i})\right] = \frac{1}{n_i^2}(n_i\sigma^2) = \frac{\sigma^2}{n_i}$$

모평균 μ_i에 대한 $100(1-\alpha)\%$의 신뢰구간

$$\left[\bar{y}_{i.} \pm t_{1-\alpha/2;\phi_E}\sqrt{MS_E / n_i}\right] \qquad E(MS_E) = \sigma^2$$

(2) 两个水平间总体均值差异的估计

$$E(\bar{y}_{i.} - \bar{y}_{i'.}) = \mu_i - \mu_{i'} \equiv a_i - a_{i'}$$

$$Var(\bar{y}_{i.} - \bar{y}_{i'.}) = Var(\bar{e}_{i.} - \bar{e}_{i'.}) = \frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_{i'}}$$

$\mu_i - \mu_{i'}$에 대한 $100(1-\alpha)\%$의 신뢰구간

$$\left[(\bar{y}_{i.} - \bar{y}_{i'.}) \pm t_{1-\alpha/2;\phi_E}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}\right]$$

15/50

2. 单因素方差分析

[例 13-2] 求前面[例 13-1]中不同水平温度下收率总体均值的95%置信区间，并绘图表示。

$$MS_E \doteq 14.1563, \ t_{0.975;(20-4)} \doteq 2.120$$

수준 1 : $[80.6 \pm 2.120\sqrt{14.1563/5}\,] \doteq [80.6 \pm 3.567] = [77.033,\ 84.167]$

수준 2 : $[85.5 \pm 2.120\sqrt{14.1563/6}\,] \doteq [85.5 \pm 3.256] = [82.244,\ 88.756]$

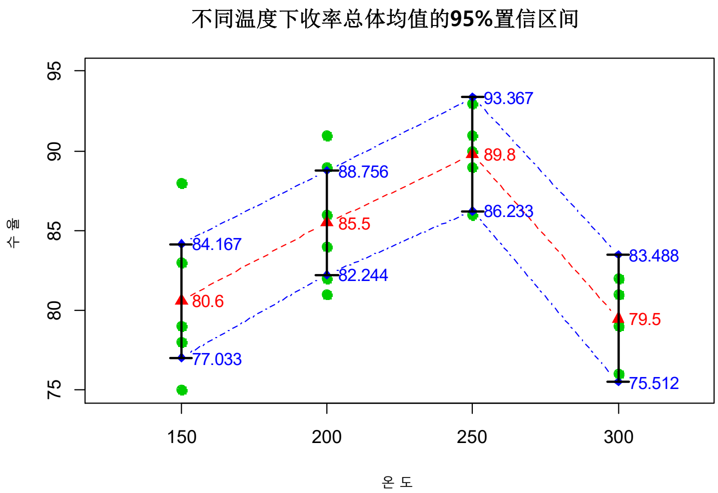
수준 3 : $[89.8 \pm 2.120\sqrt{14.1563/5}\,] \doteq [89.8 \pm 3.567] = [86.233,\ 93.367]$

수준 4 : $[79.5 \pm 2.120\sqrt{14.1563/4}\,] \doteq [79.5 \pm 3.988] = [75.512,\ 83.488]$

16/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

2. 单因素方差分析



2. 单因素方差分析

[例 13-3] 求前面[例 13-1]中两个水平温度之间收率的总体均值差异的95%置信区间，并检验总体均值差异是否显著。

$$MS_E \doteq 14.1563,\ t_{0.975;(20-4)} \doteq 2.120$$

$$\mu_1-\mu_2:\ [(80.6-85.5)\pm 2.120\sqrt{14.1563(\frac{1}{5}+\frac{1}{6})}\,] \doteq [-4.9\pm 4.830]$$

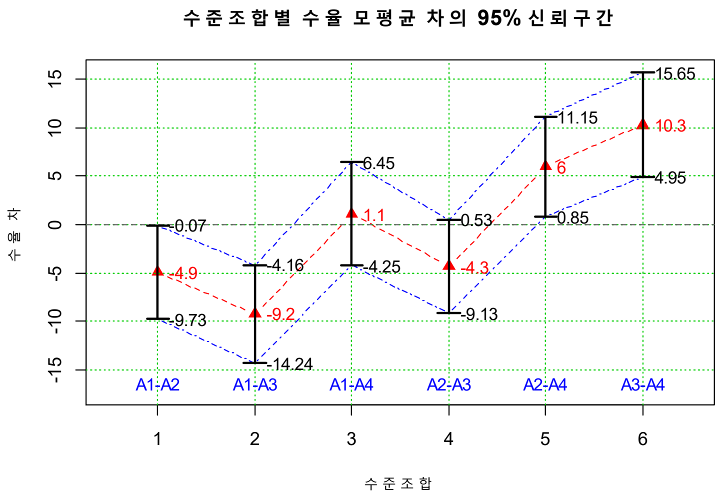
$$\mu_1-\mu_3:\ [(80.6-89.8)\pm 2.120\sqrt{14.1563(\frac{1}{5}+\frac{1}{5})}\,] \doteq [-9.2\pm 5.045]$$

$$\mu_1-\mu_4:\ [(80.6-79.5)\pm 2.120\sqrt{14.1563(\frac{1}{5}+\frac{1}{4})}\,] \doteq [1.1\pm 5.351]$$

$$\mu_2-\mu_3:\ [(85.5-89.8)\pm 2.120\sqrt{14.1563(\frac{1}{6}+\frac{1}{5})}\,] \doteq [-4.3\pm 4.830]$$

$$\mu_2-\mu_4:\ [(85.5-79.5)\pm 2.120\sqrt{14.1563(\frac{1}{6}+\frac{1}{4})}\,] \doteq [6.0\pm 5.149]$$

$$\mu_3-\mu_4:\ [(89.8-79.5)\pm 2.120\sqrt{14.1563(\frac{1}{5}+\frac{1}{4})}\,] \doteq [10.3\pm 5.351]$$



