7.　実験計画法

7.1. 実験の方法（フィッシャーの三原則）

実験計画法では管理された因子（要因 factor）水準の下での実験データから因子水準の差の推定や検定を行う。実験に際しては、(i) 反復の原則、(ii) 無作為化の原則、(iii) 局所管理の原則が重要で、これをフィッシャーの三原則という。(i) は推定値の精度向上と誤差評価に、また (ii) の無作為化は実験順序から来る条件差や評価者の慣れ等の誤差を偶然誤差に転化させるために必要である。(iii) では実験ブロック内での管理と実験順序の無作為化を行う。例えば、一日にできる実験が機材や実験結果を得るまでの時間の関係から、限定される場合がある。このような場合に実験を単調に行えば、その実験システムから結果に偏りが生じる。いま、一日に 3 回しか実験できない場合を考える。因子水準を A_i $(i = 1, 2, 3)$ とするとき、各実験日にはこれらの水準を無作為に実験中に割り付けなければならない。表 7.1 はブロック内で３水準を無作為順序に割り付けた結果である。この場合のブロックは実験日を意味し、このような実験を乱塊法 (randomized block design) という。ブロックとしては、測定者、装置および試験田等が考えられる。

表 7.1. ブロック内処理順序
の無作為割付

1 日目	A_2	A_1	A_3
2 日目	A_1	A_3	A_2
3 日目	A_2	A_3	A_1

観測データは，日常的に多くの因子とその水準が存在し，これらの影響を受けていると考えられる。本節の分散分析法（analysis of variance）は，母集団も含めて多くの因子やその水準の効果，さらにそれらの組合せ効果を推測するもので，因子効果の分析に有用な方法となる。たとえば，稲作において，種子の種類以外に，施肥の量・方法・日照・灌水など多くの因子がある。現実に管理された水田でも，これらの要因効果の推定や検定は大きな課題となる。そのほか，工場製品の品質，薬剤の臨床効果，マスメディアによる販売効果など因子効果を分析する必要性は大きい。

[例 7.1]　貯蔵された５種類の食品があり，それらの含水量を測定し、表 7.1 **の資料**を得た。これらの資料から，各食品の含水量に有意な差があるか否かの検討をしたい。この場合，食品の種類という１つの因子の５つの水準（level）A_1，A_2，A_3，A_4，A_5 について，それぞれ４回測定し，表の数値を得ている。この場合、２０回の測定を無作為順序で行う。このような実験を一元配置実験（one-way layout experiment）という。

表 7.2. 保存食品と水分量 (%)

食品	水分量				平均水分量
A_1	10	4	6	4	6.0
$\boldsymbol{A}_2$	12	10	13	7	10.5
$\boldsymbol{A}_3$	9	4	4	5	5.5
$\boldsymbol{A}_4$	14	9	10	11	11.0
$\boldsymbol{A}_5$	10	6	8	4	7.0

この例のような推測には，着目する要因について整理された形式に対応して分析法がある。この際，着目する要因は，その性格によって，　ブロック要因（block factor），　制御要因（controllable factor）の 2 種に大別される。ここに，ブロック要因とは制御不能な，たとえば特定の実験日・施設・個人などの変量的要因であり，制御要因とは選択可能でわれわれが主体的に実現できる要因，たとえば温度・圧力・薬品濃度などの因子を意味している。したがって，この分類はつねに固定化されたものではなく，分析の目的や課題により設定される環境・立場を反映して分類される。すなわち，個人という因子も，特定の人々に関しての効果を論ずる場合には母数要因であり，多くの人々を対象として薬効を論ずる場合には変量的ブロック要因である。そのほか，とくに着目する要因以外にも数多くの要因が考えられるが，これらは残差要因（residual factor）とよび，データ中の誤差とともに無作為化（randomized）して割り付けられると仮定し、残差誤差または単に誤差とよぶ。

データが整理された形式を，制御因子の数によって，一因子配置と多因子配置に分類してよぶ。それは分散分析の実験を行う立場では無作為化する因子に重要性をおくためである。ここでは，一因子実験として一元配置法と乱塊法について述べる。

7.2　一元配置法

いま，ある 1 つの制御可能な要因 A の効果，すなわち A_1, A_2, $\cdots$, A_k の K 個の水準の効果に関して，他のすべての要因の条件を固定して考える。この際，実験データにおける誤差 e は互いに独立に正規分布 $N(0,\ \sigma^2)$ とする。ここでは要因 A の各水準の効果をそれぞれの平均 μ_1, μ_2, $\cdots$, μ_K（ただし，未知である）とし，第 i 水準に対する n_i 個の無作為標本を，

$$\{X_{i1},\ X_{i2},\ \cdots,\ X_{in_i}\} \qquad (i = 1,\ 2, \cdots, K)$$

するとき，次の線形模型を考える。

$$X_{ij} = \mu_i + e_{ij} \qquad (j = 1, 2, \cdots, n_i; i = 1, 2, \cdots, K) \tag{7.1}$$

ここに，$\{e_{ij}\}$ は実験における互いに独立な誤差を示す。このとき，仮説

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

検定し，この仮説が棄却されれば k 個の水準が実験結果に与える効果は存在すると判断される。いま，

$$n=\sum_{i=1}^{k} n_i, \quad \mu=\frac{1}{n}\sum_{i=1}^{k} n_i\mu_i,$$

$\alpha_i=\mu_i-\mu \qquad (i=1,\ 2,\ \cdots,k)$

とおいて，次式を仮定する。

$$\sum_{i=1}^{k} n_i\alpha_i=0. \tag{7.2}$$

このとき，線形模型（式（7.1)）は次のように示される。

$$X_{ij}=\mu+\alpha_i+e_{ij} \qquad (i=1,\ 2,\ \cdots,\ n_i\ ;\ i=1,2,\cdots,k)$$

