

メタ分析

M1 溝川



1



1. メタ分析とは

2

メタ分析 (meta-analysis)



- 独立な研究結果の統計的な統合
- 量的レビュー
- 各研究における効果を検討した統計量 (ex. F値、t値、カイ二乗値、平均と標準偏差など) を一定の効果サイズに変換し、その効果サイズによって各研究の効果の大きさを比較したり統合する

cf. 記述的レビュー (質的レビュー)

3

分析の3つの水準 (Gene Glassによる分類)



- **一次的分析 (primary analysis)**
ー 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、もともとの統計的データの解析
- **二次的分析 (secondary analysis)**
ー データを収集した研究者以外の人間によるデータ分析、あるいは元の研究者とは異なる意図や分析戦略によるデータ分析

- **メタ分析 (meta-analysis)**
ー 複数の独立な研究結果の分析

メタ分析の対象は、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

4

メタ分析VS記述的レビュー



- **類似点**
- 以前の研究の結果をその研究領域の啓発的な要約に統合していく一般的なステップ
- 問題形成、データ収集、データ評価、分析、解釈、公の発表 (Cooper, 1984)

- **メタ分析の利点**
- 精度: 効果の大きさや有意性等の正確な指標を提供
- 客観性: 結果の抽出・重みづけの規則・基準が明確
- 再現可能性: レビュアーによって結論が異なることはない

5

分析で扱う3つの問題



- **中心化傾向 (central tendency)**
ー 典型的な研究結果は何か？
- **変動性 (variability)**
ー 典型的な研究結果の周りにどれほど変動があるのか？
< 変動: 典型的な研究結果と他の研究結果のばらつき >
- **予測性 (prediction)**
ー 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？

6

メタ分析が適用された現象

- 対人期待効果 (Rosenthal, 1968, 1969; Rosenthal & Rubin, 1971, 1978)
- 精神療法の効果性 (Smith & Glass, 1977)
- コンピューター支援学習の効果 (Dusek & Josephs, 1983)
- 援助行動における性差 (Eagly & Crowley, 1986)
- 集団におけるブレイン・ストーミングの効果 (Mullen, Johnson, Salas, 1989)
- 心筋梗塞後の心臓リハビリ (Oldridge, Gvyett, Fisher, & Rimm, 1988)

等々...

7

心理学におけるメタ分析の例(1)

- ◆ 坂元 (1999)
- 抑うつ者および高不安者はポジティブ情報よりもネガティブ情報に対する潜在記憶が優れているか？
- 対象とした論文の数: 抑うつ(9)、不安(9)
- メタ分析を用いて検討
抑うつ者は、非抑うつ者に比べて、ネガティブ情報に対する潜在記憶が優れている
高不安者は、低不安者に比べて、ある条件においてのみ、ネガティブ情報に対する潜在記憶が優れている

8

心理学におけるメタ分析の例(2)

- 論文の構成
- 問題(スライド12)
- 研究の背景
- 7つの調整変数(独立変数)
a)被験者、b)処理水準、c)教示、d)刺激タイプ、
e)呈示時間、f)フィラー課題、g)潜在記憶課題

単に抑うつと不安におけるネガティブ情報に対する潜在記憶バイアスの大きさを調べるだけでなく、研究間で異なっている条件ごとにバイアスの大きさを調べる

9

心理学におけるメタ分析の例(3)

- 方法
- 文献収集
検索方法 (PsychLit, キーワード検索)、見つけた研究の数、対象とした研究の数、分析の対象の選定基準 (スライド12)
- コード化
調整変数(独立変数)の分類(スライド12 +)
- 分析方法
変換可能な統計量が記述されていない研究は除外、効果サイズの推定値にd値を採用、不偏平均効果サイズの算出 (d値の推定バイアスを修正)、フェイル・セーフ数の算出、カテゴリー内の研究の等質性の検定、下位カテゴリー間の差の検定(スライド12)

10

心理学におけるメタ分析の例(4)

- 結果(抜粋)
- 抑うつ者の潜在記憶バイアス
・ネガティブ情報に対する潜在記憶バイアス ($d=.50, p<.001$)
- 高不安者の潜在記憶バイアス
・ネガティブ情報に対する潜在記憶バイアスなし ($d=.14, n.s.$)
・変数ごとに見ると、顕在記憶に影響されない条件でのみ潜在記憶バイアス
- 表1・2 Tests of categorical models for effect size
・潜在記憶バイアスへの抑うつ効果(表1)と不安の効果(表2)
- 考察(省略)
- 文献
- メタ分析に用いられた論文には * 印

