

岡山大学文学部紀要,
23, 107-120.

心理学研究法再考：(2)

心理学入門段階における“実験計画立案行動”の特徴と改善

長谷川芳典

心理学研究法再考

(2)心理学入門段階における“実験計画立案行動”の特徴と改善

長谷川芳典

本稿は、心理学専攻の学部2回生がみずから立案した実験計画の特徴と問題点を検討することにより、実験計画に関する入門指導のありかたを考察することを目的とする。

前回の「心理学研究法再考(1)」では心理学の研究における統計解析の誤用問題を取りあげたが(長谷川, 1994)、結果を分析する以前の問題として、実験計画段階での不備というものもかなりの程度で認められるのではないかと思う。これを改善するには、入門教育段階において、実験計画の立案について十分な指導を行なう必要があるのではないかと考える。本発表は、心理学の実習の一環として行なった“実験”シミュレーションで学生がどのような実験計画を立案したのかについて事例を紹介し、今後の入門教育の改善に役立てることをめざすものである。

さて、ひとくちに実験計画と言っても、心理学の研究では、目的や研究対象の性質に応じてさまざまなタイプの実験計画が立案される。それらは、個体間比較法と個体内比較法に大別することができる。前者は、比較的多数の被験者(被験体)を無作為にいくつかのグループに割りつけ、それぞれの群に異なる処理を施し、その結果から群間に意味のある違いが生じたかどうかをみるものである。無作為な割りつけとは、各被験体が各処理群に割りつけられる可能性を等しくするような手続きによって割りつけることである。個体間比較による実験計画は、標本の無作為な割りつけを行なうことで、コントロールできない多くの独立変数の系統的な偏りを除くことを可能にする。それゆえ、従属変数についての群間の差は実験で操作した独立変数が原因であると推定することができるのである(橘, 1986, p.35-37参照)。いっぽう個体内比較法は、比較的少数もしくは単一の個体を対象として、独立変数を系統的に操作して個体の行動に及ぶ影響を測定するものである。結論の一般化には一定の制約がある反面、個体間比較法における倫理的問題や実施上の困難を解消し、より有用な知識が引き出される場合がある(Barlow & Hersen, 1984; 岩本・川俣, 1990; 桑田, 1995)。本稿では、これらのうち、個体間比較法による実験計画の問題を扱うことにする。

個体間比較法を用いてある種の母集団における一般法則を見い出そうとする場合には、標本が母集団から無作為に抽出されていることが前提となる。また、被験者(被験体)の数は多ければ多いに越したことはない。しかし、現実の心理学実験では、母集団からの無作為な抽出はきわめて困難であるし、また予算上及び時間的な制約から被験者(被験体)の数も限定せざるをえない。たとえば、心理学専攻生が卒論研究として実験を行なう場合には、被験者はたいがい所属大学の学生であり、それも友人やサークルなどを頼って100人も集めれば上等ということになってしまう。

そこで、実験計画を実際に立案する場合には、“予算無制限、被験者数無制限、実験時間無制限”

という理論上最適ではあるが実現は不可能であるような計画を考えるよりも、限られた被験者(被験体)を対象にして、いかに効率的な計画を立てるかという現実的な対処法を指導していく必要があるように思う。

本稿は、こうした状況を考慮し、仮想被験者数が100人という制限状況のもとで、仮想の薬が記憶の保持にどのような影響を及ぼすのかを検討する実験計画を立案させることとした。

方 法

心理学研究法の授業の一環として、パソコンによる仮想実験(実験シミュレーション)を行なった。実習班 岡山大学文学部心理学専攻2回生(研究生等を含む)19名。各3名ずつ6班に分かれた(1グループのみ4名)。実験計画の立案とデータ処理方法の決定にあたっては他の班のメンバーとは相談しないように指示した。なお、実験シミュレーション上の仮想被験者と区別するため、本報告では実験計画の立案者のことはあくまで“実習班”と呼ぶことにする。

教示 4つの課題のもと、ある薬が記憶の保持にどのような効果を及ぼすか、仮想のデータに基づいて検討することを目的とした。便宜上、薬はA、B、Cの3種類あるものとし、課題1～3では、それぞれA、B、Cの薬について、また課題4は、課題1と同じ薬Aを用いるが、記憶テストを翌日ではなく当日の6時間後に実施するものとした(ただし、課題1とは異なる仮想被験者に実施するものとした)。

