

分散分析と効果量

データ解析演習

2009/05/20

教育学研究科M1 中山真孝

発表のアウトライン

- ・「平均値の差の検定には分散分析を用いる」
“平均値”なのにどうして“分散”分析？
- ・効果量って何？
- ・「効果量はメタ分析に用いられる」
レビューじゃなくても“メタ分析”

“平均値”なのにどうして“分散”分析？

の前に...

AさんとBさんがいます。Aさんはいつも必ず決まって授業開始5分前ぴったりに教室に現れます。Bさんは30分前に現れることもあるが、30分後に現れたり、5分前だったり5分後だったり、いつもばらばらです。

ところが、今月からAさんは決まって授業開始ぴったりの時間に現れるようになりました。

ところでBさんはこんなことを言っています

「先月は平均すると、授業開始時間ぴったりに来ていたけれど今月は平均すると5分前だから、よくがんばった」

“平均値”なのにどうして“分散”分析？

さて、これらについて、同じ5分の差でも(方向やもとの値はちがうとしても)こんな風に思いませんか

いつも決まって5分前だったAくんが、今月は決まってちょうどの時間に来るようになった。Aくんに関係があったにちがいない！

Bくん、平均で5分早く来たかもしれないけれど、がんばったんじゃないくて、たまたまやる！

“平均値”なのにどうして“分散”分析？

つまり、Aさんの5分差には意味があるけれど、Bさんの5分差には意味はないだろう

平均値の差が意味あるものかどうかは、その値だけでなく、そのデータがどれくらいばらついているか、どれくらい誤差をふくんでいるかによる(それが全てではないが)

このような考えにもとづいて、数学的にしたものが分散分析(統計的検定?)

「統計学の考え方の基本は、私たちが日常行なっている思考の中の良識あるものを、少し洗練して定式化したものに過ぎません。」(吉田, 1998, p268)

“平均値”なのにどうして“分散”分析？

数式でかくと

$$F = \frac{MS_A}{MS_E}$$

要因による変動(自由度で割ったもの)

誤差による変動(自由度で割ったもの)

ちなみに自由度が n-1 であったりするのは、すべての平均が0であることを制約式としておいているからで、この意味でも、平均値そのものよりも差or分散を分析している

F値は

- ・ 要因による変動(各条件の平均値差)が大きいほど
- ・ 誤差による変動(各条件の標準偏差)が小さいほど
- ・ データ数が多いほど(自由度で割るので)