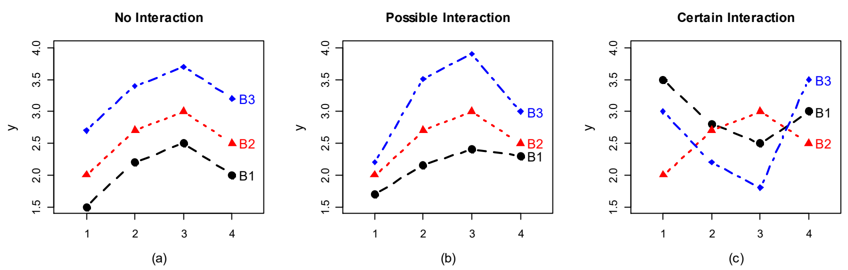
☞ 水平3处的收率与水平1、水平4呈现出显著的差异。
☞ 水平2处的收率与水平1、水平4呈现出显著的差异。
☞ 水平2和水平3处的收率在显著性水平为5%的情况下，没有表现出显著差异，但是预测在显著性水平为10%时会出现显著差异。

19/50

3. 双因素方差分析

13.3.1 交互作用

- 双因素方差分析(two-way ANOVA)：关于两个因素是否对响应值有影响的分析
- 多因素方差分析(multi-way ANOVA)：关于三个及三个以上因素是否对响应值有影响的分析
- 交互作用(interaction)：两个因素之间出现互相干涉的相互作用



20/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

3.2 数据的结构

$N = r \times s \times n$

요인 A는 r 수준, 요인 B는 s수준, 반복 n회

	A_1	A_2	$\cdots$	A_r	합계	평균
B_1	$y_{111}, y_{112}, \vdots, y_{11n}$ → $T_{11.}$, $\bar{y}_{11.}$	$y_{211}, y_{212}, \vdots, y_{21n}$ → $T_{21.}$, $\bar{y}_{21.}$	$\cdots$	$y_{r11}, y_{r12}, \vdots, y_{r1n}$ → $T_{r1.}$, $\bar{y}_{r1.}$	$T_{.1.}$	$\bar{y}_{.1.}$
B_2	$y_{121}, y_{122}, \vdots, y_{12n}$ → $T_{12.}$, $\bar{y}_{12.}$	$y_{221}, y_{222}, \vdots, y_{22n}$ → $T_{22.}$, $\bar{y}_{22.}$	$\cdots$	$y_{r21}, y_{r22}, \vdots, y_{r2n}$ → $T_{r2.}$, $\bar{y}_{r2.}$	$T_{.2.}$	$\bar{y}_{.2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_s	$y_{1s1}, y_{1s2}, \vdots, y_{1sn}$ → $T_{1s.}$, $\bar{y}_{1s.}$	$y_{2s1}, y_{2s2}, \vdots, y_{2sn}$ → $T_{2s.}$, $\bar{y}_{2s.}$	$\cdots$	$y_{rs1}, y_{rs2}, \vdots, y_{rsn}$ → $T_{rs.}$, $\bar{y}_{rs.}$	$T_{.s.}$	$\bar{y}_{.s.}$
합계	$T_{1..}$	$T_{2..}$	$\cdots$	$T_{r..}$	T	
평균	$\bar{y}_{1..}$	$\bar{y}_{2..}$	$\cdots$	$\bar{y}_{r..}$		$\bar{y}$

21/50

3. 双因素方差分析

3.2 数据的结构(接上)

$$T_{ij.} = \sum_{k=1}^{n} y_{ijk};\quad \bar{y}_{ij.} = \frac{\Sigma_{k=1}^{n} y_{ijk}}{n} = \frac{T_{ij.}}{n};\quad i=1,2,\ldots,r,\ j=1,2,\ldots,s$$

$$T_{i..} = \sum_{j=1}^{s}\sum_{k=1}^{n} y_{ijk};\quad \bar{y}_{i..} = \frac{\Sigma_{j=1}^{s}\Sigma_{k=1}^{n} y_{ijk}}{sn} = \frac{T_{i..}}{sn};\quad i=1,2\ldots,r$$

$$T_{.j.} = \sum_{i=1}^{r}\sum_{k=1}^{n} y_{ijk};\quad \bar{y}_{.j.} = \frac{\Sigma_{i=1}^{r}\Sigma_{k=1}^{n} y_{ijk}}{rn} = \frac{T_{.j.}}{rn};\quad j=1,2\ldots,s$$

$$T = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n} y_{ijk};\quad \bar{y} = \frac{\Sigma_{i=1}^{r}\Sigma_{j=1}^{s}\Sigma_{k=1}^{n} y_{ijk}}{rsn} = \frac{T}{N}$$

$$\bar{y} = \frac{\Sigma_{i=1}^{r}\Sigma_{j=1}^{s}\bar{y}_{ij}}{r \times s} = \frac{\Sigma_{i=1}^{r}\bar{y}_{i..}}{r} = \frac{\Sigma_{j=1}^{s}\bar{y}_{.j.}}{s}$$

데이터의 구조식 $y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk}$

요인 A의 주효과를 a_i, 요인 B의 주효과를 b_j,

요인 A와 B의 교호작용을 $(ab)_{ij}$

22/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

3.3 平方和的分解公式

① **총편차** $(y_{ijk}-\bar{y})$: 각각의 반응치(y_{ijk})와 전체 표본평균($\bar{y}$)과의 차이

② **요인 A의 수준간 편차** $(\bar{y}_{i..}-\bar{y})$: 요인 A의 수준별 평균($\bar{y}_{i..}$)과 전체평균($\bar{y}$)과의 차이

③ **요인 B의 수준간 편차** $(\bar{y}_{.j.}-\bar{y})$: 요인 B의 각 수준별 평균($\bar{y}_{.j.}$)과 전체 평균($\bar{y}$)과의 차이