以下に、課題1の教示内容を示す。

ある製薬会社が記憶の保持をよくする薬Aを開発した。この薬を被験者に与えることについては、最大150mgまでの範囲であれば、医学的、倫理的に何ら問題がないことが確認されている。

そこで、この薬の効果を確かめるため、100人の被験者にある記憶作業をさせ、その直後にこの薬を何gか服用させ、翌日に記憶テストを行なうこととしたい。

この薬はカプセルに入れられているので、被験者は薬の服用量を知ることができない。

上記の教示に加えて以下の点を補足した。

- ①あらかじめ実験計画を立てたうえで、データ収集にあたること。
- ②必要に応じて“予備実験”を行なってもよいが、仮想被験者の“人数”は“本実験”と合計して100名でなければならない。
- ③いかなる理由があっても、実験のやりなおしはできない。
- ④薬のカプセルの中身は見えず内容量の区別もできないものとする。したがって、偽薬を用いた何らかの対照条件を設定する場合は、空のカプセルを飲ませるだけでよい(本物の薬の量に対応して偽薬量を変化させる必要はない)。

手続

実習は、班ごとに別の実験室で実施した。仮想被験者の個々のデータは、実習班がパソコンを操作することにより数値として表示された。パソコン操作の手順及び出力は次の通りである。

- ①各実習班は、パソコン(NEC社のPC98シリーズ各種および互換機種)にインストールされている実験シミュレーションプログラムを起動する。
- ②被験者番号(1～100)が表示される。
- ③服用させる薬の量を数値入力する(0以上150以下)。
- ④その被験者の記憶テスト得点が整数値で表示される。
- ⑤被験者番号が100番になるまで②から④を繰り返す。

いずれの課題とも、薬の服用量はそれぞれ特定の関数に基づいて記憶テスト成績に影響を及ぼすものとした。

パソコンが実際に出力する得点は、この関数によって決定される値(便宜上“関数項”と呼ぶ)と、偶然に変動する値(便宜上“変動項”と呼ぶ)の和の整数部分として算出された。

得点=関数項+変動項

但し、関数項=f(服用量)

このうち変動項は、N88-Basic(MS-DOS版)のRND関数によって1回に10個の一樣乱数を発生させ、この合計値の定数倍により算出した。中心極限定理により、任意の分布に従う乱数列のn個の平均値(合計値も同様)はnが大ききときには正規分布に収束することが知られている(脇本, 1970, p.78)。そのさい、もとの分布が一樣分布である時にはn=6もあれば正規分布への近似の度合いはよくなると言われており(脇本, 1970, p.79-82)、n=10のもとで算出した変動項の値は正規分布へ十分に近似するものと考えてよいだろう。

課題1～3では、変動項は関数項と無関係に定数倍したものを用いた。実験4のみ、服用量が大きくなるに従って変動項の値及び分散が減少するように設定した。

課題1

課題1の関数項は、服用量をxmg、得点をy点とする時、 $y=0.3x$ [$0 \leq x \leq 150$] という1次関数となるように設定された。したがって関数項の値は服用量が0mgのときは0点、150mgのときは45点となる。これに加算される変動項は、平均値24.8、標準偏差4.7で近似的に正規分布する(変動項が加算されるため、服用量が0mgの時の実際の平均得点は約25点となる)。図1aには、服用量を0～150mgでランダムに1000回設定したときの得点の散布状況を示した。同じく図1bは、設定回数を本課題と同じ100回に限定した場合の散布状況を示す。

課題2

課題2の関数項は、2通りに分岐している。課題1とまったく同様に得点が服用量の1次関数となる確率が30%、それ以外の時は薬は得点に影響を及ぼさない(関数項=0となる)ように設定された。変動項は課題1とまったく同様であり、服用量が0mgの時は、平均値24.8、標準偏差4.7で近似的に正規分布する。この課題は、薬の効果が2通りに分岐するため、服用量が増加するにつれて“個体差”が必然的に増大する。その結果、服用量が150mgの時の標準偏差は18.9となる。また、服用

