佐伯胖・松原望編
『実践としての統計学』
東京大学出版会、2000

3章

因果関係を推定する

無作為配分と統計的検定

この章では，無作為配分と統計的検定の関係について説明する．因果関係を推定するために，個体差のある実験要素（人間や動物など）を使った実験を計画し，その実験要素を別々の実験条件に配分する場合には，配分は無作為に行なわなければならない．そうしないと，実験結果を統計的検定にかけても，出てくる結果は無意味なものになってしまう．

3.1 実験の「王道」

この章のテーマは「無作為配分」である．

医学や心理学などでは，人間や実験動物のように，個体差のある被験体を使って実験を行なうことが多い．そうした実験の結果から因果関係を推定しようとする場合，この「無作為配分」random assignment（「無作為割り当て」「無作為割りつけ」などと訳すこともある）は，たいへん有効な方法である．実験の「王道」とよばれることもある．

しかし，無作為配分は，それだけでは役に立たない．統計的検定と組み合わせたとき，はじめて効果を発揮するのである．

逆もまた真である．個体差のある被験体を使って実験を行ない，そのデータを統計的検定（t 検定や分散分析など）にかけることは多いが，そもそも無作為配分が行なわれていなければ，検定の結果から因果関係を正しく推定することはできないのである．いいかえると，無作為配分は，そうした実験データを統計的検定によって分析するさいの「絶対必要条件」とでもいうことになろう

か．

それほど基本的な方法であるにもかかわらず，日本の大学では，無作為配分については，あまりきちんと教えてもらえないことが多い．統計学の入門コースでは，無作為抽出の話はでてきても，無作為配分の話がでてくることは稀である．一方，統計学を利用するほうの心理学などの学問分野でも，実験計画法のテキスト（例えば，Solso & Johnson, 1989）を使った授業が行なわれている例外的なケースは別として，無作為配分のきちんとした説明をしてくれる授業は，やはり稀にしかない．あまりに基本的な方法なので，英語の実験論文にも，「無作為配分を行なった」というような記述がでてくることは滅多にない．そのため，日本では，少なからぬ数の研究者が，無作為配分についての正確な理解を欠いたまま，ときには，無作為配分なしで実験を行ない，統計的検定を誤って使用する結果になっているのである．

無作為配分そのものは，ごく単純な手続きである．複数の実験条件，例えば，実験群と統制群があるとしよう．何人もの被験者をこの二つの条件に振り分けるとき，無作為（ランダム）に振り分ける．それが「無作為配分」である．そうしてみると，別に，難しいことは何もないように見える．

しかし，無作為配分の意味を正確に理解したうえで，それを正しく利用するためには，これだけを知っていたのでは十分ではない．例えば，読者は，無作為配分についての次のような質問にどう答えるだろうか？

a) 無作為抽出――「無作為抽出」は，統計学の時間に習ったのだが，無作為配分というのは，それとは違うのだろうか？ 違うとすれば，どう違うのだろうか？ この二つはどういう関係にあるのだろうか？

b) 個人差による解釈――実験群と統制群とのあいだで，統計的に有意な平均値の差が得られた．ところが，「実験群の被験者と統制群の被験者とがもともと違っていた（例えば，実験群の被験者のほうが，たまたま，総じて頭がよかった）からではないのか？」という批判を受けた．この批判は正しいのだろうか？ こういう批判を避けるためには，どうすればよいのだろうか？

c) 組織的な配分――実験群と統制群を等しくしたいのなら，個体差を表す変数（例えば，知能）を直接測定して，同じ値をもつ被験体をそれぞれの群に振り分けたほうがよいのではないだろうか？ そうすれば，実験群と統制群は，

この変数に関して同じ値をもつことになる．無作為に振り分けたのでは，結果として，両群のあいだで個体差変数の値が違ってしまうことになりはしないだろうか？

d) 無作為配分の限界――「王道」とよばれる無作為配分は万能なのだろうか？ 無作為配分をしさえすれば，因果関係の推定はかならず正しく行なえるのだろうか？

どの質問にも確信をもって答えることができた読者は，この章を読むにはおよばない．しかし，そうでなかった読者には，難しい話ではないので，ぜひ一読を勧めたい．

3.2 方向オンチ実験――個人差の問題

3.2.1 方向オンチ実験

よその大学の学生さんから，こんな質問を受けたことがある．心理学科で，卒業論文のための実験をやろうとしている学生さんだった．

――方向オンチについて調べる実験を計画しているのだが，去年，同じような実験をやった先輩がいた．その実験では，実験群と統制群とのあいだに，統計的に有意な差が得られた．ところが，その先輩が卒論発表会で自分の研究を発表したところ，ある先生から次のような批判がでた．「実験群と統制群とのあいだには，もともと，イメージ能力の差があって，実験操作ではなく，そちらのほうがほんとうの原因となって，統計的に有意な差が生じたのではないか」という批判だった．そこで，自分の実験では，イメージ能力を統制したいのだが，イメージ能力を測るためのよい検査はないだろうか？――

だいたい，こんな内容の質問だった．

これからの話は，実験の細かい手続きとは関連がないので，おおざっぱに，次のような実験だったと考えておくことにしよう．

この学生さんは，「方向オンチの人は，自分がたどってきたルートをきちんと頭の中に残していないのではないか」という仮説を立てた．この仮説が正しいとすると，ルートを頭のなかに残しておくように努力すれば，「方向感覚」は改善されるはずである．それを調べるために，以下のような実験を行なった．

被験者は，コンピュータ・グラフィクスの動画をいくつも見る．その動画は，街を歩いていて，角を2回曲がる様子を，その歩いている人の視点から描いたものである．曲がり角の角度は，動画によって，すこしずつ変わる．2回曲がって止まったところで，被験者は，出発点の方向に腕をのばして，いわば，コンピュータ・グラフィクスの仮想空間のなかで，見えない出発点を指さす．これが課題である．実験者は，腕がさし示している方向を記録する．その方向と，出発点があるはずの方向とのあいだのズレを測ったとき，このズレが小さいほど，「方向感覚」は優れているということになる．

実験群と統制群の違いをつくりだす実験操作だが，実験群の被験者は，「動画を見ているあいだ，自分が歩いていくルートを上から（俯瞰して）見たところをイメージするように」といわれる．一方，統制群の被験者はなにもいわれない．さきほどの仮説が正しくて，自分がたどってきたルートをきちんと頭のなかに残していれば，「方向感覚」はよくなるのだとしよう．そうすると，ルートをイメージする実験群の被験者のほうが「方向感覚」がよくなって，さし示した方向のズレは小さくなるのではないか，という予測がなりたつ．

さて，この実験を行なったところ，仮説の予測どおり，「実験群の被験者は，統制群の被験者に比べてズレが小さい」という結果が出たとしよう．ズレの平均値を t 検定にかけたところ，平均値の差は統計的にも有意だったとする．

では，「仮説は支持された」と結論してよいだろうか？

この結論に疑問を投げかけたのが，さきほどの先生の批判である．――この課題の成績を上げるためには，だれでも，自発的に，ルートを上から見たイメージをつくろうとするのではないだろうか．つまり，統制群の被験者のほうも，言われなくても，イメージをつくっていたかもしれない．しかし，統制群の被験者は，たまたま，平均して，イメージをつくる能力が劣っていたのかもしれない．だとすれば，実験結果を分析したときに，統制群のズレが実験群のズレより有意に大きかったとしても，それは，たまたまイメージ能力が劣っていたことの結果にすぎない，ということになる．いいかえると，得られた有意差は，ルートをイメージしたことの効果を表しているわけではなく，イメージ能力の差を表しているにすぎないのかもしれない．とすれば，当然，「仮説が支持された」と結論することもできないことになる．

さて，この批判はあたっているのだろうか？

だが，それを考えてみるまえに，すこしまわり道になるが，実験計画の用語を使って，この問題をもうすこし一般的なかたちで捉えなおしてみることにしよう．

3.2.2　実験のロジックと個人差

実験計画の用語では，原因にあたるのではないかと推測されている変数を「独立変数」independent variable とよぶ．方向オンチ実験の場合は，「ルートのイメージをつくるかどうか」が独立変数である．独立変数は，実験者が操作する変数である．この実験では，「上から見たルートをイメージするように」という教示を与えたり，与えなかったり，という手続きによって，独立変数を操作している．

一方，因果関係の結果のほうにあたるのではないかと推測されている変数は，「従属変数」dependent variable とよぶ．方向オンチ実験では，「方向感覚」が従属変数にあたる．従属変数は，実験者が測定する変数である．この実験では，出発点があるはずの方向と腕がさし示している方向とのあいだのズレを調べる，という方法で「方向感覚」を測定している．