④ **교호작용에 의한 편차** $(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y})$: 요인 A와 요인 B의 각 수준 조합별 평균($\bar{y}_{ij.}$)과 전체평균($\bar{y}$)과의 차이에서 요인 A의 수준간 편차와 요인 B의 수준간 편차를 제한 나머지 편차

⑤ **수준내 편차 (잔차,** $y_{ijk}-\bar{y}_{ij.}$): 각 반응치(y_{ijk})와 요인 A와 요인 B의 각 수준조합별 평균($\bar{y}_{ij.}$)과의 차이

23/50

3. 双因素方差分析

3.3 平方和的分解公式(接上)

$$
\begin{aligned}
y_{ijk}-\bar{y} &= (\bar{y}_{ij.}-\bar{y})+(y_{ijk}-\bar{y}_{ij.}) \\
&= (\bar{y}_{i..}-\bar{y})+(\bar{y}_{.j.}-\bar{y})+(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y})+(y_{ijk}-\bar{y}_{ij.})
\end{aligned}
$$

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

$$SS_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk}-\bar{y})^2=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}y_{ijk}{}^2-\frac{T^2}{N} \qquad \phi_T=N-1$$

$$SS_A=sn\sum_{i=1}^{r}(\bar{y}_{i..}-\bar{y})^2=sn\sum_{i=1}^{r}\bar{y}_{i..}^2-N\bar{y}^2=\sum_{i=1}^{r}\frac{T_{i..}^2}{sn}-\frac{T^2}{N} \qquad \phi_A=r-1$$

$$SS_B=rn\sum_{j=1}^{s}(\bar{y}_{.j.}-\bar{y})^2=rn\sum_{j=1}^{s}\bar{y}_{.j.}^2-N\bar{y}^2=\sum_{j=1}^{s}\frac{T_{.j.}^2}{rn}-\frac{T^2}{N} \qquad \phi_B=s-1$$

$$SS_{A\times B}=n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y})^2=SS_{AB}-SS_A-SS_B \qquad \begin{aligned}\phi_{A\times B}&=\phi_{AB}-\phi_A-\phi_A\\&=(r-1)(s-1)\end{aligned}$$

$$SS_{AB}=n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij.}-\bar{y})^2=\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{T_{ij.}^2}{n}-\frac{T^2}{N} \qquad \phi_{AB}=rs-1$$

$$SS_E=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij.})^2=SS_T-SS_{AB} \qquad \phi_E=N-rs=rs(n-1)$$

$$\phi_T=\phi_{AB}+\phi_E=(\phi_A+\phi_B+\phi_{A\times B})+\phi_E$$

24/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

3.3 平方和的分解公式(接上)

$$SS_{AB} = n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij.} - \bar{y})^2$$

$$= n\sum_{i=1}^{r}\sum_{j=1}^{s}\left[(\bar{y}_{i..} - \bar{y}) + (\bar{y}_{.j.} - \bar{y}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y})\right]^2$$

$$= sn\sum_{i=1}^{r}(\bar{y}_{i..} - \bar{y})^2 + rn\sum_{j=1}^{s}(\bar{y}_{.j.} - \bar{y})^2 + n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y})^2$$

$$= SS_A + SS_B + SS_{A\times B}$$

$$SS_T = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk} - \bar{y})^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}\left[(\bar{y}_{ij.} - \bar{y}) + (y_{ijk} - \bar{y}_{ij.})\right]^2$$

$$= n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij.} - \bar{y})^2 + \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2$$

$$= SS_{AB} + SS_E$$

3. 双因素方差分析

[定理 13-2] 双因素方差分析平方和的分解公式

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

总平方和: $SS_T = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk} - \bar{y})^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}{y_{ijk}}^2 - \frac{T^2}{N}$

因素A平方和: $SS_A = sn\sum_{i=1}^{r}(\bar{y}_{i..} - \bar{y})^2 = sn\sum_{i=1}^{r}\bar{y}_{i..}^2 - N\bar{y}^2 = \sum_{i=1}^{r}\frac{T_{i..}^2}{sn} - \frac{T^2}{N}$

因素B平方和: $SS_B = rn\sum_{j=1}^{s}(\bar{y}_{.j.} - \bar{y})^2 = rn\sum_{j=1}^{s}\bar{y}_{.j.}^2 - N\bar{y}^2 = \sum_{j=1}^{s}\frac{T_{.j.}^2}{rn} - \frac{T^2}{N}$

交互作用平方和: $SS_{A\times B} = n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y})^2 = SS_{AB} - SS_A - SS_B$

AB总平方和: $SS_{AB} = n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{y}_{ij.} - \bar{y})^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{T_{ij.}^2}{n} - \frac{T^2}{N}$

误差平方和: $SS_E = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2 = SS_T - SS_{AB}$

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

- 双因素方差分析 自由度的分解公式

$$\phi_T = \phi_{AB} + \phi_E = (\phi_A + \phi_B + \phi_{A\times B}) + \phi_E$$

$$\Sigma_i \Sigma_j \Sigma_k (y_{ijk} - \overline{y}) = 0 \Rightarrow \phi_T = rsn - 1 = N - 1$$

$$\Sigma_i (\overline{y}_{i..} - \overline{y}) = 0 \Rightarrow \phi_A = r - 1$$

$$\Sigma_j (\overline{y}_{.j.} - \overline{y}) = 0 \Rightarrow \phi_B = s - 1$$

$$\Sigma_i \Sigma_j (\overline{y}_{ij.} - \overline{y}) = 0 \Rightarrow \phi_{AB} = rs - 1$$

$$\Rightarrow \phi_{A\times B} = \phi_{AB} - \phi_A - \phi_B$$
$$= rs - 1 - (r-1) - (s-1) = (r-1)(s-1)$$