上式で α_i は主効果（main effect) とよばれ，式（7.2）を満足するから，仮説 H_0 は次のように書き改められる。

$$H_0:\alpha_1=\alpha_2=\cdots=\alpha_k=0$$

X_{ij} は第 i 水準における第 j 観測値を表し，これらは表 7.3 に示される。

表 7.3. 一元配置法のデータ

水準	観測値				平均
A_1	X_{11}	X_{12}	$\cdots$	X_{1n_1}	$\overline{X}_{1\cdot}$
A_2	X_{21}	X_{22}	$\cdots$	X_{2n_2}	$\overline{X}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_K	X_{K1}	X_{K2}	$\cdots$	X_{Kn_K}	$\overline{X}_{K\cdot}$

この表における $\overline{X}_{i\cdot}$ は第 i 水準における標本平均を示す。すなわち，

$\overline{X}_{i\cdot}=\frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$

である。また，

$$\overline{X}_{\cdot\cdot}=\tfrac{1}{K}\sum_{i=1}^{K} n_i\overline{X}_{i\cdot}=\tfrac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{n_i} X_{ij}$$

と表記し，標本全体の平均を示す。ここに，$n=\sum_{i=1}^{K} n_i$ である。このとき，

$$E\left(\overline{X}_{i\cdot}\right)=\mu+\alpha_i, \quad E\left(\overline{X}_{\cdot\cdot}\right)=\mu$$

が成立する。この場合，主効果の実験結果への影響・効果の有無を検定するために，全変動（total sum of squared variation）

$$S_T=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{\cdot\cdot}\right)^2$$

の平方和分解を考える。

$$X_{ij}-\overline{X}_{\cdot\cdot}=\left(X_{ij}-\overline{X}_{i\cdot}\right)+\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)$$

であり，

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i\cdot}\right)\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)=\sum_{i=1}^{K}\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i\cdot}\right)$$
$$=\sum_{i=1}^{K}\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)\times 0=0$$

が成立するから，

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{\cdot\cdot}\right)^2=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)^2+\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i\cdot}\right)^2$$
$$=\sum_{i=1}^{K}n_i\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)^2+\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i\cdot}\right)^2 \tag{7.3}$$

が誘導される。ここで，

$$S_A=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)^2=\sum_{i=1}^{K}n_i\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)^2,$$
$$S_E=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i\cdot}\right)^2$$

と示せば，式（7.3）は次のような平方和の分解式が得られる。

$$S_T=S_A+S_E$$

上述の議論で，3 つの平方和が導けた。全変動 S_T は観測値 X_{ik} の標本平均 $\overline{X}$ からの残差平方和であり，標本全体の散布度の一指標である。S_A は各水準における標本平均と標本全体の平均との残差平方和で，各水準における平均 μ_i の間の差がこの統計量に反映される。この和を級間残差平方和，または級間変動という。S_E は各水準 i 内の観測値と水準 i における標本平均との差の平方和で，水準 i 内の観測誤差がこの統計量に反映される。これを，級内残差平方和，または級内変動という。

ここで，3 つの統計量 S_T，S_A，S_E の統計量の分布について述べる。まず，S_E の分布を考える．次の平方和

$$\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_{i.}\right)^2$$

は ${\chi^2}_{n_i-1}$ 分布に従い，かつ $\overline{X}_i$. と独立である。また，各水準間で標本が独立であり，χ^2 分布の再生性を考慮に入れて，

$$\frac{S_E}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\overline{X}_{i\cdot}-\overline{X}_{\cdot\cdot}\right)^2$$

が ${\chi^2}_{n-k}$ 分布することが知られる。このとき，

$$E\left(S_E\right)=(n-K)\sigma^2$$

で，さらに，$\left\{\overline{X}_1,\cdots,\overline{X}_k\right\}$ と S_E は互いに独立である。

次に，統計量 S_A の分布を考える。これは $\left\{\overline{X}_1,\ \cdots,\ \overline{X}_k\right\}$ の関数であり，S_E と互いに独立である。また，

$$S_A=\sum_{i=1}^{K}n_i\left\{\left(\overline{X}_{i\cdot}-\mu\right)-\left(\overline{X}_{\cdot\cdot}-\mu\right)\right\}^2$$
$$=\sum_{i=1}^{K}n_i\left(\overline{X}_{i\cdot}-\mu\right)^2-n\left(\overline{X}_{\cdot\cdot}-\mu\right)^2$$

となり、

$$\frac{S_A}{\sigma^2}=\sum_{i=1}^{K}\left(\frac{\overline{X}_{i\cdot}-\mu}{\sigma/\sqrt{n_i}}\right)^2-\left(\frac{\overline{X}_{\cdot\cdot}-\mu}{\sigma/\sqrt{n_i}}\right)^2$$

を得る。仮説 H_0 が成立するとき，

$$E\left(\overline{X}_{i\cdot}\right)=E\left(\overline{X}_{\cdot\cdot}\right)=\mu$$

であるから,

$$\frac{\overline{X}_{i\cdot}-\mu}{\sigma/\sqrt{n_i}} \qquad と \qquad \frac{\overline{X}_{\cdot\cdot}-\mu}{\sigma/\sqrt{n_i}}$$