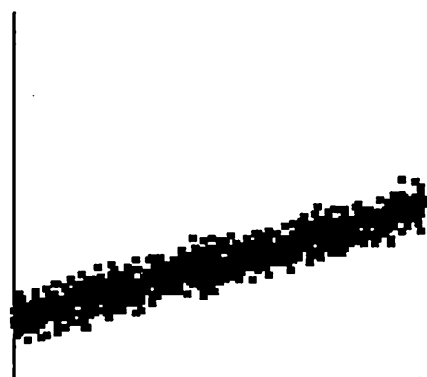


図1 a

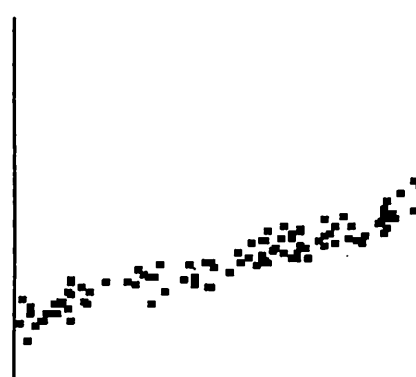


図1 b

図1 課題1で、服用量を0mgから150mgの範囲でランダムに1000回設定したときのテスト得点の散布状況。(図1 a)。同じく100回設定したときの散布状況 (図1 b)。

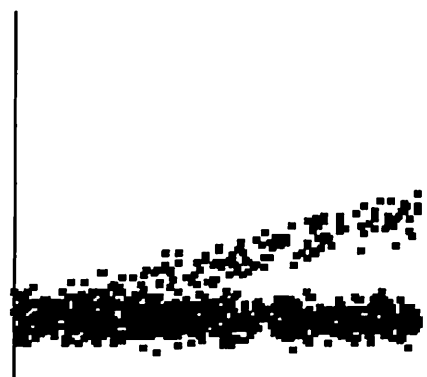


図2 a

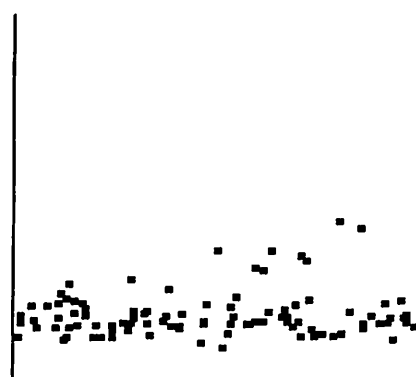


図2 b

図2 課題2で、服用量を0mgから150mgの範囲でランダムに1000回設定したときのテスト得点の散布状況。(図2 a)。同じく100回設定したときの散布状況 (図2 b)。

量が0 mg以外の条件では双頭型の分布となるため、服用量各量のもとでの得点自体は正規分布には近似されない。

この課題は日常場面にあてはめれば、“薬の効果は体質によって異なる。30%の人には有効であるが、70%にはまったく効かない”ということの意味する。このような設定は南風原 (1994) の発表にヒントを得たものである。図2 a、及び図2 bに、服用量を0～150mgにランダムに設定した場合の得点の散布状況を示した。

課題3

課題3の関数項は服用量を x mg、得点を y 点とする時、 $y=100\sin(\pi x/151)$ [$0 \leq x \leq 150$] となるように設定された。これは、“この薬はほどよく飲めばよく効くが、飲み過ぎるとほとんど効果がなくなる”という日常場面を模したものである。関数項の値は、服用量が0 mgの時には0点、75mgの

時に100点で最高となり、150mgでは2点まで減少する。また変動項は平均124.1、標準偏差23.3で近似的に正規分布する(変動項が加算されるため、服用量が0 mgの時の実際の平均得点は約124点となる)。図3 a、及び図3 bに、服用量を0～150mgにランダムに設定した場合の得点の散布状況を示した。



図3 a

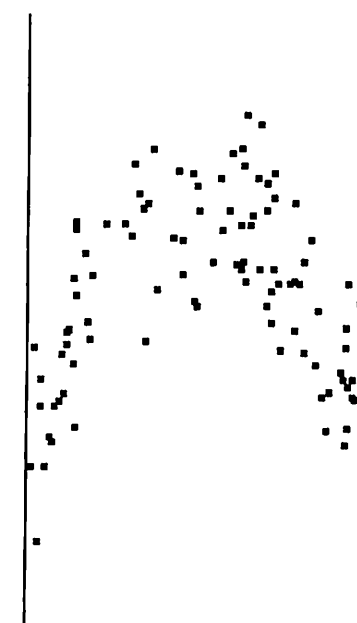


図3 b

図3 課題3で、服用量を0mgから150mgの範囲でランダムに1000回設定したときのテスト得点の散布状況。(図3 a)。同じく100回設定したときの散布状況 (図3 b)。

