創造応用 IS（稲葉）　　No.14c

2021.11.24(水 13:10 ～)

兵庫高校

A.0. 実験計画法のポイント

実験計画法の手法で求める平方和の計算式は、分散分析の平方和の計算式と同じです。

（主効果の平方和は、１元配置分散分析で繰り返しと考える。）

A.0.1. 品質管理における統計的考え方

(1) 平均値　$\overline{x} = \Sigma x_i/n \quad \Leftrightarrow \quad n\overline{x} = \Sigma x_i \quad \Leftrightarrow \quad \Sigma(x_i - \overline{x}) = 0$

(2) 平方和　$S = \Sigma(x_i - \overline{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2/n$

(3) 不偏分散　$V = \dfrac{S}{n-1} = \dfrac{\text{平方和}}{\text{自由度}}, \quad E(V) = \sigma^2$

(4) a, b :定数、X, Y ：確率変数とするとき

1)$E(aX + bY) = aE(X) + bE(Y)$

2) X, Y が互いに独立ならば、

$V(aX + bY) = a^2V(X) + b^2V(Y)$

(5) $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$（互いに独立とする。）

$\Rightarrow \overline{X} \sim N(\mu, \dfrac{\sigma^2}{n}),\ \ V = \displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$

$\Rightarrow (\overline{X} - \mu)/(\sqrt{\sigma^2/n}) \sim N(0, 1) \qquad \Rightarrow (\overline{X} - \mu)/(\sqrt{V/n}) \sim t(\phi) \quad \phi = n - 1$

［意味：分母の σ^2 にその不偏分散 V を代入すると t 分布になる。］

1) これを用いて、$H_0 : \mu = \mu_0,\ \ H_1 : \mu \neq \mu_0$ の検定を行う。

2)μ の信頼区間は、$(\overline{X} - t(\phi, 0.025)\sqrt{V/n},\ \ \overline{X} + t(\phi, 0.025)\sqrt{V/n})$

(6) 実験計画法も、やはり検定と推定である。

元のデータが正規分布に従うなら、平方和は χ^2 分布しその比は F 分布する。

$$\begin{cases} X_1, X_2, \cdots, X_m \sim N(\mu_x, \sigma^2) \\ Y_1, Y_2, \cdots, Y_n \sim N(\mu_y, \sigma^2) \end{cases} \quad \text{（互いに独立）}$$

$\Rightarrow F = V_A/V_B \sim F(m-1, n-1)$ （ $\Rightarrow$ 分散分析の 検定の原理）

A.0.2. 因子と水準：略

(1) 因子の種類（制御因子、標示因子、誤差因子、**ブロック因子**）

(2) 母数因子と変量因子

母数因子とは、設定したり制御したりできる因子。

変量因子とは、ランダムに変動する因子。誤差と良く似た振る舞いをする。

A.0.3. フィッシャーの３原則：略

(1) **局所管理**

小さなブロックを作り、その中は均一な場になるとする。１つのブロックで比較したい１通りの実験を（ランダムな順で）行なう方法を 乱塊法 という。

(2) **無作為化**

(3) 反復（と繰り返し）

繰り返しとは、同じ実験条件で複数回行なうこと。実験順序は、要因と繰り返しを混ぜて、ランダムに行なう。

反復とは、１セットの実験を複数回行なうこと。１セットの実験の中でランダムに行われる。

A.0.4. 分散分析（完全無作為化実験）と分割実験

すべての実験を、ランダムな順序で行う実験を、**完全無作為化実験（完全ランダマイズ実験）** または**分散分析**といい、２段階以上に分割して行う実験を**乱塊法**や**分割実験**という。

A.1. 分散分析

要因が１つのときには、要因の複数の水準において、複数回の実験を行い「**一元配置分散分析（一元配置）**」と言われます。要因が２つのときには、「**二元配置**」と言われます。各水準組み合わせで何回実験するかを**繰り返し**と呼び、今日は、一元配置と、二元配置の繰り返しのない場合を紹介します。

これらの手法は、**分散分析**と呼ばれます。分散分析では、すべての実験をランダムな順序で行う「完全無作為化実験」を行います。

例. 一元配置のデータ

一元配置とは、３つ以上の群の母平均のちがいについての検定と推定を行う方法である。

$$
\begin{array}{l}
x_{11}, x_{12}, \cdots, x_{1n_1} : \text{第 1 群} \\
x_{21}, x_{22}, \cdots, x_{2n_2} : \text{第 2 群} \\
\qquad\qquad\quad \cdots \\
x_{a1}, x_{a2}, \cdots, x_{an_a} : \text{第 } a \text{ 群}
\end{array}
$$