11

メタ分析の手順

- 仮説検定の定義
- 対象とする研究の抽出
- + 研究の特性の同定とコーディング
- 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

12

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. 「頭の良さ(概念的)」「WAIS得点(操作的)」
操作的に知能を定義しない限り十分に定義された一次仮説検定に取りかかることができない

- 注意すべき4つの側面
 - 1) 従属変数
 - 2) 独立変数
 - 3) サンプリングと手続き
 - 4) 統計検定(スライド15)

13

仮説検定の定義

- 操作的定義の詳細な内容への注意を怠ると...
 - 例: 12の研究セットのメタ分析の結果
 - Xが大きくなるとYも大きくなる?(図1)
実は、2つの異なる仮説検定セット(6+6)が含まれていた!(図2)

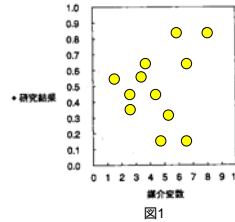


図1

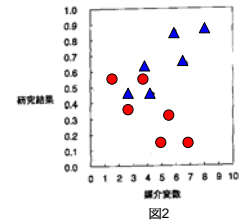


図2

(出典: Mullen, B. (小野寺訳) (2000))¹⁴

仮説検定の定義

- 4) 統計検定
 - メタ分析的統合に含めても問題のない統計量
 - 一積率相関係数(ピアソンのr、スピアマンのp、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、自由度が1のカイ二乗値
 - メタ分析統合に含められない統計量
 - 一自由度が2以上のF値、自由度が2以上のカイ二乗値
- 基本的に、2群間の比較の統計量以外は含められない
 ...3群以上の統計量では、同じF値(カイ二乗値)でも異なる分析結果のパターンを示していることがあるため

15

対象とする研究の抽出

- 研究資料の特定
 - 一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、学会で発表された論文、公表されていない資料etc...
 - 一掲載資料に限定することによるバイアス
- 研究の所在をつきとめる
 - 一起源アプローチ; 最近の論文に引用されている関連研究を探る
 - 一後続アプローチ; ある初期の研究を引用している文献を探る
 - 一要約サービス; 要約サービスでキーワードと関連した研究を探る
 - 一オンライン・コンピューター検索; 要約サービスのオンライン版

16

対象とする研究の抽出

- お蔵入り問題 (file drawer problem)
 - 一有意でない結果は発表されていないのではないか?
- 「フェイル・セーフ (Fail-Safe) 数」
 - 一メタ分析で得られた結論をひっくり返すためには、少なくともあといくつのお蔵入りしている研究が必要か?
 - 一お蔵入り問題の大きさを評価する手段

17

+ 研究の特性の同定とコーディング

- 必要な情報
 - 一サンプル数、要約統計量、発表年、発表媒体(ソース)、手法(測定装置や統計手法を含む)、標本のタイプ(調査対象の性別や年齢)など
- 媒介変数の探索の際に重要

18

共通の測定基準への変換

- 研究結果 (特定の仮説検定のために報告された統計量) をいかに統合、合併するか？

「有意水準」と「効果サイズ (effect size)」

: メタ分析の測定基準

19

共通の測定基準への変換

- 有意水準
 - 実際に差が無いとする帰無仮説が真の時に、差があるという結果が得られる見込み (確率)
 - 共通の測定基準は、Z (標準正規偏差) と p 値 (確率値)
 - Z は平均 0、分散 1 を持ち、p 値と直接結びつく (統計学のテキストの片側確率の表を参照)
- 表 A 推測統計量 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:	$Z = (r(N-2)^{1/2}) / (1-r^2)^{1/2}$ 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)

20

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、r (積率相関係数)、 Z_{FISHER} 、d (平均の標準化された差)
- 表 B 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 の Z_{FISHER} への変換式: $Z_{FISHER} = .5(\log((1+r)/(1-r)))$
- d の概念: 2つのグループの平均の差を、その測度の何らかの標準偏差の指標で割ったもの

