

# 第 10 章

## 実験計画法

**実験計画法**：あるべき要因効果を効率よく、誤らずに検出するためには、どのように実験を計画すべきかをまとめた工夫集。

**フィッシャーの三原則**：誤りのない解析結果を得るためには、どのように実験の場を配置すべきかについて整理した3つの原則。

**直交計画法**：効果の有無を検証したい要因候補が多数ある場合に、直交表を用いて実験の数を減らす効率化手法。

### 10.1 フィッシャーの三原則

第8章で学んだ分散分析は、誰にとってもわかりやすい、大変洗練された手法でした。しかし、それに用いる肝心のデータそのものが不良品では、いくら優れた手法を使ってもそこからは誤った結果しか得られません。つまり、分散分析にかけるデータは、適切に計画された実験に従って収集された、良品でなければならないのです。

一方で、効果の有無を検証したい要因の候補が多数あると、実施すべき実験の組合せは膨大になってしまいます。このような場合、何らかの方法で実験数を減らす必要が出てきます。

そうした、良いデータを効率的に収集し、正しい解析結果を得るために工夫集が**実験計画法** (experimental design) です。つまり、成功する実験を考える統計学の一分野といえるでしょう。

ですので、「実験計画法」という分析手法があるわけではありません（基本的に、実験計画法では分散分析を中心に考えます）。

実験数を減らす効率化の工夫については後半（直交計画法）で扱うことにして、まず本節では、実験を成功させるために従うべき原則から解説させていただきます。

さて、そもそも実験における「成功」とは何でしょうか？

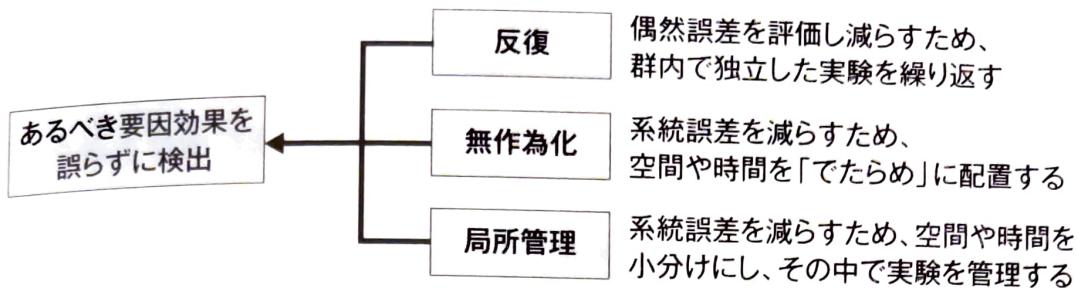


図 10.1 フィッシャーの三原則

それは失敗しないこと、つまり効果がないのにあると誤ったり、あるべき要因効果を見逃したりしないことです。これは、第一種の過誤と第二種の過誤を、どちらも極力犯さないようにすることにはかなりません。実験計画法の生みの親であるフィッシャーは、そのためには、どのように実験の場（空間や時間、検査員など）を配置すればよいのかを3つの原則として整理しました（**フィッシャーの三原則**：Fisher's three principles）。

それは、図10.1の反復の原則、無作為化の原則、局所管理の原則です。

反復とは群（水準）内で実験を繰り返すこと、無作為化とは何かの順番で実験しないこと、局所管理とは実験の場を小分けにしてその中で実験を管理することです。これらの原則に従った実験を計画・実施することではじめて精度の高い、誤差からの偏った影響を軽減させたデータを収集できるのです。

それでは1つずつ解説していきましょう。

## 10.2 原則その1：反復

**反復** (replication) とは、要因の水準（処理条件）ごとに形成される群の中で、独立した実験を繰り返し、データを複数個収集することです。

まず、第8章で学んだ、対応のない一元配置分散分析の検定統計量  $F$  を思い出してください。それは、下記のように、分子は群間変動を自由度（群数-1）で割ったもの、分母は群内変動を自由度（データ総数-群数）で割ったものでした。この統計量の式を使って反復の重要性を説明しましょう。

$$\text{分散分析の検定統計量 } F = \frac{\text{要因分散}}{\text{誤差分散}} = \frac{\text{群間変動} / (\text{群数} - 1)}{\text{群内変動} / (\text{データ総数} - \text{群数})}$$

↑ 精度が向上  
↓ 变動が大きくなる  
↑ 誤差評価に不可欠  
↑ 自由度が大きくなる

群内で実験を反復させず1回だけだと、この統計量はどうなるでしょうか？ 群内のデータが1つでは、分母の群内変動が計算できませんね。分散分析は、要因効果の有無を、誤差分散に対する要因分散の相対的な大きさで判断するので、誤差分散がわからなければ検定できません。つまり、実験結果から仮説を統計的に検証したいならば反復は欠かせないのでしょう。

また、反復の回数を増やすと良いことがあります。まず、群間変動が大きくなるためF値が大きくなり、帰無仮説の棄却域に入りやすくなります。また、群内のデータ数が多くなるため、（群内・群間変動の計算に使う）群平均の誤差が小さくなり、統計量の“精度”が向上します。それを反映して、分母の自由度が大きくなり、判定に用いる限界値が小さくて済みます（検出されやすくなります）。

いかがでしたか？ 反復がとても大切なことをご理解いただけましたでしょうか？

反復は原理的に難しいわけではないですが、ほかの2原則とそろえて圃場実験の図で説明しておきましょう。図10.2は、施肥量を3段階（無肥・少肥・多肥）設定して、穀物の収量に対する効果を検証する実験圃場を上空から見たものです。先ほどの復習になりますが、左の反復させていない実験では、いくら観測された収量が水準1<水準2<水準3となっても、処理内のバラツキ（誤差）が計算できないため分散分析を実施できません。いいかえれば、水準による収量の違いがあっても、それが誤差（偶然）の範囲のうちなのか、誤差とはいえないぐらいの大きな違いなのかが区別できないのです。ですから、右のように、それぞれの水準で実験を繰り返すのです。なお、この事例では群あたり3回ずつ反復させていますが、どのぐらい反復させるべきかについては、次に解説する検出力分析で計算できます。

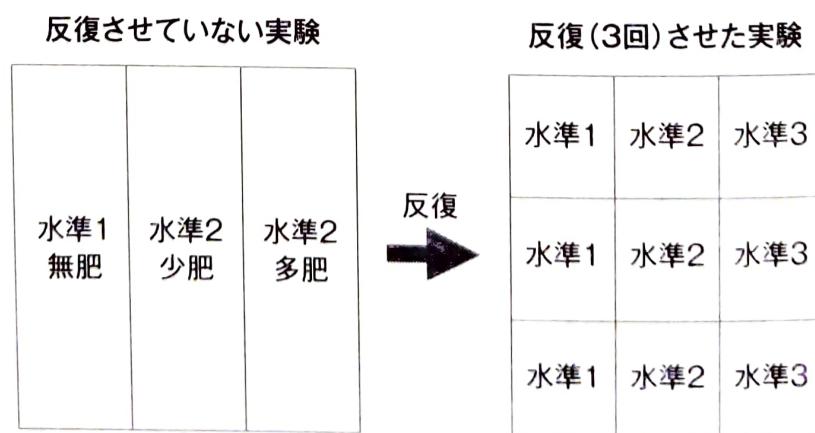


図 10.2 反復の原則（圃場の例、処理数=3）

## ● 反復数の計算（検出力分析）

群あたりどのくらい反復させれば良いのかは、総標本サイズ（反復数×群数）の決め方ですので、検出力分析（事前分析）の出番です。節6.5で学んだように、必要な標本サイズは、効果量と有意水準と検出力の3要素から逆算できます。ただし、分散分析の場合には、群数も非心度に影響を与えるため、図10.3のように、それを加えた4要素から反復数を求ることになります。

有意水準 $\alpha$ は0.05とするのが一般的で、検出力はコーベンという統計学者が0.8ぐらいは欲しいといっています。群数は処理水準数ですので、どのような実験をしたいのかが決まっていれば、自明でしょう。

となると、やはり問題は効果量です。分散分析の効果量の推定式（真の効果量は未知です）にもいろいろあるのですが、検出力分析用の無料ソフトG\*powerでは、次式で求める $\hat{f}$ が使われます。この式をよく見ると、分散分析の検定統計量Fから自由度の影響を取り除いた内容になっています。

$$\text{分散分析の効果量 } \hat{f} = \sqrt{F \text{ 値} \times \frac{\text{群数} - 1}{\text{データ総数} - \text{群数}}} = \sqrt{\frac{\text{群間変動}}{\text{群内変動}}}$$

類似の既往研究があるならば左辺のF値と自由度を使った式で、予備実験を行えたならば右辺の2つの変動を使った式で推定することになるでしょう。しかし、なんの手がかりもない場合には、仕方がないので適当な効果量を入力するしかありません。効果量 $f$ には、大きな効果量として0.4、中程度の効果量として0.25、小さな効果量として0.1が提案されています。