$$\Rightarrow \phi_E = \phi_T - \phi_{AB} = rsn - 1 - (rs - 1) = rs(n-1)$$

3. 双因素方差分析

3.4 假设检验

① 原假设: 因素A的水平变化对响应值没有影响。
对立假设: 因素A的水平变化对响应值有影响。

② 原假设: 因素B的水平变化对响应值没有影响。
对立假设: 因素B的水平变化对响应值有影响。

③ 原假设: 因素A和B之间没有交互作用。
对立假设: 因素A和B之间有交互作用。

$$(1)\begin{cases} H_0 : a_1 = \cdots = a_r = 0 \\ H_1 : \text{not } H_0 \end{cases} \quad \frac{SS_A}{\sigma^2} \overset{H_0}{\sim} \chi^2(\phi_A) \quad \Rightarrow F_0 \equiv \frac{MS_A}{MS_E} \sim F(\phi_A, \phi_E) \Big| H_0$$

$$(2)\begin{cases} H_0 : b_1 = \cdots = b_s = 0 \\ H_1 : \text{not } H_0 \end{cases} \quad \frac{SS_B}{\sigma^2} \overset{H_0}{\sim} \chi^2(\phi_B) \quad \Rightarrow F_0 \equiv \frac{MS_B}{MS_E} \sim F(\phi_B, \phi_E) \Big| H_0$$

$$(3)\begin{cases} H_0 : (ab)_{11} = \cdots = (ab)_{rs} = 0 \\ H_1 : \text{not } H_0 \end{cases} \quad \frac{SS_{A\times B}}{\sigma^2} \overset{H_0}{\sim} \chi^2(\phi_{A\times B}) \Rightarrow F_0 \equiv \frac{MS_{A\times B}}{MS_E} \sim F(\phi_{A\times B}, \phi_E) \Big| H_0$$

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

3.4 假设检验(接上)

来源	平方和	自由度	平方均值	检验统计量	否定值
A	SS_A	$r-1$	MS_A	MS_A/MS_E	$F_{1-\alpha;(\phi_A,\phi_E)}$
B	SS_B	$s-1$	MS_B	MS_B/MS_E	$F_{1-\alpha;(\phi_B,\phi_E)}$
$A\times B$	$SS_{A\times B}$	$(r-1)(s-1)$	$MS_{A\times B}$	$MS_{A\times B}/MS_E$	$F_{1-\alpha;(\phi_{A\times B},\phi_E)}$
E	SS_E	$rs(n-1)$	MS_E		
T	SS_T	$rsn-1$			

$$(1)\begin{cases}H_0: a_1=\cdots=a_r=0\\ H_1:\text{not } H_0\end{cases} \Rightarrow \text{Reject } H_0,\text{ if } \frac{MS_A}{MS_E} > F_{1-\alpha;(\phi_A,\phi_E)}$$

$$(2)\begin{cases}H_0: b_1=\cdots=b_s=0\\ H_1:\text{not } H_0\end{cases} \Rightarrow \text{Reject } H_0,\text{ if } \frac{MS_B}{MS_E} > F_{1-\alpha;(\phi_B,\phi_E)}$$

$$(3)\begin{cases}H_0: (ab)_{11}=\cdots=(ab)_{rs}=0\\ H_1:\text{not } H_0\end{cases} \Rightarrow \text{Reject } H_0,\text{ if } \frac{MS_{A\times B}}{MS_E} > F_{1-\alpha;(\phi_{A\times B},\phi_E)}$$

3. 双因素方差分析

[例 13-4] 某一化学工序，温度设为100℃, 150℃, 200℃, 250℃四个水平和压力设为气压1、气压2、气压3，按随机顺序进行两次反复试验得到如下数据。请计算平方和并填写方差分析表，以及检验在显著性水平为5%的情况下该工序中温度和压力的变化是否对收率有影响。

水平	100℃	150℃	200℃	250℃
气压1	76 79	79 81	87 91	79 82
气压2	81 79	84 86	91 94	85 84
气压3	83 85	89 88	88 86	77 76

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n} y_{ijk}^{2} = 168{,}910$$

$$\sum_{i=1}^{r}\sum_{j=1}^{s} T_{ij.}^{2} = 337{,}754$$

不同水平下因素的合计与均值

水平	100℃	150℃	200℃	250℃	合计	均值
气压1	155	160	178	161	654	81.75
气压2	160	170	185	169	684	85.5
气压3	168	177	174	153	672	84.0
合计	483	507	537	483	2010	
均值	80.5	84.5	89.5	80.5		83.75

MS Excel
=Sum( )
=Average( )
=Sumsq( )

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

$$SS_T = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n} y_{ijk}^2 - \frac{T^2}{N} = 168,910 - \frac{2,010^2}{24} = 572.5$$

$$SS_A = \sum_{i=1}^{r}\frac{T_{i..}^2}{sn} - \frac{T^2}{N} = \left(\frac{483^2+507^2+537^2+483^2}{6}\right) - \frac{2,010^2}{24} = 328.5$$

$$SS_B = \sum_{j=1}^{s}\frac{T_{.j.}^2}{rn} - \frac{T^2}{N} = \left(\frac{654^2+684^2+672^2}{8}\right) - \frac{2,010^2}{24} = 57.0$$

$$SS_{AB} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{T_{ij.}^2}{n} - \frac{T^2}{N} = \left(\frac{155^2+160^2+\cdots+153^2}{2}\right) - \frac{2,010^2}{24} = 539.5$$