さて，実験には，もうひとつ変数がある．困ったことに，この変数には，きまった呼び名がない．研究分野が変わると，呼び名も変わってしまう．「剰余変数」extraneous variable，「交絡変数」confounding variable，「二次的変数」secondary variable，「調整変数」moderator variable，「第三変数」third variable といった調子で，とても数えきれないほどである．歴史学などで使う「外生変数」exogenous variable という用語も，ほぼ同じ意味で使われることがある．統計学の「共変数」covariate という用語は，これらの用語と同じ意味でも使われるが，「独立変数」の意味でも使われるので，非常にまぎらわしい．用語が定まっていないということは，この変数についての理解がきちんと定まっていない，ということにもつながる．

それはそれとして，ここでは，「干渉変数」interfering variable という用語で統一しておくことにしよう．「独立変数と従属変数とのあいだの関係に干渉する変数」という意味の用語である．関係がある場合には，それをうち消して

しまったり，変えてしまったりする（3.6.2で具体例をあげる）．関係がない場合にも，関係があるかのように見せてしまうことがある．

しかし，ここでは，あまり細かいことには立ち入らない．干渉変数は，「独立変数とは別の変数なのだが，従属変数に影響をおよぼすのではないか，と疑われている変数」だというふうに理解しておくことにしよう．方向オンチ実験の場合は，先生の批判にあった「イメージ能力」がこの干渉変数に該当する．

干渉変数が独立変数の操作と相関するとき――例えば，実験群と統制群のあいだで干渉変数の値が異なるとき――には，実験結果の解釈を誤る可能性がでてくる．つまり，さきほどの先生が指摘したように，実験群のほうが，統制群よりイメージ能力が高かった（干渉変数の値に差があった）とすると，実験群のほうがズレが小さかった（従属変数の値に差があった）という結果のほんとうの原因は，「イメージ能力」という干渉変数のほうであって，「イメージをしたか，しなかったか」という独立変数のほうではない，という可能性がでてくるわけである．もし，干渉変数のほうがほんとうの原因なのだとしたら，「ルートをイメージすれば，方向感覚がよくなる」と結論してしまったのでは，実験結果を誤って解釈したことになる．

したがって，干渉変数は，実験者が統制しなければならない変数なのである（「統制」controlというのは，複数の実験条件のあいだで，値が違わないようにすることである）．そこで，あらかじめ，イメージ能力を測る検査をやっておき，その成績が実験群と統制群のあいだで違わないようにしておいてはどうか，という考えがでてくることになる．

先生の批判のなかに登場した干渉変数は個人差である．たまたま，実験群にイメージ能力の高い人がたくさん集まってしまった，という想定のもとに批判がなされていた．

実験で調べる人間や動物には，かならず個体差がある．その個体差のなかには，従属変数に影響をおよぼしうる変数，すなわち，干渉変数となりうる変数があるかもしれない．実験者が気づかないところで，干渉変数としての個体差の値が独立変数の操作と相関していたとすると，実験結果から，独立変数と従属変数のあいだの因果関係を正しく推定することはできなくなってしまう．したがって，個体差の統制は，実験によって因果関係を推定しようとする場合には，非常に重要な問題となるのである．

3.2.3　因果関係と実験

ここで，因果関係を推定することにどのような意味があるのか，すこし考えておくことにしよう．

というのは，「因果関係にこだわるのは無意味だ」と考える人もいるからである．そういう人たちは，「相関関係がわかれば十分だ」と主張する．

二つの変数のあいだに，何か規則的な関係があるとき，それを「相関関係」correlationという．例えば，図3.1をみてみよう．この図には，毛織物の売り上げと風邪の患者数との関係を示す，架空のデータがプロットしてある．図のなかの○や×は，それぞれ，1年のなかの一つの月を表している．つまり，ひと月ごとに，毛織物の売り上げと風邪の患者数を集計して，それをプロットしたのである．この図をみると，「毛織物の売り上げがふえると，風邪の患者が多くなる」という規則的な関係のあることが読みとれる．いいかえると，この二つの変数のあいだには，「正の相関」がある．

もし，この相関関係が因果関係だったとしたら――つまり，毛織物を着ると風邪をひきやすくなるのだとしたら――風邪で寝こんだり亡くなったりする人の数を減らすためには，毛織物の販売を禁止すればよい，ということになる．原因がなければ，結果は生じないからである．

しかし，この相関関係がたんなる規則的な関係にすぎず，因果関係を表しているわけではないとしたら，そうはいかない．

図3.1では，1年を大きく夏と冬に分けて，夏の6カ月は○で，冬の6カ

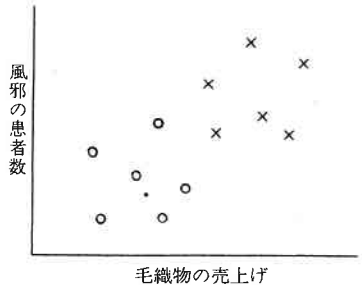

図3.1　因果関係ではない相関関係

月は × で表してある．夏と冬とに分けて調べてみると，冬は，毛織物の売り上げも多く，風邪の患者も多いことがわかる．夏は，毛織物の売り上げも少ないし，風邪の患者も少ない．

もうおわかりだろう．じつは，この相関関係を生み出しているほんとうの原因は，気温なのである．すなわち，気温が低くなると，ウールでできた暖かい服の売り上げがのびる．また，気温が低くなると，体を冷やしてしまうことが多くなって，風邪をひく人の数もふえる．つまり，毛織物と風邪とのあいだには，直接の因果関係はないのである．ただ，気温の下降が原因となって，その結果として，毛織物の売り上げも風邪の患者数も多くなるので，この二つの変数のあいだには，相関関係が現れるのである．だとすれば，毛織物の販売を禁止してみても，風邪の患者を減らすことはできない，ということになる．

因果関係がある場合は，原因のほうを操作することによって，結果のほうを変えることができる．病気をなおしたいとか，火事で燃えにくい家をつくりたいとか，なにか望ましい結果を得たい場合には，因果関係がわかっていれば，その原因のほうに働きかけることによって，望ましい結果を手に入れることができる．しかし，たんなる相関関係しかわからない場合には，いまの例からも明らかなように，一方の変数に働きかけても，かならずしも，望ましい結果を手に入れられるとはかぎらない．

現実を正確に認識しようとするとき，因果関係がもっとも適切な認識の形式なのかどうかという点については，哲学者のあいだで，こみいった議論がある．だが，それはそれとして，こういうことだけはいえるだろう．つまり，進化論的にみれば，人間が環境に適応していくためには，たんなる相関関係の認識よりは，因果関係の認識のほうがはるかに有用だったのである．

では，その因果関係を見きわめるためには，どうすればよいのだろうか？「原因ではないか」と思われる変数（つまり，独立変数）を実際に動かしてみればよい．もし，ほんとうに因果関係があるのなら，「結果ではないか」と思われる変数（つまり，従属変数）のほうも，それにつれて値が変化するはずである．もし変化しなければ，それらの変数のあいだに何か規則的な関係があったとしても，それはたんなる相関関係にすぎず，因果関係ではない，ということになる．

コラム 3.1

コレラ菌を飲む話

コレラの原因が細菌だという事実を発見したのは，ドイツの医学者コッホ (Robert Koch) である．しかし，この発見も，すぐに受けいれられたわけではなかった．コッホの説に反対していた医学界の権威ペッテンコーファー (Max J. von Pettenkofer) は，やがて，コレラ菌の存在を認めるところまでは譲歩したが，それでもなお，「コレラ菌は，コレラの唯一の原因ではない．不良な健康状態などもコレラの原因である」と主張して譲らなかった．ペッテンコーファーは，自説を証明するために，何とコレラ菌の培養液をみずから飲んでみせるという挙にでた．ところが，少しばかり腹を下しただけですんでしまい，コレラにはかからなかったのである．なぜコレラにかからなかったのかは，長らく謎とされてきたが，よく考えてみると，べつに不思議ではないのかもしれない．健康状態は干渉変数（本文を参照）として働いていたのだろう．つまり，コレラ菌の働きは，良好な健康状態によって抑えられてしまうこともありうるのである．その意味では，ペッテンコーファーは正しかったといえる．この人物は，「仕事ができなくなった人間は生きているべきではない」という自説に殉じて，ピストルで人生の幕を閉じた．読者の目には，「堅い信念の持ち主」と映るだろうか，それとも，「狂信者」と映るだろうか？

実証的研究は，大きく「実験的研究」experimental study と「相関的研究」correlational study（または，「観察的研究」observational study）とに分かれる．相関的研究の場合は，二つ（あるいは，それ以上）の変数をどちらも測定して，そのあいだに規則的な関係があるかどうかを調べる．しかし，わかるのは，相関関係があるかどうかだけである．一方，実験的研究の場合は，片方の変数（独立変数）を人為的に操作するので，因果関係があるかどうかまで調べることができる．そういうわけで，実験は，「因果関係を推定するためのもっとも強力な方法だ」といわれるのである．その実験を，観察や調査などの相関的研究から区別しているのは，独立変数の操作が含まれていることである．