それでは、G\*powerを使って反復数を求めてみましょう。対応のない一元配置分散分析は、[F tests] の中の [ANOVA: Fixed effects, omnibus, one-way]になります。事前分析を意味する [A prior: —] を選択すると、図10.4のような画面が現れますので、[Input Parameters] に4要素（パラメータ）の値を入

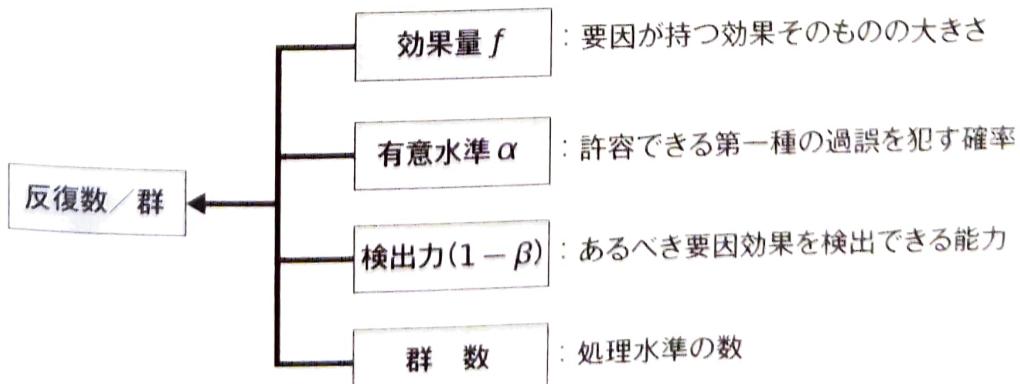


図 10.3 反復数を決める4つの要素

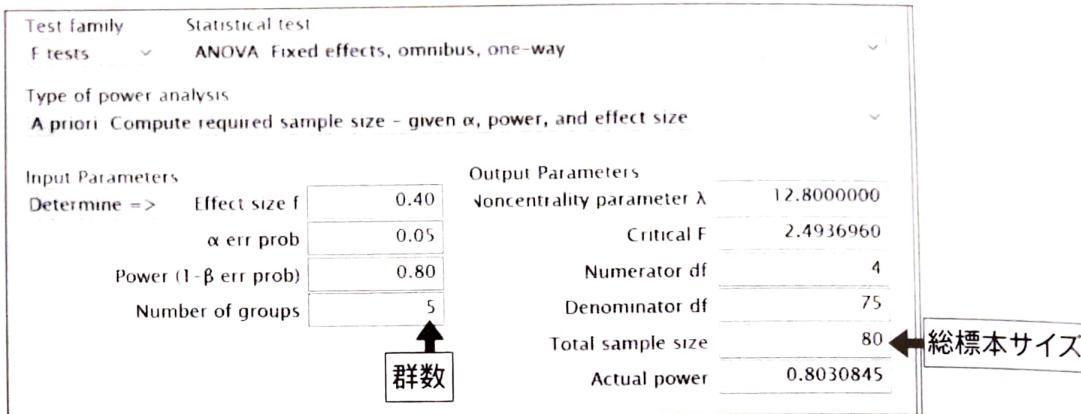


図 10.4 G\*powerによる反復数の計算（一元配置分散分析）

力します。有意水準 [ $\alpha$  err prob] には 0.05、検出力 [Power (1- $\beta$  err prob)] には 0.8 を入力して、効果量 [Effect size f] には大きな効果量が得られると想定して 0.4 を入力しておきましょう。また、群数 [Number of groups] は、5 の場合を計算してみましょう。最後に、[Calculate] を押すと、[Total sample size] は“80”と出てきます。つまり、大きい効果量が想定できるとき、有意水準  $\alpha$  が 0.05 の下で、0.8 の検出力を得るために、5 群合わせた総標本サイズとして 80 程度必要だということです。これは、群あたりでは、独立した実験を 16 回反復させなければならないことを意味しています。

いかがだったでしょうか？ みなさんが想像していた数よりも意外と多かったのではないかでしょうか？ 推定される効果量が小さければ、もっと反復させる必要が出てきます。この計算を毎回するのも大変でしょうから、有意水準が 5 % の下で検出力 80 % を得るための反復数について、表 10.1 に示しておきます。

表 10.1 必要な反復数/群の目安（対応のない一元配置分散分析）

	効果量の大きさ (有意水準 $\alpha = 0.05$ , 検出力 = 0.80 の場合)		
	小 ( $f = 0.10$ )	中 ( $f = 0.25$ )	大 ( $f = 0.40$ )
2 群	393	63	26
3 群	323	53	22
4 群	274	45	19
5 群	240	40	16
6 群	215	36	15
7 群	196	33	14

## ● 疑似反復

ここまで扱ってきた反復とは、同じ処理水準内で独立した実験を繰り返し、異なる個体から観測することです。処理を1度しか行わず、同じ個体から繰り返し観測することではありません。もし、後者を反復として扱ってしまうと(疑似反復, pseudoreplication), 本来よりも検定統計量の値が大きくなり、限界値も大きい自由度用の小さな値を使ってしまうため、本当は有意でない要因効果を有意と判定してしまう第一種の過誤確率が高くなります。

例えば図10.5は、植物の成長に関する実験を行う事例です。左は同じ区画内で4本の植物を観測して反復数を4としようとしていますが、植物の成長にもっとも影響のある土壌が同じでは独立しているとはいえないね。ある植物の成長が良い場合には、土が肥えているでしょうから、ほかの植物の成長も良くなっている可能性があります。ですから、右のように、屋外ならばコンクリートで仕切ったり、室内ならば別々のポットを使ったりして栽培しなければなりません。研究分野によって、疑似反復が発生してしまう状況は様々ですが、群内のデータ同士の関連性を考えれば判定できるでしょう。

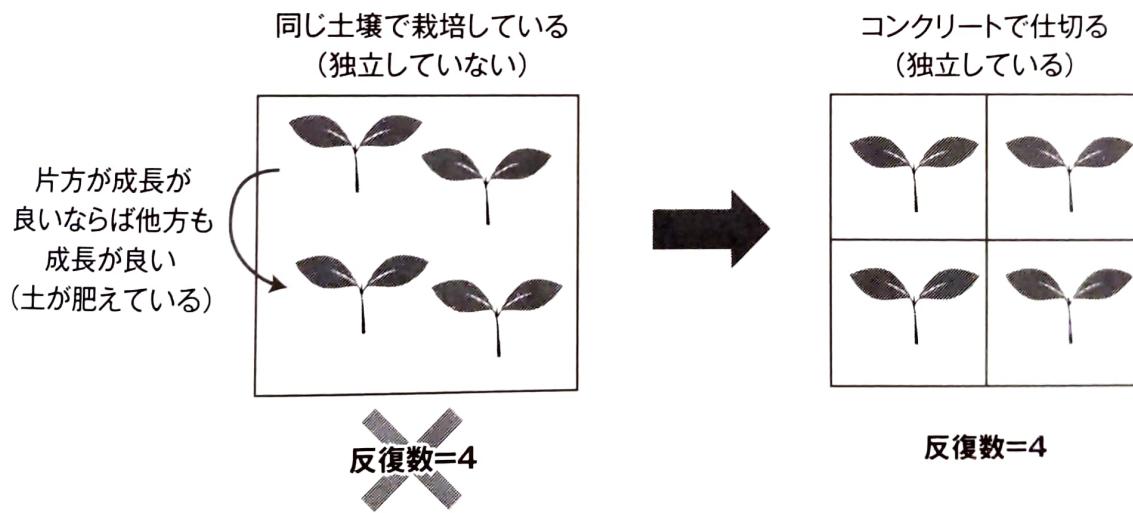


図 10.5 疑似反復の例

## 10.3 原則その2：無作為化

分散分析において検定結果を誤るとしたら、それは、実験結果に誤差の効果が“方向性を持って”(系統立つとか、偏るともいいます)入り込み、検定統計量が本来よりも大きくなる(あるいは小さく)なってしまう場合です。

本来、分散分析における誤差は、測定誤差のように、効果に方向性を持たない**偶然誤差**(節3.3)でなければなりません。実験の場(空間や時間など)の順番に伴って効果に方向性を持つような**系統誤差**は、統計量を歪めて誤った検

定結果を導いてしまうのです。わかりやすいように、反復の原則のところでも使った圃場図で解説しましょう。

図10.6左は、施肥量を水準1<水準2<水準3と設定し、穀物収量に対する効果を検証しようという実験です。一見すると反復の原則に従っていて問題なさそうですが、このままでは施肥効果がなくても“ある”と誤判定してしまう可能性が高いです（第一種の過誤確率の増大）。というのも、南側に林があるため、日照量が北側に行くに従って多くなってしまっているからです。つまり、実験目的からすれば、日照量は誤差であるにも関わらず、施肥量の効果と同じ方向性を持って結果に入り込んでしまっているのです。このような実験を実施してしまうと、観測された穀物収量データには施肥効果と日照効果が交じってしまい、後からどちらがどのくらいの効果だったのかを分離することはできません。こうした状況を**交絡**（confounding）と呼びます。