$$SS_{A\times B} = SS_{AB} - SS_A - SS_B = 539.5 - 328.5 - 57.0 = 154.0$$

$$SS_E = SS_T - SS_{AB} = 572.5 - 539.5 = 33.0$$

$$\phi_T = N-1 = 24-1 = 23$$

$$\phi_A = r-1 = 4-1 = 3$$

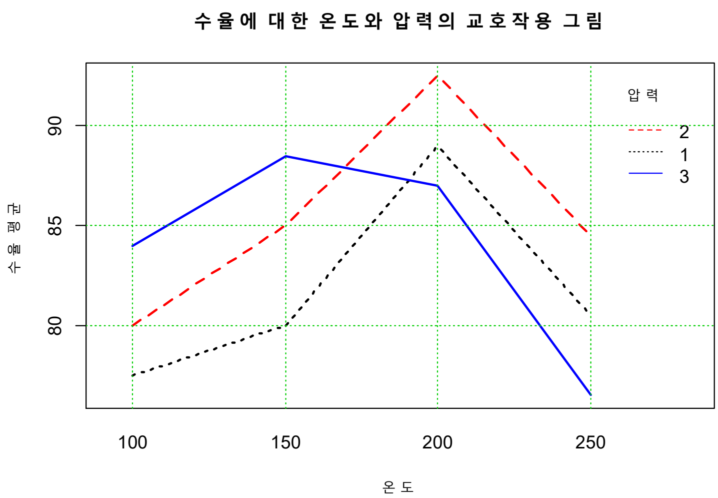
$$\phi_B = s-1 = 3-1 = 2$$

$$\phi_{A\times B} = 3\times 2 = 6$$

$$\phi_E = rs(n-1) = 4\times 3 = 12$$

来源	平方和	自由度	平方均值	F_0	$F_{0.95}$
A	328.5	3	109.5	39.818	3.490
B	57.0	2	28.5	10.364	3.885
A x B	154.0	6	25.667	9.333	2.996
E	33.0	12	2.75		
合计	572.5	23			

☞ 图中均值的连接线出现交叉，由此可知存在交互作用。
☞ 在温度水平3处，收率看起来很大，但是可以发现在压力大的情况下，温度水平2处的收率更大。



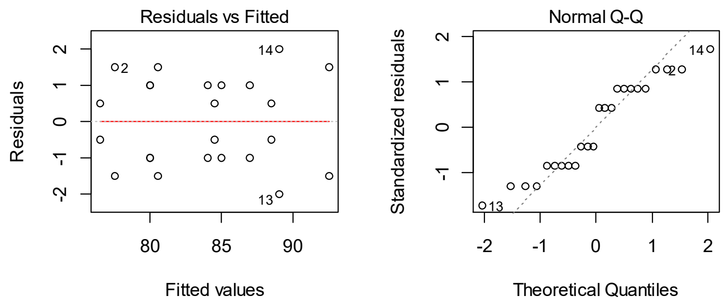
Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

```
# 诊断图表 [图 13-9]
win.graph(7,3.5); par(mfrow=c(1,2))
plot(an2, which=1:2)
```

☞ 不同水平下，残差散布没有出现大的差异，也未发现特别异常的残差，由此可以看出同方差假设没有大问题。

☞ 正态概率图没有较大脱离直线，由此可知正态性假设没有大问题。

☞ 但是，第2, 13, 14个点有些显眼，因此需对这些数据进行探讨。



3. 双因素方差分析

3.5 方差分析后的估计

(1) 不同水平下因素A的总体均值估计 $\mu(A_i) \equiv \mu + a_i$

$$\hat{\mu}(A_i) = \bar{y}_{i..} = \mu + a_i + \bar{\varepsilon}_{i..} \sim N(\mu + a_i, \frac{\sigma^2}{sn})$$

$$\Rightarrow \left[\bar{y}_{i..} \pm t_{1-\alpha/2;\phi_E} \sqrt{\frac{MS_E}{sn}} \right]$$

- 两个水平下因素A的总体均值差异估计 $\mu(A_i) - \mu(A_{i'}) = a_i - a_{i'}$

$$\hat{\mu}(A_i) - \hat{\mu}(A_{i'}) = \bar{y}_{i..} - \bar{y}_{i'..} = a_i - a_{i'} + \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{i'..} \sim N(a_i - a_{i'}, \frac{2\sigma^2}{sn})$$

$$\Rightarrow \left[(\bar{y}_{i..} - \bar{y}_{i'..}) \pm t_{1-\alpha/2;\phi_E} \sqrt{\frac{2MS_E}{sn}} \right]$$

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

(2) 不同水平下因素B的总体均值估计 $\mu(B_j) \equiv \mu + b_j$

$$\hat{\mu}(B_j) = \bar{y}_{.j.} = \mu + b_j + \bar{\varepsilon}_{.j.} \sim N(\mu + b_j, \frac{\sigma^2}{rn})$$

$$\Rightarrow \left[\bar{y}_{.j.} \pm t_{1-\alpha/2;\phi_E} \sqrt{\frac{MS_E}{rn}} \right]$$

- 两个水平下因素B的总体均值差异估计 $\mu(B_j) - \mu(B_{j'}) = b_i - b_{j'}$

$$\hat{\mu}(B_j) - \hat{\mu}(B_{j'}) = \bar{y}_{.j.} - \bar{y}_{.j'.} = b_j - b_{j'} + \bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{.j'.} \sim N(b_j - b_{j'}, \frac{2\sigma^2}{rn})$$

$$\Rightarrow \left[(\bar{y}_{.j.} - \bar{y}_{.j'.}) \pm t_{1-\alpha/2;\phi_E} \sqrt{\frac{2MS_E}{rn}} \right]$$

35/50

3. 双因素方差分析

(3) 不同水平组合下因素A, B的总体均值估计

수준조합 A_iB_j에서의 모평균의 $100(1-\alpha)\%$의 신뢰구간

①交互作用有意义的情况

$$\mu(A_iB_j) \equiv \mu + a_i + b_j + (ab)_{ij} \Rightarrow E[\bar{y}_{ij.}] = \mu(A_iB_j) \Rightarrow \hat{\mu}(A_iB_j) = \bar{y}_{ij.}$$

$$Var(\bar{y}_{ij.}) = \frac{\sigma^2}{n} \Rightarrow \left[\bar{y}_{ij.} \pm t_{1-\alpha/2;\phi_E} \sqrt{\frac{MS_E}{n}} \right]$$