とはいうものの，現実は複雑で，話はそう簡単には終わらない．

独立変数の値を操作したとき，もし，別の変数の値がいっしょに変わってしまっていたとしたらどうだろう？　従属変数の値に変化が起こったとしても，

その原因は，独立変数なのか，それとも，その別の変数（つまり，干渉変数）のほうなのか，わからなくなってしまう．したがって，実験を行なうさいには，干渉変数の値が変わらないように，きちんと統制をしなければならない．ところが，これが難題なのである．というのは，何が干渉変数になるのか，つまり，どういう変数が従属変数に影響をおよぼしうるのかは，前もって正確にわかっているとはかぎらないからである．それがすべてわかっているぐらいなら，実験をやる必要はない．そもそも，それを知るために実験をやろうとしているのだから．

しかし，個人差が干渉変数として働く場合が多いことは，過去の経験からよくわかっている（コラム 3.1 を参照）．その個人差を統制する最良の手段が無作為配分だといわれているのである．

3.3 無作為配分

3.3.1 無作為配分の効用

さきほどの学生さんの質問にもどろう．「イメージ能力を測るためのよい検査はないでしょうか？」ときかれて，私は逆にこういう質問をした．「去年同じような実験をやった先輩は，無作為配分をしたのでしょうか？」

答は，これも質問で，「無作為配分って何ですか？」

そこで，はじめの節で述べたような無作為配分のかんたんな説明をすると，「先輩は，それはやっていませんでした．無作為配分のことは，聞いたこともありませんでした」という答が返ってきた．

理由はあとで詳しく説明するが，無作為配分をやっておけば，基本的には，個体差の効果を独立変数の効果と混同する心配はなくなる（「基本的」ではない場合についても，あとで詳しく説明する）．

無作為配分は，別に難しいことではない．方向オンチ実験の場合なら，被験者を実験群と統制群に無作為（ランダム）に分ける，とただそれだけのことである．ところが，実際には，往々にして，無作為ではない配分をしてしまう人がいるのである．

例えば，実験群の被験者と統制群の被験者を全部調べてから，「有意差がなかった」ということになったりしたのでは「時間と労力が無駄になる」というので，つぎのような方法をとる人がいる．――まず，実験群の被験者だけについて実験をすませてしまい，平均値などの統計量を計算する．それから，統制群の被験者 2, 3 人について実験を行ない，その平均値を実験群の平均値と比較してみる．もし，仮説が予測したような差（例えば，「統制群のほうが平均値が大きい」というような差）がはっきりでていたら，実験をつづける．といっても，統制群の残りの被験者を一気に調べてしまうのではなく，少しずつふやしては，同じことを繰り返す．もし，予測していた結果がでそうにもないようだったら，そこで実験をうちきり，実験手続きや仮説を再検討する．

こういった方法をとると，実験がうまくいかなかった場合でも，すべての被験者を調べてしまったあとではないので，損失は少なくてすむ．そういう意味では，一見，賢い方法のようにみえるのだが，じつは，個人差の統制という問題を考えると，これは非常に危険な方法なのである．この実験のやりかたでは，実験群と統制群への被験者の配分は，明らかに，無作為ではない．はじめに実験室にやってきた被験者はすべて実験群の被験者，と決まってしまう．これがいろいろと問題をひき起こすのである．

例えば，はじめにやってくる被験者は，実験に興味をもっていて，「やる気」にあふれた人たちかもしれない．この場合，実験群のほうが平均して成績がよかったとしても，それは，「やる気」の違いを反映しているだけで，独立変数の効果ではない，という可能性を排除できない．

被験者を無作為に両群に配分すれば，どちらかの群に「やる気」の高い被験者が集まってしまう，という問題は避けることができる．これが無作為配分の効用なのである．

3.3.2 無作為配分の手続き

被験者を無作為に配分するためには，具体的にはどのようにすればよいのだろうか？

例として，10 人の被験者を半々ずつ，実験群と統制群に分ける場合を考えてみよう．まず，「1」から「10」までの数字をカードに書き，「1」から「5」までは実験群，「6」から「10」までは統制群と決める（もちろん，逆でもよい

し，奇数は実験群，偶数は統制群ということにしてもよい）．そのカードをトランプと同じ要領でよく切る．それから，カードを1枚ずつめくって，数字を順に書きとっていく．かりに，「9，10，2，8，…」という順になったとすると，最初にやってきた被験者は「9」を割り当てられるので統制群，つぎにやってきた被験者は「10」を割り当てられるので，これも統制群，そのつぎは「2」だから実験群，さらにつぎは「8」だからまた統制群……ということになる．こうすれば，10人の被験者を5人ずつ両群に無作為に配分することができる．

この方法では，カードを切ることが「無作為化」randomizationの手続きになっている．厳密にいうと，これでは，ほんとうの無作為化にはならない．なぜなら，完全な無作為，つまり，まったく何の偏りもない，という状態を実現することは非常に難しく，数学的知識に裏打ちされた，高度な技術が必要になるからである．しかし，カードを使う場合も，念入りに切りさえすれば，実用上は，さほど問題はない．

カードを切るかわりに，乱数表を使うこともできるし，コンピュータを使うこともできる．コンピュータを使う場合は，乱数発生プログラムを利用して，任意の範囲の数（例えば，「1」から「10」まで）を無作為な順序に並べる，という作業をする．厳密にいうと，コンピュータで発生させた乱数は擬似乱数で，厳密な意味では，やはり，「何の偏りもない」というわけにはいかないのだが，実用上は十分である．

3.3.3 無作為配分の対象

これまでは，人間が被験者になる実験を例にとって，説明をしてきた．動物実験の場合は，むろん，動物を無作為配分することになる．しかし，無作為配分の対象は，生き物だけに限られるわけではない．個体差のある実験要素は，すべて無作為配分の対象となるのである．

例えば，単語のリストをおぼえる記憶実験を考えてみよう．独立変数は，おぼえてからすぐに再生するか（直後条件），5分たってから再生するか（遅延条件），つまり，「遅延時間」だということにしよう．記憶力の個人差が大きいことは，だれもが知っている．そこで，直後条件と遅延条件には，別の被験者を割り当てるのではなく，同じ被験者が両方の条件で実験を行なうものとする．

こういう実験のやりかたを「被験者内計画」within-subject(s) design（または，「反復測定計画」repeated measure(ment) design）という．別の被験者を割り当てる「被験者間計画」between-subject(s) designに比べると，個人差の効果が独立変数の効果と混同される心配がないので，この点では，被験者内計画のほうが明らかに有利である．じつは，この被験者内計画は，個人差を統制するための，無作為配分とは別の，もう一つの方法なのである．ただし，次に述べるように，実験を行なった影響があとに残る場合も多く，そうした場合には，この被験者内計画は使えない．

方向オンチ実験の例で考えてみよう．はじめに実験条件で実験を行なう被験者は，「ルートを上から見たイメージをつくってください」といわれて，実験に入る．そのあと，統制条件で実験を行なうときには，イメージのことは何もいわれない．だが，課題は同じなのだから，前と同じようにイメージをつくってしまうのが普通だろう．そうなると，統制条件は実験条件と変わらなくなってしまう．かといって，この問題を避けるために，つねに統制条件のほうを先に行なうことにすると，実験条件のほうがずっと有利になってしまう．なぜなら，統制条件ですでに課題の練習をしていることになるわけだから．これでは，たとえ，実験条件のほうが成績がよかったとしても，「それはイメージをしたことの効果ではなく，たくさん練習をしたせいではないか」と批判されたとき，答えようがない．

話をもとにもどすが，こうした問題がなく，被験者内計画が使える場合にも，かならずしも無作為配分が不要になるとはかぎらない．先ほどの記憶実験の場合，直後条件と遅延条件で同じ単語をおぼえたとすると，あとでおぼえる条件のほうが圧倒的に有利になることは目にみえている．したがって，それぞれの条件では，別の単語リストをおぼえることにしなければならない．だが，だれもが知っているように，単語のおぼえやすさには差がある．「織田信長，徳川家康，……」というリストはおぼえやすいが，「酒井忠清，北畠満雅，……」というリストはおぼえにくい．二つの条件のあいだで，単語のおぼえやすさという干渉変数の値にちがいがあると，遅延時間という独立変数の効果と，その干渉変数の効果とを混同してしまうおそれがでてくる．

そこで，無作為配分を行なう．つまり，直後条件のリストと遅延条件のリス

トに，単語を無作為に配分するのである．被験者がかわるたびに，無作為配分をやり直すことにすれば，なお効果的である．

このように，個体差のある実験要素は，すべて無作為配分の対象となりうるのである．

3.4 組織的配分と無作為配分

3.4.1 なぜ無作為配分は有効なのか？

無作為配分について，いろいろと説明をしてきたが，まだ肝心の話が終わっていない．個体差のある実験要素（被験者など）を無作為に配分すると，なぜ，その個体差をうまく統制することができるのだろうか？

「むしろ，組織的に配分したほうが効果的なのではないか？」という考えが浮かんできても不思議ではない．例えば，方向オンチ実験の場合，被験者を実験条件と統制条件とに無作為に配分するより，イメージ能力をきちんと測定して，この二つの条件のあいだで，イメージ能力の平均値が等しくなるようにしたほうがよいのではないだろうか？

