7. 分散分析

（学習内容）

・分散分析の考え方を理解する

・分散分析の手順を学ぶ

7.1 分散分析の考え方

（問題）

ある化学製品の製造において、強度を高めるために 4 種類の反応温度で製造を行い、反応温度と強度の間に関係があるか調べるため次のデータを得た。これより反応温度によって強度に差があると言えるか。またどの反応温度が最適か。　　→分散分析（analysis of variance: ANOVA）

	繰り返し			計 $T_{i.}$	平均 $\bar{x}_{i.}$
A_1	2.0	2.2	2.1	6.3	2.10
A_2	2.4	2.8	2.6	7.8	2.60
A_3	2.2	2.4	2.6	7.2	2.40
A_4	2.0	1.7	2.0	5.7	1.90
				総合計	全平均
				$T = 27.0$	$\bar{\bar{x}} = 2.25$

(1) データの構造式

・特性値：対象となる特性　→強度

・因子(factor)：特性値を調べるために実験で取り上げる条件　→反応温度

・水準(level)：因子の条件変化　→A_1 A_2 A_3 A_4

$x_{ij}=\mu_i+\varepsilon_{ij}=\mu+\alpha_i+\varepsilon_{ij}$

μ_i：A_i 水準での母平均

ε_{ij}：実験誤差、ε_{ij} は互いに独立に $N(0, \sigma^2)$ に従う

μ ：一般平均（全ての母平均の平均）

α_i：A_i の母平均と一般平均の差、$\sum_{i=1}^{a}\alpha_i = 0$ （a:A の水準数）

(2) ばらつきの分解

$$x_{ij} - \bar{\bar{x}} = \underbrace{(\bar{x}_{i.} - \bar{\bar{x}})}_{\text{反応温度によるばらつき}} + \underbrace{(x_{ij} - \bar{x}_{i.})}_{\text{それ以外の誤差によるばらつき}}$$

（n:データ総数、a:A の水準数、r:繰返し数）

総平方和　$S_T = \sum_{i=1}^{a}\sum_{j=1}^{r}(x_{ij} - \bar{\bar{x}})^2$

A 間平方和　$S_A = \sum_{i=1}^{a}\sum_{j=1}^{r}(\bar{x}_{i.} - \bar{\bar{x}})^2 = r\sum_{i=1}^{a}(\bar{x}_{i.} - \bar{\bar{x}})^2$

誤差平方和　$S_E = \sum_{i=1}^{a}\sum_{j=1}^{r}(x_{ij} - \bar{x}_{i\cdot})^2$

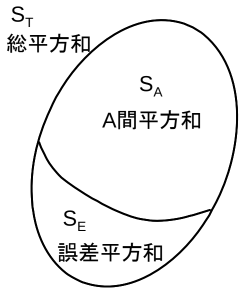
$S_T=S_A+S_E$

$$S_T = \sum_{i=1}^{a}\sum_{j=1}^{r}(x_{ij} - \bar{\bar{x}})^2 = \sum\sum\{(x_{ij} - \bar{x}_{i\cdot}) + (\bar{x}_{i\cdot} - \bar{\bar{x}})\}^2$$

$$= \sum\sum(x_{ij} - \bar{x}_{i\cdot})^2 + \sum\sum(\bar{x}_{i\cdot} - \bar{\bar{x}})^2 + 2\sum\sum(x_{ij} - \bar{x}_{i\cdot})(\bar{x}_{i\cdot} - \bar{\bar{x}})$$

ここで

$$2\sum_{i=1}^{a}\sum_{j=1}^{y}(x_{ij} - \bar{x}_{i\cdot})(\bar{x}_{i\cdot} - \bar{\bar{x}}) = 2\sum_{i=1}^{a}(\bar{x}_{i\cdot} - \bar{\bar{x}})(\sum_{j=1}^{r}(x_{ij} - \bar{x}_{i\cdot})) = 2\sum_{i=1}^{a}(\bar{x}_{i\cdot} - \bar{\bar{x}}) \times 0$$