A.1.1. 一元配置（バランスケース：各群の標本の大きさが一定の場合）($n_i \equiv n$ の場合)

データの構造式

$$x_{ij} = \mu + a_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a; \quad j = 1, 2, \cdots, n \tag{A.1}$$

$$\sum_{i=1}^{a} a_i = 0, \quad \varepsilon_{ij} \sim N(0, \sigma_E^2)$$

μ：<u>全平均</u>、a_i ：要因 A の <u>主効果</u>、ε_{ij}：<u>誤差</u>

平方和の分解

データのばらつき（<u>総平方和 S_T</u>）を、誤差的なばらつき（<u>誤差平方和 S_E</u>）と、群の違いによるばらつき（<u>要因平方和 S_A</u>）に分解するのがポイントである。

$$
\begin{aligned}
& S_T = S_A + S_E \\
& \sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \overline{T})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(\overline{A}_i - \overline{T})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \overline{A}_i)^2
\end{aligned} \tag{A.2}
$$

自由度の計算

<u>総自由度</u> ϕ_T を、<u>要因の自由度</u> ϕ_A と、<u>誤差自由度</u> ϕ_E に分解する。

$$\phi_A = a - 1, \quad \phi_E = a(n-1), \quad \phi_T = \phi_A + \phi_E \tag{A.3}$$

検定統計量の計算

$$F_0 = \frac{V_A}{V_E} \text{ である。ただし、} V_A = S_A/\phi_A,\ V_E = S_E/\phi_E \text{ と求める。} \tag{A.4}$$

棄却域

$$R : F_0 \geqq F(\phi_A, \phi_E; \alpha) \tag{A.5}$$

判定

検定統計量 F_0 が、棄却域に入ったとき、要因 A は有意水準 α で有意であったと言い、この要因が効果があったと判定する。

A.1.2. 一元配置（バランスケース：数値例）

1) 3 群のデータ

例題 A.1　臨床試験で、対照群、低用量群、高用量群の 3 群比較を行う。各群 $n = 5$ 人についての測定データを、表 A.1 に示す。

表 A.1　3 群のデータ

群	データ x_{ij}					和 T_i	データ数 n_i	平均 $\overline{A}_i$	二乗和	平方和 S_i
対照群 A_1	20	25	17	18	15	95	5	19	1863	58
低用量群 A_2	18	17	12	14	19	80	5	16	1314	34
高用量群 A_3	23	26	21	19	21	110	5	22	2448	28
計	−−−−−−					285	15	−−	5625	120

2) 平方和の分解

各群のデータ数 $n_i \equiv n$ は一定とします。

帰無仮説：$\mu_1 = \mu_2 = \cdots = \mu_a$
対立仮説：その他

まずデータのばらつきを、各群での標本平均 $\overline{A}_i$ を介して、2 つの変動に分解します。

$$\text{データの分解：} x_{ij} - \overline{T} = (x_{ij} - \overline{A}_i) + (\overline{A}_i - \overline{T}) \tag{A.6}$$

このことを、表 A.1 の 15 個のデータで表現すると、以下のようになります。

表 A.2　データの分解

$$\underset{\text{(a) データ：}x_{ij}}{\begin{array}{|c|ccccc|}\hline A_1 & 20 & 25 & 17 & 18 & 15 \\ A_2 & 18 & 17 & 12 & 14 & 19 \\ A_3 & 23 & 26 & 21 & 19 & 21 \\ \hline\end{array}} - \underset{\text{(b) 総平均：}\overline{T}}{\begin{array}{|c|ccccc|}\hline A_1 & 19 & 19 & 19 & 19 & 19 \\ A_2 & 19 & 19 & 19 & 19 & 19 \\ A_3 & 19 & 19 & 19 & 19 & 19 \\ \hline\end{array}}$$

$$= \underset{\text{(c) 群間変動：}\overline{A}_i - \overline{T}}{\begin{array}{|c|ccccc|}\hline A_1 & 0 & 0 & 0 & 0 & 0 \\ A_2 & -3 & -3 & -3 & -3 & -3 \\ A_3 & 3 & 3 & 3 & 3 & 3 \\ \hline\end{array}} + \underset{\text{(d) 群内変動：}x_{ij} - \overline{A}_i}{\begin{array}{|c|ccccc|}\hline A_1 & 1 & 6 & -2 & -1 & -4 \\ A_2 & 2 & 1 & -4 & -2 & 3 \\ A_3 & 1 & 4 & -1 & -3 & -1 \\ \hline\end{array}}$$