無作為に配分するということは，配分の結果は偶然に決まる，ということである．とすると，イメージ能力を測る検査の得点が，偶然に，実験条件と統制

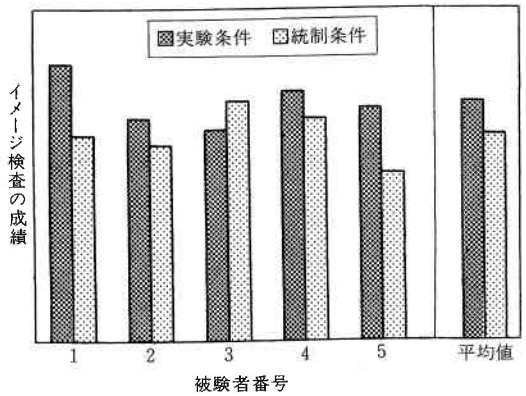

図 3.2 無作為配分の結果
「被験者番号」は，たまたま割り当てられた番号で，同じ番号の被験者のあいだに特別な関係はない．

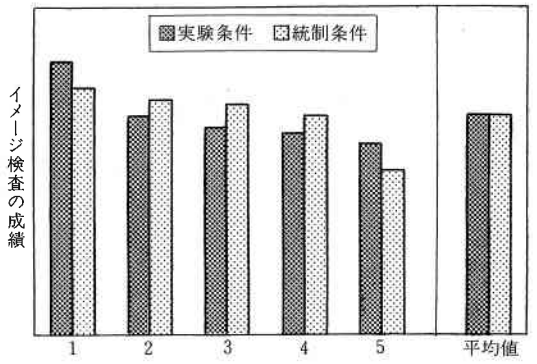

図 3.3 平均値等化法
「被験者番号」は，たまたま割り当てられた番号で，同じ番号の被験者のあいだに特別な関係はない．

条件とで違ってしまう，ということもでてくるのではないだろうか？ 図 3.2 には，10 人の被験者を無作為に分けるときに，得点がどうなるかが示してある．実験条件の平均値と統制条件の平均値は，明らかに同じにはなっていない．

配分の結果を決めるのが偶然なのだとしたら，「イメージ検査の成績の平均値が，たまたま，二つの条件のあいだで同じになる」という可能性は，むしろ，非常に小さいはずである．たいがいの場合は，アンバランスがでてくるだろう．極端な場合，成績の高い被験者が，全部，実験群に集まってしまい，統制群の被験者はみな成績の低い人ばかり，という結果になる可能性さえ，絶対にないとはいいきれない．

一方，イメージ検査の成績にもとづいて組織的な配分を行なえば，そんなアンバランスが生じることは防げる．図 3.2 と同じ 10 人の被験者を組織的に配分した結果が図 3.3 に示してある．成績の平均値は，実験条件と統制条件で等しくなっている．こうしたほうが，イメージ能力という干渉変数はきちんと統制できるのではないだろうか？

3.4.2 平均値等化法と対等化法

もっとも，平均値を等しくするというこの統制方法（「平均値等化法」mean matching）には，問題もある．

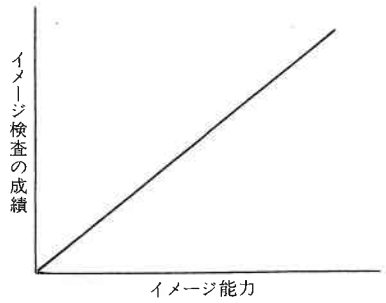

図 3.4　線形な関係

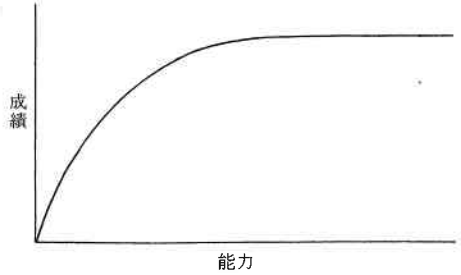

図 3.5　非線形な関係

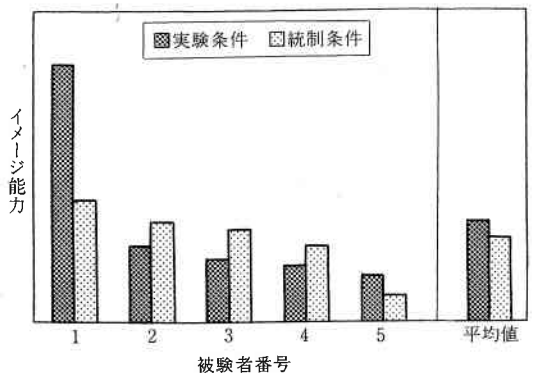

図 3.6　平均値等化法
（イメージ能力とイメージ検査成績の関係が非線形の場合）
「被験者番号」は，たまたま割り当てられた番号で，同じ番号の被験者のあいだに特別な関係はない．

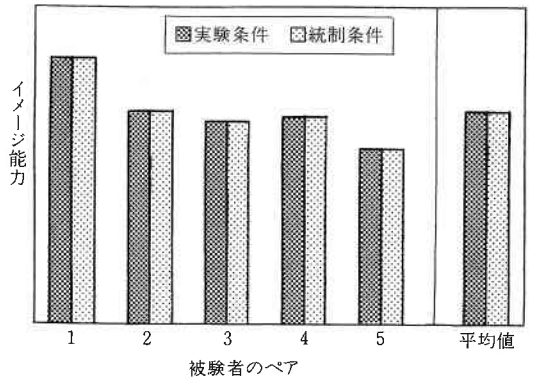

図 3.7　対等化法
（イメージ検査の成績とイメージ能力の関係が非線形の場合）
この図の被験者は，図 3.2，3.3，3.6 の被験者とは別の被験者である．図 3.6 の被験者 10 人だけを使ったのでは，うまくペアをつくることはできない．

例えば，イメージ検査の成績とほんとうのイメージ能力（実際には測定できない）との関係が「線形」linear ではないときに問題が起こる．「線形」な関係というのは，比例や反比例のように，直線（すなわち，1 次関数）で表されるような関係のことをいう（例えば，図 3.4）．

しかし，検査とそれが測っている能力とのあいだの関係は，いつも線形だとはかぎらない．例えば，学校でやった数学のテストを思い出してみよう．ひどくやさしい問題ばかりのときには，数学の学力とテストの成績との関係は図 3.5 のようになる．学力が低い生徒たちのあいだでは，成績はずいぶん違う．ところが，学力が高い生徒たちのあいだには，成績の違いがない．非常によくできる生徒も，かなりよくできる生徒も，まあよくできる生徒も，みな満点をとってしまう．

イメージ検査の成績とイメージ能力とのあいだに，図 3.5 のような「非線形」nonlinear な関係があったとしよう．そうすると，図 3.3 のように成績の平均値を等しくした場合，イメージ能力のほうは図 3.6 のようになる（イメージ能力のほんとうの値は，実際には，直接知ることはできないのだが，かりに知ることができたとしたら，こういう結果になる）．つまり，肝心のイメージ能力のほうでは，平均値が等しくなくなってしまうのである．

こうした平均値等化法の問題を避けるためには，「対等化法」pair(ed)-matching を採用すればよい．この方法では，イメージ検査の成績が同じ被験者を対にして，片方を実験条件，もう片方を統制条件に配分する．こうすれば，イメージ検査の成績について，平均値が等しくなると同時に，イメージ検査の成績とイメージ能力とのあいだにどのような非線形の関係があろうとも，イメージ能力のほうの平均値も等しくなる（図 3.7）．

さて，そういうことなら，平均値等化法のかわりに対等化法を使えば，無作為配分よりも干渉変数をうまく統制することができる，ということになるのではないだろうか？

3.4.3 組織的配分の問題点

残念ながら，対等化法のような組織的配分には，原理的にも，実際的にも，いろいろな問題があって，これだけでは，なかなか干渉変数をうまく統制することができないのである．

まず，どちらかといえば原理的な問題のほうから考えてみよう．

測定可能性　「イメージ能力を測るためのよい検査はないでしょうか？」という学生さんの質問に，私は「検査はやっても無駄でしょう」と答えた．なぜ無駄なのかというと，イメージ能力の検査は，「妥当性」validity が低いという点で悪名高いからである．イメージ能力検査はいくつもあるが，たがいに，いちじるしく相関が低い．あちらの検査で高得点をとった人が，こちらの検査ではひどく低い得点しかとれない，というケースがたくさんでてくるのである．「イメージ検査は，ほんとうにイメージ能力を測っているのか？」とか，「そもそも，『イメージ能力』はひとつの能力なのだろうか？」というような議論にすら，まだ決着がついていないような有様なのである．そんな検査の成績をもとにして，実験条件と統制条件の等質化を図ってみても，ほんとうにイメージ能力を統制できたことになるのかどうか，はなはだ心もとないではないか．