それを検定統計量の式で表したのが図10.6右です。分母にあるべき日照量が方向性を持ったために分子に来てしまったことで、本来よりも検定統計量Fが大きくなっていることを示しています。もし、施肥量が（図とは逆で）北から南に行くに従って増えるように配置した場合、今度は施肥効果（要因）と日照効果（誤差）の方向が反対になるため、検定統計量が本来よりも小さくなり、あるべき施肥効果を見逃す可能性が高くなります（第二種の過誤確率の増大）。

さて対処法としては、系統誤差を発生させている原因である林を全て伐採するのがもっとも良いでしょうが、そうもいかないでしょう。そこで、交絡している日照量と施肥量のうち、どちらかの区画の配置（並び方）をでたらめにするのです。「でたらめに」というと、ちょっと聞こえが悪いので、作為性を無

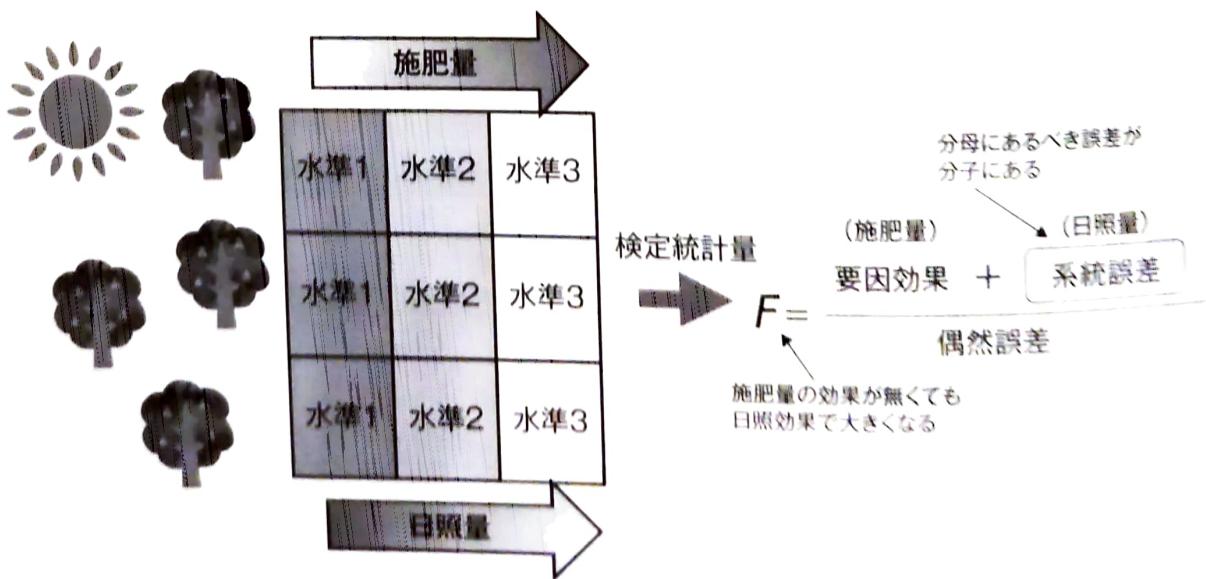


図 10.6 無作為化していない圃場実験（施肥水準数=3）

くすという意味で**無作為化**（randomization）と呼びましょう。日陰の入り方を無作為に並び替えることはできないので、この事例では施肥水準がその対象となります。図10.7左のように、施肥水準の区画の配置を無作為化すれば、日照効果は各施肥水準に対して方向性を持たず、均一に入り込むため、交絡は生じません。

図10.7右の検定統計量で説明すると、無作為化は、 $F$ 値が歪んだ原因である分子の系統誤差を、本来あるべき分母の偶然誤差に**転化**しているのです。 $F$ 値の歪みが修正されれば、分散分析で結果を誤ることはなくなり、**正確さが向上する**というわけです。

さて、事例では圃場の実験区画という「空間」を使って説明しましたが、時間や人（作業者・検査官など）の順番も対象になります。例えば、冬ならば暖房を使うため、時間が経つごとに実験室の温度も徐々に上がっていくでしょう。その中で、pH（要因）と酵素反応速度（特性）を一元配置実験する場合、時間とともに変化する室温は誤差なのにもかかわらず反応速度に影響してしまいます。そこで、（室温を並べ替えるのは難しいため）pHの順番を無作為化すれば、室温の反応速度への影響が平均化（系統誤差から偶然誤差へ転化）され、pHと反応速度との関係のみを正しく検証することができるのです。

このように、順番を入れ替えられる要因については、常に無作為化することを心がけるようにしてください。

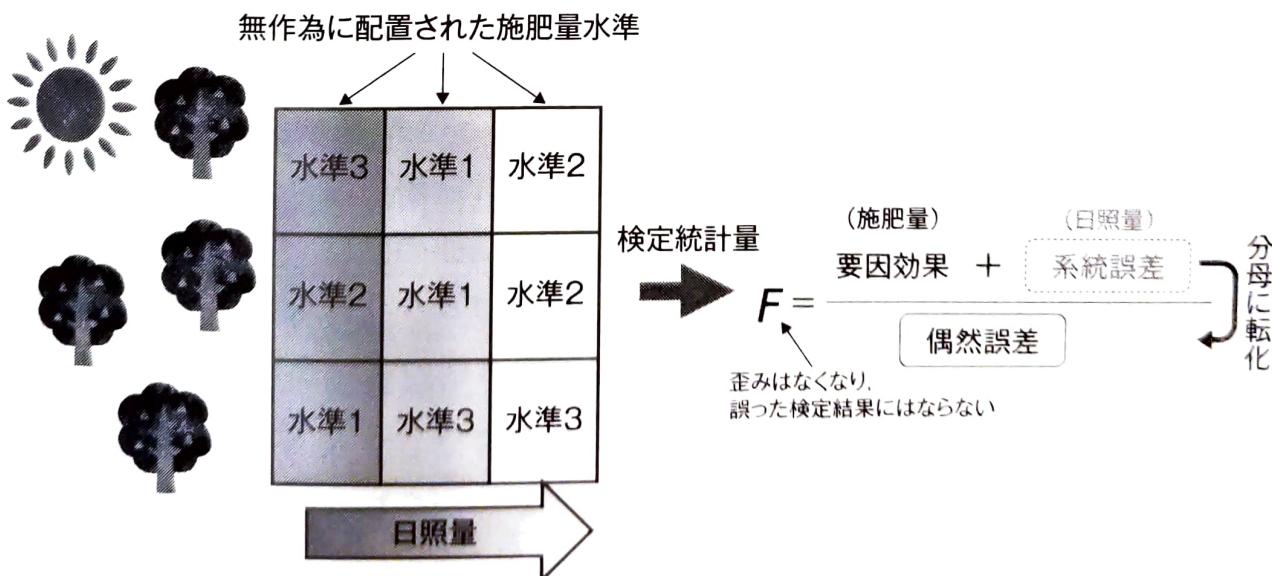


図 10.7 無作為化した圃場実験

## 10.4 原則その3：局所管理

**局所管理** (local control) は、無作為化と同様、系統誤差の対策ですが、より積極的にその影響を取り除きます（偶然誤差の軽減にも役立ちます）。

無作為化は、系統誤差を偶然誤差に転化することで、検定統計量の歪みを正す方法でしたが、次の2つの場面では、上手な系統誤差の対処法とはいえません。

1つは、(当たり前ですが) 交絡している要因水準のどれもが無作為化の難しい場合です。工場ならば製造ライン、農業ならば収穫日など、系統誤差を生むとわかっていても、順番を入れ替えるのが困難な要因は意外と多いものです。たとえ無作為化が可能な要因でも、反復数が多くなると、水準の取替え引っ越し替えを繰り返すのは容易なことではなくなるでしょう。

もう1つは、系統誤差の効果がとても大きい場合です。系統誤差が大きいと、たとえ偶然誤差へ転化できても、検定統計量の分母に来る誤差が大きくなりすぎて、要因効果を検出できなくなってしまいます ( $F$  値が小さくなってしまうため)。農業ならば圃場区画、聞き取り調査ならば調査員や訪問地域などがその典型でしょう。また、反復数の多い実験を計画すると、大きな実験の場（空間や時間）が必要になりますが、それに伴い様々な系統誤差が出てきてやはり無作為化では対応し切れなくなってしまいます。人の感覚を用いる官能検査でいえば、何度も実験すると、疲労や慣れが徐々に大きくなって、系統誤差となることは容易に想像できますね。