②交互作用没有意义的情况(交互作用 A X B并入误差项的情况)

$$\mu(A_iB_j) \equiv \mu + a_i + b_j = (\mu + a_i) + (\mu + b_j) - \mu$$

$$\Rightarrow E[\bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}] = \mu + a_i + b_j \Rightarrow \hat{\mu}(A_iB_j) = \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}$$

$$Var(\bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}) = \sigma^2 \left(\frac{1}{sn} + \frac{1}{rn} - \frac{1}{rsn} \right) = \sigma^2 \frac{(r+s-1)}{N}$$

$$\Rightarrow \left[(\bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}) \pm t_{1-\alpha/2;\phi_E} \sqrt{MS_E' \frac{(r+s-1)}{N}} \right]$$

36/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

(3) 不同水平组合下因素A, B的总体均值估计(接上)

- Pooling: 没有意义的交互作用并入误差项内

$$SS_E' = SS_E + SS_{A\times B}$$
$$\phi_E' = \phi_E + \phi_{A\times B} = rs(n-1) + (r-1)(s-1)$$
$$MS_E' = \frac{SS_E + SS_{A\times B}}{rs(n-1)+(r-1)(s-1)}$$

$$Var(\bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}) = Var(\bar{y}_{i..}) + Var(\bar{y}_{.j.}) + Var(\bar{y})$$
$$+2Cov(\bar{y}_{i..}, \bar{y}_{.j.}) - 2Cov(\bar{y}_{i..}, \bar{y}) - 2Cov(\bar{y}_{.j.}, \bar{y})$$
$$= \sigma^2\left(\frac{1}{sn} + \frac{1}{rn} + \frac{1+2-2-2}{rsn}\right) = \sigma^2\left(\frac{1}{sn} + \frac{1}{rn} - \frac{1}{rsn}\right)$$

37/50

3. 双因素方差分析

[例 13-5] 求前面[例 13-4]中收率提高最大的温度和压力水平组合，并计算在该水平组合下收率总体均值的95%置信区间。

- 因素在不同水平组合下的均值

水平	100℃	150℃	200℃	250℃
气压1	77.5	80.0	89.0	80.5
气压2	80.0	85.0	92.5	84.5
气压3	84.0	88.5	87.0	76.5

收率提高最大的温度和压力水平组合为 A_3B_2

$MS_E = 2.75,\ t_{0.975;12} \approx 2.179$

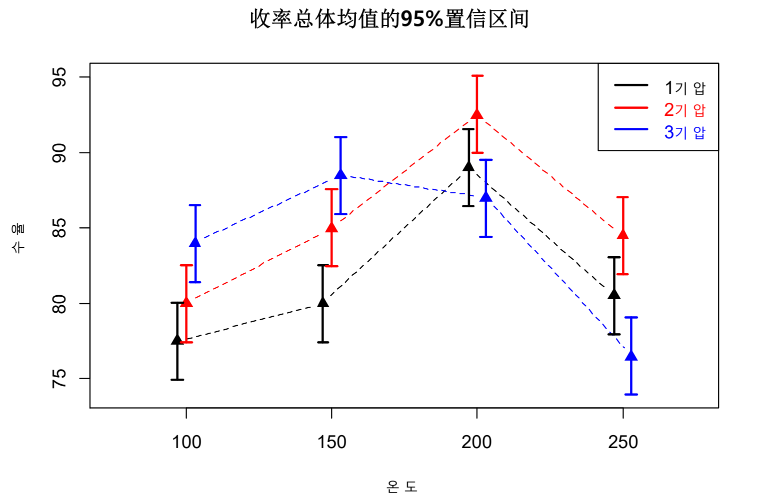
$\mu(A_3B_2)$: $[92.5 \pm 2.179\sqrt{2.75/2}\,] \doteq [92.5 \pm 2.555] = [89.945,\ 95.055]$

```
# 求不同水平组合的均值 ⇒ 使用tapply( )函数
ym <- tapply(收率, list(压力, 温度), mean); ym
   100  150  200  250
1 77.5 80.0 89.0 80.5
2 80.0 85.0 92.5 84.5
3 84.0 88.5 87.0 76.5
```

38/50

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析



3. 双因素方差分析

3.6 在不重复试验的情况下

$y_{ij} = \mu + a_i + b_j + (ab)_{ij} + e_{ij}$ → 无法求交互作用

交互作用和试验误差混淆(confound)

(1) 平方和计算

요인 A는 r 수준, 요인 B는 s수준 $N = r \times s$

	A_1	A_2	$\cdots$	A_r	合计	均值
B_1	y_{11}	y_{21}	$\cdots$	y_{r1}	$T_{.1}$	$\bar{y}_{.1}$
B_2	y_{12}	y_{22}	$\cdots$	y_{r2}	$T_{.2}$	$\bar{y}_{.2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_s	y_{1s}	y_{2s}		y_{rs}	$T_{.s}$	$\bar{y}_{.s}$
合计	$T_{1.}$	$T_{2.}$	$\cdots$	$T_{r.}$	T	
均值	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\cdots$	$\bar{y}_{r.}$		$\bar{y}$

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

$$T_{i.} = \sum_{j=1}^{s} y_{ij};\quad \bar{y}_{i.} = \frac{\Sigma_{j=1}^{s} y_{ij}}{s} = \frac{T_{i.}}{s};\quad i=1,2\ldots,r$$

$$T_{.j} = \sum_{i=1}^{r} y_{ij};\quad \bar{y}_{.j} = \frac{\Sigma_{i=1}^{r} y_{ij}}{r} = \frac{T_{.j}}{r};\quad j=1,2\ldots,s$$

$$T = \sum_{i=1}^{r}\sum_{j=1}^{s} y_{ij};\quad \bar{y} = \frac{\Sigma_{i=1}^{r}\Sigma_{j=1}^{s} y_{ij}}{rs} = \frac{T}{N}$$

$$\bar{y} = \frac{\Sigma_{i=1}^{r}\Sigma_{j=1}^{s} y_{ij}}{r\times s} = \frac{\Sigma_{i=1}^{r}\bar{y}_{i.}}{r} = \frac{\Sigma_{j=1}^{s}\bar{y}_{.j}}{s}$$