百歩ゆずって，かりに，イメージ検査には妥当性がある――つまり，「イメージ能力」はひとつの能力で，イメージ検査はそれを測っている――のだとしても，なお，「信頼性」reliability の問題が残る．信頼性が低い，つまり，測定誤差が大きいと，検査の成績が能力を表しているとみなすことはできなくなる．同じ成績でも，能力は違うかもしれないのだから，対等化法もうまく機能しなくなる．成績の平均値は同じでも，能力の平均値には大きな差がでてしまうかもしれないのである．

一般に，妥当性や信頼性の高い測定方法が，すべての干渉変数について，かならず手に入るとはかぎらない．干渉変数が正確に測定できなければ，組織的配分によって干渉変数を厳密に統制することは，とうてい望めない．

統制し残した干渉変数　今度は，千歩ゆずって，イメージ検査は，妥当性も信頼性も非常に高いと仮定しよう．ところが，それでもなお，組織的配分の問題は解消しないのである．

「方向感覚」に影響する変数はイメージ能力だけではないかもしれない．例えば，記憶力がよいと，歩いてきた道筋を正確におぼえておけるので，「方向感覚」はよくなるかもしれない．もしそうだとすると，対等化法によってイメージ能力を「実験条件＝統制条件」としても，干渉変数をきちんと統制したことにはならない．記憶力については，たまたま「実験条件＞統制条件」となっていて，そのために，実験条件の被験者のほうが，「方向感覚」課題でよい成績を上げたのかもしれない．

「それなら，記憶力も測ればよいではないか」という意見もあるだろう．しかし，「方向感覚」に影響するかもしれない変数は，イメージ能力と記憶力だけとはかぎらないのである．知能も影響するかもしれない．あるいは，知らない場所をどれだけ歩きまわったことがあるか，という過去の経験も影響するかもしれない．さらに，移動速度は一定ではない場合が多いので，移動速度と所要時間をもとに，移動した距離を割り出す能力が重要なのかもしれない．

こういう具合に，変数がふえてくればふえてくるほど，そのなかに，正確に測定できない変数が入ってくる可能性が高くなってくる．つまり，対等化法がうまく働かなくなってくるのである．

万歩ゆずって，それらすべての変数が正確に測定できたとしても，「思いもよらない変数が干渉変数になっているかもしれない」という可能性はつねに残る．そういう変数は測定していないわけだから，当然，統制もできないことになる．

すべての干渉変数を漏れなく特定するためには，従属変数に因果的な影響を

およぼしうる変数をすべて知っている必要がある．しかし，それがわからないからこそ，目星をつけた変数を独立変数として，因果関係を確かめる実験をやってみようというのである．そう考えてみると，干渉変数を漏れなく特定することは，そもそも原理的に不可能なのだ，と言わざるをえない．

組織的配分に由来する干渉変数　組織的配分の手続きそれ自体が干渉変数を導き入れてしまうこともある．

例えば，こういう手続きで対等化法を行なったとする．――はじめにやってきた5人の被験者は，すべて実験条件の被験者とする．実験を始めるまえにイメージ検査をやって，成績を記録しておく．6人目以降は，すべて統制条件ということになるが，イメージ検査の成績が実験条件の被験者のだれとも一致しない人には，実験は行なわずに，イメージ検査だけでおひきとりいただく．実験条件の被験者と同じ成績を上げた人だけを統制条件の被験者とし，実際に実験を行なう．実験条件の被験者と同じ成績の被験者が統制条件でもそろったところで，実験は終わりになる．

こういう手続きをとった場合，実験条件の被験者は，すべて，統制条件の被験者よりも先に実験を行なうことになる．だが，被験者の募集を始めてすぐに実験室にやってくる被験者は，「やる気」にあふれた人たちかもしれない．だとすると，「方向感覚」課題で，たとえ実験条件の被験者のほうが成績がよかったとしても，それは「やる気」の差を反映しているにすぎず，独立変数の効果を反映しているわけではないかもしれない．

もちろん，「実験を行なう順序が干渉変数として働くかもしれない」と，前もってわかっているのなら，実験条件と統制条件のあいだで順序の差がでないような手続きを工夫すればよいのである．だが，対等化法の手続きのどこに干渉変数が潜んでいるのか，いつもかならず予測できるわけではない．知らぬまに干渉変数を導き入れてしまうことがない，とはいいきれないのである．

以上述べてきたような，どちらかといえば原理的な問題のほかに，組織的配分には，もっと実際的な問題もある．

経済性　対等化法を使うには，例えば，イメージ検査で同じ成績をとった人のペアをいくつもつくらなければならない．「10人の人にイメージ検査をやったら，うまく5組のペアができた」というようなうまい話には，まずならな

いだろう．5組のペアをつくるには，50人にイメージ検査をやってもらわなければならないかもしれないし，100人にやってもらわなければならないかもしれない．方向オンチ実験のような実験では，ふつう，各条件25人ぐらいか，それ以上の被験者が必要になる．25組のペアをつくるとなると，いったい何百人，あるいは何千人にイメージ検査をやらなければならなくなるのか，見当もつかない．つまり，対等化法は，非常に不経済なのである．

そこで，ふつうは，ぴったり同じ成績の人ではなく，だいたい同じぐらいの成績の人をペアにすることが多い．だが，これでは，当然のことながら，きちんとした統制ができるわけはない．

そこへいくと，無作為配分の場合は，10人の被験者を，ただ無作為に，実験条件と統制条件に振り分ければいいのだから，じつに経済的である．各条件25人必要なら，50人を無作為に振り分ければよい．そもそも，イメージ検査をやる必要がないという点でも，はるかに経済的だといえるだろう．

組み合わせ可能性　さきほどの話にあったように，統制しなければならない干渉変数は，一つだけとはかぎらない．イメージ能力のほかに，記憶力，知能，過去の経験，……とたくさんあるかもしれない．それらの干渉変数をすべて統制するためには，全部の変数を測定しなければならないのだから大変である．かりに，全部，正確に測定できたとしても，さらに難しい問題が待ちかまえている．

対等化法で統制を行なうためには，どの干渉変数についても同じ値をとる，というペアをつくらなければならない．しかし，イメージ能力も同じ，記憶力も同じ，知能も同じ，過去の経験も同じ，……というようなペアを，はたしていくつつくることができるだろうか？　一つでもできれば僥倖で，何万人もの人に検査をしてみても，一つもできないかもしれない．統制しなければならない干渉変数の数がふえればふえるほど，ペアをつくるのは困難になるだろう．

一方，無作為配分のほうはどうかというと，無作為配分をすれば，干渉変数となりうる個体差は，いわば一網打尽に統制することができるので，組み合わせの心配は，まったくする必要がないのである．

そういわれると，当然，疑問がわいてくるにちがいない．――では，なぜ，無作為配分をやるだけで，個体差に由来する干渉変数を一網打尽に統制するこ

とが可能になるのだろうか？

3.5 統計的検定と無作為配分

3.5.1 統計的検定の役割

無作為配分を行なうと，統計的検定の威力を最大限に利用して，干渉変数の統制を行なうことができるから――というのが，その答である．

ごくおおざっぱにいうと，統計的検定というのは，偶然がもつ数学的な性質を利用して，偶然と必然とを見分ける方法だ，ということができる．

例えば，方向オンチ実験で，「方向感覚」を調べる課題を行なった結果，実験条件のほうが統制条件より平均値が少し低かった（すなわち，ズレが小さかった）とする．t 検定で分析したところ，確率値 p の値は 0.04 となったので，「統計的に有意である」と結論した．――ということは，どういうことなのかというと，「この平均値の差が偶然に得られる確率は 0.04，つまり，4% かそれ以下しかない」ということなのである．いいかえると，「この平均値の差は，偶然ではなくて必然，すなわち，独立変数の効果だという可能性が高い」ということになる．

つまり，統計的検定を使って，「偶然か必然か」という判断をしているのである（あくまでも確率的な判断ではあるが）．それができるのは，偶然がもつ数学的な性質を利用して，統計的検定が偶然の働き（もうすこし正確にいうと，偶然によるデータの変動）を取り除いてくれるからである．

では，その偶然の変動は何によって生み出されるのかというと，データ，つまり従属変数に影響しうる要因でなければならないのだから，「干渉変数だ」ということになる．イメージ能力，記憶力，知能，過去の経験，……といった，個体差に由来する干渉変数から，騒音や気温など，個体差とは別の干渉変数にいたるまで，無数の干渉変数が作用した結果，偶然の変動が生み出されたのだ，と考えることができる．偶然の変動であるかぎり，統計的検定によって取り除くことができる．

「被験者を無作為配分する」ということはどういうことかというと，「干渉変数となりうる個人差の値を無作為に配分するということだ」と考えることができる．いいかえれば，実験条件と統制条件とのあいだで，個体差の値がどうなるかは，偶然に決まる，ということなのである．結果的に，実験条件のほうが，イメージ能力の平均値が高くなったとしても，それは偶然の結果なのである．偶然の結果である以上，干渉変数におけるそうした平均値の差の影響は，統計的検定によって取り除くことができる．取り除いたあとで，偶然的ではない，独立変数による必然的な効果が残るかどうかを調べることができるのである．

