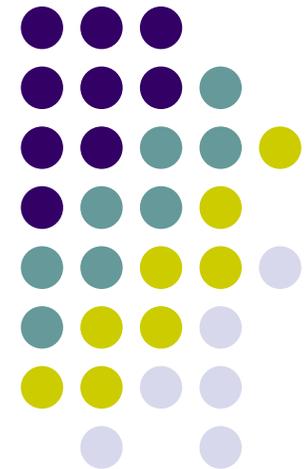


メタ分析(続編)

2008/01/16

M1 唐牛祐輔





メタ分析 Meta-Analysis

例)「いろんな研究を合わせて、全体的に見ると、どうなるのかが知りたい」 メタ分析

■ メタ分析

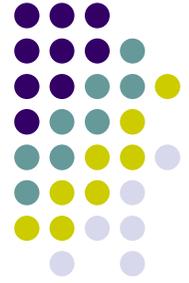
- 独立な研究結果の統計的な統合
 - 既存研究の量的レビュー cf.記述レビュー
 - メタ分析は一次水準の仮説検定の基本的なストラテジーを記述レビューの目標に応用したもの
- メタ分析とは、各研究結果を共通の測定基準に変換し、さまざまな研究結果を要約、統合、検証するための統計的テクニック



研究結果の“統計的統合”とは？

- 研究によって、測定方法はさまざま
例)「術前の心理教育的介入による、術後結果の改善の効果を知りたい」
 - 術後の痛み(痛み止めの量)
 - ナースステーションへの苦情
 - 評定尺度による自己報告
- 研究結果を比較・統合するためには、共通の測定基準に変換することが必要

効果量



効果量 (Effect Size)

- 標準化した平均値の差 (d)

(The standardized mean difference statistic)

- 1標準偏差あたりの平均値の差

$$d = \frac{\overline{X}_e - \overline{X}_c}{SD}$$

\overline{X}_e = 実験群の平均値

\overline{X}_c = 統制群の平均値

SD = 両群のプールされた標準偏差

プールされた標準偏差

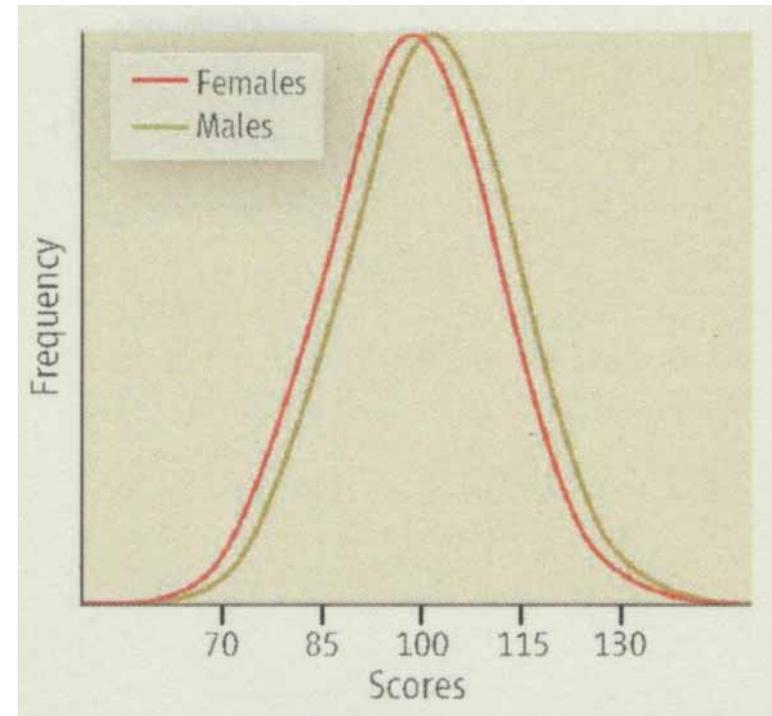
$$SD = \sqrt{\frac{(n_e - 1)SD_e^2 + (n_c - 1)SD_c^2}{n_e + n_c - 2}}$$



効果量 (Effect Size)

- 効果の大きさを示す指標として使われる
 - 効果量の目安 (慣例的なものであり、あくまで目安)
 - 0.11 - 0.35 … 小
 - 0.36 - 0.65 … 中
 - 0.66 - 1.00 … 大
 - 例) 効果量0.20の場合
 - 2つの群に, 14.7%の分布のずれがあることを表す (85.3%の分布の重なりがある)

Fig.1 効果量0.20のときの分布図
(出典: Hyde & Linn, 2006)

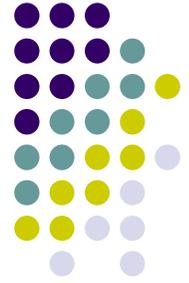




メタ分析を用いた研究例

- Hyde(2005) “The Gender Similarities Hypothesis”
 - さまざまな心理学的変数における男女差は見られるのか？
 - この仮説を検証するために、メタ分析を用いて検討