課題4

課題4の関数項は服用量を x mg、得点を y 点とする時、 $y=\sin(\pi(x+100)/502)$ [$0 \leq x \leq 150$] となるように設定された。この課題を単独にとらえると、これは、“この薬はたくさん飲むほど効果が上がるが、量が増えるに従い効き目の増え方は限界に近づく”という日常場面を模したものとも言える。しかし、ここでは、“課題4は課題1と同様の薬Aを用いるが、記憶テストを翌日ではなく当日の6時間後に実施するものとする。”というように教示内容の一部を変更し、天井効果の可能性を気づかせることを主目的とした。つまり、記憶テストが翌日ではなく当日に行なわれることによって、テストは相対的に易しくなり、多数の者が満点に近い得点をとった可能性を指摘させることにあった。

この課題の関数項の値は、服用量が0 mgのときは-66点、150mgのときは100点となる。また変動項は、服用量が0 mgのときは平均124.1点で標準偏差は23.3点であるが、服用量の増加に伴って直線的に減少し、150mgのときは平均49.6点で標準偏差は9.3点となる。服用量各量のもとでの得点は正規分布に近似するようになっている。なお、服用量が0 mgの時の関数項の値はマイナスとなっているが、変動項の加算により、実際には平均値が58点、また服用量が150mgの時には150点になるよう

に設定されている点に留意していただきたい。図4 a、及び図4 bに、服用量を0~150mgにランダムに設定した場合の得点の散布状況を示した。

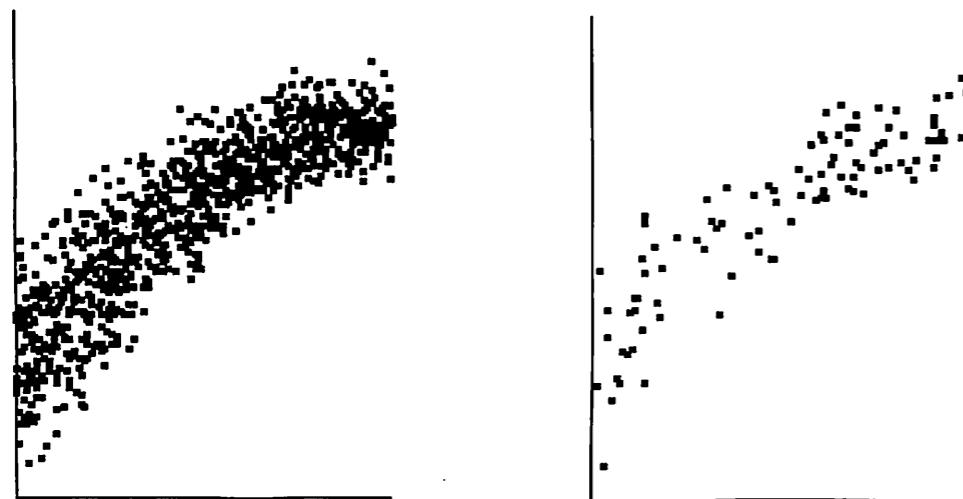


図4 a

図4 b

図4 課題4で、服用量を0mgから150mgの範囲でランダムに1000回設定したときのテスト得点の散布状況。(図4 a)。同じく100回設定したときの散布状況 (図4 b)。

報告

実験計画の内容と実験シミュレーションの結果および考察について、班別に報告書を提出させた。さらに、各課題の設定内容や各報告書の問題点等を解説したのちに、全員に個別にレポートを提出させた。

結果

全体的特徴

いずれの班も、薬の服用量が異なる条件をいくつか設定し、分散分析と多重比較に基づく分析を行なった。以下、6つの班を、A班~F班と呼ぶ。このうち、A班のみは、予備実験を行わず、25名×4群(0、50、100、150mg)でいきなり比較した。残りの5つの班は、原則として20~44名の仮想被験者を対象にまず予備実験(または第1実験)を行なって全体的な特徴を把握したのち、条件を絞って本実験(または第2実験)で最適量を把握する方略をとった。

B班は、薬の効果があらかじめ5つのパターンのいずれかであるとの仮説を立てて予備実験を行なった。5つのパターンとは、①薬の服用量が多くなるほど薬の効果があがる、②薬の服用量が一定量を超えるとその効果が頭打ちになる、③薬の服用量がある一定量を超えないとその効果が現れない、④薬の服用量に最適範囲がある、⑤薬の服用量と効果に関係が見いだせない。このうち、①及び④の予測パターンは、課題1及び課題3で設定した関数項とほぼ同一であった。