また，無作為に配分される個人差は，イメージ能力だけに限られるわけではない．被験者を無作為に配分することによって，被験者に付随する個人差は，全部，無作為に配分されることになる．したがって，無作為配分と統計的検定を組み合わせれば，すべての個人差を同時に統制することができるのである．

これが，無作為配分が「王道」とよばれる所以なのである．

無作為配分を行なったかどうかは，実験結果を解釈するうえで非常に重要なので，無作為配分を行なっていない実験は，「準実験」quasi-experiment とよんで，無作為配分を行なった真正の実験とは区別することが多い．

無作為配分を行なった実験では，有意差がでた場合，それが「個人差による偶然の効果なのではないか」という心配はしなくてもすむ．もう少し正確にいうと，そういう可能性は，確率値 p で表される程度のわずかなものでしかない，と考えることができる．方向オンチ実験の場合も，無作為配分をやっていれば，「イメージ能力の違いがこの有意差のほんとうの原因なのではないか」という先生の批判には，「個人差は，イメージ能力テストなどよりもずっと信頼のおける無作為配分によって統制してあるので，その心配はほとんどありません」と答えることができたのである．

ここで，念のために，干渉変数の影響を「取り除く」partial out という表現の意味について，注釈を加えておくことにしよう．これは，もちろん，個体差そのものを「取り除く」という意味ではない．無作為配分を行なったからといって，当然のことながら，被験者のイメージ能力がみな同じになるわけではない．個体差の影響を「取り除く」というのは，それが独立変数の効果とは混同されないようにする，ということなのである．

もう少しだけ詳しく説明しよう．例えば，実験群の平均値と統制群の平均値を比べる場合，その差が独立変数の効果を表しているのかどうかを判断するた

めに，統計的検定では，その平均値の差を，推定された「偶然の効果」と比較する（数学的には，比をとる）．「偶然の効果」よりもはるかに大きければ，その平均値の差が偶然に得られた可能性は小さいので，「その差は，独立変数の効果を表している」と判断することになる．観測された平均値には，もちろん個体差の影響が含まれている．したがって，平均値の差も，個体差の影響で生まれた可能性がある．しかし，その個体差の影響が偶然に生じたものならば，それに由来する平均値の差は，「偶然の効果」とそう変わらない大きさのものになるだろう．もし，実際に得られた平均値の差が「偶然の効果」よりずっと大きなものだったとすると，その平均値の差には，個体差に由来する偶然的な差以外に，独立変数の効果が含まれていた可能性が高くなる．つまり，検定の結果は「有意」となる．一方，もし，無作為配分を行なっておらず，個体差の影響は偶然に生じたものではないということになれば，個体差の影響に由来する平均値の差だけでも，「偶然の効果」よりずっと大きくなってしまう可能性がでてくる．この場合には，平均値の差と「偶然の効果」とを比較することにより，独立変数の効果を正確に判断することはできなくなってしまうのである．

もうひとつだけつけ加えておくと，無作為配分によって個人差の影響を「取り除く」ことができたということは，個人差の影響がないことを立証したということではない．例えば，「イメージ能力が方向感覚には影響しない」ということが立証できたわけではないのである．イメージ能力の影響を調べるためには，イメージ能力を独立変数とした別の実験を行なってみなければならない．

3.5.2 無作為配分の重要性

統計的検定は，偶然の効果を取り除くことができる．しかし，取り除くことができるのは，偶然の効果，すなわち，「無作為誤差」random error（あるいは「偶然誤差」）だけである．

実験条件と統制条件のあいだに，何か偶然ではない違い，すなわち，「バイアス」bias（あるいは，「恒常誤差」constant error，「系統誤差」systematic error）があった場合には，統計的検定では，それを取り除くことはできない．例えば，先に実験室にやってきた被験者をみな実験条件の被験者にした結果，実験条件では，被験者の「やる気」が統制条件より高くなっていたとする．この「やる気」の差は，偶然生じたものではない．実験手続きの必然的な結果である．したがって，これは無作為誤差ではなく，バイアスである．統計的検定は，このバイアスを取り除くことはできないのである．したがって，検定の結果，「有意差あり」という結論がでたとしても，従属変数における有意差のほんとうの原因は，独立変数ではなく，このバイアスかもしれない，という可能性が排除できないことになる．

統計的検定から正しい結論を導き出すためには，データの変動（ばらつき）は，すべて，偶然の効果，すなわち，無作為誤差でなければならない．いいかえると，すべてのバイアスを取り除けるように，注意深く，実験手続きを決めておく必要がある．これは，統計的検定を行なう前にやっておかなければならないことなのである．

個体差のある実験要素（被験者など）を使って，被験者間計画で実験を行なう場合，バイアスの重要な源泉は個体差である．被験者内計画の場合も，前にでてきた記憶実験のなかのおぼえなければならない単語のように，個体差のある実験要素を使うときには，個体差がバイアスの源泉になりうる．このバイアスを取り除く，いちばん確かな方法が無作為配分なのである．したがって，個体差のある実験要素を使う実験においては，無作為配分は，統計的検定を使うための大前提なのだといっていい．

だが，現実には，無作為配分なしに実験を行ない，その結果を統計的検定にかけて結論を引き出している研究が少なくない．たしかに，どのようなデータでも，統計的検定にかけることはできる．形式的な条件さえ整っていれば（例えば，t 検定の場合，平均値・標準偏差・標本数という数値のセットが二組そろっていさえすれば），検定の機械的な手続きを実行することは，つねに可能である．「1% 水準で有意」「5% 水準で有意差なし」というような，もっともらしい結論も，かならずでてくる．しかし，無作為配分を行なっていないのであれば，その結論は信用には値しない．肉挽き機にバナナをいれても，挽肉はできないのである．

方向オンチ実験の場合も，無作為配分をやっていない以上，はじめに紹介した学生さんの先輩が先生から指摘されたような問題からは，逃れることができない．つまり，有意差のほんとうの原因は，独立変数ではなく，イメージ能力

の違いかもしれないし，あるいは，記憶力の違いかもしれないし，ほかの予想もしなかった個人差変数の違いかもしれない．すなわち，統計的検定では取り除くことのできない，バイアスかもしれないのである．

3.5.3 組織的配分と無作為配分の組み合わせ

有意差がでた場合の話をしてきたが，こんどは，有意差がでなかった場合について，少し考えてみよう．

無作為誤差も，あまり大きくなりすぎると，統計的検定では取り除くことができなくなってしまう．そうなると，ほんとうは，独立変数の効果があるのに，「有意差なし」という間違った結論を下してしまうことになる．いいかえると，統計学の用語でいう「第2種のあやまり」type II error を犯してしまうことになる．

もし，ほんとうは差があるのだとすると，こういう結果がでたのは，「検定力」statistical power が足りないからである．「検定力」というのは，「ほんとうは差があるとき，正しく『有意差がある』と結論できる確率」のことをいう．

検定力を上げるためには，被験者の数を増やす，という方法がある．これは，検定のさいの「自由度」degrees of freedom を大きくすることにつながる．

しかし，もうひとつ，誤差を小さくするという方法もある．被験者の数を増やすと，実験にかかる時間と労力（それに，ときには費用）がかさむので，もし誤差を小さくする方法があるのなら，そちらを使ったほうが経済的である．

従属変数に大きな影響をおよぼすことがはっきりと予測できるような個体差がある場合，たんに無作為配分をしたのでは，その個体差がみな無作為誤差になり，検定力を低下させてしまう．そこで，無作為配分と組織的配分を組み合わせるという方法を考えることになる．

方向オンチ実験の場合を例にとろう．過去の研究から，知能が重要な個人差干渉変数になりそうだ，ということが予測できたとする．その場合，あらかじめ知能を測っておき，知能に関しては対等化法を適用する．つまり，同じ知能指数をもつ被験者のペアをつくる．どのペアについても，どちらを実験条件，どちらを統制条件に配分するかは，こんどは，無作為に決める．

こういう方法で組織的配分（対等化法）と無作為配分を組み合わせると，組織的配分によって，知能の個人差については，実験条件と統制条件のあいだで無作為誤差を減らすことができる．その結果，検定力が高くなる．しかし，ペアの配分については無作為配分を行なっているので，統計的検定を正しく適用し，正しい結論を得ることができるのである．

前に，実験条件の被験者だけに，全部，先に実験をやってもらい，イメージ検査の成績が同じ人だけに統制条件の被験者になってもらう，という方法は危険な方法だ，と述べた．しかし，そういうやりかたをとらず，まず，統制したい干渉変数についての測定（イメージ検査や知能検査）を行ない，その得点にもとづいてペアをつくってから，ペアのメンバーを無作為配分するのであれば，バイアスをつくりだす危険は避けることができる．

