メタ分析

2010/5/26
M1　古見文一

メタ分析 (meta-analysis)とは？

- 独立な研究結果の統計的な統合
- 既存研究の量的レビュー　cf.記述的レビュー (質的レビュー)
- 以前の研究結果を用いて理論的な問題を要約、統合、検証するという問題に対する一般的な概念アプローチを具体化したもの

Gene Glassによる3つの水準

- 一次的分析
  収集されたデータをそのままに、そのデータを収集した研究者が行なう
- 二次的分析
  データを収集した研究者以外の人間によるデータ分析、あるいはもとの研究者とは異なる意図や分析ストラテジーによるデータ分析
- メタ分析
  複数の独立な研究の結果の分析
  (メタ分析の統計的テクニックは分析単位が一次的水準の統計的仮説検定で構成されたデータベースに対して開発されてきたもの)

記述レビューVSメタ分析

- 共通点
- 最初の疑問から最後の結論まで解析者を動かす一般的な一連のステップ
- 問題形成、データ収集、データ評価、分析、解釈、公の発表（Cooper, １９８４）
- 明確な疑問の形成、概念的に焦点を絞ること（探索的対仮説検証的）、研究の選択、関連した研究特性への結果の一般化、結果（Light & Pillemer, １９８４）

記述レビューVSメタ分析

- メタ分析の利点とは
- 精度
  メタ分析による研究領域の要約と統合は検討中の現象の効果の大きさ（効果量）、有意性の正確な指標、さらにこれらの研究結果の変動の正確な指標を提供する
- 客観性
  メタ分析では結果の抽出、結果に重みづけをする事に対する規則や基準が明確である
- 再現可能性
  ある研究領域のメタ分析で得られた結論は、同じ研究を含めて同じ規則に従い、誰かが研究結果を統合するなら、同じになる

一次水準の仮説検定VSメタ分析

- 一次水準の研究とメタ分析の間の違い
- メタ分析の分析単位は研究結果である（一次水準仮説検定は被験者）
- メタ分析に含まれる統計的仮説検定は異なるサンプル分散をもつ可能性が高く、分散の等質性の仮定は満たされない
- 分析方法が研究間で異なっていた場合、異なる統計量をいくつかの共通の測定基準に変換するためのテクニックがメタ分析には存在する

3つの分析的問題

- 中心化傾向 (central tendency)
  典型的な反応を表現する方法 (平均値や中央値、最頻値など)
- 変動性 (variability)
  研究結果にはばらつき (変動) はどの程度あるのか？ばらつきは同一の研究母集団からのものかどうか (範囲や標準偏差)
- 予測性 (prediction)
  研究結果のばらつき、つまり変動は何によって生じるのか
  (関心対象の測度と他の意味ある測度との相関、異なる下位集団における反応者の測定平均の差)

メタ分析の手順

仮説検定の定義

↓

対象とする研究の抽出

↓

共通の測定基準への変換

↓

有意水準と効果サイズの結合

↓

有意水準と効果サイズの比較

仮説検定の定義

- 十分に定義された仮説検定は概念的な定義を超えて有用な操作的定義へと導く（ex.情報処理能力→WAIS得点）
- 例の場合、操作的に「情報処理能力」といった概念的に定義された知性からWAIS得点のように操作的に知性を定義しない限り十分に定義された一次仮説検定にとりかかることができない
- 注意すべき４つの側面

  ・従属変数

  ・独立変数

  ・サンプリングと手続き

  ・統計検定

仮説検定の定義

- 注意

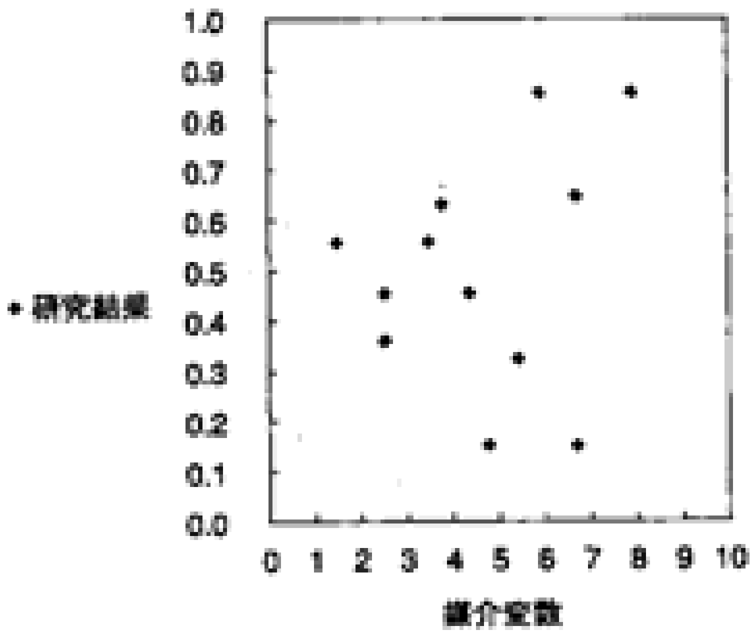


図1

この図はどう読み取れるでしょうか？

(出展：Bullen.B(小野寺訳) ,2000)

仮説検定の定義

- さっきのスライドの図はYに対するXの平均効果は約0.5で、研究結果と媒介変数の間に明らかな変動はない($r(10)=.190$)

だけど...