局所管理は、そのような（無作為化が難しいときや系統誤差がとても大きい）ときに導入することで、系統誤差を検定統計量から取り除く方法です。

局所管理の説明をする前に、いろいろと導入場面の話ばかりしてしまいましたが、一言でいえば「実験の場を小分けに（ブロック化）して、その小分けにした局所の中で実験を一通り管理する（全ての水準の実験を行う）こと」です。こちらも圃場実験の事例で解説しましょう。

図10.8の左を見てください。広い圃場全体を使って実験を管理しようとするから、場所によって日照量が違う（林による系統誤差が発生する）のです。しかし、例えば南の3区画を1つのブロック、同じように真ん中の3区画と北側の3区画をそれぞれ1つのブロックと、圃場を小分けすれば、同一ブロックの中の日照量はそれほど変わりませんよね。つまり、各ブロックの中で一通りの実験をこなせば、系統誤差の発生を抑えられる（正確さが向上する）というわけです。また、環境が均一になるということは、ブロックに付随した様々な

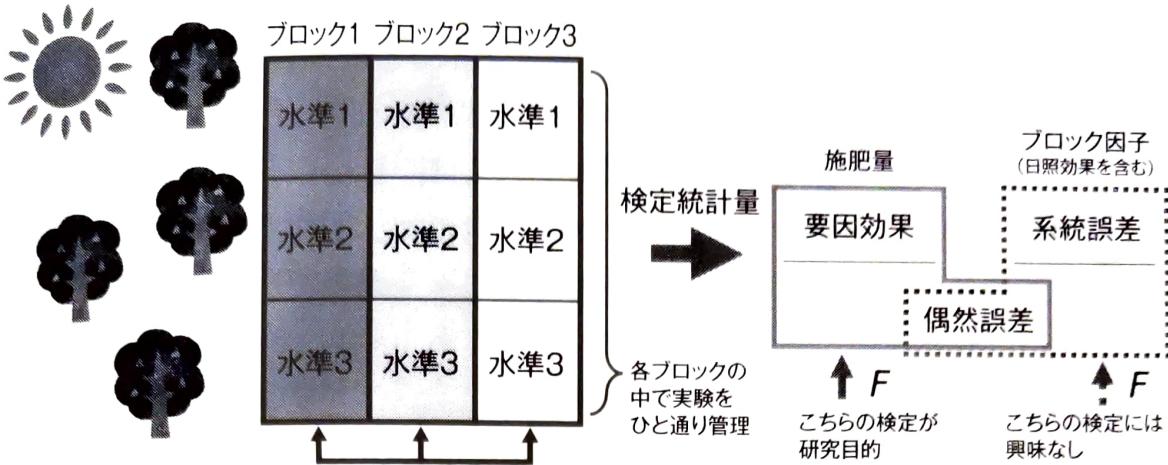


図 10.8 局所管理した圃場実験

偶然誤差の削減にもつながるため精度も向上します。

それでは局所管理を取り入れた実験データは、どのような分散分析にかけられよいのでしょうか？ 実は、第8章で学んだ「対応のある一元配置分散分析」(Excel分析ツールだと「繰り返しのない二元配置分散分析」)です。というのも局所管理とは、図10.8の右のように、ブロック差という系統誤差を1つの要因(ブロック因子)として捉えているからです。171ページの表8.3でいえば、AさんとBさん、それぞれがブロックということになります(個人差を1つの要因と考えます)。ですから、図8.8と図10.8右も同じことを意味しています。具体的には、図8.8の(自由度で割った)被験者間変動が図10.8右の系統誤差、同様に個人差が除去された群内変動が偶然誤差になります。

なお、ブロック因子の効果も検定することができますが、あくまで系統誤差の影響を取り除くのが局所管理の目的で、それを検出することには興味はないので、実際には不要でしょう(分析ツールで繰り返しのない二元配置分散分析を使うと、自動的に検定されてしまいますが……)。また、ブロック因子には、要因(事例でいえば施肥量)との間に交互作用のないものを設定します。もし、交互作用がありそうな場合には、もはやブロック因子ではなく普通の要因として設定し、繰り返しのある二元配置分散分析で検定する必要があります(もちろん反復させなければなりません)。

## ● ブロック化の対象

以上のように局所管理は、無作為化のように系統誤差の影響を平均化していくのではなく、ブロック因子として積極的に実験に取り

込むことで要因効果から分離させようという、強力な系統誤差の対処法です。

ここまで、局所管理を系統誤差の対処法の1つとして説明してきましたが、実験の場を小分けに（ブロック化）することは、方向性を持たない偶然誤差を減らすことにも役立ちます。圃場の事例でいえば、日陰の問題（系統誤差）以外にも、圃場全体で実験しようとすると、土壌の肥沃度や水はけの良さ、農道からの距離の違いなど、統制するのが困難な様々な誤差が出てきます。そしてそれらは方向性がある場合もあればない場合もあるでしょう。しかし、小分けにすることで、それらの影響は全て軽減されます（ブロック内ではそれほど条件が違わないため）。ですから、方向性はなくともブロック因子として効果が大きそうならば、空間や時間に限らずブロック化の対象としてください（ブロック因子が複数となっても構いません←後述の「ラテン方格法」）。

このように局所管理はとても優れた考え方ですが、なんでもかんでもブロック化の対象とすると実験や分析が複雑になってしまうので、影響の大きい順に2つ程度にして、残りは無作為化にまかせるべきでしょう。

また、ブロック因子を減らすテクニックとして、1つのブロック因子に複数の誤差要因をかぶせる方法があります。例えば圃場で、日照量のほかに担当作業者の違いによる影響が大きいと考えられるならば、ブロックごとに担当作業者を割り振れば、日照量の違いと担当作業者の違いを合わせて1つのブロック因子として扱えますね（これらは誤差なので交絡していても構いません）。

最後に、ブロック化の対象として、どのようなものが考えられるかを挙げておきましょう。

- **官能検査**：検査員
- **工場実験**：製造ライン、原料ロット、日、作業者、出荷ロット、作業時間帯
- **農場実験**：圃場の区画、植物工場の棚、果樹の個体、播種日、収穫日
- **アンケート（聞き取り調査）**：調査員、訪問地域、回答日

## 10.5 いろいろな実験配置

### ● 完全無作為化法と乱塊法

ここまでフィッシャーの三原則を解説してきましたが、実験の場を配置するときは必ず3つの原則全てに従わなければならないわけではありません。

もちろん、反復は誤差分散の計算に不可欠なので、データを分散分析にかけるならば実施しなければなりません。しかし、無作為化は、実験の順番が

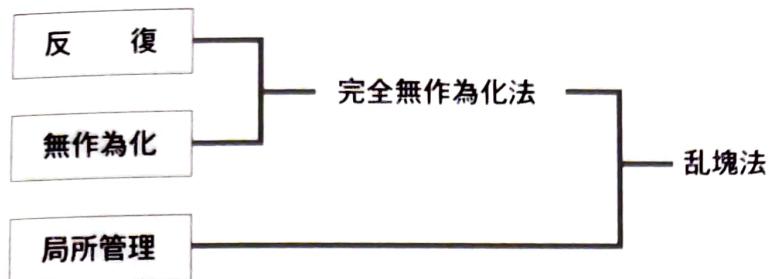


図 10.9 フィッシャーの三原則と実験配置法

(要因効果と同じ／反対の) 方向性を持っていないと断言できるならば不要です。ただし、空間にしろ、時間にしろ、そのままの順番で実験すると、少なからず系統誤差を生じる恐れをはらんでいるものです。ですから無作為化が可能ならば、系統誤差予防という意味でも実施しておいた方が良いでしょう。このように、反復と無作為化の2つの原則に従った実験管理を**完全無作為化法** (completely randomized design) と呼び、もっとも基本的な実験計画としてよく用いられています（図10.9）。

圃場実験の例だと図10.10左のように、圃場全体で実験を管理するのが完全無作為化法です。この完全無作為化法に、局所管理を取り入れた（つまり三原則全てに従う）実験計画を**乱塊法** (randomized block design) と呼びます。図10.10の右のように、ブロック単位で実験を管理しています（反復は圃場全体、無作為化はブロック単位）。乱塊法は、無作為化だけでは対処できない規模の大きい（反復の多い）実験や、個人差・実験日など大きな系統誤差が生じる可能性のあるときに向いています。注意しなければならないのは、大した系統誤差でないのに小分けにすると、誤差分散の自由度が小さくなってしまうため、むしろ検出力が弱まってしまうことです。