C班も、一定の仮説を立てて予備実験を行なった。それは、①薬の量と効果が比例する、②一定

量に達するとそれ以上の効果はない、③一定量以上では効果が減少する、のいずれかであろうというものであった。

F班は、予備実験および本実験という構成ではなく、第1実験から第2実験というステップをとり、第2実験では、第1実験で得点が最高値を示した条件と、服用量0mgの対照条件のみを、仮想被験者を多めに確保して比較する方略をとった。

表1~4に、各班の実験計画および結論の概略を課題別に示す。

表1 課題1における各班の実験計画および結論の概略

班	予備実験 (第1実験)	本実験 (第2実験)
A	行なわず。	25名×4群(0、50、100、150mg) 分散分析の結果主効果あり。多重比較により各条件間にすべて有意差あり。薬が多いほど効果大であり、150mgが最適と結論。
B	5名×4群(0、50、100、150mg) 薬が多いほど効果大であると予測。	20名×4群(120、130、140、150mg) 分散分析の結果主効果あり。多重比較により150mgが最適と結論。
C	2名×16群(10mg刻みで0~150mgの16条件) 一定量に達するとそれ以上の効果なし、という仮説を採用。	11or12名×6群(0、30、60、90、120、150mg) 68名 予備実験から立てた仮説とは異なり、分散分析の結果主効果あり。下位検定により、60mg以上服用すれば記憶がよくなる効果があると結論。
D	5名×4群(0、50、80、100、150mg) 薬は有効であり150mgが最適であろうと予測した。	20名×4群(120、130、140、150mg) 分散分析の結果主効果あり。多重比較により150mgが最適と結論。
E	4名×5群(0、40、80、120、150mg) 薬の量を多くすると得点も上がると見通し。	20名×5群(0、40、80、120、150mg) 予備実験のデータを込みにした。 分散分析の結果主効果あり。多重比較により各条件間にすべて有意差あり。150mgまでなら、服用量に比例して効果が高まると結論。
F	4名×11群(0、15、30、45、60、75、90、105、120、135、150mg) 分散分析および多重比較の結果、主効果が有意であり150mgがもっとも有効と結論。	28名×2群(0、150mg) t検定の結果有意差あり。 150mg服用の有効性を追認。

表2 課題2における各班の実験計画および結論の概略

班	予備実験（第1実験）	本実験（第2実験）
A	行なわず。	25名×4群（0、50、100、150mg） 分散分析の結果主効果あり。多重比較により0mgと50mgの間を除く各条件間にすべて有意差あり。50mgより多い量であれば、服用量が多くなるほど効果大であると結論。
B	5名×4群（0、50、100、150mg） 薬が多いほど効果大であるようだが、バラツキが大きい。	20名×4群（0、50、100、150mg） 分散分析の結果主効果あり。多重比較により150mgが最適と結論。
C	2名×16群（10mg刻みで0～150mgの16条件） 少量では効果ないが、100mgあたりで効果が現れると予測。	11or12名×6群（0、30、60、90、120、150mg）68名 分散分析の結果主効果なし。薬Bは効果なしと結論。
D	5名×4群（0、50、80、100、150mg） 分散が大きく、顕著な差は認められず。	5名×16群（0～150mg、10mg刻み） 分散分析の結果主効果なし。薬Bは効果なしと結論。
E	4名×5群（0、40、80、120、150mg） 80mgのあたりで最も効果があると予測。	10名×8群（0、40、60、80、100、120、135、150mg） 分散分析の結果主効果なし。何mg服用してもしない場合と変わらないと結論。
F	4名×11群（0、15、30、45、60、75、90、105、120、135、150mg）で比較。分散分析で主効果有意。多重比較の結果135mg群他のいくつかの群より有意に大。	28名×2群（0、135mg） t検定の結果有意差あり。 135mg服用の有効性を追認。