“A によるばらつき”が”誤差によるばらつき”より大きければ、A によるばらつきを認める

平均平方（分散）で比較する

・自由度：平方和を構成する独立な成分の数

・平均平方(mean square)：平方和を自由度で割ったもの、分散

$V_A=S_A/\phi_A$　　　$V_E=S_E/\phi_E$

S_Tの自由度　ϕ_T=n−1

S_Aの自由度　ϕ_A=a−1

S_Eの自由度　ϕ_E=a(r−1)

$\phi_T=\phi_A+\phi_E$

$F_0=V_A/V_E$　を自由度（ϕ_A, ϕ_E）の F 分布に従って検定する　←（母分散の比に関する検定）

7.2 一元配置分散分析

一元配置：一つの因子についてそれが特性値に影響を与えているかどうか調べる実験解析方法

A 因子以外の条件の変動は全て実験誤差とみなす

完全無作為法によりランダムにサンプリングする

(1) 分散分析の手順

1) 平方和を求める

総平方和　$S_T = \sum_{i=1}^{4}\sum_{j=1}^{3}(x_{ij} - \bar{\bar{x}})^2$ =1.110

A 間平方和　$S_A = \sum_{i=1}^{4}\sum_{j=1}^{3}(\bar{x}_{i\cdot} - \bar{\bar{x}})^2 = 3\sum_{i=1}^{4}(\bar{x}_{i\cdot} - \bar{\bar{x}})^2$ =0.870

誤差平方和　$S_E = \sum_{i=1}^{4}\sum_{j=1}^{3}(x_{ij} - \bar{x}_{i\cdot})^2$ =0.240

2) 自由度を求める

S_Tの自由度　ϕ_T=n−1　=12−1=11

S_Aの自由度　ϕ_A=a−1　=4−1=3

S_Eの自由度　$\phi_E=\phi_T-\phi_A$=a(r−1)　=4*2=8

3）分散分析表をまとめる

要因	平方和 S	自由度 ϕ	平均平方 V	F_0
A E	S_A　0.87 S_E　0.24	ϕ_A　3 ϕ_E　8	$V_A=S_A/\phi_A$　0.29 $V_E=S_E/\phi_E$　0.030	V_A/V_E　9.7 **
計	S_T　1.11	ϕ_T　11		

帰無仮説　H_0:A の水準による分散と誤差による分散は等しい　$\sigma_A=\sigma_E$

対立仮説　H_1:A の水準による分散の方が誤差による分散より大きい　$\sigma_A \geqq \sigma_E$　（片側検定）

有意水準　α=0.05, 0.01

4）$F_0 \geqq F(\phi_A, \phi_E; \alpha)$ なら有意水準 α で有意

F_0=9.7 ≧ F(3,8;0.05)=4.07

F_0=9.7 ≧ F(3,8;0.01)=7.59

有意水準 1%で有意　（分散分析表に**を付ける）

(2) 分散分析後の推定

最適水準　$\bar{x}_{i.}$ を比較する　A_2 水準

強度の推定　←（母平均に関する推定）

1）点推定　$\mu_i = \bar{x}_{i.}$　=2.60

2）区間推定

$$-t(\phi_E, \alpha) \le (\bar{x}_{i.} - \mu_i)/\sqrt{V_E / r} \le t(\phi_E, \alpha)$$

$$\bar{x}_{i.} - t(\phi_E, \alpha)\sqrt{V_E / r} \le \mu_i \le \bar{x}_{i.} + t(\phi_E, \alpha)\sqrt{V_E / r}$$

$$2.60 - t(8, 0.05)\sqrt{0.030/3} \le \mu_i \le 2.60 + t(8, 0.05)\sqrt{0.030/3}$$

2.37≦ μ_i ≦2.83

（$x_{i1}, x_{i2}, \cdots, x_{ir}$ だけから ϕ=r−1 として計算した分散 V を使うより、ϕ_E, V_E を使う方が精度が高い）

（課題）

次のデータについて分散分析を行え

	繰り返し			
A_1	1.0	2.0	2.0	1.0
A_2	1.0	1.0	2.0	0.0
A_3	2.0	3.0	4.0	3.0