- **メタ分析の大まかな流れ**
 - PsychINFOなどのデータベースやキーワード検索を用いて対象となる研究を取り出す
 - 統計量を抽出して効果サイズ(d)を算出
 - 重み付け平均効果サイズを算出して、全研究を統合した際の全体的な結果の方向、マグニチュードを得る
 - グループ内の効果サイズの均質性の検定 分割



メタ分析を用いた研究例

□ 方法

- 対象: 46のメタ分析研究(124の効果サイズを抽出)
- 6つのカテゴリー(認知的変数、コミュニケーション、社会的・個人的変数、心理学的幸福、運動行為、その他)に関して検討

➤ 効果量: Cohen(1988)の d を使用

$$d = \frac{\overline{X}_M - \overline{X}_F}{SD}$$

\overline{X}_M = 男性の平均値
 \overline{X}_F = 女性の平均値

男性優位の場合に正の値をとる



メタ分析を用いた研究例

- メタ分析に含めた研究の一覧 (出典: Hyde(2005)より一部抜粋)

Table1. 研究一覧表の一例

Major Meta-Analyses of Research on Psychological Gender Differences

Study and variable	Age	No. of reports	<i>d</i>
Cognitive variables			
Hyde, Fennema, & Lamon (1990)			
Mathematics computation	All	45	-0.14
Mathematics concepts	All	41	-0.03
Mathematics problem solving	All	48	+0.08
Hedges & Nowell (1995)			
Reading comprehension	Adolescents	5*	-0.09
Vocabulary	Adolescents	4*	+0.06
Mathematics	Adolescents	6*	+0.16
Perceptual speed	Adolescents	4*	-0.28
Science	Adolescents	4*	+0.32
Spatial ability	Adolescents	2*	+0.19
Hyde, Fennema, Ryan, et al. (1990)			
Mathematics self-confidence	All	56	+0.16
Mathematics anxiety	All	53	-0.15
Feingold (1988)			
DAT spelling	Adolescents	5*	-0.45
DAT language	Adolescents	5*	-0.40
DAT verbal reasoning	Adolescents	5*	-0.02
DAT abstract reasoning	Adolescents	5*	-0.04
DAT numerical ability	Adolescents	5*	-0.10
DAT perceptual speed	Adolescents	5*	-0.34
DAT mechanical reasoning	Adolescents	5*	+0.76
DAT space relations	Adolescents	5*	+0.15



メタ分析を用いた研究例

- 男女差を発達の観点から検討
 - 抽出した効果サイズを、検討したい観点から統合

Table2. サンプルの年齢層ごとの比較 (出典: Hyde, 2005)
Selected Meta-Analyses Showing Developmental Trends in the Magnitude of Gender Differences

Study and variable	Age (years)	No. of reports	<i>d</i>
Hyde, Fennema, & Lamon (1990) Mathematics: Complex problem solving	5-10	11	0.00
	11-14	21	-0.02
	15-18	10	+0.29
	19-25	15	+0.32
Kling et al. (1999) Self-esteem	7-10	22	+0.16
	11-14	53	+0.23
	15-18	44	+0.33
	19-22	72	+0.18
	23-59	16	+0.10
	>60	6	-0.03
Major et al. (1999) Self-esteem	5-10	24	+0.01
	11-13	34	+0.12
	14-18	65	+0.16
	19 or older	97	+0.13
Twenge & Nolen-Hoeksema (2002) Depressive symptoms	8-12	86	-0.04
	13-16	49	+0.16
Thomas & French (1985) Throwing distance	3-8	-	+1.50 to +2.00
	16-18	-	+3.50



メタ分析を用いた研究例

□ 結果

- 研究結果の78%が $d=0 - 0.35$ の範囲に含まれることが示された
- 数学や科学などでは効果サイズ d は小さい(詳細はHyde & Linn, 2006)

□ 結論

- 領域によっては男女差は小さく、統計的に有意な差が見られるとしても、その効果サイズは概して小さい
- 男女差の大きさを示す指標として効果量を用いて、その大きさを問題とすることで、男女差が一般に言われるほどには大きくないことを示した



メタ分析を用いた研究例

■ Jamieson & Harkins(2007)

- ワーキング・メモリー容量の指標として、抗サッケード課題 (antisaccade task) を用いる
- (問題-introduction-において)この課題を用いることの妥当性を示すために、“ワーキング・メモリー容量不足時に抗サッケード課題成績が低下する”という仮説を検証した実験について、メタ分析を用いて実験結果を統合



メタ分析を用いた研究例

- 4つの論文から、抗サッケード課題とワーキング・メモリー容量の関連を検討した8つの実験結果を抽出
 - 重み付け平均相関係数 r
 - Combined Z (有意水準の結合)
 - p 値を報告
- ◆ 実験結果を統合して、中程度の有意な相関が見られた
 - ワーキング・メモリー容量の指標として抗サッケード課題を用いることの妥当性を示す
 - レビュー論文ではなくても、これまでの先行研究のまとめとして、メタ分析を用いることも可能