(c) より、$S_A = n\sum_{i=1}^{a}(\overline{A}_i - \overline{T})^2 = 5\{0^2 + (-3)^2 + 3^2\} = 90$

(d) より、$S_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \overline{A}_i)^2 = 1^2 + 6^2 + \cdots + (-1)^2 = 120$

3) 分散分析表による検定
分散分析の結果を、表にまとめます。

表 A.3　分散分析表

要因	平方和	自由度	平均平方	F_0 値	限界値
因子 A	S_A	$\phi_A = a-1$	$V_A = S_A/\phi_A$	$F_0 = V_A/V_E$	$F(0.05)$
誤差 E	S_E	$\phi_E = a(n-1)$	$V_E = S_E/\phi_E$	$--$	$--$
計　T	S_T	$\phi_T = an-1$	$--$	$--$	$--$

例題 **A.2**　表 A.1 のデータについて、一元配置分散分析を行え。

表 A.4　分散分析表

要因	平方和	自由度	平均平方	F_0 値	限界値
因子 A	90	2	45.0	4.50	3.89
誤差 E	120	12	10.0	$--$	$--$
計　T	210	14	$--$	$--$	$--$

結論：$F_0 = 4.50 \geqq 3.89 = F(2, 12; 0.05)$ より、有意水準 5% で、要因 A によって母平均には違いがあるといえる。

4) 投与群での母平均の推定
母平均 μ_i の **点推定値** は $\overline{A}_i$ です。この点推定量の分布は、n 個の平均ですから、

$$\overline{A}_i \sim N\left(\mu_i, \sigma^2/n\right) \tag{A.7}$$

となります。ここで、$S_E/\sigma^2 \sim \chi^2(\phi_E)$ であることを利用すると、(A.7) 式を標準化して、分母の σ^2 に V_E を代入することで、自由度が ϕ_E の t 分布が導かれます。そこで、μ_i の 95% 信頼限界は、次のようになります。

信頼上限：$\overline{A}_i + t(\phi_E, 0.025)\sqrt{V_E/n}$
信頼下限：$\overline{A}_i - t(\phi_E, 0.025)\sqrt{V_E/n}$

例題 **A.3**　表 A.1 のデータについて、高用量群の母平均の推定を行え。
解答：点推定値は、$\widehat{\mu}_3 = \overline{A}_3 = 22.0$, 95% 信頼限界は、$\overline{A}_3 \pm t(12, 0.025)\sqrt{10.0/5} = 22.0 \pm 2.179\sqrt{2} = 22.0 \pm 3.1 = 18.9,\ 25.1$ となる。

データ数が揃っている場合を バランスケース、揃っていない場合を アンバランスケース といいます。アンバランスケースは、次節で議論します。

A.1.3. 一元配置（アンバランスケース：各群の標本の大きさが異なる場合）

各群での平均と、総平均は、次の式で推定します。

$$\overline{A}_i = \sum_{j=1}^{n_i} x_{ij}/n_i, \quad \overline{T} = \sum_{i=1}^{a}\sum_{j=1}^{n_i} x_{ij}/N, \quad \left(\text{ただし、} N = \sum_{i=1}^{a} n_i \right)$$

1) アンバランスケースの平方和の分解

要因（群間）平方和と誤差（群内）平方和は、次の式に修正します。

$$\text{要因平方和：} S_A = \sum_{i=1}^{a} n_i (\overline{A}_i - \overline{T})^2 \tag{A.8}$$

$$\text{誤差平方和：} S_E = \sum_{i=1}^{a}\sum_{j=1}^{n_i} (x_{ij} - \overline{A}_i)^2 \tag{A.9}$$

例題 A.4　臨床試験で、対照群、低用量群、中用量群、高用量群の 3 群比較を行う。$n_1 = 4,\ n_2 = 5,\ n_3 = 3,\ n_4 = 4$ で、総人数 $N = 16$ 人についての測定データを表 A.5 に、また、データの分解を表 A.6 に示す。

