和光大学人間関係学部紀要　3号（1998）

実践的・探索的研究の効果測定における「効果偏差値」の提案

伊藤武彦

偏差値についてはイメージを持っているが、統計的検定については敬して遠ざかる人が多い。実践的研究の効果の評価では、被験者を十分多く集めることができない場合があり、効果量の提示の方が統計的有意差検定よりも重要である。本論文では、まず、実践的研究と探索的研究において、効果量が有効な統計値であることを述べる。次に、偏差値について説明をおこなう。以上をふまえ、効果量を偏差値で表した、効果偏差値の概念を提案し、それが実践的・探索的研究の評価とその普及に有効であることを述べる。

1　実践的・探索的研究と統計

心理学の研究においては、仮説検定のために統計学的分析をおこなうことが通例となっている。統計学の数学的基礎を充分に理解した上で、統計的な分析をおこなうことが望ましいことはいうまでもない。しかし、心理学を専攻している学生や研究者は、統計的原理を十分に理解せずに統計処理や検定を行っている場合が少なくない。

その一方で、心理学研究における統計処理は、従来、実験的方法による実験計画法に基づく厳密な条件統制下での比較的少数の標本による t 検定や分散分析などによる検定と、多人数からの標本抽出法に基づく心理測定的な方法による因子分析などの多変量解析を利用した研究が行われてきた。近年は、共分散構造分析（構造方程式モデル）の手法が、潜在変数を扱えること、モデルの適合性が検討できること、モデル構成が自由であることの利点があり、着目されている（伊藤、1997）。しかし、共分散構造分析は被験者数が少なくとも100名はいないと実用にならない。

心理学研究の方法として、現場に密着して行うタイプの実践的研究の結果を数量化して統計的に検討するための大きな問題は、第1に、被験者数が限られていることである。被験者数が少ないと、検定力powerが小さくなり、効果が実感として認められても、統計的検定を行うと統計的に有意な差が出ない。そうすると本来効果があるにも関わらず「効果がなかった」と誤って結論づけてしまうことになる。その成果が、論文の形で日の目をみずに、せっかくの努力が報われないという残念な結果になってしまう。現場や状況によっては、簡単に被験者を増やすことができない場合が多いのである。このような研究の不成功は、統計的には、被験者が少ないことが、検定力を小さくし、実際は帰無仮説が棄却されるべきである（＝差がある）にもかかわらず、誤ってそれを採択してしまう（＝差がないと誤判断する）という「第2種の誤り」を犯してしまうことにあたる。

現場研究での統計利用上の第2の大きな問題は、統計量や検定の結果の解釈において、従来の推測統計に基づく有意差検定の論理がわかりにくく、他専攻の専門家や非専門家に説明することも難しい問題である。統計的検定の問題をとってみても、

いとう　たけひこ　本学部助教授　（1998年12月22日　提出）

Everitt & Hay（1992）の指摘のように、専門的訓練を受けた心理学者でも誤解をしている人が多いという問題がある。ましてや非専門家に説明することは大変難しい。

本研究では、地域や現場に対して、ある種の介入を行いその効果を数量的に評価することを実践的研究の効果測定とよぶ。それを、統計的な解釈の正しさを保ちながら、しかもわかりやすい形で評価し、さらに現場の他の人々と共有するために、「効果量」と「偏差値」をキーワードとして、「効果偏差値」という概念を提案しその可能性と有効性を検討する。

2　効果量effect size

効果量は被験者数に左右されずして効果の大きさを示す統計量である。ただし、日本ではその概念自体もあまり知られていない。効果量を用いて諸研究の個々の結果から全体的な傾向や治療・介入の効果を検討することをメタ分析（Smith, & Glass, 1977）とよんでいる。日本ではメタ分析の論文は、南風原（1997）や根建ほか（1995）などわずかしかない。効果量がどの程度あるかを示せば、探索的な研究としては十分であると考える。効果量は、治療群と対照群あるいは事前テストと事後テストの各々の平均値と標準偏差があれば計算できる。筆者は、効果の検討をおこなう統計量として探索的研究で行う場合は効果量を用い、確認的研究では、効果量と併せて検定をおこなうことを提案する。被験者の数を獲得することの困難な実践的効果研究は（統計的見地からすれば）、探索的研究として位置づけることにより、検定力不足からくる第2のエラー、すなわち、効果があるのに有意差が出ず帰無仮説を採択してしまうという誤りからのがれるために、検定結果を保留し、効果量で治療・実践・介入を評価することが可能となる。