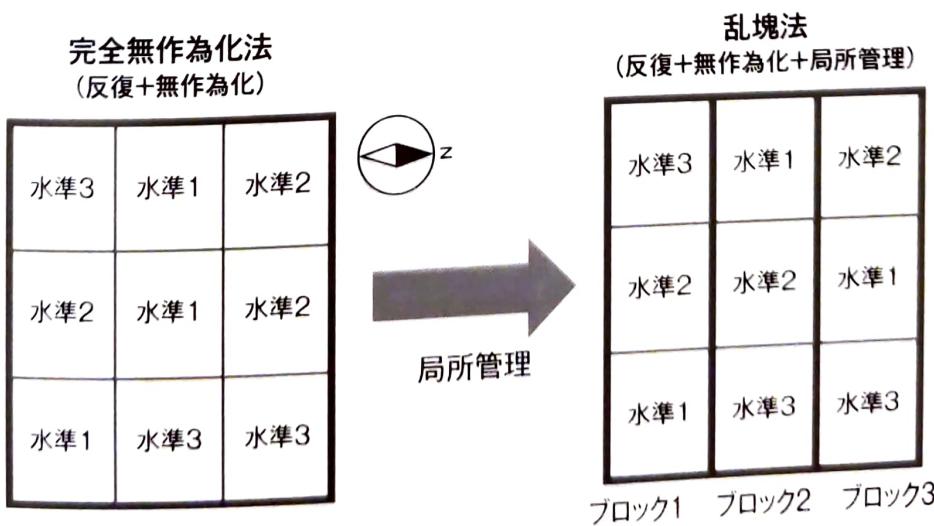


図 10.10 完全無作為化法と乱塊法

## ● ラテン方格法

乱塊法を発展させ、2種類のブロック因子を導入した実験配置を**ラテン方格法** (Latin square design) と呼びます。ラテン方格とは、 $n$  行× $n$ 列の表の各行列に、 $n$  個の異なる数字や記号が1回だけ現れるようにした表のことです。少し前に「数独」という鉛筆パズルが流行しましたが、あれこそラテン方格の応用例です。また、次節で解説する直交表もラテン方格を発展させたものです。

このラテン方格を使うと、2種類のブロック因子を導入できます。

図10.11の左は、圃場の南北（行）方向に系統誤差（例えば日照量）が入る状況において、東西（列）方向のブロック因子を1つ導入した乱塊法の圃場実験（図10.10右と同じ）です。図の右は、南北に加えて東西（列）方向にも系統誤差が入る状況において、南北（行）方向のブロック因子を導入したラテン方格法です。列のブロック内だけでなく、行のブロック内でも無作為に施肥水準が並び替えられているのがわかりますでしょうか（どの行と列を見ても同じ施肥水準の数字はありませんね）？なお、事例ではどちらのブロック因子も区画という空間を対象としていますが、片方だけ時間（作業日など）とすることなどもできます。また、ラテン方格は行列の数が同じでなければならないので、水準数と同じ数だけのブロックが必要となる点には注意してください。

ラテン方格法による実験で収集されたデータは、1つの要因に加えてブロック因子が2つですので三元配置分散分析となります（ただし、検定結果に興味があるのは要因効果のみです）。ですから、Excel分析ツールでは扱えませんが、Rコマンダーならば多元配置分散分析というメニューで分析可能です。

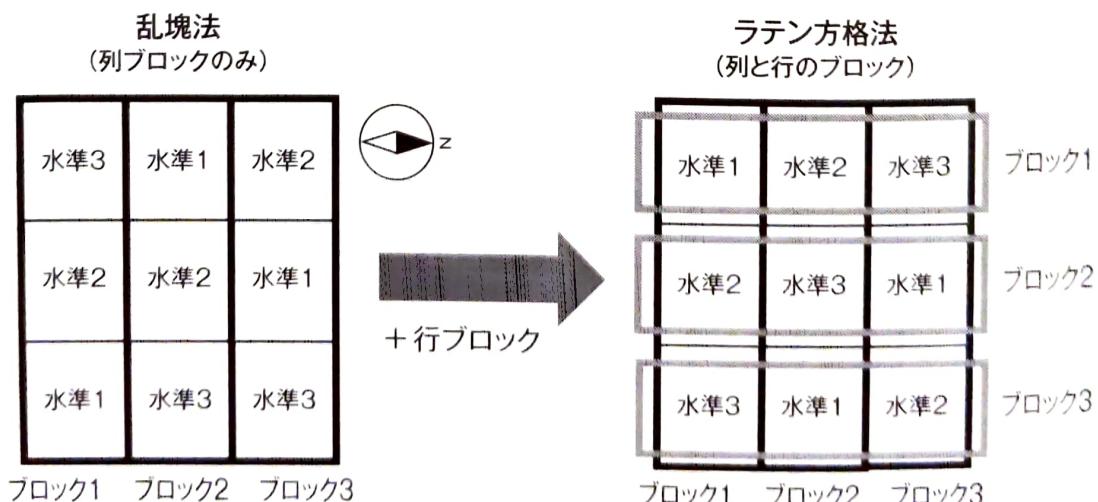


図 10.11 ラテン方格法

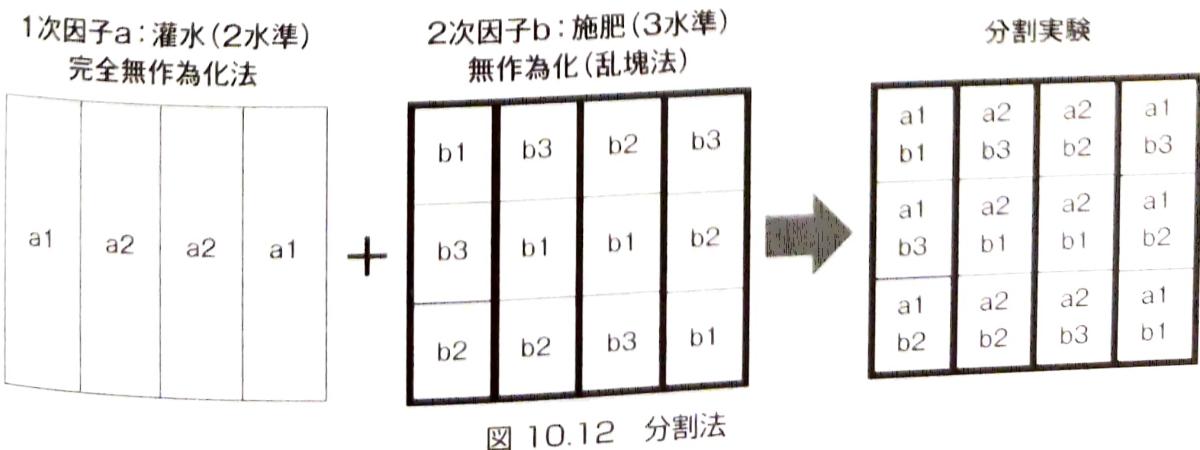
## ● 分割法

実験で検証したい要因が複数ある場合、全要因を同時に無作為化して配置するのは難しい場合があります。そのようなとき、水準を入れ替えるのが面倒な要因（**1次因子**）から配置して、次に1次因子の各水準内で、水準入れ替えが容易なほかの要因（**2次因子**）を配置すると効率的です。このように、実験配置を一度にではなく、段階的に分割して行う方法を**分割法**（あるいは分割区法：split-plot design）と呼びます。

図10.12は、屋外の圃場で、灌水（水管理）と施肥という2つの要因が収量という特性に与える効果を検証する実験です。施肥量の水準を狭いブロックの中で変化させるのは大した手間ではないですが、灌水はそうはいきません（大変な労力・時間・費用がかかります）。そこで、最初に灌水（1次因子）の水準配置を圃場全体で行ってしまいます。この図では、2水準を無作為に2反復させて完全無作為化法となっていますが、乱塊法やラテン方格法を採用することもあります。

次に水準変更が容易な施肥要因（2次因子）を、1次因子の各水準内で無作為に配置します。このように、2次因子の配置は単なる無作為化が多いのですが、1次因子の水準がブロックとなるので、実質的には乱塊法になります（この段階でブロック因子をもう1つ導入してラテン方格とすることもあります）。図10.12を見ていただければわかる通り、2要因同時に無作為化する場合には灌水水準を何度も変更しなければならなかったのに対して、分割法を使ったことで2回で済みました（ $a_1 \rightarrow a_2$  と  $a_2 \rightarrow a_1$ ）。なお、この事例では2次までですが、必要ならば3次因子や4次因子……と設定します。

乱塊法と何が異なるのかわかり難かったかもしれません。基本的に乱塊法は実験要因と系統誤差を生じるブロック因子とを同時に配置する方法なのに対して、分割法は複数の実験要因を段階的に配置する（1次の実験要因の水準を



2次ではブロックと見なし、検定も行う) 方法であるということです。

さて、分割法で計画された実験（**分割実験**）で観測されたデータの分析ですが、1次と2次で工程が全く異なる実験であることを考えれば、1次因子の配置で発生する誤差と2次因子の配置において発生する誤差とは異なるため、本来は1次と2次を分けて分散分析にかけなければなりません。ただし、計算が複雑になるので（本書でも扱いません）、図10.12の一番右のデータに対する簡易的な分析として、単なる「繰り返しのある二元配置分散分析」を実施してもよいでしょう（水準組合せあたりの繰り返し数は2回）。

