

英語教育研究法セミナー 2 B

1. p 値の意味すること、しないこと

浦野 研 (urano@ba.hokkai-s-u.ac.jp)
 北海学園大学

発表資料およびその他補足情報は、ウェブ上にも公開します。以下の URL をご利用ください。
<http://www.urano-ken.com/research/seminar/>

0. 要旨

実証研究のうち、何らかの数値をデータとして扱ういわゆる量的研究では、英語教育に関わる研究論文や学会発表においても、「 $p < .05$ 」や「有意差が見られた」といった表現がよく使われる。また、研究を行う側も読む（聴く）側も、統計処理の過程を見ずに p 値のみを確認し、それに基づいて結果の解釈・議論・批評を行うことが多い。ところが、実際の研究には被験者数 (sample size) が極端に少ないものや多いものがあり、p 値だけを見て研究結果の解釈を行うことには問題がある。そこで本発表では、統計手法の中でも比較的イメージしやすい t 検定と相関を用い、p 値がどのように導き出され、それが何を意味して、何を意味しないのかを解説したい。主に p 値と被験者数の関係に注目し、効果量 (effect size) や検定力 (power) といった概念を紹介しながら、被験者数が特に少ない（多い）ときの結果解釈における注意点や、先行研究の結果を比較・分析する方法（メタ分析）も紹介する。

1. 有意性（有意差）とは

1.1. 相関と有意性

次の 2 つの散布図を比べてください。

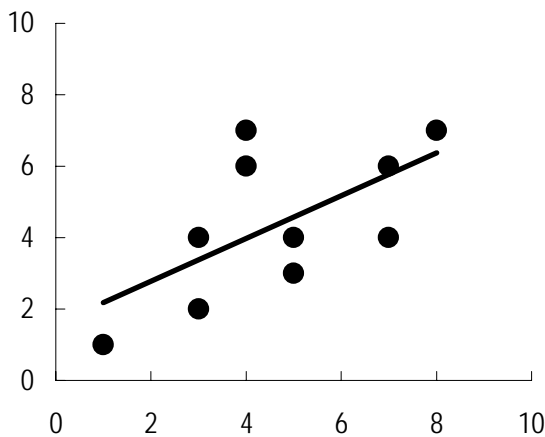


図 1 (N = 10)

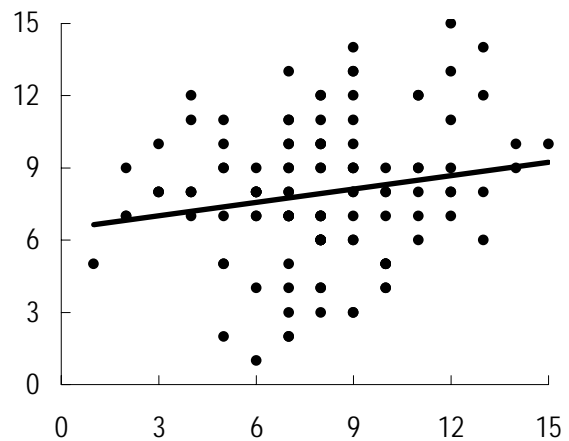


図 2 (N = 119)

問 1. はっきりしたパターンを示しているのはどちらですか？

図 1. $r = .627$

図 2. $r = .184$

問 2. 統計的に有意なのはどちらですか？

図 1. $p = .052$

図 2. $p = .045$

表1. ピアソンの相関係数の有意性検定における r の臨界値 (抜粋)

N	3	4	5	10	20	30	40	50	100	200	500	1000
両側検定で有意水準5%の場合	.997	.950	.878	.632	.444	.361	.312	.279	.197	.139	.088	.062