実際には，すでに記したように，厳密に得点が等しいペアをつくることは難しいので，「乱塊法」randomized blocks design とよばれる，つぎのような方法をとることが多い．まず，全部の被験者を得点の順にならべる．つぎに，得点が多い順に，実験条件の数だけ被験者を選ぶ．もし，実験条件が五つあるのなら，5人の被験者を選ぶ．この手続きを，被験者がいなくなるまで続ける．そうすると，5人の被験者からなるグループがいくつもできる．さいごに，それぞれのグループの中で，被験者をそれぞれの実験条件に無作為に配分するのである．この乱塊法で得られたデータは，被験者内計画の場合と同じ統計的検定によって分析できる（Edwards, 1985; 森・吉田，1990 参照）．

3.6 無作為抽出と無作為配分

3.6.1 相違点

統計学の本をみると，「統計的検定を行なうためには，標本は母集団から無作為に抽出しなければならない」と書いてある．無作為抽出を説明するとき，例としてよく使われるのは，選挙の予測である．全有権者を「母集団」population とし，そこから，「標本」sample を無作為に選ぶ．その標本となった人たちに，どの党に投票するかをたずねる．標本を無作為に選ぶ，すなわち，無作為抽出を行なうことによって，標本を調査した結果から，母集団の特性，すなわち，有権者全体のうち，どのぐらいの割合の人がどの党に投票しようとし

ているかについて，正確な推定を行なうことが可能になるのである．

無作為配分は，この無作為抽出とよく混同される．両者のあいだには，あとに述べるような共通点もあるのだが，はっきりとした相違点もある．無作為配分を正しく理解するためには，まず，その相違点を正確に認識しておかなければならない．

無作為抽出は，母集団から標本を選ぶ手続きである．一方，無作為配分は，標本（被験者など）を複数の実験条件に振り分ける手続きである．したがって，手続きとしては，全然ちがう．

手続きがちがう以上，目的もちがう．無作為抽出の目的は，標本を母集団と等質にすることである．一方，無作為配分の目的は，複数の実験条件のあいだで，標本同士を等質にすることである．

しかし，どちらの等質化も，いわば，目先の目的にすぎず，等質化の最終的な目的は別にある．その最終的な目的も，無作為抽出と無作為配分とでは違っている．その違いを理解するためには，クックとキャンベル（Cook & Campbell, 1979）が提唱した「内部妥当性」internal validity と「外部妥当性」external validity という区別が役にたつ．

内部妥当性があるということは，「研究結果が現実を正しく反映している」ということを意味する．実験の場合，干渉変数の効果を独立変数の効果と取り違えたりすることなく，独立変数と従属変数とのあいだに因果関係があるのかどうか，正しく判断することができれば，その実験には，「内部妥当性がある」ということになる．

無作為配分は，この内部妥当性を確保するための方法の一つなのである．被験体の配分を無作為に行えば，個体差干渉変数の効果を独立変数の効果と取り違える可能性は，非常に小さくすることができる．

一方，外部妥当性があるということは，「研究結果を一般化できる」ということを意味する．実験の結果が，その実験で使われた被験者や材料など以外にもあてはまるのであれば，「外部妥当性がある」ということになる．

無作為抽出は，この外部妥当性を確保するための方法なのである．母集団から標本を無作為に抽出すれば，標本から母集団への一般化が可能になる．つまり，標本でわかったことが母集団にどの程度あてはまるのか，推計学的手法をもちいて，確率的に推定することができるのである．

一方，無作為配分は外部妥当性を保証してくれないし，無作為抽出は内部妥当性を保証してくれない．

無作為配分を行なえば，「独立変数と従属変数とのあいだには因果関係がある」と強い確信をもって結論することができる．しかし，この結論には，厳密にいえば，「その実験に参加した被験者にかぎって」という但し書きをつけなければならない．例えば，四国のある女子短期大学の2年生48人から得られた実験結果が，はたして，同じ大学の他の2年生にもあてはまるのか，日本の大学生全体にもあてはまるのか，ひいては，人類全体にもあてはまるのか，ということになると，いかなる確率をもってしても，推定はできないのである．つまり，外部妥当性は保証されない．

逆に，選挙予測のために，標本を無作為に抽出したとしても，極端な話，調査員として雇われた学生アルバイトが，手間を惜しんで，調査用紙の一部に自分で記入してしまったとしたら（こういうことは，実際によく起こるらしいが），調査結果が，標本となった人たちの投票傾向を正確に反映しているとは，もちろん，いえないだろう．すなわち，「無作為抽出はしたが，内部妥当性はない」ということになる．

さらに，もともと，調査結果が正しくないものだとすると，それを母集団に一般化して，正しい結論が得られるはずはない．つまり，内部妥当性があることは，外部妥当性があることの前提条件になっているわけである．その内部妥当性を確保するための基本的な方法の一つが，無作為配分なのである．

3.6.2　共通点

このように，無作為配分と無作為抽出は，重要な点で異なっているのだが，他方，本質的な共通点もある．それは，推計学的な手法が扱うことのできる偶然のプロセスをつくりだす，という点である．

まず，無作為抽出の場合を考えてみよう．母集団から標本を無作為抽出すると，標本と母集団の関係は，偶然のプロセスによって支配されるものとなる．その偶然のプロセスがもつ数学的な性質を利用すると，標本の特性（割合や分散など）から母集団の特性（同じく，割合や分散など）を推定することができ

コラム 3.2

心理学の研究はみな無意味？

統計の入門コースで無作為抽出の話を聞いた学生のなかから，こんなことをいいだす人がときどき出てくる――「心理学では，無作為抽出をやったという研究はみたことがない．にもかかわらず，大多数の研究は統計的検定を使っている．無作為抽出をやっていなければ，検定は無意味なはずだ．とすると，心理学の研究は，大多数が無意味だということになるのではないだろうか？」たいへん鋭い指摘で，こんなことを思いつく学生は，よほど優秀な学生にちがいない．しかし，本文で述べたとおり，実際には，統計的検定の前提は無作為抽出だけではないのである．無作為抽出を行なっていなくても，無作為配分を行なっていれば，検定は，十分に意味のあるものになる．検定の結果にもとづいた結論も，やはり意味のあるものになる．というわけで，幸いなことに，かならずしも，大多数の心理学的研究が無意味だということにはならないのである．無意味な研究がないというわけではないにしても……．

なる．その独立変数ともとの実験の独立変数とのあいだに，「交互作用」interaction があるかどうかを調べることになるわけである．これはもう，科学的探究そのものにほかならない．

つまり，現実に可能な一般化とは，科学的探究の進展とともに，徐々にその範囲が明らかになっていくようなものなのである．「無作為抽出の手続きから自動的に確定する」というものでは，けっしてない．

そうしてみると，個々の科学的探究で使われる無作為配分は，ある意味では，無作為抽出より，はるかに重要な手続きだといってもよいのかもしれない．にもかかわらず，統計学の教科書では，ふつう，「全有権者」のような具体的な母集団からの無作為抽出だけが解説してあって，そのあとすぐに，統計的検定の話になってしまう．無作為配分の解説がのっている例は，ほとんどないのではないだろうか．そのため，「無作為抽出を行なわない実験では，無作為配分が，統計的検定を正しく行なうための必須の前提になる」という重要な事実がよく理解されずに終わってしまい，無作為配分なしの誤った統計的検定が横行する結果になっているのである．

3.7　無作為配分の限界

無作為配分の重要性については，これ以上強調する必要はないだろう．しかし，「王道」といわれる無作為配分も，けっしてオールマイティではなく，いくつもの限界をもっている．最後に，その限界を確認しておくことにしよう．

無作為配分は，いつもできるとはかぎらない．これは確かに大きな限界である．

例えば，新しい教授法の効果を調べる実験授業を考えてみよう．新しい教授法で教える実験群のクラスと，従来の教授法で教える統制群のクラスは，生徒を無作為配分して編成したいところだが，かならずしも，そこまで学校側の協力が得られるとはかぎらない．このように，現実的な制約のために無作為配分ができない，というケースはしばしばでてくる．

一方，原理的に無作為配分ができないという場合もある．例えば，「日本語や中国語には欠陥があるせいで，日本人や中国人は高度の思考能力が劣っている」というアメリカの研究者の説を検証するために，日本人とアメリカ人を比較する実験を計画したとする（高野，1995）．この場合，言語の違い以外の干渉変数はできるだけ統制したいわけだが，かといって，被験者を日本人とアメリカ人に無作為に配分するというわけにはいかない．こうした「現場研究」field study では，どうしても既存のグループを比較せざるをえないのである．その場合，実験は，無作為配分を含まない準実験となる．無作為配分による個人差変数の統制ができない分，準実験においては，因果関係の推定が難しくなる．しかし，準実験においても，因果関係の推定をできるだけ確実なものにしようと，さまざまな技法が開発されている（Cook & Campbell, 1979 参照）．無作為配分ができない場合には，そうした技法を上手に活用することが必要になってくるだろう．ただ，それをしたとしても，無作為配分を行なった場合に比べて，因果関係の推定が不確かになることは避けられないのである．