メタ分析のその他の活用法

- メタ分析をレビュー以外に利用できないか？
 - 自分がこれから行いたい実験・調査と類似の手続きの先行研究の結果を統合
 - 効果量のおおよその予測を立てる
 - この効果量から、G * Powerなどを用いて、サンプルサイズを決定する(詳しくは布井くんの回を参照)
 - 自分の一連の研究結果をメタ分析
 - 結果が曖昧な実験に関して、おおよその結果の傾向をつかみ、次の研究のヒントとする
 - 単一では結果がやや弱い研究結果でも、それを統合することで、より強力な(説得力のある)結果を示す



メタ分析をしてみよう

■ 初級実習【目撃証言】の結果を統合してみよう

- 毎年、結果が有意だったり、有意にならなかつたり…
- 本当に効果があるのかを調べ、実習の実験として妥当なのか検証

□ 対象とする実験

- 2002年度～2007年度の初級実習【目撃証言】

□ 研究の抽出

- ロッカーに眠っている資料を漁ったり、以前に初級実習でTAをやっていた先輩方にあたってみたりして収集

□ 統合対象とする実験仮説

- 誤導された項目は、統制項目に比べて、(実際には見たのに)「見た(正)」と報告される数が少なくなる



メタ分析をしてみよう

□ 実験の内容(手続き)

- 3分間のビデオを視聴
- ビデオの内容について記した(誤情報を含む)文章を読む
- ビデオ内容についての各質問事項(誤導された項目 / 統制項目)に、「正」 / 「誤」で答える

□ 効果量

- 平均値、SDから、Cohenのdを算出

$$d = \frac{\bar{X}_c - \bar{X}_m}{SD}$$

\bar{X}_c = 統制項目について「正」と報告した平均値
 \bar{X}_m = 誤導項目について「正」と報告した平均値

- 誤導されて「正しい」と報告する数が少なくなると、正の値を取る(「正しい」と答えるのが正解)

メタ分析をしてみよう



Table3. 各初級実習の統計量

No.	実施年度	参加者数	誤導項目 (<i>SD</i>)	統制項目 (<i>SD</i>)	<i>p</i> 値	<i>F</i> 値	効果量 (<i>d</i>)
1	2002	29	1.38 (1.22)	2.45 (0.93)	$p < .01$	11.33	0.66
2	2003	23	1.44 (1.17)	2.96 (1.04)	$p < .05$	29.49	1.13
3	2004	19	1.53 (1.31)	2.74 (1.21)	$p < .05$	5.75	0.55
4	2005	23	1.87 (1.39)	2.09 (1.06)	<i>n.s.</i>	0.50	0.15
5	2006	21	2.19 (0.91)	2.43 (0.79)	<i>n.s.</i>	0.92	0.21
6	2007	26	1.92 (1.17)	2.65 (1.04)	$p < .05$	5.87	0.48

今回の実験では、独立の値ではなく、被験者内計画であるため、プールされたSDは、*F*値から算出したものを用いています。そのため、*d*値の計算結果が、やや異なるかと思えます。



メタ分析をしてみよう

- 重み付け平均効果量 d を算出

$$d = \frac{\sum \omega_j d_j}{\omega_j}$$

ω_j = 仮説検定 j に割り当てられる重み付け
(ここでは N を用いた)

d_j = 仮説検定 j の効果量 d

- 平均効果量 $d=0.53$
- (初級実習における) 誤情報による誤導効果は、中程度の効果量であることが示された

実際は、統合には、 d 値ではなく、 Z_{FISHER} を用いる。これを用いると、統合した Z_{FISHER} は 0.22 であり、ここから d 値を求めると、 $d=0.44$ となる。本来はこの値を用いるが、今回は便宜的にこの値を用いることとする。



メタ分析をしてみよう

- G * Powerで、サンプルサイズを計算してみよう
 - 先程算出された d 値を使用
 - $d=0.53$
 - 有意水準は.05に設定(両側検定)
 - 検出力は0.50に設定
 - 結果: サンプルサイズ $N=58$
 - 初級実習にはあまり向いてない…?
 - ◆ ただし、G * Powerでは被験者内計画は実行してくれないので、あくまで上記のサンプルサイズは被験者間計画
 - 被験者内ならもっと少なくて済むはず
 - 被験者間として効果量を算出すると、 $d=0.75$ であり、この場合のサンプルサイズは $N=30$ 十分実現可能?



参考資料

- 資料・Web

- Shadish, W. R. (2001). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. (pp.417-455). Houghton Mifflin College Div.
- 「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>
- 「効果量・検出力・サンプルサイズ」(布井雅人) 2007年度心理データ解析演習発表資料 <http://www.educ.kyoto-u.ac.jp/cogpsy/personal/Kusumi/datasem07/nunoi.pdf>
- 「日常記憶(目撃証言)」 2007年度初級実習レジюме(唐牛・前原・志波・平・常深)

- 引用文献

- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, 581-592.
- Hyde, J. S., & Linn, M. C. (2006). Gender similarities in Mathematics and Science. *Science*, 314, 599-600.
- Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, 93, 544-564.