表 A.5　4 群のデータ

群	デ ー タ x_{ij}					和 T_i	データ数 n_i	平均 $\overline{A}_i$	二乗和	平方和 S_i
対照群 A_1	18	23	18	17		76	4	19	1466	22
低用量群 A_2	16	19	20	12	13	80	5	16	1330	50
中用量群 A_3	19	18	23			60	3	20	1214	14
高用量群 A_4	24	23	20	21		88	4	22	1946	10
計	−−−−−−					304	16	−−	5956	96

表 A.6　データの分解

(a) データ：x_{ij}

A_1	18	23	18	17	
A_2	16	19	20	12	13
A_3	19	18	23		
A_4	24	23	20	21	

−

(b) 総平均：$\overline{\overline{x}}$

A_1	19	19	19	19	
A_2	19	19	19	19	19
A_3	19	19	19		
A_4	19	19	19	19	

=

(c) 群間変動：$\overline{A}_i - \overline{T}$

A_1	0	0	0	0	
A_2	−3	−3	−3	−3	−3
A_3	1	1	1		
A_4	3	3	3	3	

+

(d) 群内変動：$x_{ij} - \overline{A}_i$

A_1	−1	4	−1	−2	
A_2	0	3	4	−4	−3
A_3	−1	−2	3		
A_4	2	1	−2	−1	

(c) より、$S_A = \sum_{i=1}^{a} n_i(\overline{A}_i - \overline{T})^2 = 4 \times 0^2 + 5(-3)^2 + 3 \times 1^2 + 4 \times 3^2 = 84$

(d) より、$S_E = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \overline{A}_i)^2 = (-1)^2 + 4^2 + \cdots + (-1)^2 = 96$

ここで、総平方和を次の式に変更すると、平方和の分解 $S_T = S_A + S_E$ が成り立ちます。

$$\text{総平方和}: S_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \overline{T})^2 \tag{A.10}$$

2) アンバランスケースでの検定

自由度は、$\phi_A = a - 1,\ \phi_E = \sum_{i=1}^{a}(n_i - 1) = N - a$ となります。以下、平均平方や F_0 値、棄却限界値を求めて要因に意味があるかどうかの検定ができます。これらを分散分析表にまとめます。

例題 A.5　表 A.5 のデータについて、一元配置分散分析を行え。

表 A.7　分散分析表

要因	平方和	自由度	平均平方	F_0 値	限界値
因子 A	84	3	28.0	3.50	3.49
誤差 E	96	12	8.0	−−	−−
計　T	180	15	−−	−−	−−

結論：$F_0 = 3.50 \geqq 3.49 = F(3, 12; 0.05)$ より、有意水準 5% で、要因 A によって母平均には違いがあるといえる。

3) アンバランスケースでの母平均の推定

母平均 μ_i の **点推定値** は $\overline{A}_i$ です。この点推定量の分布は、n_i 個の平均ですから、

$$\overline{A}_i \sim N\left(\mu_i, \sigma^2/n_i\right) \tag{A.11}$$

となります。ここで、$S_E/\sigma^2 \sim \chi^2(\phi_E)$ であることを利用すると、(A.11) 式を標準化して、分母の σ^2 に V_E を代入することで、自由度が ϕ_E の t 分布が導かれます。そこで、μ_i の 95% 信頼限界は、次のようになります。

信頼上限：$\overline{A}_i + t(\phi_E, 0.025)\sqrt{V_E/n_i}$
信頼下限：$\overline{A}_i - t(\phi_E, 0.025)\sqrt{V_E/n_i}$

例題 A.6　表 A.5 のデータについて、中用量群の母平均の推定を行え。
解答：点推定値は、$\widehat{\mu}_3 = \overline{A}_3 = 20.0,\ $ 95% 信頼限界は、$\overline{A}_3 \pm t(12, 0.025)\sqrt{8.00/3} = 20.0 \pm 2.179\sqrt{8/3} = 20.0 \pm 3.6 = 16.4,\ 23.6$ となる。

2021.11.24(水 13:10 ～)

兵庫高校

A.2. 二元配置（繰り返しなし）

多群の母平均の違いを、２つの要因に対して、検定と推定を行う方法である。

データ

要因 A が第 i 水準で、要因 B が第 j 水準のデータを x_{ij} とする。

データの構造式

$$x_{ij} = \mu + a_i + b_j + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a; \quad j = 1, 2, \cdots, b$$

$$\sum_{i=1}^{a} a_i = 0, \quad \sum_{j=1}^{b} b_j = 0, \quad \varepsilon_{ij} \sim N(0, \sigma_E^2)$$