はともに正規分布 $N\left(0,1\right)$ する。よって,仮説 H_0 の下で S_A/σ^2 は ${\chi^2}_{k-1}$ 分布に従い,かつ $\overline{X}_{\cdot\cdot}$ と互いに独立であることがわかる。このことから,仮説 H_0 の下で,

$$E\left(S_A\right)=\left(K-1\right)\sigma^2$$

である。また,

$$S_A=\sum_{i=1}^{K} n_i\left(\overline{X}^2_{i\cdot}-2\overline{X}_{i\cdot}\overline{X}_{\cdot\cdot}+\overline{X}^2_{\cdot\cdot}\right)=\sum_{i=1}^{K} n_i\overline{X}^2_{i\cdot}-n\overline{X}^2_{\cdot\cdot}$$

であり,一般に,

$$E\left(\overline{X}^2_{i\cdot}\right)=Var\left(\overline{X}^2_{i\cdot}\right)+E\left(\overline{X}_{i\cdot}\right)^2=\frac{\sigma^2}{n_i}+\mu_i^2,$$

$$E\left(\overline{X}^2_{\cdot\cdot}\right)=Var\left(\overline{X}^2_{\cdot\cdot}\right)+E\left(\overline{X}_{\cdot\cdot}\right)^2=\frac{\sigma^2}{n}+\mu^2$$

が成立する。このことから,

$$E\left(S_A\right)=\left(K-1\right)\sigma^2+\sum_{i=1}^{K} n_i\alpha_i^2 \qquad (7.4)$$

を得る。したがって,$\alpha_1,\alpha_2,\cdots,\alpha_k$ の中で少なくとも 1 つが 0 に等しくなければ,$E(S_A)>(k-1)\sigma^2$ である。ここで統計量

$$F_0=\frac{S_A/(k-1)}{S_E/(n-K)} \qquad (7.5)$$

を考える。この統計量は級間残差平方の平均 $V_A=S_A/(k-1)$ と級内残差平方の平均 $V_E=S_E/(n-k)$ の比であり,V_E は σ^2 の一致かつ不偏推定量である。H_0 が真であれば,V_A は σ^2 の一致かつ不偏推定量である。このとき,F_0 は自由度対 $(k-1,\ n-k)$ の F 分布 $F(k-1,\ n-k)$ に従う。また,H_0 が不成立のときは,S_A は式 (7.4) から,

$$\sigma^2+\sum_{i=1}^{K} n_i\alpha_i^2/(K-1)$$

の一致かつ不偏推定量であり,この統計量は自由度対 $(k-1,\ n-k)$ の非心 F 分布に従い,同じ自由度対の F 分布に対し,大きな値をとる傾向をもつことが知られる。そこで,F_0 を理論的な同じ自由度の F 分布と比較して,その出現の様子を検討する。もし,実現値 F_0 の値が著しく大きな値として得られれば,データに対して仮説 H_0 が成立していることは認め難い,他方,F 分布から,F_0 が得られたものと考えられる値であれば仮説 H_0 は採択される。いま,有意水準を α としたとき,仮説 H_0 の検定は次の手順で示される。

（1）　$P(F \geqq F(K-1, n-K, \alpha) = \alpha$ となる定数 $F(k-1, n-k, \alpha)$ を求める。

（2）　式 (7.5) の統計量 F_0 の実現値を求め，その値が $F(k-1, n-k, \alpha)$ に等しいか，または大きいとき仮説 H_0 を棄却し，その他の場合 H_0 を採択する。

分散分析を行う場合の便宜上のために，分散分析表（table of analysis of variances,table of ANOVA：表 7.4）をあげておく。

表 7.4. 一元配置実験の分散分析表

要因	平方和	自由度	不偏分散	F
級間変動	$S_A = \sum_{i=1}^{K} n_i \left(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot}\right)^2$	$K-1$	$V_A = S_A/(K-1)$	V_A/V_E
級内変動	$S_E = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_{i\cdot}\right)^2$	$n-K$	$V_E = S_E/(n-K)$	
全変動	$S_T = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_{\cdot\cdot}\right)^2$	$n-1$		

実際に一元配置法を応用するときは，S_A と S_T を直接計算し，

$$S_T = S_A + S_E$$

を利用して S_E を導出する。表にも示しているように，自由度についても平方和と同様な分解ができる。すなわち，

$$n-1 = (k-1) + (n-k)$$

である。この関係式は分散分析において重要な関係式の 1 つである。分散分析の具体的計算は次の手順で行う。

（1）　$T_i = \sum_{j=1}^{n_i} X_{ij} (i = 1,\ 2,\ \cdots,\ k),\ T = T_1 + \cdots + T_k$ の計算をする。

（2）　$C = \dfrac{T^2}{n}$ を算出する。

（3）　$S_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i} {X_{ij}}^2 - C$ を算出する。

（4）　$S_A = \sum_{i=1}^{k} \dfrac{{T_i}^2}{n_i} - C$ を算出する。

（5）　$S_E = S_T - S_A$ を算出する。

例 7.1 において，この一連の計算を行えば次のようになる。

（1）　$T_1 = 24,\quad T_2 = 42,\quad T3 = 22,\quad T_4 = 44,\quad T_5 = 28,\quad T = 160,$