表3 課題3における各班の実験計画および結論の概略

班	予備実験（第1実験）	本実験（第2実験）
A	行なわず。	25名×4群（0、50、100、150mg） 分散分析の結果主効果あり。多重比較により0mgと50mgの間、100mgと150mgの間に有意差あり。薬Cは記憶の保持に有効であり、50mgが適量であると結論。
B	5名×4群（0、50、100、150mg） 最適量は50～100mgのあいだにあると予測。	8条件（60、70、80、90、100、120mg；60と80条件のみ20名他は10名） 分散分析の結果主効果あり。多重比較により60mgが最適と結論。
C	2名×16群（10mg刻みで0～150mgの16条件） 薬が増えると効果が上がるが、100mgあたりからは効果が下がると予測。	11or12名×6群（0、30、60、90、120、150mg）68名 分散分析の結果主効果あり。多重比較により60mgぐらいが最適で、100mg以上は効果が減少し、150mgでは飲んでも意味がないと結論。
D	5名×4群（0、50、80、100、150mg） 0mgと150mgにはほとんど差がなかった。 有効服用量は50mg～150mgの間にあると予測。	16名×5群（25、50、75、100、125mg） 分散分析の結果主効果あり。多重比較により50～100mgの服用が最適であると結論。
E	4名×5群（0、40、80、120、150mg） 120mgのあたりで最も効果があり、それより多くなると効果が下がると予測。	10名×8群（0、20、40、60、80、105、125、150mg） 分散分析の結果、主効果あり。多重比較の結果、13対において有意差あり。この薬は20mg程度で効果を現し、80mg前後で最大、120mgを超えると逆に効果が下がると結論。
F	4名×11群（0、15、30、45、60、75、90、105、120、135、150mg） 分散分析の結果有意差あり。105mgの時に得点が最高となった。	28名×2群（0、105mg） t検定の結果有意差あり。 105mg服用の有効性を追認。

表4 課題4における各班の実験計画および結論の概略

班	予備実験（第1実験）	本実験（第2実験）
A	行なわず。	25名×4群（0、50、100、150mg） 分散分析の結果主効果あり。多重比較の結果ほとんどの群間に有意差があったが、100mg条件と150mg条件では差なし。100mgで十分であると結論。
B	5名×4群（0、50、100、150mg） 薬が多いほど効果大であるようだが100mg以上ではそれほど伸びていない。最適量は50～100mgのあいだにあると予測。	20名×4群（90、110、130、150mg） 分散分析の結果主効果あり。多重比較により110mgが最適と結論。
C	2名×16群（10mg刻みで0～150mgの16条件） 薬が増えると効果が上がるが、ある一定量以上からは変化なしと予測。	11or12名×6群（0、30、60、90、120、150mg）68名 分散分析の結果主効果あり。多重比較の結果90mgまでは薬が多いほど効果大、それ以上はあまり変化なしと結論。
D	5名×4群（0、50、80、100、150mg） 50mg～150mgのあいだでは効果ありと推定。	20名×4群（90、110、130、150mg） 分散分析の結果主効果あり。多重比較の結果90mg条件は他群より有意に小、他群間では差なし。90mgまでは薬が多いほど効果大、それ以上はあまり変化なしと結論。
E	4名×5群（0、40、80、120、150mg） 120mgのあたりで最も効果があり、その後は下がると予測。	10名×8群（0、40、60、80、100、120、135、150mg） 分散分析の結果主効果あり。多重比較の結果14対で有意差が認められたが、120mg以上の条件間では差がなく、服用は120mgで十分であると結論。
F	4名×11群（0、15、30、45、60、75、90、105、120、135、150mg） 135mgで最も効果大の傾向あり。	28名×2群（0、135mg） t検定の結果有意差あり。 135mg服用の有効性を追認。

考 察

課題1について

課題1は得点が服用量の1次関数となるように設定されたものであった。すべての班が“薬が多いほど効果大”との正しい結論に達した。

課題2について

この課題は、南風原（1994）の発言にヒントを得て設定したものであった。南風原（1994）の提起は“食塩の摂取量が増えると高血圧になる体質の人は20%、食塩を取りすぎても血圧に影響のない人は80%である”というような発見は分散分析を主体とした伝統的な統計検定法からは生まれてこないというような内容であったが、ここでは心理学に関係をもたせるために“記憶に影響を及ぼす薬”とし、また2つの体質の比率は少数者のデータがより反映するように2：8から3：7にと変更したものであった。