## 10.6 直交計画法 — 実験の効率化 —

### ● 直交（配列）表

ここまででは、あるべき要因効果を誤らずに検出するためには、どのように実験の場を配置すべきかを学んできました。しかし、実験計画法にはもう1つ目的があります。実験の効率化（実験を小さくすること）です。

例えば、いちから手探りの初期実験の場合、主効果の存在を分散分析で検証したい要因候補がたくさんあるでしょう。しかし、たとえ2水準しかない要因でも7つあれば、組合せ数は $2^7$ ですから、全部で128回（種類）の実験を実施しなければなりません。手軽に実施できる実験なら問題ないでしょうが、1回の実験に結構な時間や費用、労力、スペースを必要とする場合、全てを実施するのは極めて困難です。

そこで、**直交表**（orthogonal array）という魔法の表を使って、一部の必要な実験だけを実施するようにするのが**直交計画法**（orthogonal design）です。つまり、実験をスクリーニング（選別）して効率化を図るのです。

直交表（直交配列表ともいいます）にもいろいろあるのですが、例えば表10.2を使うと、2水準の要因が7つある場合でも、なんと8種類の実験で済んでしまうのです。なぜ、このようなことが可能なのでしょうか？

表10.2は、正式には $L_8(2^7)$ 直交表といいます（以降、 $L_8$ 表と略します）。名称の読み方から説明すると、Lはラテン方格の頭文字で、添字の8は水準の組合せ（実験の種類）の数、括弧の中の2は水準数、指数の7は直交表に割り付けられる要因（誤差も含みます）の数です。つまり直交表の列は要因、表中の数字は水準、行は水準の組合せ（実験の内容）を表しています。例えば実験番号1の（行の）内容は、7つの要因の水準が全て1となります。

表 10.2  $L_8(2^7)$  直交表

実験番号	[1]	[2]	[3]	[4]	[5]	[6]	[7] ← 要因の割付 (誤差含む)
実験の種類 (水準の組合せ)	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1 ← 水準
8	2	2	1	2	1	1	2

水準ごとに4回反復している  
ので列の変動を計算できる

The diagram shows the fractional factorial design  $L_8(2^7)$ . It consists of the letter  $L$ , the number  $8$ , and the expression  $(2^7)$ . Four arrows point to different parts of the expression: one from the text "水準の組合せ (実験の種類) の数" (Number of factor combinations (number of experiments)), one from the text "列(要因)の数" (Number of columns (factors)), one from the text "ラテン方格" (Latin square), and one from the text "水準の数" (Number of levels).

列[1]や[2]を（縦に）見てみると、いずれも水準1の実験が4回、水準2の実験が4回反復されているため、各列に割り付けられた要因の変動を、実験で観測されたデータから計算できるというわけです（反復数が異なると、水準間の分散が異なったり、後述の直交を保てなくなったりします）。

さて、肝心の直交表を使うと実験数を間引ける理由ですが、それは「大半（あるいは全部）の交互作用の検証をあきらめている」からです。

（あるいは全部）の交互作用の検証

例えば、 $L_8$  表は、もともと 2 水準の要因が 3 つの三元配置が基本です。要因を  $a, b, c$  とすると、交互作用は全部で  $a \times b, a \times c, b \times c, a \times b \times c$  の 4 つです。もし、主効果に加えて交互作用の存在を全て検証したいならば、3 要因です。2 水準の全組合せ、つまり  $2^3$  で 8 種類の実験を実施しなければなりません。しかし、そもそも交互作用は、実験の目的からすると全てが重要というわけでは

ないですし、むしろあると困る場合が多いでしょう。また $a \times b \times c$ などという高次の交互作用は、あったとしても極めて小さくなります。そこで、それほど重要ではない交互作用や高次の交互作用の検証は潔くあきらめる（存在しないと仮定する）のです。そうすれば、水準の組合せのうちのいくつかは不要となり、実験が間引けるというわけです（実際には、あきらめた交互作用の代わりに第4、第5の要因を割り付けます）。

## ● いろいろな直交表

本書では、2水準の要因だけを配列させた**2水準系直交表**の代表格であるL<sub>8</sub>表しか扱いませんが、ほかにも、もっと多くの要因を割り付けられるようにした大きな表や、3水準の要因だけを配列させた**3水準系直交表**、2水準と3水準の要因を混ぜて配列させた**混合系直交表**があります。

4水準や5水準の表はないの？と思われるかもしれません、大抵の主効果は、水準1から水準2になるに従って特性値が上がるとか下がるとかの単純な傾向であることが多いため、2水準だけでも十分役に立つのです。さらに3水準ならば、上がって下がる、あるいは下がって上がるなどの複雑な効果まで捉えることができます。ただし、どうしても4水準の要因を導入したいという場合には、2水準系の2列を統合して1つの列として割り付ける方法（**多水準法**）もあります。

さて、こうした直交表は自作することができますが、大きな表になると、直交するように数字を配列するのはなかなか難しいので、よく使われる直交表は公開されています。本書でも次に挙げる代表的な直交表を付録Ⅷ～Xに掲載していますので、検証したい要因数や交互作用の有無に合わせて選ぶことができます。ただし、2水準系や3水準系では一部の交互作用を検証できますが、混合系では全くできない点に注意が必要です。

- **2水準系直交表** : L<sub>4</sub>(2<sup>3</sup>), L<sub>8</sub>(2<sup>7</sup>), L<sub>16</sub>(2<sup>15</sup>)
- **3水準系直交表** : L<sub>9</sub>(3<sup>4</sup>), L<sub>27</sub>(3<sup>13</sup>)
- **混合系直交表** : L<sub>18</sub>(2<sup>1</sup> × 3<sup>7</sup>), L<sub>36</sub>(2<sup>11</sup> × 3<sup>12</sup>)

## ● 要因の割付

直交表によって交互作用を検証できるものとできないものがあるのは不思議に思われるかもしれません、それは交互作用が現れる列を配列できるかどうかという、単に技術的な問題です。つまり、2水準系や3水準系の場合には、特定の列に交互作用が現れるように配列できるのですが、混合系の場合にはそれ

ができないのです（交互作用は各列に均等に配分されます）。

ですから、直交表に要因を割り付けて実験を計画する際、混合系は自由に割り付けられるのですが、2水準系や3水準系の場合は交互作用が現れる列を避けなければなりません。L<sub>8</sub>表を事例に、どのように要因を割り付ければよいのかを解説しましょう。

既述したように、L<sub>8</sub>表は2水準の要因が3つの三元配置から出発しています。図10.13の一番上がその基本形です。なお、要因が複数ある場合、「この主効果の存在は必ず検証したい」というものから「割り付けられるならば、とりあえず検証しておくか」というものまでいろいろでしょう。ですから、ここではaを一番重要な要因として、b、cと行くに従ってその順位は下がるものとします。

普通はもっとも検証したい重要な要因を最初に（左の列から）割り付けるでしょうから、列[1]にaを、列[2]にbを割り付けることになります。次に列[3]ですが、要因cよりも、まずはaとbの交互作用が見たいだろうということで、列[1]と列[2]の交互作用が現れるように配列されています。表10.2で、水準1を-1、水準2を1に置き換えてみれば、列[3]が列[1]と列[2]を乗じた列であることがわかると思います。よって、列[3]には要因は割り付けず（a×b用に残しておきます）、列[4]に要因cを割り付けます。同様に、列[5]には列[1]と列[4]の交互作用（a×c）、列[6]には列[2]と列[4]の交互作用（b×c）、列[7]には列[1]と列[2]と列[4]の交互作用（a×b×c）が現れるように配列されています。

ただし、このままでは要因は3つしか割り付けられていないので、実験はスクリーニングされていません。そこで、検証不要な交互作用を決めて、その列に第4の要因、第5の要因……を割り付けて行くのです。また、分散分析を実

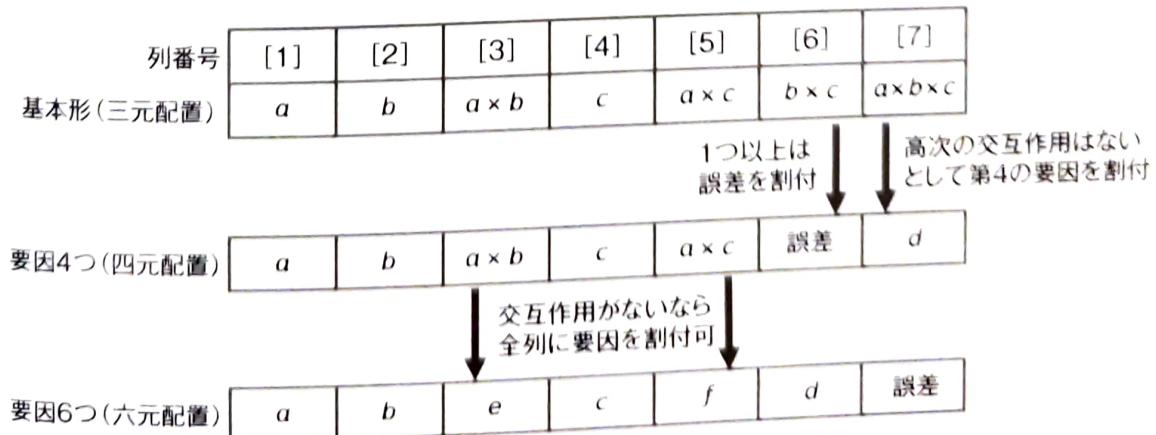


図 10.13 L<sub>8</sub>(2<sup>7</sup>)直交表への割付事例

施するには誤差（検定統計量  $F$  の分母）が必要なので、最低1列を誤差（群内変動）の計算用に残しておく必要があります（2列以上を誤差としても問題ありません）。もちろん、予備実験や類似の既往実験から誤差を計算できる場合には、7列全てに要因や交互作用を割り付けられます。