仮説検定の定義

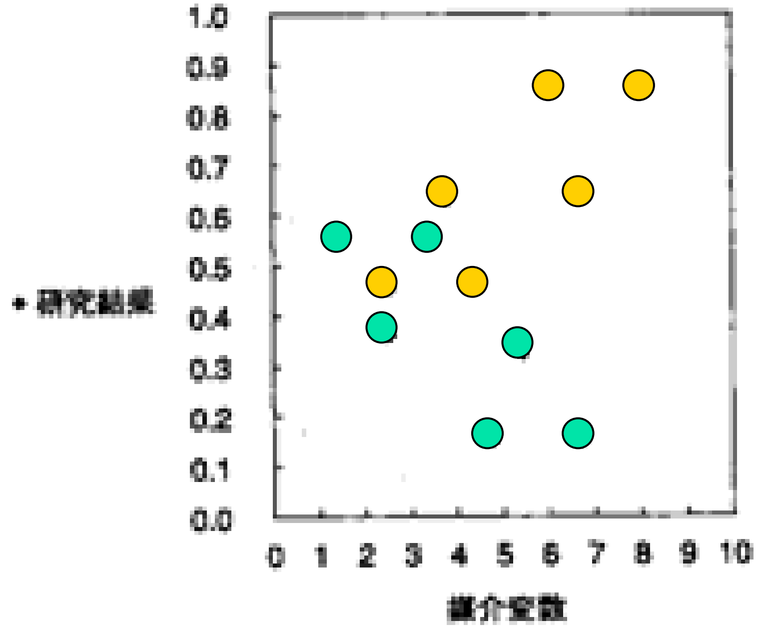


図2 (出展：Bullen.B(小野寺訳) ,2000)

実はこのように2つの明確に区別できる仮説検定セットが含まれているかも。もしそうなら、Yに対するXの平均効果は緑印だと0.35、黄印だと0.65、さらに黄印だと正の関連(*r*(4)=.717)だけど、緑印だと負の関連(*r*(4)=-.717)

仮説検定の定義

- 統計検定

・メタ分析的統合に含めても問題のない統計量

積率相関係数（ピアソンのr,スピアマンのρ、φ、点双列相関係数）、t検定、Z得点、片側p値、分子の自由度が1のF値、自由度1のカイ2乗値

・メタ分析的統合に含められない統計量

分子の自由度が2以上のF値、自由度2以上のカイ2乗値

基本的に2群の比較の統計量しか含められない

3群以上の場合、同じF値(カイ2乗値)でも異なる分析結果のパターンを表していることがある

対象とする研究の抽出

- 研究資料の特定
専門ジャーナルに印刷・掲載された論文、書籍、卒論、学位論文、地方学会や全国学会で発表された論文、公文書、公表されていない資料など
- 掲載資料は質が高い研究という仮定に基づいてそれだけに頼ると、バイアスが生じているかもしれない(Atkinson, Furlong, & Wampold, 1982他)

- 研究の所在をつきとめる

起源アプローチ…最近の論文に引用されている関連研究の所在をつきとめる

後続アプローチ…初期の論文が引用されている関連研究の所在をつきとめる

要約サービス…要約サービスでキーワードと関連した研究の所在をつきとめる

オンライン・コンピューター検索…要約サービスのオンラインバージョン

図書館のデータベース http://edb.kulib.kyoto-u.ac.jp/gakunaidb.html

見えない大学…同じ問題に取り組んでいる科学者のネットワーク

拾い読み…周辺的な関連学問分野の印刷物を通しての拾い読み

対象とする研究の抽出

- お蔵入り問題…有意でない結果は公表されていないのではないか？

- フェイル･セーフ数…メタ分析で得られた結論をひっくり返すのに必要な、お蔵入り問題の最少数

  ★お蔵入り問題の大きさを評価

共通の測定基準への変換

- さまざまな種類の推測統計量をいくつかの共通の測定基準に変換しなくてはならない

- 有意水準と効果サイズ：メタ分析の測定基準

共通の測定基準への変換

- 有意水準...実際に差がないとする帰無仮説が真のときに、観測された結果が得られる見込み
  有意水準に対する共通した測定基準はZ（標準正規偏差）、p値(確率値)
  Zは平均0、分散1をもち、特定のp値と直接に結びつく。