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

$$\Rightarrow SS_T = SS_A + SS_B + SS_E \quad (SS_E = SS_{A\times B})$$

3. 双因素方差分析

[定理 13-3] 双因素方差分析 平方和的分解公式 (不重复)

$$SS_T = SS_A + SS_B + SS_E \quad (SS_E = SS_{A\times B})$$

总平方和: $SS_T = \sum_{i=1}^{r}\sum_{j=1}^{s}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} y_{ij}^{\ 2} - \frac{T^2}{N}$

$\phi_T = N-1 = rs-1$

因素A平方和: $SS_A = s\sum_{i=1}^{r}(\bar{y}_{i.}-\bar{y})^2 = s\sum_{i=1}^{r}\bar{y}_{i.}^2 - N\bar{y}^2 = \sum_{i=1}^{r}\frac{T_{i.}^2}{s} - \frac{T^2}{N}$

$\phi_A = r-1$

因素B平方和: $SS_B = r\sum_{j=1}^{s}(\bar{y}_{.j}-\bar{y})^2 = r\sum_{j=1}^{s}\bar{y}_{.j}^2 - N\bar{y}^2 = \sum_{j=1}^{s}\frac{T_{.j}^2}{r} - \frac{T^2}{N}$

$\phi_B = s-1$

误差平方和: $SS_E = \sum_{i=1}^{r}\sum_{j=1}^{s}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y})^2 = SS_T - SS_A - SS_B$

$\phi_E = \phi_T - \phi_A - \phi_A = (r-1)(s-1)$

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

(2) 方差分析与假设检验

① 原假设：因素A的水平变化对响应值没有影响。

② 原假设：因素B的水平变化对响应值没有影响。

$H_0:\ a_1 = a_2 = \cdots = a_r = 0$ $\quad H_0:\ b_1 = b_2 = \cdots = b_s = 0$

$\Rightarrow F_0 \equiv \dfrac{MS_A}{MS_E} \sim F(\phi_A, \phi_E) \mid H_0$ $\quad \Rightarrow F_0 \equiv \dfrac{MS_B}{MS_E} \sim F(\phi_B, \phi_E) \mid H_0$

$\Rightarrow$ Reject H_0, if $\dfrac{MS_A}{MS_E} > F_{1-\alpha:(\phi_A,\phi_E)}$ $\quad \Rightarrow$ Reject H_0, if $\dfrac{MS_B}{MS_E} > F_{1-\alpha:(\phi_B,\phi_E)}$

来源	平方和	自由度	平方均值	检验统计量	否定值
A	SS_A	$r-1$	MS_A	MS_A/MS_E	$F_{1-\alpha;(\phi_A,\phi_E)}$
B	SS_B	$s-1$	MS_B	MS_B/MS_E	$F_{1-\alpha;(\phi_B,\phi_E)}$
E	SS_E	$(r-1)(s-1)$	MS_E		
T	SS_T	$rs-1$			

3. 双因素方差分析

[例 13-6] 某一化学工序，温度设为100℃, 150℃, 200℃, 250℃四个水平和压力设为气压1、气压2、气压3，按随机顺序各进行一次试验后得到了如下数据。求计算平方和 → 填写方差分析表，并检验在显著性水平为5%的情况下温度和压力的变化是否对收率有影响。

水平	100℃	150℃	200℃	250℃	合计	均值
气压1	77.5	80	89	80.5	327	81.75
气压2	80	85	92.5	84.5	342	85.5
气压3	84	88.5	87	76.5	336	84.0
合计	241.5	253.5	268.5	241.5	1005	
均值	80.5	84.5	89.5	80.5		83.75

$\sum_{i=1}^{r}\sum_{j=1}^{s} y_{ij}^{2} = 84,438.5$

$$SS_T = \sum_{i=1}^{r}\sum_{j=1}^{s} y_{ij}^2 - \frac{T^2}{N} = 84,438.5 - \frac{1,005^2}{12} = 269.75$$

$$SS_A = \sum_{i=1}^{r}\frac{T_{i.}^2}{s} - \frac{T^2}{N} = \left(\frac{241.5^2+253.5^2+268.5^2+241.5^2}{3}\right) - \frac{1,005^2}{12} = 164.25$$

$$SS_B = \sum_{j=1}^{s}\frac{T_{.j}^2}{r} - \frac{T^2}{N} = \left(\frac{327^2+342^2+336^2}{4}\right) - \frac{1,005^2}{12} = 28.5$$

$$SS_E = SS_T - SS_A - SS_B = 269.75 - 164.25 - 28.5 = 77.0$$

Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

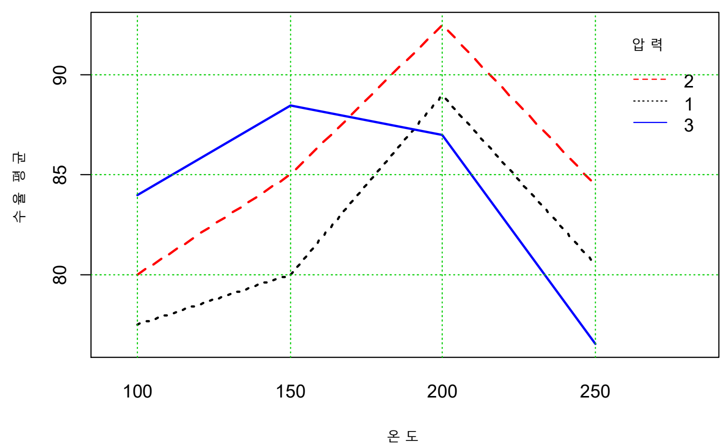
3. 双因素方差分析

来源	平方和	自由度	平方均值	F_0	$F_{0.95}$
A	164.25	3	54.75	4.266	4.757
B	28.5	2	14.25	1.110	5.143
E	77.0	6	12.833		
계	269.75	11			