21

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
 - 各 Z に重みづけ (重み w) 標本サイズ、方法論の質
 - 有意水準の結合: $Z = \frac{\sum w_j Z_j}{\sqrt{\sum w_j^2}}$
 - (w_j = 仮説検定 j の結果に割り当てられる重みづけ)
 - (Z_j = 仮説検定 j の有意水準に関する Z)
- フェイル・セーフ数の産出
 - $Nfs(p=.05) = \left(\frac{\sum Z_j}{1.645}\right)^2 - k$
 - (Z_j = 仮説検定 j の有意水準に関する Z)
 - (k = メタ分析に含まれる仮説検定の数)

22

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
 - r から求められた Z_{FISHER} に重みづけ (例えば、「1」や「N_j」)
 - 効果サイズの結合: $Z_{FISHER} = \frac{\sum w_j Z_{FISHER_j}}{\sum w_j}$
 - (w_j = 仮説検定 j の結果に割り当てられる重みづけ)
 - (Z_{FISHER_j} = 仮説検定 j の効果サイズに関する Z_{FISHER})
- フェイルセーフ数の算出
 - $Nfs(Z_{FISHER} = 0.100) = \frac{k(\bar{Z}_{FISHER} - Z_{FISHER_c})}{Z_{FISHER_c}}$
 - (k = メタ分析に含まれる仮説検定の数)
 - ($\bar{Z}_{FISHER} = k$ 個の仮説検定に対する平均 Z_{FISHER})
 - (Z_{FISHER_c} = 特定の基準に関する Z_{FISHER})

23

有意水準と効果サイズの比較

- (1) 中心化傾向
 - 典型的な研究結果は何か？
- 複数の研究の有意水準の結合 (標準化正規偏差 Z) と、効果サイズの結合 (r の Z_{FISHER} 変換)
- スライド 22・23

24

有意水準と効果サイズの比較

(2) 変動性: 拡散比較

- 典型的な研究結果の周りにどれほど変動があるのか?

- 拡散比較

— 仮説検定の有意水準や効果サイズが有意に異なるかどうかを調べる(なぜ異なるかは明らかにしない) cf. 焦点比較

- 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: 変動に関する重要な側面を検出する手助け

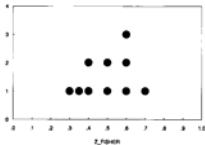


図3 効果サイズの度数分布の例 (出典: Mullen, B. (小野寺訳) (2000))

25

有意水準と効果サイズの比較

(3) 予測性: 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?

- 研究結果の予測変数

— 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

- 焦点比較

— 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(特定され、方向性を持ち、十分に定義された説明) cf. 拡散比較

2つ以上の予測変数の組み合わせ効果を見ることが出来る

26

メタ分析への批判

- お蔵入り問題 (file drawer problem)

- リンゴとオレンジを混ぜている

— 異なる種類 (サンプリング・手続き等) の仮説検定の結果を統合?

— メタ分析の焦点を狭くすると、目的は明確になるが対象とする研究の数は少なくなる

— 焦点を広くする際には、「リンゴ」と「オレンジ」の個別効果も調べられるように準備しておく

- Garbage In Garbage Out

— データベースに含まれた仮説検定の質が低い

— 「リンゴとオレンジ」の特殊ケース

27

一般的なルール

- 厳密に、批判に耐える形で、注意深く実行する

— 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つの分析的問題に答えること

- 2つ以上のアプローチを行なう

— 補足的な分析として他のアプローチも行い、結果に違いがあるなら慎重に注意を払う

- 全ての情報の報告

— 批判的な読み手が知りたいと思うことを全て提供する

(関連した情報の表、批判に頑健でなく関連がない分析結果も示す)

28

2. メタ分析ソフト「統合」

29

「統合」実行方法

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」

30

サンプルデータ (Mullen, B. (小野寺帆) (2000) を参考に作成)

著者	出版年	統計量 (df)	N	効果の方向	Xの強度
A	1988	$z = 23(1)$	110	+	7
B	1990	$t = 6.5(98)$	100	+	5
C	1992	$t = 6.0(63)$	66	+	9
D	1994	$r = .335(78)$	80	+	5
E	1996	$r = .535(68)$	70	+	10
F	1998	$Z = 3.891(-)$	70	+	3
G	2000	$F = 15.0(1,88)$	90	+	8
H	2002	$F = 10.25(1,63)$	65	+	4
I	2004	$p = .00001(-)$	80	+	8
J	2006	$p = .01(-)$	60	+	2