他方，無作為配分ができたとしても，それだけで因果関係の推定が完璧になるわけではない．

第一に，無作為配分は，個体差を統制する手続きにすぎない．個体差以外の干渉変数を統制するためには，無作為配分とは別に，それぞれの干渉変数に応

じた，適切な統制の手続きをとらなければならない．

第二に，その個体差干渉変数も，無作為配分では統制しきれない場合がある．実験の最中に，実験条件のあいだで異質化が生じた場合である．例えば，被験体の「損耗」attrition が起こったとしよう．「損耗」というのは，被験体の脱落のことである．ネズミを使った実験では，酷使しすぎて，ネズミが何匹か途中で死んでしまうことがある．人間の場合も，あまり大変な実験だと，死んでしまう，ということにはならなくても，何人かの被験者が途中で実験を投げ出してしまうことがある．こうした損耗の割合が独立変数と相関していた場合には，はじめに無作為配分を行なっていたとしても，個体差干渉変数の問題がふたたび浮上してくることになる．

例えば，実験群のほうが課題が難しかったため，統制群より損耗率が高かったとすると，実験が終わったとき，実験群では，強靭な被験体の割合が統制群より多くなっているかもしれない．従属変数の測定結果が得られるのは，その残った被験体についてだけである．したがって，たとえ実験群のほうが優れた成績を示していたとしても，それは独立変数の効果ではなく，この個体差干渉変数，すなわち強靭さの効果だ，という可能性も考えなければならないことになる．

損耗の場合は，問題が生じたことがわかるだけ，まだましかもしれない．というのは，実験者の気づかないところで個人差が生じてしまう場合もあるからである．例えば，新しい教授法が従来の教授法より優れているかどうかを調べるために，ひとつのクラスでは新しい教授法を，別のクラスでは従来の教授法を使って，1年間，授業を行なったとしよう．学校が実験に協力的で，クラス分けをするときに，無作為配分をしてくれたとする．だが，途中で，よい先生のいる塾が近くにでき，新しい教授法で教えているクラスのほうでは，友だち同士が誘いあって，その塾に通う生徒が多くなったとしたらどうだろう．学力の個人差にはバイアスがかかってしまうことになる．はじめに無作為配分をしたにもかかわらず，二つのクラスのあいだにある学力の差は，もはや無作為に生じたものとはいえない．こうなると，学年末のテストで，たとえ新しい教授法で教えられたクラスのほうが成績がよかったとしても，それを教授法の効果なのだと結論してよいかどうかは，はなはだあやしくなってくる．ほんとうは塾の効果だったのかもしれない．しかし，塾のことを実験者が知らずにいたとすると，確信をもって誤った結論を下してしまうことになりかねない．

このように，実験の最中に起こった個体差干渉変数の異質化については，無作為配分はまったく無力なのである．

第三に，無作為配分の統制力は，統計的検定に由来しているので，統計的検定の限界は，そのまま，無作為配分の限界になる．統計的検定は，偶然がもつ数学的性質を利用して推定を行なうために，いくつもの仮定をもうけていることが多い．「分析するデータにはその仮定があてはまらない」という場合には，統計的検定の結果は，あまり信用できないことになる．例えば，「二つの母集団の分散が等しい」という仮定を考えてみよう．母集団の分散は，直接には知りようがないので，この仮定が正しいのかどうか，正確に判断することはできない．分散が等しいかどうかを調べる検定もあるが，これも検定である以上，何か別の仮定を使っている．したがって，いわば，問題を先送りしているだけで，最終的な解決にはならないのである．もし，「分散が等しい」という仮定が，ほんとうは正しくなかったとすると，この検定の結果は誤ったものになってしまう．検定が誤っているとすれば，当然，無作為配分によって個体差干渉変数をうまく統制することもできないことになる（統計的検定にまつわる問題については，本書の第4章および橘（1986）などを参照）．

無作為配分のこうした限界を考えると，きちんと無作為配分をしたからといって，その実験だけから，因果関係について最終的な判断を下すことは危険だということがわかる．

すでに述べたように，妥当性や信頼性のあやしいイメージ能力検査は，個人差干渉変数を統制する方法としては，あまりあてにはできない．一般に，個人差干渉変数を個別に統制しようとする方法は，たとえ妥当性や信頼性が高い場合でも，いろいろな問題を抱えている，という話もした．無作為配分ができる場合には，無作為配分がもっとも強力な統制の方法なのである．

とはいえ，その無作為配分も万能ではない．実際には無作為化の手続きが不正確で，統計的検定の数学的モデルが扱えるほんとうの無作為なプロセスは，実現できていなかったかもしれない．あるいは，実験者の気づかないところで，実験中に被験者の異質化が起こっていたかもしれない．統計的検定が依存して

いる数学的モデルの仮定が実際には満たされていなかったか，またそれとは別の理由から，数学的モデルによる推論が正しく働かなかったかもしれない．いずれにしても，無作為配分を行なったからといって，個人差干渉変数の統制が完璧になるという保証はないのである．当然，因果関係の推定のほうも，完璧だという保証はない．

こうした無作為配分の限界に対処するための最善の方法は，追試である．世界中のいくつもの実験室で，慎重に行なわれた実験において，同じような結果が繰り返し観察されたのなら，その結果は信用に値する．ある実験室でたまたま働いていた干渉変数が，ほかのいくつもの実験室でもたまたま同じように働く，という確率はきわめて低いからである．物理学では，一時期，「常温核融合」という現象が世界中にセンセーションをまき起こしたが，追試の失敗が続いた結果，近年では，その存在自体が疑問視されるようになってきた(Taubes, 1993)．

実際の研究では，ただ追試をしてみるというだけではなく，「干渉変数の効果が独立変数の効果と混同されているのではないか」というはっきりとした疑いをもって実験をする場合も多い．目星をつけた干渉変数を独立変数として，それがある値のときには前の実験と同じ結果が得られるが，別の値のときにはそういう結果は得られない，ということを示そうとするのである．

さきほど触れた，「日本語や中国語に欠陥があるせいで，日本人や中国人は高度の思考能力が劣っている」と主張するアメリカ人が行なった実験では，中国人や日本人とアメリカ人を比較し，「中国人や日本人のほうが劣っている」という結果を得ていた．その原因は言語の違いだと推定していたのである．ところが，私が日本人の大学生で追試を行なってみたところ，文系の学生はアメリカの学生と同じ結果のパターンを示すことがわかった．もとの実験を行なったアメリカ人の研究者に問い合わせてみた結果，もとの実験では，アメリカ人の被験者は文系の学生ばかりだったが，中国人や日本人の被験者の中には，文系と理系の両方の学生がいたらしい，ということもわかった．そこで私は，もとの実験の中に見つかった手続きのミスも考え合わせて，「関数についての知識のちがい（干渉変数）が言語のちがいと混同されていたのではないか」という仮説をたてた．この仮説からは，「アメリカの理系の学生を被験者にして実

験を行なってみれば，中国人や日本人，とくに理系の学生とそっくりな結果のパターンが得られるはずだ」という予測がでてくる．実際にそういう実験を行なったところ，まさしく，この予測どおりの結果になったのである（高野，1995 参照）．これは，無作為配分ができない準実験の例だが，無作為配分ができる場合にも話はかわらない．

実験というものは，「絶対に確かな手続きがあって，それを使えば，絶対に確かな結論が得られる」というものではないのである．どれほど精緻な実験を行なっても，誤りの可能性はつねに残る．因果関係についての推定を少しでも確かなものにしていくためには，いろいろな視点からいろいろな実験を続けていくことが不可欠なのである．新たな視点は，研究者のあいだの相互批判から生まれる．その相互批判が，結果として，より確かな推定に近づくための協同作業となるのである．

ただし，いうまでもないことだが，個々の実験は，完璧ではないにせよ，できるかぎり確実な方法にもとづいて行なわなければならない．そうした方法の一つが無作為配分だというわけである．

引用文献

Cook, T. D., & Campbell, D. T. 1979 *Quasi-experimentation : Design and analysis issues for field settings*. Houghton Mifflin.

Edwards, A. L. 1985 *Experimental design in psychological research*, 5th ed. Harper & Row.

森　敏昭・吉田寿夫(編)　1990　心理学のためのデータ解析テクニカルブック．北大路書房．

Solso, R. L., & Johnson, H. H. 1989 *An introduction to experimental design in psychology : A case approach*, 4th ed. Harper & Row. 浅井邦二(監訳)　1999　改訂心理学実験計画入門．学芸社．

橘　敏明　1986　医学・教育学・心理学にみられる統計的検定の誤用と弊害．医療図書出版社．

高野陽太郎　1995　言語と思考．大津由紀雄(編)　認知心理学 3　言語．東京大学出版会．pp. 245-259.

Taubes, G. 1993 *Bad science : The short life and weird times of cold fusion*. Random House. 渡辺　正(訳)　1993　常温核融合スキャンダル．朝日新聞社．