この課題は、“体質により薬が効く人と効かない人がいる”という設定にはなっていたが、基礎的な統計解析の知識でこれを見破るには、①分散の大きさの変化を把握すること、②個々のデータをグラフにあらわして散布状況を把握すること、が必要条件となるだろう。しかし実習班のうちA班とB班は散布状況をグラフに表したにもかかわらず、“データのばらつきが大きいと、薬を飲む量が多いほど記憶力がよくなる”（B班）という結論にとどまった。また、これらの班を含めて、最終的に“薬の量が多いときには効果あり”とした班が3班、逆に“効果なし”とした班が3班とまちまちな結論に至った。南風原（1994）の指摘どおり、少なくとも入門段階では、このタイプの現象は把握しにくいことが裏付けられたと言ってよいだろう。

課題3について

課題3は、“この薬はほどよく飲めばよく効くが、飲み過ぎるとほとんど効果がなくなる”という日常場面を模したものであった。実験シミュレーションでは、すべての班がほぼ妥当な結論を導き出した。

課題4について

課題4は、課題1と関連づけて、天井効果の可能性を気づかせることであったが、実際には“たくさん飲むほど効果が上がるが、量が増えるに従い効き目の増え方は限界に近づく”という関連づけなしの結論にとどまった。また“効果が同じなら薬は少ないほうがよい”という“節約の原理”がはたらき、90mg～120mg程度の服用量で十分であろうとの判断がなされた。

課題2と同様に本課題では散布度の違いに注目することが重要なポイントであった。単に薬の効き目が限界に近づくのであれば分散は減少しないはずである。150mg服用条件における分散の減少は、むしろテスト側の難易度にかかわる天井効果と推定すべきものであろう。しかし、実際にはそのような指摘をした班はなかった。

全体的特徴について

まず第1に、いずれの班も、平均の違いにばかり着目し、散布度の大きさの違いにはほとんど目を向けなかった。確かに、“薬の量と記憶テスト得点の平均値とのあいだにどのような関係があるか”

ということは主要な関心事ではある。しかし、“薬の量は得点のバラツキにどのような影響を及ぼすか”という効果も見逃すわけにはいかない。たとえば（記憶の薬というより情動に影響を与える薬の話になるが）、好不調の激しい人にある薬を与えると作業のムラがなくなったとしよう。その薬は、作業量の平均値を増やす効果はなくても、バラツキを減らす効果があるという点で重宝されるであろう。あるいは（薬とは関係ないが）、ある学校の2つのクラスの平均点は同じであるが分散が異なっていたとしよう。教師の側から見れば、分散の小さいクラスのほうが教えやすいことは言うまでもない。石田（1990, p.19）は、“現代心理学の研究のほとんどは、そのデータの分析における検討の対象として代表値に注目し、散布度は、その時の補助的な統計量としてのみ利用されている傾向が強い。”と指摘した上で、“…散布度自体も心理学的に有意義である場合が多々あるので、このことをもう少し強く認識して分析を行なうべきであると考えられる”と指摘しているが、今後の入門教育の中でも、散布度に注意を向けさせる指導を強めていく必要があると思う。

統計解析では、分散の等質性を前提とする検定法も多い。たとえば、（対応がない場合の）t検定は分散の等質性を前提としており、母分散が等しいと仮定できないときにはウェルチの検定などを用いる必要がある（分散分析における分散の等質性の問題については、小牧（1995, p.111-115）を参照されたい）。しかし実際に、パソコンソフトを用いてt検定をする場合には、あらかじめ等分散の検定を行なった上で、通常のt検定を用いるかウェルチの検定を用いるかを自動的に振り分けてしまうことが多い（例えば、脇本・垂水・田中, 1984）。そのため、ユーザーは、検定の途中過程に注意を払わず、まして分散が等しくない場合にその理由まで考えようとせず、もっぱら最終結果の有意差の有無のみに関心を示す。こうした統計ソフトの安易な利用も、散布度に目をむけない一因になっているかもしれない。

第2に、いずれの班も、薬の量に関して比較的少数の条件を設定し、分散分析を主体に検討するという方法をとった。この方法自体は決して誤りではないが、図1bや図2bで示したように、特にグループ分けせず、個体ごとになるべく多様な摂取量を割り当て、相関係数と個々のデータの散布状況から薬の効果を検討するという方法も考えられる。少なくとも、“個体間比較は即、分散分析”という画一的な立案は改める必要があるだろう。

第3に、今回の課題は“薬の効果を検討する”という内容であったため、私たちが日常生活のなかで薬の効果についていただいている“素朴な仮説”が影響を与えた点も無視できない。その代表的なものは、B班が想定した5つのパターンであった。こうした想定は、スムーズに実験を行なうメリットがある反面、想定外のパターンに対しては、誤った結論や思いこみによる解釈をもたらしやすいように思う。そうした危険性があるにもかかわらず今回結論上の誤りがごく少数であったのは、課題2を除いて関数関係の設定が比較的単純であったためであると考えられる。