効果の方向: + は期待された方向、- は期待されない方向
Xの強度: 独立変数Xの操作の強度

31

メタ分析結果1

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*****	1988	1	23	1	110	4.796	0.494	0.457	0.209	0	1.028	7
2	B*****	1990	1	6.5	98	100	5.912	0.617	0.549	0.301	0	1.313	5
3	C*****	1992	1	6	64	66	5.323	0.693	0.6	0.36	0	1.5	9
4	D*****	1994	1	0.335	78	80	3.037	0.348	0.335	0.112	0.001	0.711	5
5	E*****	1996	1	0.535	68	70	4.771	0.597	0.535	0.286	0	1.266	10
6	F*****	1998	1	3.891	68	70	3.891	0.504	0.465	0.216	0	1.051	3
7	G*****	2000	1	15	88	90	3.711	0.402	0.382	0.146	0	0.826	8
8	H*****	2002	1	10.25	63	65	3.069	0.393	0.374	0.14	0.001	0.807	4
9	I*****	2004	1	0	78	80	4.892	0.614	0.547	0.299	0	1.307	8
10	J*****	2006	1	0.01	58	60	2.576	0.346	0.333	0.111	0.005	0.705	2

基本統計量

平均と標準偏差	発行人	人数	Z値	FisherのZ	相関	d	変数強度
平均	1998	79.1	4.1978	0.5008	0.4577	1.0514	6.1
S D	4.899	15.5335	1.0467	0.1189	0.0927	0.2698	2.5475

32

メタ分析結果2

相関行列

相関係数

	発行人	人数	Z値	FisherのZ	相関	d	変数強度
発行人	1.0000	-0.3364	-0.6097	-0.4898	-0.4836	-0.4931	-0.3045
人数	-0.3364	1.0000	0.4817	0.117	0.1361	0.1061	0.2247
Z値	-0.6097	0.4817	1.0000	0.9182	0.9237	0.9144	0.5923
FisherのZ	-0.4898	0.117	0.9182	1.0000	0.9992	0.9998	0.6094
相関	-0.4836	0.1361	0.9237	0.9992	1.0000	0.9981	0.6091
Cohenのd	-0.4931	0.1061	0.9144	0.9998	0.9981	1.0000	0.6104
変数強度	-0.3045	0.2247	0.5923	0.6094	0.6091	0.6104	1.0000

33

メタ分析結果3

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 = 13.333
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 = 0.504
人数による重みづけあり結合相関 r = 0.465
人数による重みづけあり結合相関の2乗 = 0.216225
人数による重みづけあり結合効果サイズ d = 1.05
人数による重みづけなしの結合 Z = 13.275
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 = 0.501
人数による重みづけなしの結合相関 r = 0.463
人数による重みづけなしの結合相関の2乗 = 0.214369
人数による重みづけなしの結合効果サイズ d = 1.044

34

メタ分析結果4

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)=641.196
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)=315.705
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=40.335
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 2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となります。
一方、有意でないなら、各研究は有意に異質ではないこととなります。
統計学のテキストなどの 2の表を参照して下さい。
Zと平均Zによる拡散比較 2値(DF=9)=10.9558536
FisherZと平均FisherZによる拡散比較 2値(DF=9)=9.98280164

35

メタ分析結果5

分解クラスター

共通サンプル・サイズの平均平方推定値 = 8.68

もとの研究番号とU値

研究番号(1)のU値=4.288,(2)のU値=5.356,(3)のU値=6.015,(4)のU値=3.021,(5)のU値=5.182,
(6)のU値=4.375,(7)のU値=3.489,(8)のU値=3.411,(9)のU値=5.330,(10)のU値=3.003

ソート後のUと隣り合うUの差

U値=3.021	.391
U値=3.411	.078
U値=3.489	.799
U値=4.288	.087
U値=4.375	.807
U値=5.182	.148
U値=5.330	.026
U値=5.356	.860
U値=6.015	-3.012
U値=3.003	

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 Independent Studies. Psychological Bulletin, 1983, Vol. 93, No.3, 563-573. p.565
Table 1より数値を引用

36

文献・web



- Mullen, B. (小野寺訳) (2000) 基礎から学ぶメタ分析 ナカニシヤ出版
- 坂元桂 (1999) 抑うつ者および高不安者のネガティブ情報に対する潜在記憶バイアス - メタ分析による検討 - 性格心理学研究, 7, 57-65
- 「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-analysis/index.html>
- 「メタ分析」(小川絢子) 2003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/datasem03/ogawa.files/frame.htm>