μ：全平均、a_i ：要因 A の主効果、b_j ：要因 B の主効果、ε_{ij}：誤差

平方和の分解

データのばらつき（平方和）を、誤差的なばらつきと、群の違いによるばらつきに分解するのがポイントである。

$$S_T = S_A + S_B + S_E$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \overline{T})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{A}_i - \overline{T})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{B}_j - \overline{T})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(x_{ij} - \overline{A}_i - \overline{B}_j + \overline{T})^2$$

自由度の計算

$$\phi_A = a - 1, \quad ,\phi_B = b - 1, \quad \phi_E = (a-1)(b-1), \quad \phi_T = ab - 1$$

検定統計量の計算

$$F_0(A) = \frac{V_A}{V_E}, \ F_0(B) = \frac{V_B}{V_E} \text{ である。}$$

ただし、$V_A = S_A/\phi_A, \ V_B = S_B/\phi_B, \ V_E = S_E/\phi_E$ と求める。

棄却域

$$R : F_0(A) \geqq F(\phi_A, \phi_E; \alpha)$$

$$R : F_0(B) \geqq F(\phi_B, \phi_E; \alpha)$$

判定

検定統計量 $F_0(A)$ が、棄却域に入ったとき、要因 A は有意水準 α で有意であったと言い、この要因が効果があったと判定する。

検定統計量 $F_0(B)$ が、棄却域に入ったとき、要因 B は有意水準 α で有意であったと言い、この要因が効果があったと判定する。

A.3. 分散分析の概要

1) 一元配置分散分析

(1) データの構造

$x_{ij} = \mu + a_i + \varepsilon_{ij}$ （μ：一般平均、a_i：A_i の効果で $\sum a_i = 0,\quad \varepsilon_{ij}$：誤差）
$(i = 1, 2, \cdots, a;\ \ j = 1, 2, \cdots, n)$

(2) 平方和の分解と別表現

$S_T = S_A + S_E$

$$S_T = \sum_i \sum_j (x_{ij} - \overline{T})^2,\ \ S_A = \sum_i \sum_j (\overline{A}_i - \overline{T})^2 = \sum_i \frac{T_{Ai}^2}{n} - \frac{T^2}{an}$$

（S_A は、A の各水準での平均と全体の平均の差の２乗和）

(3) 自由度　$\cdots$　$\phi_T = \phi_A + \phi_E$，$\phi_T = an - 1$，$\phi_A = a - 1$

(4) 繰返し数が異なる場合

平方和は、本来の計算式（２乗和の形）で考えれば良い。

(5) 平方和の分解の理由：証明のポイントは、「（各群内の）偏差の和が０」

$$\sum_{i=1}^{a} \sum_{j=1}^{n} (x_{ij} - \overline{A}_i)(\overline{A}_i - \overline{T}) = \sum_{i=1}^{a} (\overline{A}_i - \overline{T}) \left[\sum_{j=1}^{n} (x_{ij} - \overline{A}_i) \right] = 0$$

2) 二元配置分散分析（繰り返しなし）

(1) データの構造

$x_{ij} = \mu + a_i + b_j + \varepsilon_{ij} \qquad (i = 1, \cdots, a; j = 1, \cdots, b)$

制約条件：$\sum a_i = \sum b_j = 0,\quad \varepsilon_{ij} \sim N(0, \sigma_E^2)$

(2) 平方和の分解と別表現

$S_T = S_A + S_B + S_E$

$$S_T = \sum_i \sum_j (x_{ij} - \overline{T})^2,\ \ S_A = \sum_i \sum_j (\overline{A}_i - \overline{T})^2 = \sum_i \frac{T_{Ai}^2}{b} - \frac{T^2}{ab}$$

(3) 自由度

$\phi_T = \phi_A + \phi_B + \phi_E$，$\phi_T = ab - 1$，$\phi_A = a - 1$，$\phi_E = (a - 1)(b - 1)$

(4) 交互作用の検出は出来ない。（もし、存在したら誤差と分離出来ない。）

3) 二元配置分散分析（繰り返しあり）

(1) データの構造

$x_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk} \qquad (i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, n)$

制約条件：$\sum a_i = \sum b_j = \sum_i (ab)_{ij} = \sum_j (ab)_{ij} = 0,\quad \varepsilon_{ijk} \sim N(0, \sigma_E^2)$