では、効果量にはどのようなものがあるのだろうか。日本心理学会（1991）では投稿論文に、特に効果量の記述を指定していない。しかし、アメリカ心理学会の論文投稿マニュアル（APA, 1994）には、有意差検定での p 値では標本数（被験者数）などに左右されて効果の重要性（大きさ）や関係の強さそのものがわからないとして、標本数に依存しない測度を用いることを勧めている（p.18）。そこで紹介されている効果量、関連の強さを表す指標は、r^2（決定係数）、η^2（イータ自乗）、ω^2（オメガ自乗）、R^2、ϕ^2（ファイ自乗）、Cramér（クラメール）のV、Kendall（ケンドール）のW、Cohen（コーエン）のd、Cohen（コーエン）のκ（カッパ）、Goodman & Kruskalのλ（ラムダ）、Goodman & Kruskalのγ（ガンマ）、Jacobson & Truaxのclinical significance（臨床有意性）、RoyのΘ（シータ）、Pillai-BartletのV、の14の統計量である[注1]。

これらのうちで本論文で最も重要なのがCohenのdという指標である。これについては、まず、偏差値を説明してから検討したい。

3　偏差値

偏差値という統計量は日本固有のものであるといわれている。かなりの大学生が偏差値についての具体的なイメージをもっている。功罪は別として日本の大学入試のシステムが偏差値をもとにして成り立っているからである。

あるテストの偏差値は(1)式のように計算できる。

(1)の式のうち、分数で表現されている部分を標準得点または標準化得点またはz得点という。標準得点とは各人の点数を変換して、全体の平均値が0、標準偏差が1になるように加工した値である(2)。偏差値というのは、標準得点を、平均値を50にして、標準偏差を10倍して加工した値だといえる(3)。偏差値と標準偏差と全体の平均値が分かっていると、もとの得点（素点）が復元できる(4)(5)。

$$\text{本人の偏差値} = \frac{\text{本人の得点} - \text{全体の平均}}{\text{標準偏差}} \times 10 + 50 \qquad (1)$$

$$\text{標準得点 } z = \frac{\text{本人の得点}-\text{全体の平均値}}{\text{全体の標準偏差}} \quad (2)$$

$$\text{偏差値} = (\text{標準得点}) \times 10 + 50 \quad (3)$$

$$\text{本人の素点} = \frac{\text{本人の偏差値}-50}{10} \times \text{全体の標準偏差} + \text{全体の平均値} \quad (4)$$

$$= \text{本人の標準得点} \times \text{全体の標準偏差} + \text{全体の平均値} \quad (5)$$

池田（1993）にしたがって偏差値という統計量について説明しよう。

「偏差値は難しさが異なるテストを共通の尺度の上で比較するのに便利である。それはいつも、平均値が50、そして得点の広がりを示す標準偏差が10になるように調整されているからである。ただし、その場合、受験者は同一、またはそれと同等の人たちでなければならない。

素点のままでは同じ国語の60点と英語の60点が同一の学力水準を表すとは考えにくい。しかし、偏差値に換算すると両者の分布はほぼ重なり、偏差値60はどちらのテストでもほぼ同じ位置づけを表すことが分かる。その結果、偏差値による学力表示は次のような利点をもつことができる。

①受験者が同一であれば（もしくは同等であるとみられる十分な根拠があれば）、異なったテストの間の比較ができる（素点のままでは、難易度やばらつきが異なるのでできない）。

②もし、２つの異なるテストの間で、得点の分布が同一であれば、偏差値の同じものは順位も同じになる。

③さらに得点の分布が正規分布であれば、偏差値を知るだけで、大体の順位を知ることができる。

④受験者集団の規模あるいは範囲が広ければ広いほど、共通尺度としての偏差値の適用範囲が広がる。

すなわち、クラス単位の偏差値があればクラスの中だけ、学校単位の偏差値であれば学校の中だけ、さらに全国的な偏差値であれば全国規模で比較可能である。

これらのことが、広範囲の受験者を対象とする偏差値が受験界で広く用いられるいちばん大きな理由である。しかし、そのこととテスト自身の品質がよいテスト（すなわち、信頼性や妥当性のあるテスト）かどうかとは別であることに注意しなければならない。信頼性や妥当性の低いテストでもその得点を偏差値に換算することはできるし、偏差値に換算したからといって、そのテスト得点の信頼性や妥当性が高まるわけではない」（池田, 1993, pp.56-57）。