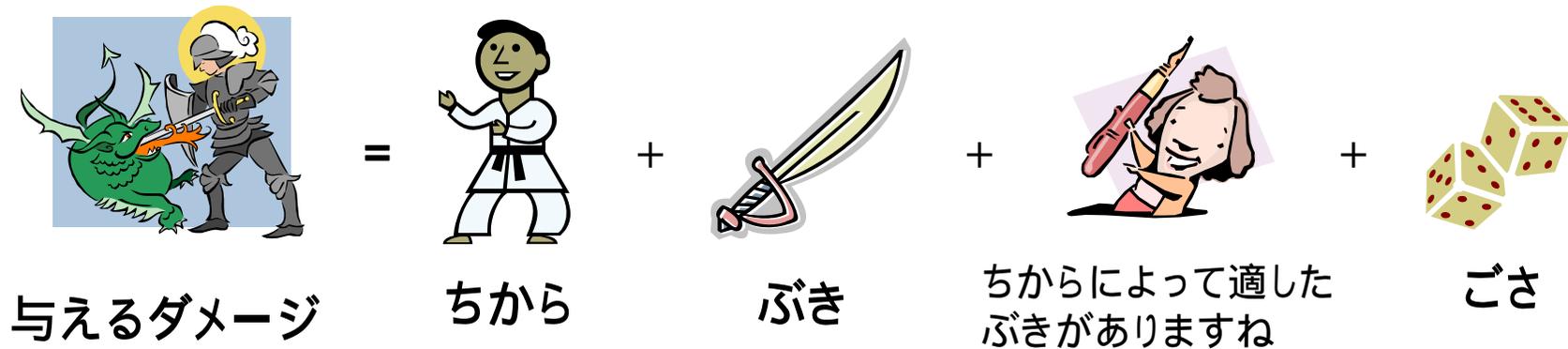
大きくなる

MSA などについてももう少し詳しく

$$\begin{aligned} x_{ijk} &= \mu + a_j + b_k + e_{ijk} \\ x_{ijk} &= \mu + a_j + b_k + (ab)_{jk} + e_{ijk} \end{aligned}$$

あるセルの値 Aの要因効果 交互作用効果 誤差

イメージ的には



構造式のように平方和を分解

$$SS_A = \sum_{i=1}^n \sum_{j=1}^p \sum_{k=1}^q \hat{a}_j^2 = nq \sum_{j=1}^p \hat{a}_j^2 = nq \sum_{j=1}^p (\bar{x}_{.j.} - \bar{x})^2, \dots$$

$$SS_T = SS_A + SS_B + SS_{A \times B} + SS_e$$

自由度で割ると平均平方

$$MS_A = \frac{SS_A}{df_A}, MS_e = \frac{SS_e}{df_e} \dots$$

効果量って何？

いきなり数式から

$$d = \frac{(\text{実験群の平均} - \text{統制群の平均})}{\sqrt{\frac{\text{実験群の標準偏差}^2 + \text{統制群の標準偏差}^2}{2}}}$$

ちなみに t 値は分母のルートの中を df で割ったもの

とてもおおざっぱに言えば、

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

要因の効果を誤差 (それ以外の効果)

で割ったもの

$$\eta^2 = \frac{\text{ある要因の平方和 } (SS_{\text{effect}})}{\text{全体平方和 } (SS_{\text{total}})}$$

$$\text{partial } \eta^2 = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{\text{error}}}$$

イータと読む

効果量って何？

ところで。。。

$$F = \frac{MS_A}{MS_e}$$

要因による変動 (自由度で割ったもの)

とてもおおざっぱに言えば、

誤差による変動 (自由度で割ったもの)

要因の効果を**誤差で割ったもの**

ちなみに自由度が1であったら平均値そのものが0であることを制約式としておいているからで、この意味でも、平均値そのものよりも差or分散を分析している

F値は

- ・ 要因による変動 (各条件の平均値差) が大きいほど
- ・ 誤差による変動 (各条件の標準偏差) が小さいほど
- ・ データ数が多いほど (自由度で割るので)

大きくなる

効果量って何？

効果量は目新しいものではなく、分散分析
でいえば F 値と同じ考え方に基づく指標

効果量は(基本的に)

- ・要因による変動(各条件の平均値差)が大きいほど
- ・誤差による変動(各条件の標準偏差)が小さいほど
大きくなる

効果量って何？

効果量は実質的な差の標準化された値
これを用いれば単位が異なるもの同士も
ある程度比較することができる
メタ分析も

では効果量と F 値とはどう違うのか？なぜ分散分析
で効果量ではなく F 値, p 値を用いるのか？？

効果量って何？

分散分析の手順 (F 値の後に...)

自由度をもとに、i.i.d.でその F 値が得られる確率 = p 値を求め、 (通常0.05) 以下なら有意とする

有意かどうかは自由度 サンプルサイズの影響も受ける。また、検定しているのは差がないという帰無仮説

そもそも F 値自体サンプルサイズの影響を受ける

i.i.d = 独立同一分布 帰無仮説であり、同一ではないと主張したい

(独立性は手続き等で保証しないといけない)