さて、図10.13の真ん中が、第4の要因  $d$  を導入するとき（四元配置）の典型的な割付パターンです。列[7]の  $a \times b \times c$  という高次の交互作用は存在しても極めて小さいでしょうし、あっても使いものになりませんので、代わりに要因  $d$  を割り付けるのです。残るは列[5]と[6]なので、どちらかを誤差とします。普通は  $b \times c$  よりも  $a \times c$  の方に興味があるでしょうから、列[5]を  $a \times c$  として、列[6]を誤差とするのが一般的でしょう。

なお、どのような交互作用も存在しないと仮定できるならば、図10.13の一番下のように、誤差用に1列を残しておけば、そのほかの6列を全て要因に割り付けることができます。ただし、交互作用があったら、割り付けた要因の主効果と交絡する危険があります。

## ● 直交計画と分散分析

図10.13真ん中の典型的な四元配置の割付パターンで実験を計画してみましょう。表10.3は、 $L_8$ 表から列[1][2][4][7]を取り出し、それぞれに要因  $a$  から  $d$  を割り付けたものです。列[3][5][6]は交互作用と誤差の列ですので、検定統計量の計算には使いますが実験の計画には使いません。また、最右列の  $x_1 \sim x_8$  は、8種類の実験を実施して観測したデータで、 $\bar{x}$  は総平均です（どちらにも具体的な値が入ります）。

表 10.3  $L_8$  表を使った四元配置実験

列番号 実験番号	[1] a	[2] b	[4] c	[7] d	データ $x_1$ $x_2$ $x_3$ $x_4$ $x_5$ $x_6$ $x_7$ $x_8$	$\bar{x}$
1	1	1	1	1	$x_1$	
2	1	1	2	2	$x_2$	
3	1	2	1	2	$x_3$	
4	1	2	2	1	$x_4$	
5	2	1	1	2	$x_5$	
6	2	1	2	1	$x_6$	
7	2	2	1	1	$x_7$	
8	2	2	2	2	$x_8$	$\bar{x}$

復習：群間変動は、  
反復数 ×  $\sum (\text{群平均} - \text{総平均})^2$

[群平均  $\bar{x}_1 - \text{総平均 } \bar{x}$ ]  $^2$

これらの和に4(反復数)を  
乗すれば群間変動

[群平均  $\bar{x}_2 - \text{総平均 } \bar{x}$ ]  $^2$

水準ごとにまとまっている列[1]を事例として、割り付けた要因 $a$ の主効果の“群間変動”を計算してみましょう。第8章の復習になりますが、群間変動は、反復数× $\sum(\text{群平均} - \text{総平均})^2$ です。ですので、水準1の群平均 $\bar{x}_1$ から総平均 $\bar{\bar{x}}$ を引いた偏差の2乗と、水準2の群平均 $\bar{x}_2$ から総平均 $\bar{\bar{x}}$ を引いた偏差の2乗を足し合わせたものに、反復数の4をかければ群間変動となります。自由度は水準数-1ですから“1”となり、(2水準系の場合は) 群間変動がそのまま要因分散(検定統計量Fの分子)となります。

一方、検定統計量Fの分母となる誤差分散も、誤差を割り付けた列[6]を使って、全く同じように計算できます(列[1]と違って水準1と水準2が飛び飛びなので面倒です)。後は、それらの比であるF値を限界値(第1、第2自由度とも1なので、5%有意水準で161.45となります)と比較して、効果がないという帰無仮説を判定します。以上、この作業を誤差以外の全ての列(主効果と交互作用)で繰り返します。

このように、直交表を使って計画した実験データを分散分析にかけるのは手間がかかります。ですので、実際には直交表の分散分析がそのままできるソフトウェアを使うことになるでしょう。ソフトウェアでは、海外のJMPやSPSS Conjointなどが有名ですが、日本製(社会情報サービス)のエクセル統計でもL<sub>8</sub>表のみ分散分析できます。ただし、2水準系ならば、Excel分析ツールでも要因の効果を検定できます(3水準系や混合系もダミー変数という技を用いれば可能ですが、ちょっと面倒です)。分析ツールの「回帰分析」を選択し、[入力Y範囲]にデータ列を、[入力X範囲]に直交表の誤差以外の列(交互作用を含む)を指定すれば、出力される各変数(列)のp値から、各主効果や交互作用が有意かどうかを判定できます(回帰分析は、第12章で学びます)。

## 10.7 直交計画法の応用分野

### ● 品質工学

直交表は、近年、品質工学(パラメータ設計)とマーケティング(コンジョイント分析)の2分野で発展的に利用されています。それぞれが1冊のテキストになる内容ですので、概要と直交表の利用場面だけ紹介させていただきます。

**品質工学**(quality engineering)は、技術開発や新製品開発を効率的に行う技法体系のことで、田口玄一(トピックス⑦)が独りで築き上げました。品質工学は、図10.14のように3つの分野から構成されています。

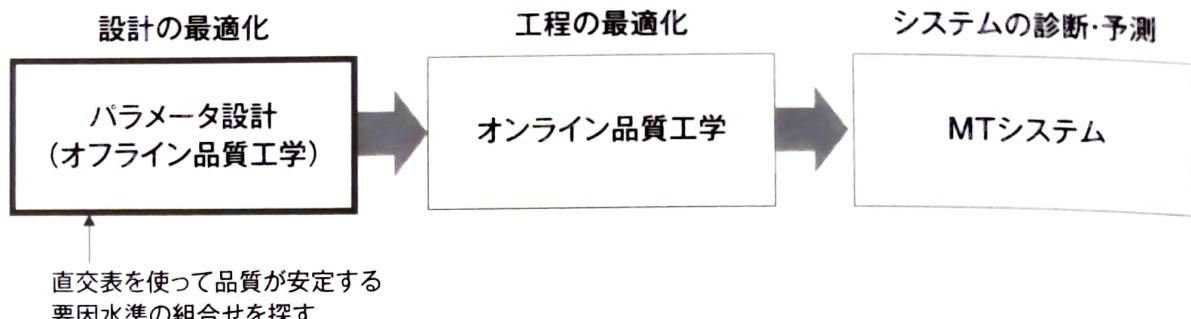


図 10.14 品質工学

- ① パラメータ設計（オフライン品質工学）：製造品質のバラツキを抑えるためのパラメータ（制御因子）を求める。
- ② オンライン品質工学：調整や管理のタイミングなど、製造段階の工程条件を費用とのかねあいで最適化する。
- ③ MTシステム：多変量解析を発展させたマハラノビス距離によって、システム異常や不良品の発生を予測する。

このうち、①のパラメータ設計（parameter design）において直交表が用いられます。

実験計画法や分散分析の目的は、特性（結果）が最大になる要因水準の組合せを見つけることでした。いいかえると、高い性能を發揮するパラメータをひたすら追い求めていたのです（穀物収量を上げるために施肥量や給水量をどうすればよいかなど）。しかし、消費者（ユーザー）にとっての「品質の良さ」とは、実際に製品を使用する様々な環境の下でも安定して性能を発揮することであることに田口博士は気付いたのです。

例えば、ある特定の気候の下でしか収量が高くならない施肥量や給水量の組合せを発見しても、ほとんどの農家にとって使い物になりません。やはり、どのような気候の下でも収量を高くできる組合せこそ、全ての農家にとってありがたい栽培法（パラメータ）ですよね。

そこで、たくさんの要因（制御できなければなりません）を直交表に割り付けて、多様な環境（気候などの制御できない要因、つまり誤差）の下で実験を観測し、特性のバラツキがもっとも小さくなる水準の組合せを見つけるのがパラメータ設計です（SN比という安定性指標を用います）。なお、パラメータ設計では、 $L_{18}(2^1 \times 3^7)$ 直交表など、交互作用が各列に分散された混合系が用いられます。市場で未知の交互作用があっても負けない、実用性の高い製品を開発するためです。