表１　推測統計量のZへの変換

t の場合:	$Z=(df(log(1+(t^2/df))))^{1/2}(1-(1/(2df)))^{1/2}$
$F(1,df)$ の場合:	$Z=(df(log(1+(F/df))))^{1/2}(1-(1/(2df)))^{1/2}$
$\chi^2(1)$ の場合:	$Z=\sqrt{\chi^2}$
r の場合:	$t=(r(N-2)^{1/2})/(1-r^2)^{1/2}$ それから
	$Z=(df(log(1+(t^2/df))))^{1/2}(1-(1/(2df)))^{1/2}$

（ここでの対数は自然対数（すなわち，log_{10} ではなくて log_e）であることに注意)

共通の測定基準への変換

- 効果サイズ(effect size)
  ある特定の原因から生じた効果の大きさ
  共通の測定基準は積率相関係数r、FisherのZ、平均の標準化された差d

表2　推測統計量のrへの変換

t の場合:	$r = [t^2/(t^2 + df)]^{1/2}$
$F(1, df)$ の場合:	$r = [F/(F + df)]^{1/2}$
$\chi^2(1)$ の場合:	$r = (\chi^2/N)^{1/2}$
Z の場合:	$r = (Z^2/N)^{1/2}$

rに対するFisherのZ変換.. $Z_{FISHER} = .5(log[(1 + r)/(1 - r)])$

dとは2つのグループの平均の差を、その測度のなんらかの標準偏差の指標で割ったもの

有意水準と効果サイズの結合

- 有意水準の結合
  各研究からの有意水準Zに重みづけ(w)
  (サンプル数が大きかったり方法論の質が高い研究がより貢献するように)
- 有意水準を結合するメタ分析的手続き (公式)

$$Z = \frac{\Sigma w_j Z_j}{\sqrt{\Sigma w_j{}^2}}$$

$$\left(\begin{array}{l} w_j = \text{仮説検定}\ j\ \text{の結果に割り当てられる重みづけ} \\ Z_j = \text{仮説検定}\ j\ \text{の有意水準に関する}\ Z \end{array} \right)$$

有意水準と効果サイズの結合

- 有意水準に対するフェイル･セーフ数の公式(Rosenthal.1979)

$$Nfs(p=.05)=\left(\frac{\Sigma Z_j}{1.645}\right)^2-k$$

$$\begin{pmatrix} Z_j = \text{仮説検定}\ j\ \text{の有意水準に関する}\ Z \\ k = \text{メタ分析に含まれる仮説検定の数} \end{pmatrix}$$

有意水準と効果サイズの結合

- 効果サイズの結合
  各研究からの積率相関係数rからフィッシャーのZを求め、それに重み付け
- 効果サイズの結合のためのメタ分析的テクニック（公式）

$$Z_{FISHER} = \frac{\Sigma w_j Z_{FISHER_j}}{\Sigma w_j}$$

$$\left(\begin{array}{l} w_j = \text{仮説検定 } j \text{ の結果に割り当てられる重みづけ} \\ Z_{FISHER_j} = \text{仮説検定 } j \text{ の効果サイズに関する } Z_{FISHER} \end{array} \right)$$

有意水準と効果サイズの結合

- 効果サイズに対するフェイル・セーフ数の公式(Orwin,1983)

$$Nfs(Z_{FISHER}=0.100)=\frac{k(\bar{Z}_{FISHER}-Z_{FISHER_c})}{Z_{FISHER_c}}$$

$$\left(\begin{array}{l} k=\text{メタ分析に含まれる仮説検定の数} \\ \bar{Z}_{FISHER}=k\text{個の仮説検定に対する平均}\ Z_{FISHER} \\ Z_{FISHER_c}=\text{特定の基準に関する}\ Z_{FISHER} \end{array}\right)$$

有意水準と効果サイズの比較

- ３つの分析的課題（スライド7）

(1)中心化傾向

「典型的な研究結果は何か」「平均効果は何か」

☆複数の有意水準と効果サイズの結合

有意水準と効果サイズの比較

- 拡散比較と焦点比較

拡散比較

研究結果の変動タイプに関してかなり拡散し、無指向性で不特定である。単に結合研究結果、あるいは平均研究結果の周りにかなりの量の散らばりがあるという程度を示すだけであり、なぜ研究結果がお互いに異なるのかに関して直接には何も明らかにしない。変動性という第二の一般的な分析上の疑問を指向する。

焦点比較

どれほど研究結果が異なっているかに対して、特定され、方向性を持ち、十分に定義された説明に集中している。予測と説明という第三の一般的な分析上の疑問を指向する。

有意水準と効果サイズの比較

- 変動に対する効果サイズの分解クラスター分析
  - 隣り合った順位のU（平均平方推定値）を検討
- 度数分布
  - 典型的な研究結果の周りの変動に関する重要な側面を検出する手助けになる