(2) 平方和の分解と別表現

$S_T = S_A + S_B + S_{A\times B} + S_E$

$$S_T = \sum_i \sum_j \sum_k (x_{ijk} - \overline{T})^2,\ \ S_A = \sum_i \sum_j \sum_k (\overline{A}_i - \overline{T})^2 = \sum_i \frac{T_{Ai}^2}{bn} - \frac{T^2}{abn}$$

（S_A は、A の各水準での平均と全体の平均の差の２乗和）

$$S_{AB} = \sum_i \sum_j \sum_k (\overline{A_i B_j} - \overline{T})^2 = \sum_i \sum_j \frac{T_{AiBj}^2}{n} - \frac{T^2}{abn}$$

（S_{AB} は、$A_i B_j$ の各水準での平均と全体の平均の差の２乗和）

$S_{A\times B} = S_{AB} - S_A - S_B$

(3) 自由度

$\phi_T = \phi_A + \phi_B + \phi_{A\times B} + \phi_E$，$\phi_T = abn - 1$，$\phi_A = a - 1$，$\phi_{A\times B} = (a - 1)(b - 1)$

2021.11.24(水 13:10 〜)

兵庫高校

A.4. 平方和のまとめ（手計算の時の計算式）

$$\text{総平方和}：S_T = \sum_{i,j} (\text{データ})^2 - \frac{(\text{データ総和})^2}{\text{総データ数}}$$

$$\text{要因平方和}：S_A = \sum_{i} \frac{(A_i\text{水準のデータ和})^2}{A_i\text{水準のデータ数}} - \frac{(\text{データ総和})^2}{\text{総データ数}}$$

$$2\text{要因平方和}：S_{AB} = \sum_{i,j} \frac{(A_iB_j\text{水準のデータ和})^2}{A_iB_j\text{水準のデータ数}} - \frac{(\text{データ総和})^2}{\text{総データ数}}$$

$$\text{交互作用平方和}：S_{A\times B} = S_{AB} - S_A - S_B$$

$$\text{誤差平方和}：S_E = S_T - (\text{各因子と交互作用の平方和の和})$$

A.5. プーリングの目安

(1)「<u>$\boldsymbol{F_0}$ **の値が 2 以下**</u>」ないしは「有意水準 20% 程度で有意でない」ならプールする。

(2) 因子をある程度絞り込んだ後の実験（例えば、一元配置法や二元配置法等）では、主効果はプールせず、交互作用のみをプーリングの考慮の対象とする。

(3) 因子を絞り込もうとする実験（多元配置法・直交配列表・分割法等）では、主効果をプールしてもよい。ただし、交互作用をプールしない場合は、対応する主効果はプールしない。例えば、$A\times B$ をプールしないなら、A も B もプールしない。

A.6. 前回の演習問題９について

疑問点：なぜ、区間推定は、ｔ分布か？

まず、データの構造式は、以下の通りです。

$$x_{ij} = \mu + a_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a; \quad j = 1, 2, \cdots, n \tag{1}$$

これは、言い換えると、$x_{ij} \sim N(\mu_i, \sigma^2),\ \mu_i = \mu + a_i$ と考えられます。

ここで、ある水準 A_i におけるデータ n 個の平均を考えると、

$$\overline{A}_i \sim N(\mu_i, \sigma^2/n) \tag{2}$$

が導かれます。これを標準化すると、

$$X = \frac{\overline{A}_i - \mu_i}{\sqrt{\sigma^2/n}} \sim N(0, 1) \tag{3}$$

が得られます。

一方、各群における平方和 S_i を誤差分散 σ^2 で割った量は $\dfrac{S_i}{\sigma^2} \sim \chi^2(n-1)$ です。各群同士が互いに独立ですから、再生性によって、これらの和である誤差平方和 S_E を誤差分散で割った量は、

$$Y = \frac{S_E}{\sigma^2} \sim \chi^2(a(n-1)) \tag{4}$$

が分かります。

ちなみに、ｔ分布の定義は、

$$X \sim N(0, 1),\ Y \sim \chi^2(k),\ X \perp\!\!\!\perp Y \quad \text{のとき、} t = \frac{X}{\sqrt{Y/k}} \sim t(k) \tag{5}$$

ですから、(5) 式に、(3),(4) 式を、$k = a(n-1) = \phi_E$ を代入すると、以下の式を得ます。

$$t = \frac{\overline{A}_i - \mu_i}{\sqrt{V_E/n}} \sim t(\phi_E) \tag{6}$$

この (6) 式を用いて、検定や、区間推定が行われます。

理由は、以上です。