偏差値は大学などの入学試験の入試難易ランキング表、すなわち、入試の合格レベルの予想にも使われる。たとえば、ある予備校（代々木ゼミナール,1998, p.334）では私立大学難易ランキング表において、各大学、各学部を偏差値においてランクづけている。数値は予備校の主催する公開模試の入試科目偏差値の平均であり、入試科目の少ない大学は高めに出る傾向がある。ランクの偏差値に達した場合の合格可能性は概ね55％であるが、高競争率の大学では50％前後になることもあるとしている。

もちろん、偏差値教育が受験中心の教育の代名詞的意味を持ち、偏差値が日本の教育をゆがめてきたとする批判がある。例えば、村越（1988）は次のように述べている。「偏差値は今日、教育の世界では学力測定の際の数値として利用され、しばしばその功罪が議論されている。ある一定の学力テストを行い、１人１人の生徒の成績の順位でみる偏差値は、得点（素点）の多少にかかわらず、そ

の生徒集団の中での学力の相対的位置を確定するうえできわめて便利なものである。しかし、それが現実の教育のなかで、とくに生徒の進学校の決定の手段に使われるようになると問題化してくる。1970年代、偏差値が教育界に入り込んだのは、中学校の教師が客観的で科学的な資料をもとにして生徒の進学を決定するようになってからであるといわれている。中学校在籍中に何回か業者による学力テストを実施し、その大量のデータをもとに生徒の学力の相対的な位置を判断すれば、それまで経験と勘に頼っていた進路指導をより客観的におこなうことができる。しかし、これは同時に高校の序列化の傾向をもたらした。」

この批判はもっともである。しかし、本研究ではいわゆる偏差値教育の問題は取り扱わない。偏差値それ自体は、教育問題と独立した統計量として扱いたいからである。豊田（1998）は、通学時間の偏差値の例を出し、偏差値が高いから「良い」と必ずしもいえないことを、わかりやすく説明している。

4　効果量と効果偏差値

Smith & Glass（1977）は、Psychological AbstractsおよびDissertation Abstractsの２つの抄録雑誌を中心に、精神療法の効果研究を1,000件収集し、そのうち、治療群と対照群の比較、または異なる治療を受けた群の間の比較をおこなっている357件の研究をメタ分析の対象とした。

これらの研究では、群間の比較の結果がさまざまな形で報告されているが、Smith & Glass（1997）は研究間の比較と統合のために共通の指標を考え、個々の研究で報告された比較結果をすべてこの共通の指標の値に変換した。この指標は効果量(Effect Size:以下*ES*とする）とよばれるもので、適応検査の得点など治療の効果測定に用いられた変数における治療群と対照群の平均値の差を対照群の標準偏差で割ったもの、つまり(6)のように、

$$\text{効果量}ES = \frac{\text{治療群の平均}-\text{対照群の平均}}{\text{対照群の標準偏差}} \qquad (6)$$

となる（ただし、不安検査の得点など、治療群の平均のほうが低くなることが期待される変数の場合は、*ES*の値の正負の符号を反転させる）（南風原, 1997）。

これは(2)(3)の式とよく似ている。標準得点と偏差値は、個人の得点だけでなく、あるグループについても、表現することができる。あるグループの平均値がわかれば、それをそのグループの点数として代表することにして、(2)の式の表現を変えて、(7)と考えてみよう。そうすると、全体に対するそのグループの平均点に対応する偏差値が(1)の式と同様に、(8)のように計算できる。

$$\text{あるグループの標準得点 } z = \frac{\text{あるグループの平均値}-\text{全体の平均値}}{\text{全体の標準偏差}} \qquad (7)$$

$$\text{あるグループの偏差値} = \frac{\text{あるグループの平均値}-\text{全体の平均値}}{\text{全体の標準偏差}} \times 10 + 50 \qquad (8)$$