おわりに

以上、実験計画の立案行動の特徴について検討を加えたが、そもそも何の前提条件ももたない最適解というものには実験計画には存在しない、という点を確認しておきたい。その第1の理由は、タイプ1のエラーとタイプ2のエラーの重みづけが検討対象によって異なるからである。例えば、救命器具の安全性を検査するような状況では、安全な器具を欠陥品と判定する誤りは許容されても、

欠陥品を安全であると判定する誤りは人命にかかわることなので極力避けなければならない。これに対して、装飾品に欠陥品があった場合には、後日、代替品と取り替えれば落着く。つまり欠陥品を完成品であると判定する誤りをおかしてもそれほど深刻にはならない。今回の場合も、“効力のない薬を有効と判定する誤り”と“効力のある薬を無効と判定する誤り”のどちらをどの程度重要視するかについて何らかの前提条件がなければ“最適”計画を立案するわけにはいかないのである。

薬の効果についてどういう情報を必要とするかによっても最適な計画の内容は変わってくる。服用量と得点との関数関係を知りたいということであれば、服用量の条件はなるべく細かく変化させなければならない。いっぽう、単に薬が有効かどうかだけに関心があるのであれば、条件はなるべく少数に設定して、同一条件内の個体数をなるべく増やすほうが望ましいであろう。今回の実験シミュレーションでは、求める情報内容については特に教示を行なわなかった。そのため、有効性の確認のみを重視した班もあれば、薬の最適量を探ろうとした班もあった。

今回、実験計画を立案したのは文学部心理学専攻の2回生であり、すでに学んだ統計解析法は、t検定、分散分析、多重比較、ピアソン相関係数など、ごく初歩的なものに限られていた。実験計画は、どのような統計解析を行なうのかを予定して立案されるのが一般的であろう。それゆえ知っている統計法のレベルに応じて制約を受けるという点も考慮しておく必要があるだろう。

いずれにせよ、“最適な実験計画はどのようなものか”を論じることは本稿の目的ではない。あくまで、心理学の入門指導の段階で、学生がどのような部分にとらわれやすいか、どのような部分を見落としやすいか、という視点から“実験計画立案行動”の特徴を把握することを目的とするものであった。この観点からみて、特に強調しておきたいのは、①散布度の違いに注意を払わない傾向がある、②もっぱら平均値で現象をとらえようとし個体ごとの生のデータを軽視する傾向がある、の2点である。今回は、個体間比較における実験シミュレーションを行なったが、個体内比較におけるシミュレーション場面で学生がどのような部分にとらわれやすいか、どのような部分を見落としやすいか、という点についても、今後検討したいと考えている。

引用文献

- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). Pergamon Books. [バーロー・ハーセン(著) 高木俊一郎・佐久間徹(監訳) (1993). 叢書・現代の心理学 別巻1 一事例の実験デザイン——ケーススタディの基本と応用——. 二瓶社]
- 南風原朝和 (1994). 一般化可能性の評価と有意性検定. [尾見康博(企画)・南博文(司会) 日本心理学会シンポジウム2 心理学の方法論見直し——新しい表現法の可能性をめぐる——. 日本心理学会第58回大会発表論文集 (S11)]
- 長谷川芳典 (1994). 心理学研究法再考: (1)基礎的統計解析の誤用をなくすための30のチェック項目. 岡山大学文学部紀要, 21, 47-59.
- 石田潤 (1990). 第1章: データを数値で表現する方法. [森俊昭・吉田邦夫(編). 心理学のためのデータ解析テクニカルブック (pp.1-42). 北大路書房.]
- 岩本隆茂・川俣甲子夫 (1990). シングル・ケース研究法. ——新しい実験計画法とその応用. 頸草書房.
- 小牧純爾 (1995). データ分析法要説. ナカニシヤ出版.
- 桑田繁 (1995). 新しい実験計画としての単一被験者法の紹介 (III). 全日本鍼灸学会雑誌, 44, 213-220.

橋敏明 (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.

脇本和昌 (1970). 初等情報処理講座 5 乱数の知識. 森北出版.

脇本和昌・垂水共之・田中豊 (1984). パソコン統計解析ハンドブック I 基礎統計編. 共立出版.