## トピックス⑦

### 直交計画と品質工学

実は、直交配列表を作成するのは大変難しいのです（というよりも一貫した論理がありません）。そのため、直交計画という考えは昔からありました、なかなか普及しませんでした。そうしたなか、線点図という考え方によって直交配列表を使いややすく改良し、この世に普及するきっかけを作った日本人がいます。現代の技術開発や新製品開発には欠かせない品質工学（海外では、タグチメソッドと呼ばれています）という分野を構築した田口玄一（1929～2012）です。田口先生は若い頃からデミング賞（日本ではもっとも権威のある賞の1つです）を受賞するなど活躍されていましたが、意外にも世界的に注目されたのは1980年代に入ってからでした。その後、自動車会社フォードがタグチメソッドを取り入れて復活したのを皮切りに、他のアメリカの主力産業も次々と取り入れ、現在では、「アメリカを蘇らせた男」として認知されています。最近になって実験計画法の復刻版が出版されましたので、興味のある方は読んでみてください（ただしちょっと難しいです）。『実験計画法上・下 第3版 復刻版』（各8,800円、丸善）

### ● コンジョイント分析

**コンジョイント分析** (conjoint analysis) は、消費者がどのような製品やサービスを好むのかを調べる「マーケティング・リサーチ（市場調査）」のための手法です。

例えば、みなさんがスマートフォン・メーカーに就職し、主婦をターゲットとした新製品開発をまかされたとしましょう。スマホの製品属性には、画面サイズ（6インチ／5インチ）、ストレージ容量（512 GB／256 GB）、非接触型ICカード（有／無）、OS（Android／iOS）の4つがあるとします。みなさんは開発に先立ち、どのような属性水準を組み合わせた製品（プロファイル）を主婦が好むのかを知るため、アンケートなどで市場調査を実施することになるでしょう。

そこでみなさんは、各属性について1つずつ、どの水準が好きかを聞いていくべきだらうと思うかもしれません。しかし、消費者は、スマホ売り場で「1番重要なのはOSで、Androidがいいな」、「2番目に重要なのは画面サイズで、コンパクトな5インチがいいな」などと製品属性と水準を1つずつ部分評価して購入するわけではありません（開発者としてはその情報が欲しいのです）。実際には、直感的に「製品Aが1番欲しいな」とか「製品Bは全然欲が……」。実際には、直感的に「製品Aが1番欲しいな」とか「製品Bは全然欲しくないな」というように、製品を属性と水準のかたまりとして総合評価していくのです。

いるのです（conjointとは結合という意味です）。

そこで、アンケートにおいても、実際の売り場のように、いくつかの仮想製品を並べて、順位付けさせたり1つだけ選択させたりするのです。そして、回帰分析などを使って総合評価を部分評価に分解し、新製品の開発に生かすというわけです（図10.15）。

そのとき消費者に提示する仮想製品のプロファイルを作成するのに、直交表を利用するのです。スマホの例ですと、2水準の属性が4つですから、全て組み合わせると32種類のプロファイルができてしまい、とてもアンケートで聞ける数ではなくなってしまいます。しかし、 $L_8$ 表を使えば8種類で済むため、回答者への負担を大幅に軽減できるというわけです。

ただし、ほかの属性と関連性が高い「価格」などを属性にすると、直交性が崩れてしまったり、不合理なプロファイル（安いのに性能が高い製品など）ができたりしてしまうため、注意が必要です。

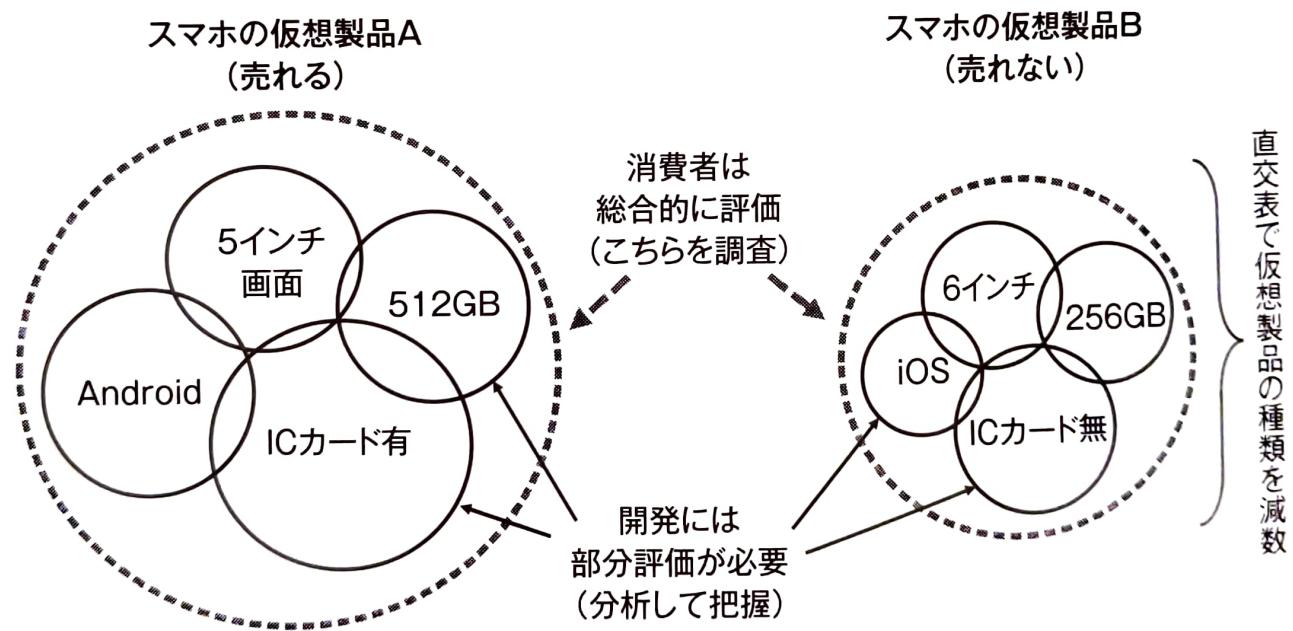
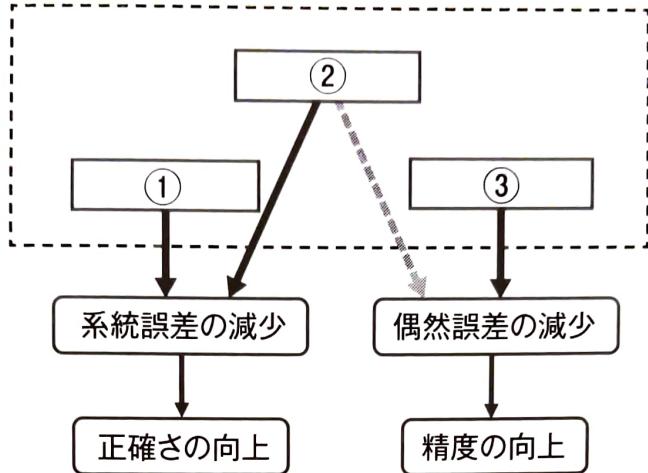


図 10.15 コンジョイント分析の概念（円の大きさが重要度）

## 章末問題

**問1** 下の図は、フィッシャーの三原則とその役割について整理したものである。①～③に入る原則の名称を答えなさい。

フィッシャーの三原則



**問2** 次は、フィッシャーの著書『実験計画法』で紹介された「紅茶の実験」を要約したものである。a～dについて答えなさい。

- **目的**：ある婦人が、ミルクティーを味わえば、ミルクと紅茶のどちらを先にカップに注いだかを識別することができると主張した。そこでフィッシャーは、本当に婦人が識別能力を有しているのかどうかを検証する実験を計画した。
- **実験**：合計8杯のミルクティーのうち、4杯はミルクを先に、残りの4杯は紅茶を先に注いで、それを無作為の順序で婦人に供する（婦人はどちらも4杯ずつであることを知っている）。
- **判定**：婦人が8杯のミルクティーを、それが受けた処理と一致するように、4個ずつの2組に分けることができれば、婦人の主張は正しいといえる。

- a. 帰無仮説と対立仮説はどのような内容になるか？
- b. なぜ同じ処理の紅茶を1杯ずつではなく4杯ずつ入れたのか？ それぞれの識別能力がなくても偶然に当たる確率を比較して答えなさい。  
※ヒント：4杯ずつのときの偶然当たる確率は、異なる8杯の中から4杯を（順序を考えずに）取り出す「組合せの数」を使って計算します。
- c. なぜ無作為の順番で、紅茶を婦人に供したのか？

- d. このような実験計画（実験の場の配置）を何と呼ぶか？また、婦人のお腹が膨れないように、味見を1日に各処理1杯ずつにして（供する順番は日によって無作為に変える）、実験を4日に分けるとしたら、どのような実験計画になるだろうか？

問3 次はL<sub>8</sub>表に（図10.13真ん中と同じ）典型的な四元配置の割付パターンを用いて計画して実験した結果である。4つの主効果と2つの交互作用を検定しなさい（有意水準 $\alpha = 0.05$ ）。できれば手計算が望ましいが、Excel分析ツールの「回帰分析」を用いても良い。