効果量って何？

いくら効果量は実質的な差の標準化された値だといっても、あくまでそれはあるサンプルから得られた値でしかない

それが真の効果量であるかどうかはわからない

cf 小数の法則

サンプルサイズを考慮しないとイケない

分散分析はそれを考慮して有意性を検定している

サンプルサイズが大きくなれば得られた値も真値に近づく

cf 大数の法則

効果量って何？

つまり p 値は

- ・効果量が大きいほど

- ・データ数が多いほど

小さくなる

有意になったのはどちら(どれ)によるものかを考えなくてはならない

効果量、 p 値に影響を与えるものとして…

・誤差による変動(各条件の標準偏差)が小さいほど
小さくなる

誤差を統制することが重要

だが、無作為化が多く用いられ、誤差の統制の重要性
は認識されていない？ (吉田, 2002)

同じ効果量でも、誤差(他の要因)による影響が大きい
のか当該要因による変動(効果)が大きいかを考えた
方がよいかもしれない

2乗と偏 2乗(偏相関比)について

$$\eta^2 = \frac{\text{ある要因の平方和 } (SS_{effect})}{\text{全体平方和 } (SS_{total})}$$

ある要因の平方和が大きいと大きくなる

が、全体平方和が大きいと小さくなる

要因数がふえると小さくなる

$$partial \eta^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$$

ある要因の平方和のみをとりだして、誤差の大きさと比べる(誤差が大きいと小さくなる)

効果量を実際に扱うために

水本・竹内(2008)には各分析における効果量の指標と大きさ判断の目安がまとめられ、効果量を計算してくれるエクセルファイル(第一著者のページ)も紹介されている

表 1
検定(分析)の種類ごとに見る代表的な効果量の指標と大きさの目安

使用される検定(分析)	対象と注意	効果量の指標	効果量の目安		
			小 (Small)	中 (Medium)	大 (Large)
(1) 相関分析		r	.10	.30	.50
(2) 重回帰分析		R^2	.02	.13	.26
		f^2	.02	.15	.35
(3) t 検定 (t -test)	対応あり・なしとも同じ	r	.10	.30	.50
		d	.20	.50	.80
(4) 一元配置分散分析 (One-way ANOVA)	全体の差の検定	η^2	.01	.06	.14
		partial η^2	-	-	-
		ω^2	.01	.09	.25
	多重比較	f	.10	.25	.40
		r	.10	.30	.50
(5) 二元配置分散分析 (Two-way ANOVA)	主効果	η^2	.01	.06	.14
		partial η^2	-	-	-
		ω^2	.01	.09	.25
	多重比較	η^2	.01	.06	.14
		partial η^2	-	-	-
多元配置分散分析* (Multi-way ANOVA) *三元配置以上の分散分析	交互作用	ω^2	.01	.09	.25
	多重比較	r	.10	.30	.50

検定(分析)	対象と注意	効果量の指標	効果量の目安		
			小 (Small)	中 (Medium)	大 (Large)
(6) 共分散分析 (ANCOVA)	共変量の影響を取り除いて分析し、主効果、交互作用、多重比較の効果量は(4)や(5)と同じ	multivariate η^2 (multivariate R^2)	-	-	-
(7) 多変量分散分析 (MANOVA)	多変量検定	multivariate partial η^2	-	-	-
多変量共分散分析 (MANCOVA)	従属変数ごとの分散分析	主効果、交互作用、多重比較の効果量は(4)や(5)と同じ			
(8) カイ2乗検定 (χ^2 test)	2×2の分割表	$\phi (= W)$.10	.30	.50
	2×2以外	Cramer's V	.10	.30	.50
(9) マン・ホイットニーの U 検定 ウィルコクソンの符号順位和検定 クラスカル・ウォリスの順位和検定 フリードマン検定	検定統計量を Z に変換して r を求める	r	.10	.30	.50