（2）　$C = 160^2/20 = 1280,$

（3）　$S_T = 10^2 + 4^2 + \cdots + 4^2 - C = 1482 - 1280 = 202,$

（4）　$S_A = \dfrac{24^2 + 42^2 + \cdots + 28^2}{4} - C = 1386 - 1280 = 106,$

（5）　$S_E = 202 - 106 = 96.$

上の計算結果は，分散分析表として表 7.5 に整頓される。

表 7.5. 表 7.2 のデータの分散分析表

要因	平方和	自由度	不偏分散	F
級間変動	106	4	26.5	4.14
級内変動	96	15	6.4	
全変動	202	19		

有意水準 $\alpha = 0.05$ の下で，F 分布表から，$F(4,\ 15,\ 0.05) = 3.06$ を得る。いま，統計量 F_0 の実現値は 4.14 であるから，仮説 $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ は棄却される。すなはち，要因水準間に有意差は認められる。

水準 k での平均の推定値は $\overline{X}_{k\bullet} = X_{k\bullet}/n_k$ であるので、その分散は σ^2/n_k となり、$100(1-\alpha)$ %信頼区間は

$$\overline{X}_{k\cdot} \pm t\,(N-K, \alpha/2)\,\sqrt{V_E/n_k}$$

で与えられる。ここに、$t\,(N-K, \alpha/2)$ は自由度 $N-K$ の t 分布の上側確率 $\alpha/2$ に対する点（上側 $50\alpha\%$ 点）である。また、水準 k と l の平均の差の信頼区間は

$$\overline{X}_{k\cdot} - \overline{X}_{l\cdot} \pm t\,(N-K, \alpha/2)\sqrt{\left(\tfrac{1}{n_k} + \tfrac{1}{n_l}\right) V_E}$$

となる。

[例 7.2] 例 7.1 で各水準の平均値の 95 %信頼区間を構成する。表 7.5 から $V_E = 6.4,\ N = 20,\ K = 5$ で、

$$\overline{X}_{1\cdot} = T_1/4 = 24/4 = 6.0,\ \overline{X}_{2\cdot} = 10.5,$$
$$\overline{X}_{3\cdot} = 5.5,\ \overline{X}_{4\cdot} = 11.0,\ \overline{X}_{5\cdot} = 7.0$$

ある。また、$t\,(15, 0.025) = 2.131$ となるので、$\mu_i\ (i = 1, 2, 3, 4, 5)$ の 95% 信頼区間はそれぞれ

$$6.0 \pm 2.131\sqrt{6.4/4} = 6.0 \pm 2.70 = (8.70, 3.30)\,,$$
$$10.5 \pm 2.70,\quad 5.5 \pm 2.70,\quad 11.0 \pm 2.70,\quad 7.0 \pm 2.70$$

である。

例 7.1 のように水準による応答の差が認められたとき、どの水準間に差があるかを比較検定することになる。もし、前もって注目し比較したい水準間がなければ、全ての水準対に対する比較を行うことになる、例 7.1 の場合は 10 通りの対比較を行うことになるので、すべての水準対に関する個々の検定の有意水準を α とすると、10 個の比較を同時に行うための有意水準は最大で 10α となる。このことは検定を多重に繰り返すことから来るもので、このような比較を多重比較 (multiple comparison) という。このとき、ボンフェロニ

法では個々の比較に対する有意水準を $\alpha/10$ にすることを提案している。このように個々の比較での有意水準を厳しくすることで、対象とする全ての比較検定に対する有意水準を α 以下にすることができる。また、実験によっては、水準 A_1 が標準となっていて、この標準との比較に意味がある場合が考えられる。この場合の比較は A_1 と A_i $(i = 2, 3, 4, 5)$ の 4 組になるので、4 組の同時比較で結論を得る場合には個々の検定での有意水準 とすればよい。この場合、次の対立仮説が考えられる。

H_{11}: $\mu_1 \neq \mu_i$ $(i = 2, 3, 4, 5)$ の少なくとも一つが成立する。
（両側検定）

H_{12}: $\mu_1 > \mu_i$ $(i = 2, 3, 4, 5)$ の少なくとも一つの不等式が成立する。
（片側検定）

H_{13}: $\mu_1 < \mu_i$ $(i = 2, 3, 4, 5)$ の少なくとも一つの不等式が成立する。
（片側検定）

[例 7.3]　例 7.1 で A_1 を標準として、標準に対するその他の食品に品質を比較する場合を考える。このとき、水分量が多いことが保存食品の品質劣化に関係し、対立仮説は次のように片側で考えるのが妥当である。比較は 4 対になり、帰無仮説 H_0: $\mu_k = \mu_1$ $(k = 2, 3, 4, 5)$ を対立仮説 H_1: 「少なくとも一つの k に対して $\mu_k > \mu_1$ が成立する」をボンフェロニ法で検定する。有意水準を高々0.05 とすると、個々の検定を H_{k0}: $\mu_k = \mu_1$ を対立仮説 H_{k1}: $\mu_k > \mu_1$ に対して有意水準 0.0125 $(= 0.05/4)$ で行う。ここで、検定統計量を

$$t_k = \frac{\overline{X}_{k\cdot} - \overline{X}_{1\cdot}}{\sqrt{\left(\frac{1}{4}+\frac{1}{4}\right)V_E}} \quad (k = 2, 3, 4, 5)$$