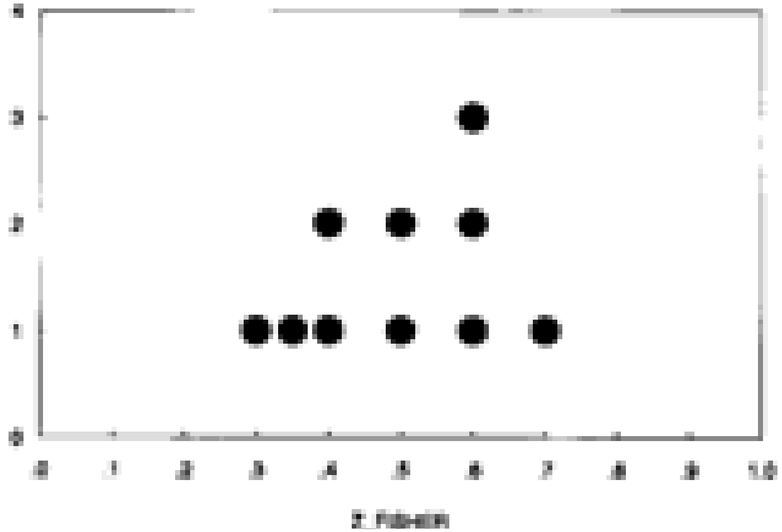


図3．度数分布の例(出展：Bullen.B(小野寺訳),2000)

有意水準と効果サイズの比較

- 典型的な反応は何か→典型的な反応の周りにはどれほど変動があるか→「なぜ？」
- 焦点比較
  変動性は特定の予測変数によってどの程度説明されるか？
- 直接コード化、評価者の評定、事後理論指標、記録文献・歴史文献資料から得られる研究結果の予測変数を用いて研究結果がどれほど異なるかを調べる

メタ分析についてのよくある質問

- お蔵入り問題(fire drawer problem)
  研究者はポジティブなデータを得られたときのみ発表する傾向があり、編集者は有意でないとリジェクトしたりする
  Rosenthalのフェイル・セーフ数も一種の方法ではあるが問題の解決ではない
- リンゴとオレンジを足しているのではないか？
  その通りであるが、果物という枠組みであれば一緒にできる。すなわちどの階層を分析の対象とするかが問題である
- Garbage In - Garbage Out
  リンゴとオレンジの特殊ケース
  ダメ研究がデータベースに含まれている結果、結局ダメ分析になってしまうかもしれない

一般的なルール

- 批判に耐えうるかたちを考慮しながら現象を解明すること
  仮説検定を選択する際の基準、予測変数の選択と導出、研究結果の適切な共通測定基準への変換、3つの一般的な分析的問題に答える
- どれも2つのやり方でやること
  補足的な分析として他のアプローチも続行し、もし結果に違いがあるなら慎重に注意を払う
- すべてを報告すること
  分析を追試したいと望む批判的な読み手に、彼らが知りたいと望むものすべてを提供する
- 注意深くあること

一次水準の研究者が自分の研究手続きに誤差が入り込むのを最小にしようと努力するのと同じように、メタ分析の解析者も自らの研究の厳密さ、そしてそれが意味を持ち、再現可能であることを保障することに常に関心を抱くべきである

メタ分析はどんな研究で使われている？

- 対人期待効果（Rosenthal,1968,1969;Rosenthal ＆Rubin, 1971, 1978)
- 行動研究における自発的参加者の特性(Rosenthal & Rosnow, 1975)
- 精神療法の効果性(Smith & Glass, 1977)
- 態度と成績に及ぼすクラス規模の効果(Glass & Smith, 1979)
- 欺瞞的コミュニケーションの特徴(DePaulo, Zuckerman, & Rosenthal, 1980)
- 大学進学前の数学教育における電卓の効果(Hembree & Dessart, 1986)
- 社会的影響における性差(Eagly & Carli, 1981)
- 人種差別廃止の効果(Israel, 1981; Miller & Davidson-Podgorny, 1987)
- 血清脂質とリポ蛋白に対する運動の異なる効果(Tran & Weltman, 1985)
  など

メタ認識には長い歴史がある
では、心理学ではどんな研究に使われているのか？

メタ分析はどんな研究で使われている？(2)

- Milligan, K., Astington,J.W.,& Dack,L.A. (2007). Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-belief Understanding. *Child Development, 78,* 622-646

この論文を紹介します

問題

- 言語と誤信念課題のパフォーマンスの高さの関係を調べた結果のデータは多量にある。
- しかしこの２つの関係を強いと述べている研究もあれば、弱いと述べている研究もある。
- だから、これらの不一致を統合し、これらの結果のばらつきをメタ分析のアプローチで調査し、説明する必要がある。

方法

- 104の研究についてメタ分析
- ピアソンの積率相関係数*r*によって言語能力と誤信念課題のパフォーマンスの相関を研究横断的に調べてみた

結果

- 誤信念課題と言語能力の間には相関が見られた（r=.19, p=.05）
- 言語能力の種類による効果サイズは .35≦M≦.66
- 誤信念課題の種類による効果サイズは .35≦M≦.52