偏差値の数式と効果量の数式とを比較し、対照群あるいは母集団を基準にして治療群の効果量を偏差値に換算すると、表1のような対応が得られる。ここで、効果量に対応する偏差値を効果偏差値とよぶことにする。効果偏差値と効果量との関係は次の式(9)で表される。

効果偏差値＝効果量×10＋50　　(9)

ここで、効果偏差値は効果量を10倍して50を足した値である。これは、次式の偏差値と標準化得点（z得点）との関係と同じである(10)。

偏差値＝ 10× z得点＋50　　(10)

このような関係を利用して、効果偏差値によって検定力分析を表現することもできる。検定力の分析とは、たとえば、平均値の差の検定において

表1　効果偏差値と効果量

効果量d	効果偏差値
0	50
.1	51
.2	52
.3	53
.4	54
:	:
.9	59
1.0	60
1.1	61
:	:
2.0	70
:	:

2群間の母平均の差が0である（等しい）という帰無仮説が偽である、すなわち、2群間には差があるのに、差がないと誤って結論づけてしまう第2のエラー（β）を1から引いた値である(11)。

検定力Power＝1-β　　(11)

なお、2群間の母平均の差がないという帰無仮説が真であるのに、誤って有意差があると判断してしまう確率、すなわち帰無仮説を誤って棄却してしまうことを第1のエラーとよび、その確率を危険率αで表す。表2に両者の関係を示す。

検定力Powerは1からβを引いたものであり被験者数を予め設定することが可能な場合は、.80に設定することが慣習化されつつある。実際には群間差があるにもかかわらず"統計的に有意差なし"と誤って判断する確率（β）が小さければ小さいほど検定力は高くなる。ところでPowerは、効果量とサンプル数と危険率αと検定の種類（両側検定か片側検定か）によって規定される(12)。

検定力Power
=f（効果量，サンプル数，α，検定の種類）(12)

検定力は、1）効果量が大きいほど（母分散が小さいほど＋母平均の差が大きいほど)、2）サンプル数が多いほど、3）αを大きく設定するほど(例：危険率が1％水準よりも5％水準の方が)、4）両側検定よりも片側検定の方が、一般に高くなる。

この式のうち実験や治療の実施をおこなう前に、何人サンプル数が必要であるかは、予想される効果量とα水準と検定力によって算出することができる(13)。

表2　第1のエラーと第2のエラー

エラー	危険率	よく使われる値	説明
第1	α	.05（5%水準）	誤って帰無仮説を棄却する割合
第2	β	.20（20%水準）	誤って帰無仮説を採択する割合

必要なサンプル数=
f（期待される効果量，α 水準，検定力）　(13)

ここで α 水準= .05、検定力=.80と固定した場合には、(14)のようになる。

必要なサンプル数=f（期待される効果量）　(14)

期待される効果量を予め設定する（たとえばd=.5、すなわち効果偏差値=55）ことにより、検定力が十分な実験をおこなうためのサンプル数を割り出すことができる。そのために検定力分析の統計パッケージが開発され利用されている。

ところで、Cohen（1988）は行動科学における検出力分析のために効果量の３つの基準を推奨している。すなわち、小効果量、中効果量、大効果量の３つである。これはおおまかな基準であるが、効果量 ES、効果偏差値で表すと表３になる。

相関係数 r は初学者にとってイメージのつかみにくい統計量である。相関係数よりも、それを自乗した決定係数 r^2 を用いて２変量間の関係の強さをパーセンテージでイメージすることを推奨したい。すなわち、決定係数とは、２変量の間で、ある変量がもうひとつの変量を何パーセント予測できるかを表す、すなわち、分散（ばらつき）の割合（パーセント）をあらわす指標だからである。

同じように、効果量は数値としては初学者や部外者にはリアリティのない、実感を伴わないイメージのしにくい指標である。「この治療法の効果量は.6」であったと表現するよりも偏差値50を基準にして「治療の効果偏差値は56であった」とした方がイメージがつかみやすいのではないだろうか。

5　まとめ

実践的・探索的な研究では、被験者数が少ないという問題点があるばかりに、いわゆる有意差検定で、有意な結果がでない場合が多い。しかし、これまで述べてきた効果量effect sizeを導入することにより、被験者数が少ないことにより、検定力が低いままでも、実験結果についての一応の目安を得ることができる。