とすると

$$t_2 = \frac{10.5-6.0}{\sqrt{\left(\frac{1}{4}+\frac{1}{4}\right)6.4}} = 2.52 \quad (P = 0.0118)$$
$$t_3 = \frac{5.5-6.0}{\sqrt{\left(\frac{1}{4}+\frac{1}{4}\right)6.4}} = -0.28 \ (P = 0.608)$$
$$t_4 = \frac{11.0-6.0}{\sqrt{\left(\frac{1}{4}+\frac{1}{4}\right)6.4}} = 2.80 \quad (P = 0.0067)$$
$$t_5 = \frac{7.0-6.0}{\sqrt{\left(\frac{1}{4}+\frac{1}{4}\right)6.4}} = 0.56 \quad (P = 0.292)$$

を得る。ボンフェロニ法での多重比較で水準 A_2 と A_4 の水分が標準 A_1 に対して有意に多いことが検出される。

例 7.3 で水準間の差 $\mu_k - \mu_1$ の同時信頼区間を考える。水準 A_1 を標準としたときの、個々の 95 % 信頼区間は

$\mu_2 - \mu_1$ の信頼区間：$10.5 - 6.0 \pm 2.490\sqrt{\left(\frac{1}{4}+\frac{1}{4}\right)6.4} = 4.5 \pm 4.4542$

$\mu_3 - \mu_1$ の信頼区間：$5.5 - 6.0 \pm 4.4542 = -0.5 \pm 4.4542$

$\mu_4 - \mu_1$ の信頼区間：$11.0 - 6.0 \pm 4.4542 = 5.0 \pm 4.4542$

$\mu_5 - \mu_1$ の信頼区間：$7.0 - 6.0 \pm 4.4542 = 1.0 \pm 4.4542$

である。しかし、4つの水準差の同時信頼区間を構成するには、それぞれの信頼区間を厳しくする必要がある。95 % 同時ボンフェロニ信頼区間は母数を含む 95 %以上の確率であり、

$\mu_2 - \mu_1$ の信頼区間：$10.5 - 6.0 \pm 2.490\sqrt{\left(\frac{1}{4}+\frac{1}{4}\right)6.4} = 4.5 \pm 4.454$
$\mu_3 - \mu_1$ の信頼区間：-0.5 ± 4.454
$\mu_4 - \mu_1$ の信頼区間：5.0 ± 4.454
$\mu_5 - \mu_1$ の信頼区間：1.0 ± 4.454

になる。この結果から減点 0 を含まない区間 $\mu_2 - \mu_1$ と $\mu_4 - \mu_1$ に対する水準の効果が有意となる。このことは例 7.3 の検定結果と同等である。

注意 7.1. 多重比較では実験に先立って設定する対立仮説に関係し、全ての対比較を行う場合にボンフェロニ法は厳しすぎるので、Tukey 法が提唱されている。また、例 7.3 のように対照群と処置群の比較のみに限定して、検出力を向上させる場合は Dunnett 法が提唱されている。Dunnett 法での棄却限界値は 2.36 である。Dunnett 法は数値表を用いいるか、統計ソフトで計算できる。大事なことは、実験に際しての対立仮説である。

ここでは要因水準で誤差分散が等しいことを仮定して議論をしている。分散の一様性の検定は二標本検定の場合と同様に F 検定や t 検定を行う前提の確認にも用いられる。誤差分布が正規分布に従うとして水準 i $(i = 1, 2, ..., K)$ における誤差分散を σ_i^2 とするとき、仮説

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$

を検定することになる。検定法としては標本分散に基づく (i) Batrtlet の検定、(ii)Hartley の検定、(iii) Chochran の検定および (iv) 範囲を用いた検定が提案されてている。(i) 反復数が多い場合の漸近的な検定で、水準間で反復数が不ぞろいの場合にも適用できる。(ii), (iii) と (iv) は反復数が全ての水準で同じときに適用される。統計解析ソフトを用いる場合は、等分散性の検定を行った方が良い。もし、この仮定が棄却されれば、適当なデータの変換を行った後で、分散分析を行う必要がある。

問 7.1. 5 個の異なる条件の下である作物を生産し、それらの生産量 (kg) への効果を調べるために次の結果を得た。このとき、生産条件が生産量に与える効果があるかどうかの検定を意水準は 0.05 で行え。ただし、全ての生産条件におけるデータ分布は正規分布で分散は等しいとする。

表 7.6. 作物の生産量

要因	測定値			
A_1	1334	1324	1120	842
A_2	1952	1922	1818	1325
A_3	1206	1378	1005	708
A_4	1576	1278	1613	1358
A_5	1802	1400	1566	1756

7.3. 乱塊法（無作為層別配置法）

一元配置法はただ 1 つの制御要因の水準に関し，実験データを無作為にとる配慮ですんだが，実際には他に無視できない要因もある。一元配置法では,このような要因水準に関して無作為化を通じ，すべて誤差としたわけである。しかし，実験日や実験地区などのように，無視できないブロック要因に関して無作為化できない場合がある。このような際には，各ブロック内で無作為データをとる乱塊法 (randomized block design）がある。

いま，ある処理要因 A の水準を A_1，A_2，$\cdots$，A_p，ブロック要因 B の水準を B_1，B_2，$\cdots,Bq$ とし，各データ X_{ij} は処理水準 A_i の全水準について各ブロック水準 B_j の中で無作為化して実験しデータをとることとする。ここでの主要な目的は，処理要因 A に対する効果の推測であり，ブロック要因に関するものではない。このときのデータの構造模型は次式

$$X_{ij} = \mu + \alpha_i + b_j + e_{ij} \quad (i = 1,\ 2,\ \cdots,\ p\ ;\ j = 1,\ 2,\ \cdots,\ q)$$