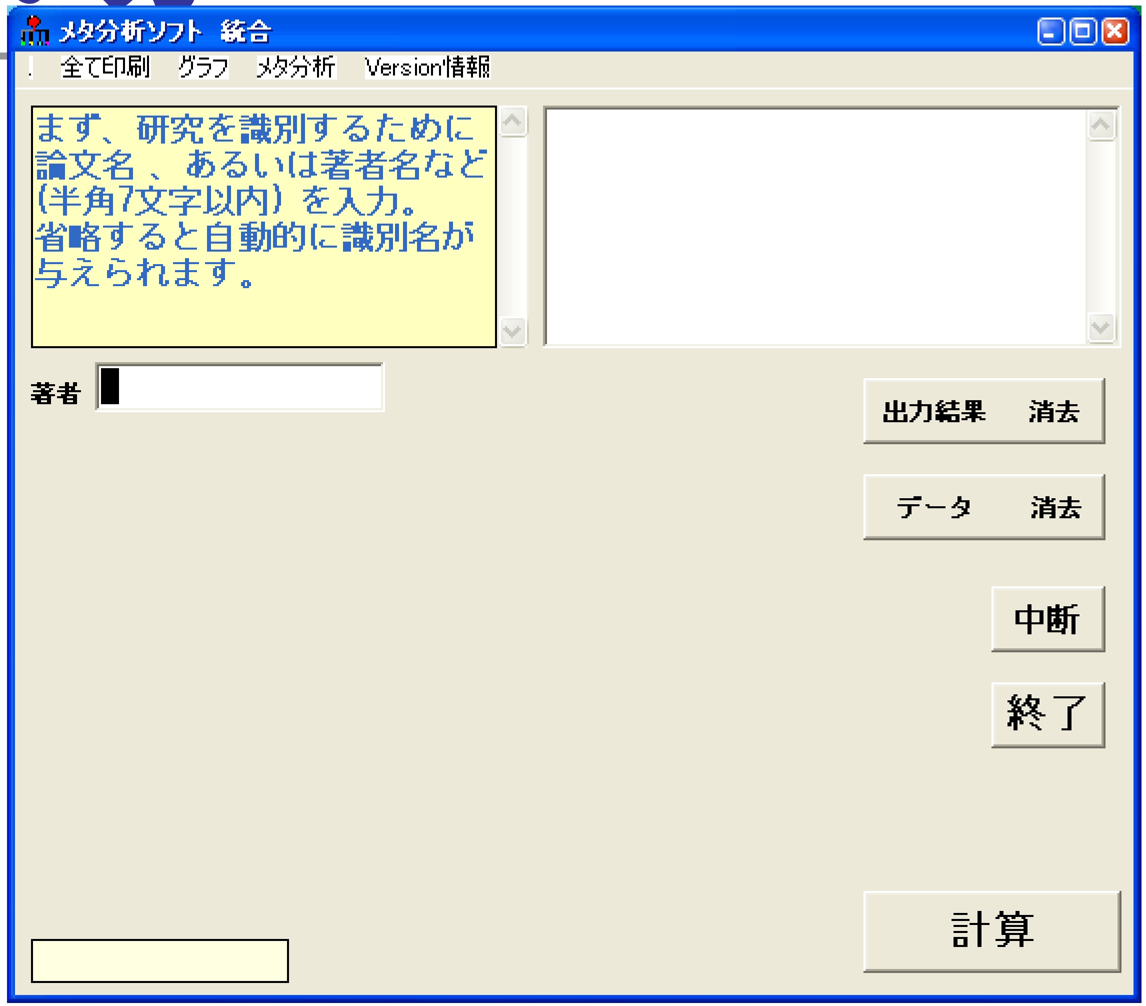
考察

- 誤信念課題と言語能力の間の相関は、年齢の影響によるということは昨今の研究でも示唆されていることだが、新たに分かったことがある
- 誤信念課題の成績を語彙は12%、意味は23%、文法は29%、補語は44%、一般的な言語は27%説明できた

やってみようメタ分析
メタ分析ソフト「統合」

統合の使い方

1. 「著者名」⇒「発行年」⇒「西暦」の入力
2. 「統計量」の入力（半角数字）
3. 「研究間で共通する変数強度」の入力
4. 「仮説支持方向（DOE）」（デフォルト＝支持方向）
5. 「計算」
6. （対象とする全論文について1～5の手順を繰り返す）
7. 「メタ分析」



やってみようメタ分析

サンプルデータ（Mullen, B., 小野寺訳(2000)を参考に作成）

著者	出版年	サンプルサイズ	統計量（df）	Xの強度	仮説支持方向
A	1981	110	$\chi^2(1)=23$	7	+
B	1984	100	$t(98)=6.5$	5	+
C	1985	90	$F(1, 88)=15.0$	8	+
D	1982	80	$r(78)=.335$	5	+
E	1983	80	$p=.0000001$	8	+
F	1988	70	$Z=3.891$	3	+
G	1990	60	$p=.01$	2	+
H	1986	65	$F(1, 63)=10.25$	4	+
I	1987	70	$r(68)=.535$	10	+
J	1989	65	$t(63)=6.0$	9	+