効果量の最も代表的なCohenの d を算出するために、平均値と標準偏差(SD)（または分散でもよい：分散の平方根が標準偏差である）を、統計的有意差のあるなしに関わらず論文中に明示すべきである。そうすれば式(6)により、効果量の値が算出される。被験者数の確保の難しい実践的研究や、探索的な目的で行われる被験者数の少ない研究では、むしろSDを平均値とともに論文に明記することが、t 検定などにより有意水準（p 値）を示すことよりも重要である。

しかし、Cohenの効果量 d の値の大きさを感覚的につかむことは難しい。統計に習熟していない者は標準得点の値の感覚はつかみにくいが、偏差値は（事の善し悪しとは別に）なじみがあり、直感的に理解できる統計量である。効果量に対して、「標準得点vs偏差値」と同等の関係を持つような効果偏差値を用いることを本論文で提案した。

標準得点：偏差値＝効果量：効果偏差値　(15)

という関係である。(15)をみると、標準得点と効果量は表現が簡潔であるけれども、偏差値と効果偏差値のほうが直感的に日本人の読者にとって理解しやすい。

被験者確保の難しい実践的研究や被験者数が少

表３　効果の大中小の３つの大きさレベル

効果の大きさ magnitude	効果量 $ES(d)$ effect size	効果偏差値 h	相関係数 r	決定係数 r^2
小	.2	52	.1	1%
中	.5	55	.3	9%
大	.8	58	.5	25%

ない探索的研究においては、群間差の大きさ・効果の大きさが偏差値に換算されることにより、その効果についての強さの感覚がアナロジカルにつかめると思われるのである。しかも、メタ分析的に、他の研究の効果量・効果偏差値と比較することができ、相対的な効果の強さを把握することができる。

また、卒業論文レベルの論文指導や統計教育においても、被験者の人数に関わらず、効果偏差値を算出することは学生の直感的理解をうながすことができると思う。効果偏差値を示すことにより、卒論を書いている本人が統計的習熟の度合いが低くても、データから、リアリティのある群間差や効果の感覚がつかめるのではないだろうか。

【注】

（注１）

表Ａにこれらの指標のおおまかな分類と解説を載せておく。このうち、（５）のCohen(コーエン)のdが通常「効果量effect size」とよばれている最も代表的な統計量である。

表Ａ　効果量・関連の強さの指標と統計的検定との関係　(APA, 1994, p18を伊藤が分類・解説)

指標、対応する統計量、　関連する検定、文献　　等
（１）２変量の単相関（量的変数×量的変数）の大きさ　（分散の割合） r^2（決定係数：分散説明率）は相関係数rを自乗したもの ある変数でもう一つの変数を予測するときに説明できる分散の割合を表す 関連の度合いを0～100%で表せる。$0 \leqq r^2 \leqq 1$
（２）クロス表での２変数の関連の強さ（質的変数×質的変数） ϕ^2　（ファイ係数の自乗）2×2のクロス表：ϕ係数（四分点相関係数）を自乗したもの $\phi^2 = \chi^2$／被験者数　$0 \leqq \phi^2 \leqq 1$ CraméerのV（クラメールの連関係数）　$k \times n$のクロス表　$0 \leqq V \leqq 1$　森・吉田（1990, pp.240-241.） KendallのW（ケンドールの一致係数）　$k \times n$のクロス表（n≧8の時、$\chi^2 = k(n-1)W$）　$0 \leqq W \leqq 1$　武藤（1995, pp.466-469.） Goodman & Kruskalのλ（ラムダ）（グッドマン・クラスカルの予測係数）　$0 \leqq \lambda \leqq 1$　森・吉田（1990, pp.241-242.） Goodman & Kruskalのγ（ガンマ）（グッドマン・クラスカルの順序連関係数） $-1 \leqq \gamma \leqq 1$森・吉田（1990, pp.243-246.） ※　2×2表の場合、$r = \phi$　$\lvert \phi \rvert = V$、なので、$r^2 = \phi^2 = V^2$（分散の割合を表す、　r^2に準じる）
（３）回帰分析の独立変数（量的変数）で目的変数を説明できる指標（分散の割合） R^2　決定係数　回帰分析で独立変数で従属変数の分散を説明できる割合, $0 \leqq R^2 \leqq 1$
（４）ある独立変数が説明する分散の割合の指標　　分散分析 ω^2　（オメガ自乗）　ω^2=要因の平方和（分散）/　全体の平方和（分散） Keppel(1991, pp.64-68.) η^2　（イータ自乗）　η^2 =要因の平方和（分散）/　全体の平方和（分散） ※　η^2もω^2 も、（注２）の表Bの分散の割合PVに該当する値である。 $\eta^2 = v1F/(v1F+v2)$の式で、イータ自乗は自由度$v1v2$とF値から計算できる。 また、t検定の場合は$\eta^2 = t^2/(t^2+v2)$の式で、イータ自乗は自由度$v2$とt値から計算できる。 η^2はr^2を一般化したものである。
（５）平均値の差の大きさを表現する指標 Cohen(コーエン)のd　　平均値　　t検定、分散分析　効果偏差値= d×10＋50