ここに，α_i は A_i の処理効果，b_j は B の影響効果を表す確率変数，また $\{e_{ij}\}$ は誤差を示し，互いに独立で正規分布 $N(0,\ \sigma^2)$ に従うとする。一般性を失わず母数 $\{a_i\}$ に関しては,

$$\sum_{i=1}^{p}\alpha_i = \alpha. = 0$$

とし，変量 $\{\beta_i\}$ は互いに独立な正規分布 $N(0, \sigma_B^2)$ に従い、かつ誤差 $\{e_{ij}\}$ と独立とする。この実験観測のデータの配置を表 7.7 に示す。この際の全変動 S_T は次のように分解される。

$$\begin{aligned} S_T &= \sum_{i=1}^{p}\sum_{j=1}^{q}\left(X_{ij} - \overline{X}_{..}\right)^2 \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q}\left(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}_{..} + \overline{X}_{i\cdot} - \overline{X}_{..} + \overline{X}_{\cdot j} - \overline{X}_{..}\right)^2 \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q}\left(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}_{..}\right)^2 + q\sum_{i=1}^{p}\left(\overline{X}_{i\cdot} - \overline{X}_{..}\right)^2 + p\sum_{j=1}^{q}\left(\overline{X}_{\cdot j} - \overline{X}_{..}\right)^2 \\ &= S_E + S_A + S_B \end{aligned} \tag{7.6}$$

ここに

$$\begin{aligned} S_E &= \sum_{i=1}^{p}\sum_{j=1}^{q}\left(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}_{..}\right)^2, \\ S_A &= q\sum_{i=1}^{p}\left(\overline{X}_{i\cdot} - \overline{X}_{..}\right)^2, \\ S_B &= p\sum_{j=1}^{q}\left(\overline{X}_{\cdot j} - \overline{X}_{..}\right)^2 \end{aligned}$$

とし、S_A は要因 A の効果に関する平方和、S_B はブロック要因 B に関する平方和、そして S_E は誤差 e に関する平方和である。

表 7.7. 乱塊法におけるデータ

	B_1	B_2	$\cdots$	B_q
A_1	X_{11}	X_{12}	$\cdots$	X_{1q}
A_2	X_{21}	X_{22}	$\cdots$	X_{2q}
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
A_p	X_{p1}	X_{p2}	$\cdots$	X_{pq}

次に，統計量 S_T, S_E, S_A , S_B の分布特性を考える。まず，それぞれの無作為化に対する期待値を考える。

$$E\left(\overline{X}_{i\cdot}\right) = \mu + \alpha_i, \quad E\left(\overline{X}_{\cdot j}\right) = \mu, \quad E\left(\overline{X}_{\cdot\cdot}\right) = \mu$$

であり，各平方和に期待値については，

$$E\left(S_E\right) = E\left\{\sum_{i=1}^{p}\sum_{j=1}^{q}\left(e_{ij} - \overline{e}_{i\cdot} - \overline{e}_{\cdot j} + \overline{e}_{\cdot\cdot}\right)^2\right\} = \left(pq - p - q + 1\right)\sigma^2,$$

$$E\left(S_A\right) = E\left\{q\sum_{i=1}^{p}\left(\alpha_i + \overline{e}_{i\cdot} - \overline{e}_{\cdot\cdot}\right)^2\right\} = \left(p-1\right)\sigma^2 + q(p-1)\sigma_A^2,$$

$$E\left(S_B\right) = E\left\{p\sum_{j=1}^{q}\left(\overline{e}_{\cdot j} + b_j - \overline{e}_{\cdot\cdot} + \overline{b}_{\cdot}\right)^2\right\} = \left(q-1\right)\sigma^2 + p(q-1)\sigma_B^2$$

を得る。ここに，

$$\sigma_A^2 = \frac{\sum_{i=1}^{p}\alpha_i^2}{p-1}$$

とおいている。上式から $V_E = (S_T - S_A - S_B)/(p-1)(q-1)$ は σ^2 の不偏推定量であり，S_E/σ^2 は要因 A に関する仮説に無関係につねに自由度 $(p-1)(q-1)$ の χ^2 の分布に従う。また，式 (7.6) から S_A/σ^2 は自由度 $p-1$ をもち，非心率 $q(p-1){\sigma_A}^2/\sigma^2$ の非心 χ^2 分布に従い，S_E/σ^2 と独立である。したがって，$V_A = S_A/(p-1)$ として，仮説 $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ の下で $F_0 = V_A/V_E$ は自由度対 $(p-1,\ (p-1)(q-1))$ の F 分布に従う（表 7.8）。

表 7.8. 乱塊法における分散分析表

要因	平方和	自由度	不偏分散	F 値	$E\,(V)$
A	S_A	$p-1$	$V_A = \frac{S_A}{p-1}$	$F_0 = \frac{V_A}{V_E}$	$\sigma^2 + q\sigma_A^2$
B	S_B	$q-1$	$V_B = \frac{S_B}{q-1}$	$\frac{V_B}{V_E}$	$\sigma^2 + p\sigma_B^2$
E	S_E	$(p-1)(q-1)$	$V_E = \frac{S_E}{(p-1)(q-1)}$		σ^2
T	S_T	$pq-1$			