※「Xの強度」＝独立変数Xの操作の強度

結果の出力

メタ分析結果出力

以下の研究数 10 の研究を統合したメタ分析結果

2010年4月28日 13:29:56

分析対象となった研究データ（FZ = Fisher の Z）

No	著者	年	支持	統計	DF	N	Z	FZ	r	r^2	p	d	強度
1	A*******	1981	1	23	1	110	4.796	0.494	0.457	0.209	0	1.028	7
2	B*******	1984	1	6.5	98	100	5.912	0.617	0.549	0.301	0	1.313	5
3	C*******	1985	1	15	88	90	3.711	0.402	0.382	0.146	0	0.826	8
4	D*******	1982	1	0.335	78	80	3.037	0.348	0.335	0.112	0.001	0.711	5
5	E*******	1983	1	0	78	80	5	0.631	0.559	0.312	0	1.348	8
6	F*******	1988	1	3.891	68	70	3.891	0.504	0.465	0.216	0	1.051	3
7	G*******	1990	1	0.01	58	60	2.576	0.346	0.333	0.111	0.005	0.705	2
8	H*******	1986	1	10.25	63	65	3.069	0.393	0.374	0.14	0.001	0.807	4
9	I*******	1987	1	0.535	68	70	4.771	0.597	0.535	0.286	0	1.266	10
10	J*******	1989	1	6	63	65	5.315	0.698	0.603	0.364	0	1.512	9

印刷　クリップボードへコピー　終了

結果の出力

メタ分析結果出力

基本統計量

平均と標準偏差

	発行年	人数	Ｚ値	FisherのZ	相関	d	変数強度
平均	1985.5	79	4.2078	0.503	0.4592	1.0567	6.1
ＳＤ	2.8723	15.6205	1.0535	0.1214	0.0944	0.2759	2.5475