Note. Cohen (1998: 1992), Field (2005), Tabachnick and Fidell (2006) などを基に作成。効果量の大きさはあくまで目安であるので研究分野によって変わる。(3) d , (4) f , (8) W についての詳細は, Cohen (1988) を参照のこと。 η^2 の大きさの目安は文献によっては, r を2乗した r^2 に合わせて, $\eta^2 = .01$ (効果量小), $\eta^2 = .09$ (効果量中), $\eta^2 = .25$ (効果量大)としているものもある。また, partial η^2 の効果の大きさの基準は明確なものがない。multivariate η^2 と multivariate partial η^2 の値は従属変数(dependent variable)の数によって変わるため, 効果量の目安は Cohen (1998) を参照。

「効果量はメタ分析に用いられる」

効果量は実質的な差の標準化された値。これを用い
れば単位が異なるもの同士もある程度比較す
ることができる

「効果量はメタ分析に用いられる」

しかし、実際それで論文を書こうと思うと、たくさ
んの論文を読んでレビューとして書かなけれ
ばならない。。。

効果量 メタ分析 レビュー 敷居が高い なイメージ

レビューじゃなくても“メタ分析”

“メタ分析”をそのまま、“分析についての分析” or
“1つ高次なところから分析”と考えると実験論文
でも、効果量を用いて“メタ分析”はできる or した
ほうがよい！

例えば、Larsen & Baddley(2003)では、先行
研究で関連あるといわれる変数をシステマ
ティックに操作し、複数の仮説を検討してい
る レビュー的な要素は強いが

TABLE 1
 Mean proportion of letters in the correct serial position:
 Effects of a single concurrent token at equal intervals

<i>Concurrent task</i>	<i>Letter set</i>			
	<i>Similar</i>		<i>Phonologically dissimilar</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Control	.61	.17	.83	.15
Irrelevant speech	.62	.17	.81	.16
Tapping	.53	.19	.73	.22
Articulatory suppression	.46	.17	.46	.12

TABLE 4
 Summary of the magnitude of the effect of each variable manipulated across Experiments 1-3

<i>Condition</i>		<i>Experiment</i>					
		<i>1</i>		<i>2</i>		<i>3</i>	
		<i>Single token at equal intervals</i>	<i>Syncopated single token</i>	<i>Multiple tokens</i>	<i>Effect</i>	η^2	<i>Effect</i>
Irrelevant speech	Decrement	2	.062	11	.467	15	.539
	Similarity effect	19	.532	12	.462	11	.287
Manual tapping	Decrement	10	.226	37	.844	11	.388
	Similarity effect	20	.587	0	.004	24	.642
Articulatory suppression	Decrement	37	.827	41	.881	42	.878
	Similarity effect	0	.000	-3	.053	-3	.060

Note: Magnitude of decrement from the control condition for the dissimilar letters is given, in percentages, together with the effect size (eta squared). The effect of acoustic similarity (the difference between similar and dissimilar letters) and effect size (eta squared) is also shown for each condition.

レビューじゃなくても“メタ分析”

ふつうの実験論文でも、総合考察の部分では“メタ分析”を行っている or 行うべきはず
そのようなときに効果量について積極的に言及した方が根拠が明確になる

「差がない」ことを主張する際にも根拠となる
もちろん、サンプルサイズには気をつけたほうがいいが

もっと、効果量という指標を使いましょう！

参考文献

狩野裕 (2007) 心理・教育測定論講義資料

Larsen, J., & Baddeley, A. D. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology*, **56A**, 1249–1268.

水本篤・竹内理 (2008) 研究論文における効果量の報告のために 基礎的概念と注意点 *英語教育研究* 31 57-66.

永田靖 (2000) 入門実験計画法 日科技連出版社

吉田寿夫 (1998) 本当にわかりやすい すごく大切なことが書いてある ごく初歩の統計の本 北大路書房

吉田寿夫 (2002) 第3章 研究法に関する基本姿勢を問う: 本来の姿ないし基本に戻ろう 下山晴彦・子安増生 (編) 心理学の新しいかたち 方法への意識 誠信書房