この配置法の目的は，処理要因 A の効果，すなわち要因 A の水準間の有意差の検定であり，上の統計量 F_0 によって，仮説 H_0 の検定が次のようにできる。

(i)　$F_0 \leqq F(f_1, f_2, \alpha)$ のとき，仮説 H_0 を採択する。
(ii) $F_0 > F(f_1, f_2, \alpha)$ のとき，仮説 H_0 を棄却する。

ここに，

$$f_1 = p - 1, \quad f_2 = (p-1)(q-1).$$

処理効果の推定は，式 (7.130) から，
$\hat{\mu} = \overline{X}_{..}, \quad \hat{\alpha}_i = \overline{X}_{i.} - \overline{X}_{..} \quad (i = 1,\ 2,\ \cdots,\ p)$ で与えられ，これが不偏推定量であることも明らかである。乱塊法ではブロック間の誤差変動を分離し、要因効果の推定および検定精度を向上させることができる。

[**例**7.4] ある合成反応において，反応温度 A の 4 水準が合成物質の収量（gr）に及ぼす効果を調査したい。1 日（ブロック要因）に 4 回の実験を行い，反応温度を変化させ 3 日かけて**表**7.9 の効果を得，この表から**表**7.10 の分散分析表を得た。表 7.10 から $F_0 = V_A/V_E = 10.94 > F(3, 6, 0.05) = 4.76$ であり，合成物質の収量に与える 4 水準の温度効果に有意差が認められる。

表 7.9. 反応温度と合成物質の量

温度	15 °C	20 °C	25 °C	30 °C
1 日目	62.1	66.8	73.5	64.7
2 日目	67.3	68.3	74.8	70.2
3 日目	63.4	67.6	69.7	68.9

表 7.10. 分散分析表

要因	平方和	自由度	不偏分散	F 値	$E(V)$
A	107.19	3	35.73	10.94 ($P = 0.008$)	$\sigma^2 + 2\sigma_A^2$
B	27.44	2	13.72	4.195 7 ($P = 0.07$)	$\sigma^2 + 3\sigma_B^2$
E	19.60	6	3.27		σ^2
T	154.23	11			

7.4. ノンパラメトリック法

前節まではデータの分布は特定の母数（パラメータ）を用いて表現されている。このときの検定法は分布の仮定の崩れに対して、頑健でない面を持っている。ここでは、データの分布を特定の型に限定しない仮説検定の方法について述べる。このような検定法はノンパラメトリック法 (nonparametric method) と総称されている。考え方を説明するために二母集団の位置母数 (location parameter) の検定（二標本検定）を考える。位置母数とは平均、中

央値、モードのように分布の位置を定める母数をいう。いま、二母集団間の位置母数の比較を考え、一方を処置群、他方を対照群とし、それぞれの分布関数を $F(x)$ と $G(x)$ で示す。このとき、

$$F(x) = G(x - \Delta) \qquad (7.7)$$

を仮定し、仮説

$$H_0 : \Delta = 0$$

の検定を考える。式 (7.7) の関係は２つの分布の形が相似であることを示している。ここに、Δ は処理効果を意味し、ここでの仮定は処理効果の加法性である。ノンパラメトリック法では、観測値の順序やある条件を満たす観測値の個数を母集団情報として、検定や推定を行う。

[例 7.5] ある年齢層から無作為に抽出された男性 5 人と女性 6 人の血清コレステロール値を測定し、次のデータを得た。

男性: 167, 208, 225, 200, 182;

女性: 222, 168, 198, 186, 150, 180 (mg/l)

平均値に関する男女間の差の検定を行う場合に、データの正規性が崩れていれば、前節で述べた t 検定は有効でない。□

上の例での問題点を解決する方法としては、値を大きさの順に並べた時の順位を検定の情報として用いる。ここでは、ウィルコクソンの順位検定 (Wilcoxon test) について述べる。２つの母集団からの標本をそれぞれ、$\{X_1, X_2, \ldots, X_m\}$ と $\{Y_1, Y_2, \ldots, Y_n\}$ で示す。ここにとする。このデータを合併し、大きさの順に小さいほうから並べたときの X_i の順位を R_i とするとき、

$$W = \sum_{i=1}^{m} R_i$$

をウィルコクソンの順位統計量という。上の例で男性の順位統計量は

$$W = 2 + 9 + 11 + 8 + 5 = 35$$

である。検定の考え方としては、この順位和に顕著な偏りが見られるとき、２つの母集団が同等であるという仮説を棄却することになる。連続分布では同じ値が複数回観測されることは理論上ない。しかし、観測のときの丸め誤差によって同一値として記録される場合がある。このでは同一の値が観測値として得られていないものとして議論する。同一の値がある場合は修正する方法があり、汎用的な統計ソフトでは自動的に補正される。データ X と Y の分布関数が (7.7) の関係がある場合には、統計量 W による検定法は有効性である。このような関係のとき、Y の平均値は X の平均値より Δ だけ大きくなり、ウィルコクソン統計量はこの影響を受ける。帰無仮説 $H_0 : \Delta = 0$ の下では二標本の分布は一致し、ウィルコクソンの順位統計量の分布についての分布表が作られている。分布は標本数 m と n の関数である。ウィルコクソン検定は次の手順を説明する前にウィルコクソン統計量の性質を述べる。