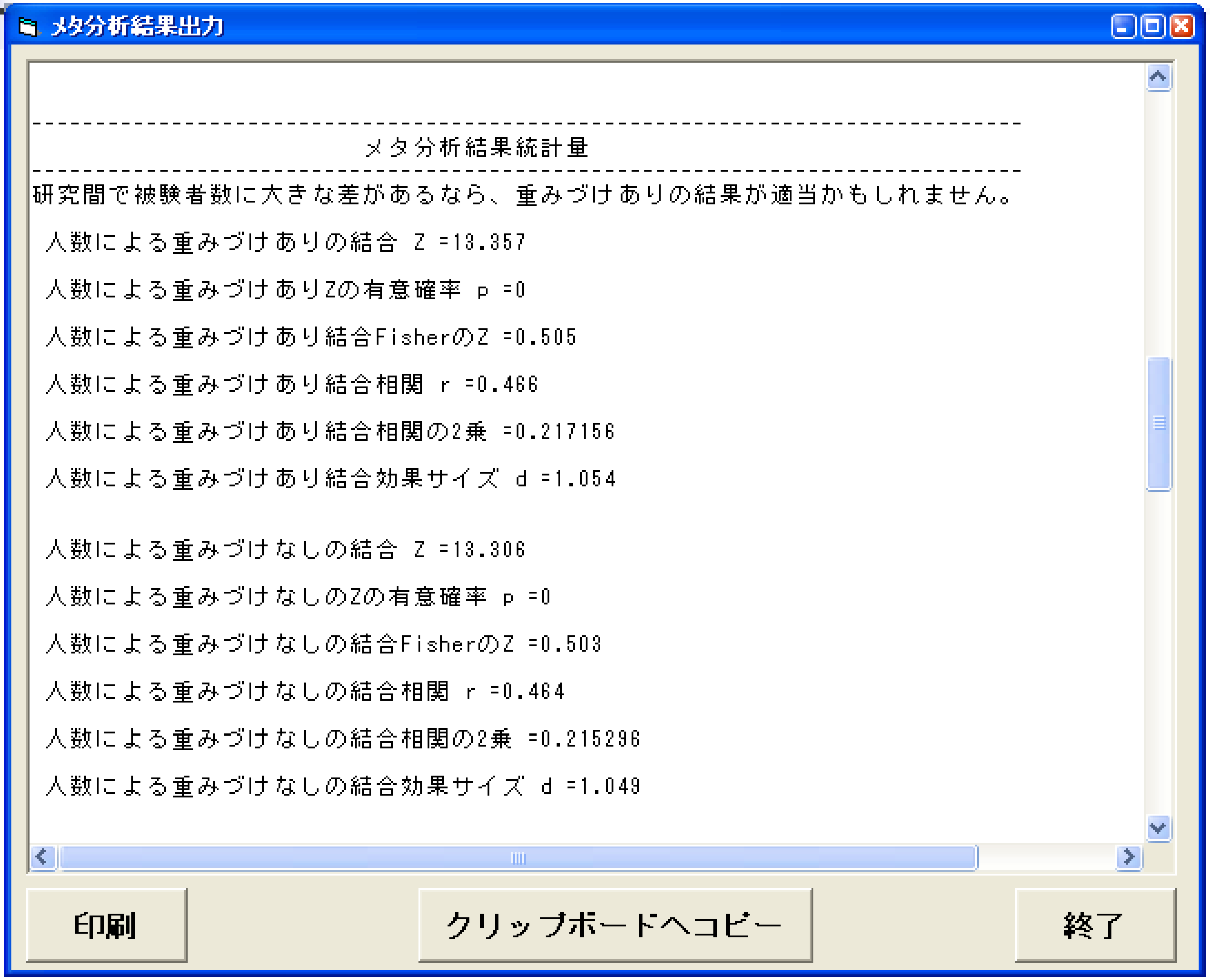
相関行列

相関係数

	発行年	人数	Z値	FisherのZ	相関	d	変数強度
発行年	1.0000	-0.8024	-0.2255	0.0594	0.0443	0.0666	-0.2118
人数	-0.8024	1.0000	0.4704	0.101	0.1212	0.0898	0.2161
Z値	-0.2255	0.4704	1.0000	0.9167	0.923	0.9124	0.5952
FisherのZ	0.0594	0.101	0.9167	1.0000	0.9992	0.9998	0.6119
相関	0.0443	0.1212	0.923	0.9992	1.0000	0.998	0.6114
Cohenのd	0.0666	0.0898	0.9124	0.9998	0.998	1.0000	0.6129
変数強度	-0.2118	0.2161	0.5952	0.6119	0.6114	0.6129	1.0000