（注２）効果量と他の統計量の関係を表Ｂに示す。

表B　効果量と統計量との関係　（Murphy, K. R., & Myors, B. 1998　p. 29, p57. より作成）

統計	効果量dで表す	自由度 v1	自由度 v2
相関係数	$d = 2r / \sqrt{(1-r^2)}$		
平均値の差の t 検定	$d = 2t / \sqrt{v2}$	-	N-2

統計	分散の割合PVで表す	自由度 v1	自由度 v2
平均値の差の t 検定	$PV = t^2 / (t^2 + v2)$		
分散分析	$PV = v1 \times F / (v1 \times F + v2)$	$v1$	$v2$
効果量　d	$PV = d^2 / (d^2 + 4)$		

統計	F値で表す	自由度 v1	自由度 v2
平均値の差のt検定	$F(1, \mathrm{v2}) = t^2$	-	N-2
相関係数	$F(1, \mathrm{v2}) = (r^2 \times v2)/(1 - r^2)$	-	N-2
重回帰分析のR^2	$F(\mathrm{v1}, \mathrm{v2}) = R^2 \times v2 / (1 - \mathrm{R}^2) v1$	p	N-p-1
カイ自乗	$F(\mathrm{v1}, \mathrm{v2}) = \chi^2 / v1$	$v1$	∞
効果量ES（Cohenのd）	$F(1, \mathrm{v2}) = \mathrm{d}^2 \times \mathrm{v2} / 4$	-	N-2
効果量ES（繰り返しのある測度のd）	$F(1, \mathrm{v2}) = d^2 \times v2 / 4\sqrt{(1 - r_{ab})}$		N-2

【参考・引用文献】

American Psychological Association 1994 *Publication manual (4th ed.)* American Psychological Association. p.18.

Cohen, J. 1988 *Statistical power and analysis for the behavioral sciences. (2nd ed.)* Hillsdale, N.J. : Lawrence Erlbaum Associates, Inc.

Everitt, B. S. & Hay, D. F. 1992 *Talking about statistics : A psychologist's guide to data analysis.* London : Edward Arnold

南風原朝和　1997　メタ分析による精神療法の効果研究の統合　精神療法, 23, 131-136.

池田　央　1993　心理測定法　放送大学教育振興会

伊藤武彦　1997　計算統計学の最近の動向：共分散構造分析の理論と応用　東西南北１９９７：和光大学総合文化研究所年報, 128-137.

森　敏昭・吉田寿夫 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

村越邦男　1988　偏差値　　世界大百科事典（第25巻）平凡社

Murphy, K. R., & Myors, B. 1998 *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests.* Mahwah, NJ: Lawrence Erlbaum Associates.

武藤真介　1995　統計解析ハンドブック　朝倉書店

根建金男・市井雅哉・関口由香, 他　1995　認知行動療法は効くか？：メタアナリシスと個人差要因の視点から　カウンセリング研究, 28, 87-103.

日本心理学会　1991　執筆・投稿の手引き　日本心理学会

Smith, M. L. & Glass, G. V. 1977 Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.

豊田秀樹　1998　調査法講義　朝倉書店

代々木ゼミナール　1998　'98年度大学入試・代ゼミデーターリサーチVol.3　JEC日本入試センター