定理 7.1. ウィルコクソン統計量 W の平均と分散は、帰無仮説 $H_0 : \Delta = 0$ の下で次のようになる。

$$E(W) = \frac{m(N+1)}{2}$$
$$Var(W) = \frac{mn(N+1)}{12}$$

ここに、$N = m + n$ である。

証明　仮定の下でデータ X_i の順位 R_i は同様な確かさで順位 1 から N までをとるので

$$E(R_i) = \sum_{i=1}^{N} i \times \frac{1}{N} = \frac{N+1}{2}$$

である。このことから

$$E(W) = \sum_{i=1}^{m} E(R_i) = \frac{m(N+1)}{2}$$

を得る。また、

$$E(R_i^2) = \sum_{i=1}^{N} i^2 \times \frac{1}{N} = \frac{(N+1)(2N+1)}{6}$$
$$E(R_iR_j) = \left\{ \left(\sum_{i=1}^{N} i \right)^2 - \sum_{i=1}^{N} i^2 \right\} \times \frac{1}{N(N-1)} = \frac{(N+1)(3N+2)}{12} \quad (i \neq j)$$

であるので

$$Var(W) = E(W^2) - \left\{E(W)^2\right\} = E\left\{ (\textstyle\sum_{i=1}^{m} R_i)^2 \right\} - \frac{m^2(N+1)^2}{4}$$
$$= m\frac{(N+1)(2N+1)}{6} + m(m-1)\frac{(N+1)(3N+2)}{12} - \frac{m^2(N+1)^2}{4} = \frac{mn(N+1)}{12}$$

を得る。

定理 7.2. ウィルコクソン統計量 W は、帰無仮説 $H_0 : \Delta = 0$ の下で平均に関して対称である。

証明　順序 R_i に対して、逆順位 $R_i' = N + 1 - R_i$ を順位とするウィルコクソン統計量 W' を考えると、

$$W' = \sum_{i=1}^{m}(N - R_i) = m(N+1) - W$$

を得る。このことから、定理が示される。

以上の性質を基にして、ウィルコクソン検定の手順は次のようにまとめられる。

(1) 対立仮説 $H_1 : \Delta > 0$ のとき

手順 1) 標本 $\{X_1, X_2, \ldots, X_m\}$ と $\{Y_1, Y_2, \ldots, Y_n\}$ を合併し、$m + n$ 個の順序を決める。

手順 2) 与えられた有意水準 α に対して、次式を満たす最小の整数 w を分布表から求める。

$$P\left(W \geq w\right) \leq \alpha$$

手順 3) ウィルコクソン統計量 W の標本値 w_0 を計算値、$w_0 \geq w$ なら仮説 H_0 を棄却し、$w_0 < w$ なら仮説 H_0 を採択する。

(2) 対立仮説 $H_1 : \Delta > 0$ のとき

手順 1)　(1) と同じ

手順 2)　与えられた有意水準 α に対して、次式を満たす最大の整数 w を分布表から求める。

$$P\left(W \leq w\right) \leq \alpha$$

手順 3) ウィルコクソン統計量 W の標本値 w_0 を計算値、$w_0 \leq w$ なら仮説 H_0 を棄却し、$w_0 > w$ なら仮説 H_0 を採択する。

(3) 対立仮説 $H_1 : \Delta > 0$ のとき

上の例で有意水準 0.05 の両側検定を考えれば、棄却域は $W \leq 18, W \geq 42$ である。 結果は $W = 35$ であるので、帰無仮説は棄却されず、コレステロール値に関する男女間差は無いものと統計的に判断できる。

連続分布に従う観測値の場合は、実際には観測値は四捨五入によって適当な精度の下で観測され、同じ値の観測値が記録されることになる。この場合の取り扱いは、同じ値の観測値に平均順位を当てて、補正することになる。多くの統計用のソフトでは、この補正に関する処理ができる。

問 7.2. ネズミ 12 匹を 6 匹ずつの 2 群に分けて、餌 A, B を与えて飼育した。7 週間後の体重の増加は次の通りであった。

A: 156, 183, 120, 113, 138, 145

B: 130, 148, 117, 133, 140, 142 (g)

このデータから餌の違いがネズミの発育に影響があるかどうか、ウィルコクソン検定を行え。

標本数 m と n が大きいとき、ウィルコクソン統計量 W は、帰無仮説 $H_0 : \Delta = 0$ の下で正規分布 $N\left(\frac{m(N+1)}{2}, \frac{mn(N+1)}{12}\right)$ に漸近的に従う。このことから、統計量

$$Z = \frac{W - \frac{m(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}}$$

が $N\left(0, 1\right)$ に従うことを利用して検定することができる。

ウィルコクソン統計量 W は次のように変形できる。関数

$$u(t) = \begin{cases} 0 & (t < 0) \\ 1 & (t \geq 0) \end{cases}$$

を考えると

$$W = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{m} u(X_i - X_j) + \sum_{j=1}^{n} u(X_i - Y_j) \right\}$$

となる。ここで

$$\sum_{i=1}^{m}\sum_{j=1}^{m} u(X_i - X_j) = \frac{1}{2}m(m+1)$$

であるから

$$W = \sum_{i=1}^{m}\sum_{j=1}^{n} u(X_i - Y_j) + \frac{1}{2}m(m+1)$$

を得る。このとき

$$U = \sum_{i=1}^{m}\sum_{j=1}^{n} u(X_i - Y_j)$$

をマンーウィットニー (Mann-Whitney) 型のウィルコクソン統計量という。

3 群以上の比較を場合の一元配置実験で、正規性が認められない場合のノンパラメトリック法とし、クラスカルーワリス検定が考えられている。これはウィルコクソン検定の拡張であり、統計解析ソフトには組み込まれている。適応の前提は処理の加法性が満たされるときである。