印刷　クリップボードへコピー　終了

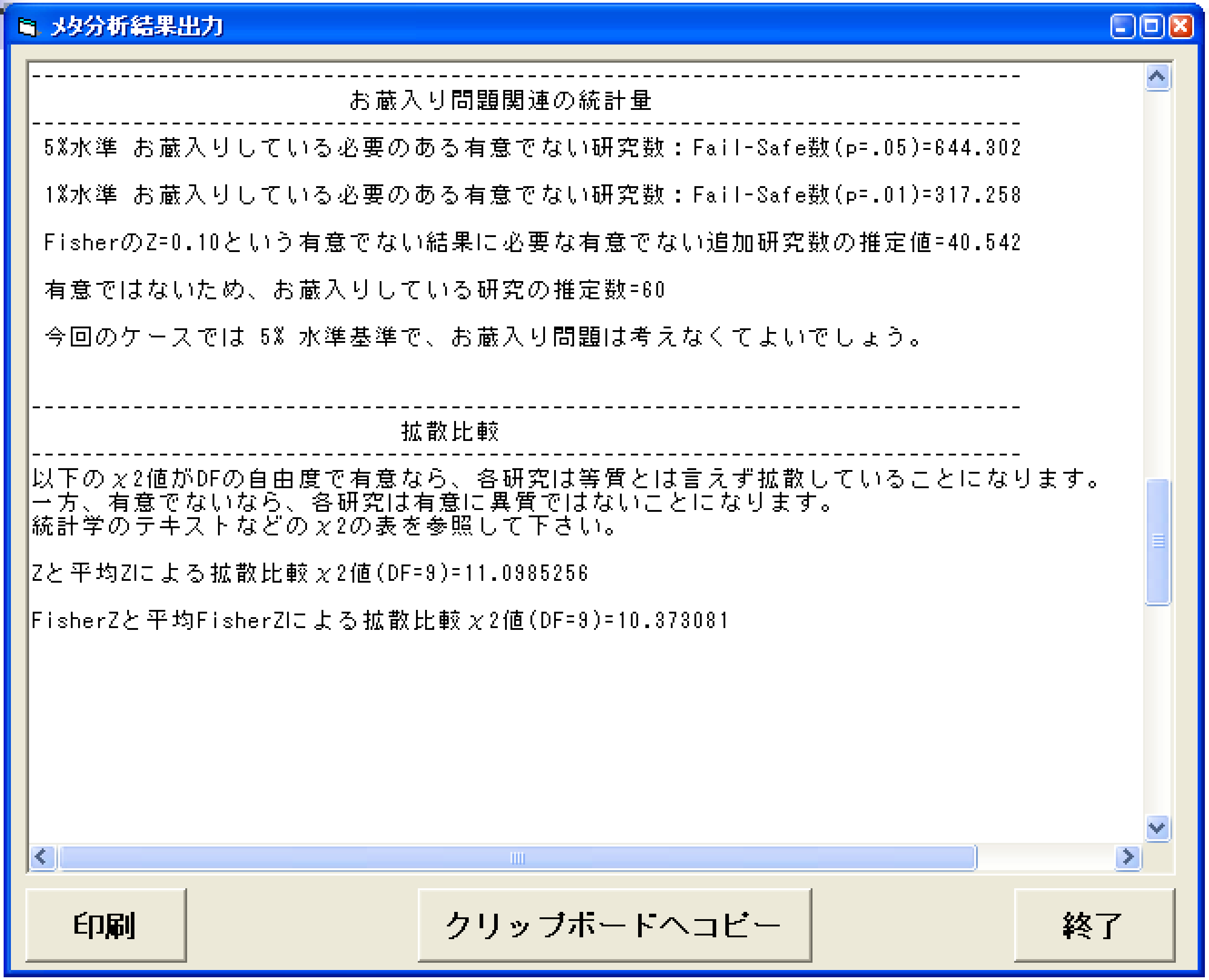
結果の出力


メタ分析結果出力

メタ分析結果統計量

研究間で被験者数に大きな差があるなら、重みづけありの結果が適当かもしれません。
人数による重みづけありの結合 Z =13.357
人数による重みづけありZの有意確率 p =0
人数による重みづけあり結合FisherのZ =0.505
人数による重みづけあり結合相関 r =0.466
人数による重みづけあり結合相関の2乗 =0.217156
人数による重みづけあり結合効果サイズ d =1.054
人数による重みづけなしの結合 Z =13.306
人数による重みづけなしのZの有意確率 p =0
人数による重みづけなしの結合FisherのZ =0.503
人数による重みづけなしの結合相関 r =0.464
人数による重みづけなしの結合相関の2乗 =0.215296
人数による重みづけなしの結合効果サイズ d =1.049
印刷
クリップボードへコピー
終了

結果の出力


メタ分析結果出力
お蔵入り問題関連の統計量
5%水準 お蔵入りしている必要のある有意でない研究数：Fail-Safe数(p=.05)=644.302
1%水準 お蔵入りしている必要のある有意でない研究数：Fail-Safe数(p=.01)=317.258
FisherのZ=0.10という有意でない結果に必要な有意でない追加研究数の推定値=40.542
有意ではないため、お蔵入りしている研究の推定数=60
今回のケースでは 5% 水準基準で、お蔵入り問題は考えなくてよいでしょう。
拡散比較
以下のχ2値がDFの自由度で有意なら、各研究は等質とは言えず拡散していることになります。
一方、有意でないなら、各研究は有意に異質ではないことになります。
統計学のテキストなどのχ2の表を参照して下さい。
Zと平均Zによる拡散比較χ2値(DF=9)=11.0985256
FisherZと平均FisherZによる拡散比較χ2値(DF=9)=10.373081
印刷
クリップボードへコピー
終了

結果の出力

メタ分析結果出力

分解クラスター

共通サンプル・サイズの平均平方推定値＝8.67

もとの研究番号とＵ値

研究番号(1)のＵ値=4.283
研究番号(2)のＵ値=5.349
研究番号(3)のＵ値=3.485
研究番号(4)のＵ値=3.017
研究番号(5)のＵ値=5.471
研究番号(6)のＵ値=4.370
研究番号(7)のＵ値=3.000
研究番号(8)のＵ値=3.407
研究番号(9)のＵ値=5.176
研究番号(10)のＵ値=6.052

印刷　クリップボー

メタ分析結果出力

ソート後のＵと隣り合うＵの差

Ｕ値=3.000
├ .017
Ｕ値=3.017
├ .390
Ｕ値=3.407
├ .078
Ｕ値=3.485
├ .798
Ｕ値=4.283
├ .087
Ｕ値=4.370
├ .806
Ｕ値=5.176
├ .173
Ｕ値=5.349
├ .121
Ｕ値=5.471
├ .581
Ｕ値=6.052

差が研究数(k=10)でみて、Gapより大きければ、そこがクラスターの境界です(p=.05)。

k	3	4	5	6	7	8	9	10	11	12	13	14	15	20	50
Gap	2.48	2.28	2.13	2.02	1.93	1.85	1.81	1.76	1.72	1.69	1.66	1.64	1.61	1.46	1.14

Hedges, L. V. & Olkin, I. (1983). Clustering Estimates of Effect Magnitude From Indepen
Psychological Bulletin, 1983, Vol. 93, No.3, 563-573. p.565 Table 1より数値を引用

印刷　クリップボードへコピー　終了

資料・Web

- Milligan, K. Astington,J.W.,&Dack, L.A. (2007). Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-belief Understanding. *Child Development, 78,* 622-646
- Mullen, B. 小野寺訳(2000). 基礎から学ぶメタ分析　ナカニシヤ出版
- 「統計学自習ノート・メタアナリシス」(青木繁伸@群馬大学) http://aoki2.si.gunma-u.ac.jp/lecture/meta-analysisu/index.html
- 「メタ分析」(唐牛祐輔)　2007年度心理データ解析演習発表資料 http://kyoumu.educ.kyoto-u.ac.jp/cogpsy/personal/Kusumi/datesem07/karouji.pdf
- 「メタ分析」(溝川藍)　2006年度心理データ解析演習発表資料 http://kyoumu.educ.kyoto-u.ac.jp/cogpsy/personal/Kusumi/datesem06/mizokawa.pdf