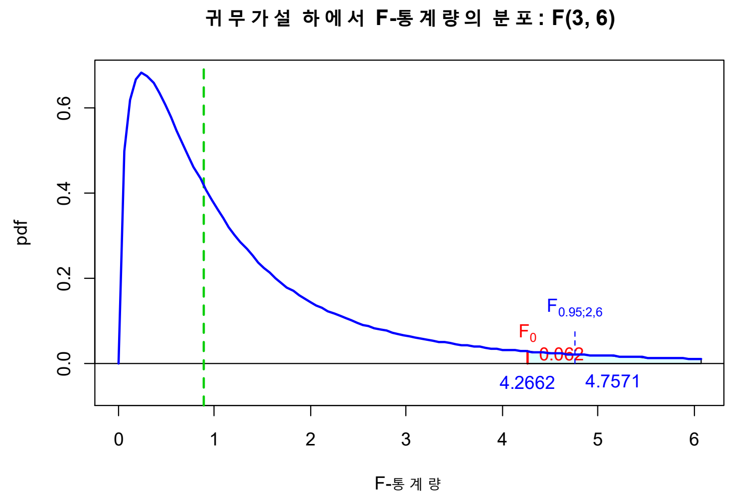
- ➢ 没有充足的证据表明该工序中4个水平的温度变化对收率有影响。
- ➢ 没有充足的证据表明该工序中3个水平的压力变化对收率有影响。
- ✓ 此例题是对[例 13-4]的重复数据进行求均值而人为制造的。
- ✓ 虽然[例 13-4]中因子A和B的效果都是显著的，但是在该例题中都表现为不显著。
- ✓ 其中最主要的原因是交互作用的效果被混淆为误差项中，导致误差平方和变大。
- ✓ 在预测交互作用显著的情况下，设计试验时需考虑重复进行。

☞ 图中均值的连接线出现交叉，由此可知存在交互作用。
☞ 在温度水平3处，收率看起来很大，但是可以发现在压力大的情况下，温度水平2处的收率更大。

수율에 대한 온도와 압력의 교호작용 그림

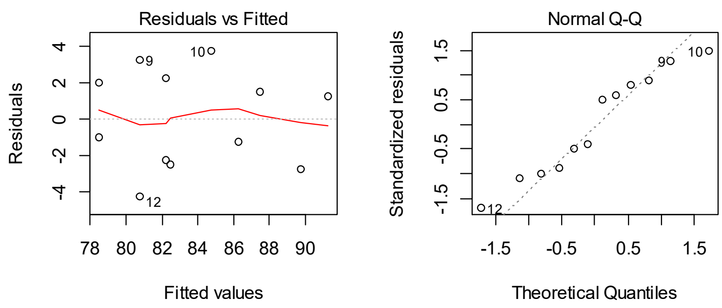


Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org



3. 双因素方差分析

# 诊断图表[图 13-12]
win.graph(7,3.5); par(mfrow=c(1,2))
plot(an2, which=1:2)
☞ 残差的范围相比[图 13-9]，增加了近两倍。
☞ 正态概率图没有大大脱离直线，由此可知正态性假设没有大问题。
☞ 但是，第9, 10, 12个点有些显眼，因此需要对这些数据进行探讨。



Kim, Dae Ho (Mokwon University)
mis@mokwon.ac.kr / mis@acm.org

3. 双因素方差分析

(3) 方差分析后的估计

① 요인 A의 수준 i에서의 모평균에 대한 $100(1-\alpha)\%$의 신뢰구간

$$\left[\bar{y}_{i.} \pm t_{1-\alpha/2;\phi_E}\sqrt{\frac{MS_E}{s}}\right]$$

② 요인 A의 수준 i와 i'에서의 두 모평균 차에 대한 $100(1-\alpha)\%$ 신뢰구간

$$\left[(\bar{y}_{i.}-\bar{y}_{i'.}) \pm t_{1-\alpha/2;\phi_E}\sqrt{\frac{2MS_E}{s}}\right]$$

③ 요인 B의 수준 j에서의 모평균에 대한 $100(1-\alpha)\%$ 신뢰구간

$$\left[\bar{y}_{.j} \pm t_{1-\alpha/2;\phi_E}\sqrt{\frac{MS_E}{r}}\right]$$

④ 요인 B의 수준 j와 j'에서의 두 모평균 차에 대한 $100(1-\alpha)\%$ 신뢰구간

$$\left[(\bar{y}_{.j}-\bar{y}_{.j'}) \pm t_{1-\alpha/2;\phi_E}\sqrt{\frac{2MS_E}{r}}\right]$$

⑤ 수준조합 A_iB_j에서의 모평균의 $100(1-\alpha)\%$의 신뢰구간

$$\left[(\bar{y}_{i.}+\bar{y}_{.j}-\bar{\bar{y}}) \pm t_{1-\alpha/2;\phi_E}\sqrt{MS_E\frac{(r+s-1)}{rs}}\right]$$

3. 双因素方差分析

[例 13-7] 求前面[例 13-6]中收率提高最大的温度和压力的水平组合，并计算该水平组合下收率总体均值的95%置信区间。

$$\max\{\bar{y}_{i.}\} = \bar{y}_{3.} = 89.5, \quad \max\{\bar{y}_{.j}\} = \bar{y}_{.2} = 85.5$$

온도와 압력의 수준조합은 A_3B_2

$$MS_E \doteq 12.833, \ t_{0.975;6} \doteq 2.447$$

$$\mu(A_3B_2) : [(89.5+85.5-83.75) \pm 2.447\sqrt{12.833(\frac{1}{3}+\frac{1}{4}-\frac{1}{12})}\,]$$
$$\doteq [91.25 \pm 6.198] = [85.052, 97.448]$$

☞ 由于未进行反复试验，与**[例 13-5]**相比，置信区间的误差增大了两倍以上。