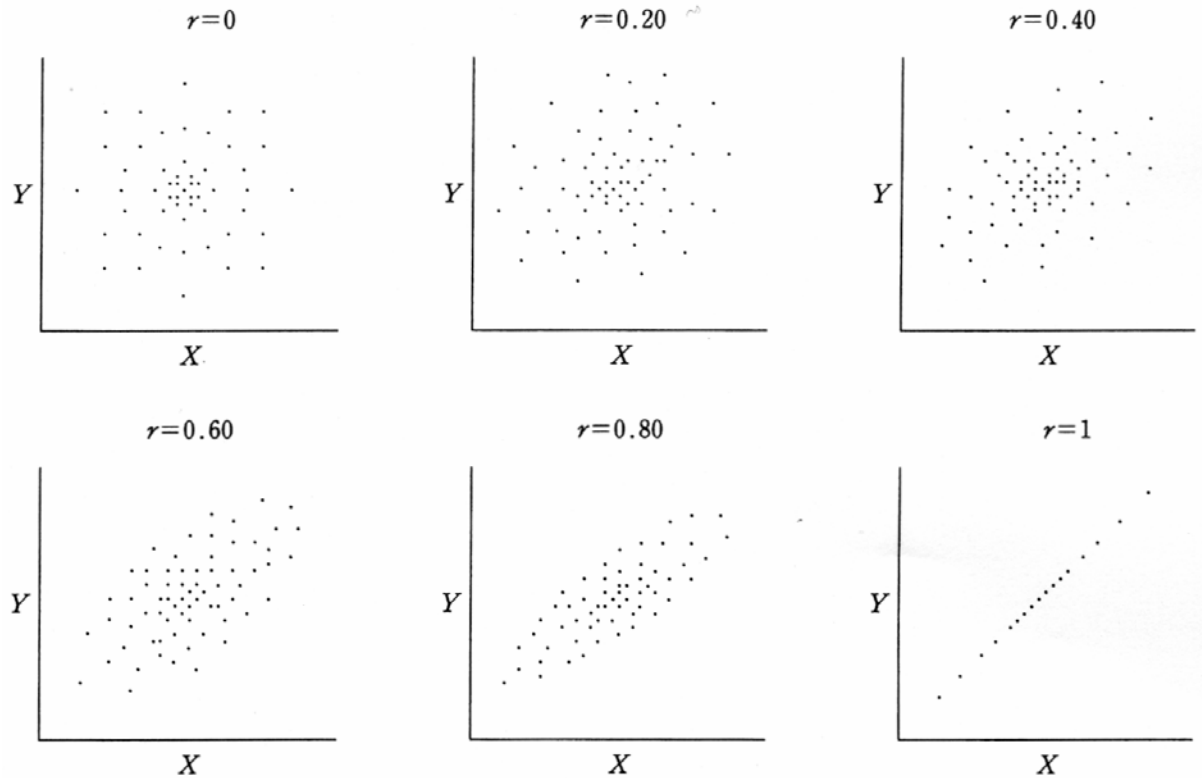


図3. 相関係数の値と相関図の様相の対応関係 (吉田, 1998, p. 75)

1.2. 推測統計 (inferential statistics)

実際に集めたデータ (標本) から、より大きな全体像 (母集団) の分布を推測する。

例：全国から無作為に抽出した有権者 3,000 人を対象に支持政党の調査をし、その結果から日本の有権者全体の政党支持率を推測する。

1.3. 正規分布

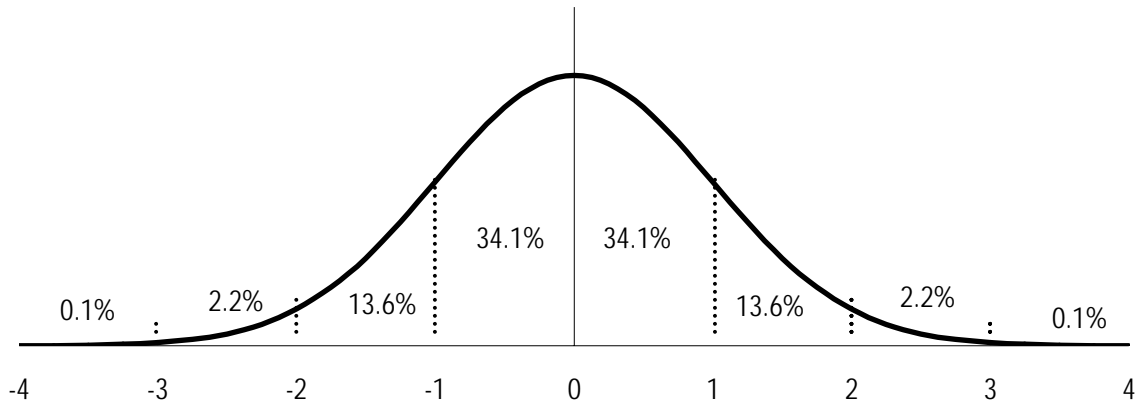


図4. 標準正規分布 (平均=0、標準偏差=1)

1.4. t 検定と有意差

表2. t の臨界値 (抜粋)

df (= n-1)	1	2	3	4	5	10	20	50	100	200
両側検定で有意水準5%の場合	12.71	4.30	3.18	2.78	2.57	2.23	2.09	2.01	1.98	1.97

t 検定の算出方法 (2つの条件のデータ数が同じ場合) :

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{SD_1^2 + SD_2^2}{n-1}}}$$

← (1) 2条件の平均の差
 ← (2) 2条件の分散 (標準偏差の2乗) の和
 ← (3) データ数 - 1 = 自由度 (degree of freedom: df)

上の式から言えること :

- | | | |
|---|---|-----------|
| (1) 2つの条件の平均の差が大きいほど
(2) 各条件の標準偏差が小さいほど
(3) 各条件のデータ数が多いほど | } | t 値が大きくなる |
|---|---|-----------|

(1)、(2) は研究者がコントロールできないが、(3) については研究者が自由に増減できる。

有意な差 (または、有意な相関) が得られやすい研究をするためのきわめて有効な方法があります。それは、とにかく多くのデータを集めることです。なぜならば... 統計的検定の結果はデータ数が多いほど有意になりやすいからです。そのため、データ数を増やしさえすれば、きわめて小さな差でも“(統計的には) 有意である”という結果になる可能性が高まります。例えば、 $N = 1000$ の場合には、 $r = 0.062$ というきわめて小さな相関係数 (すなわち、非常に弱い関係) でも有意になります...。このように、統計的推定には、“データ数という、研究者が任意に決められる要因によって結果が左右されてしまう”という根本的な問題があります。 (吉田, 1998, p. 232; 下線は浦野による)

1.5. ここまでのまとめ

データ数 (被験者数) が有意性に与える影響は大きい :

- (1) データ数が少ないときは、実際よりも差 (関係) が矮小化されてしまう。
 - 本当は差があっても、データからは有意な差が見出されない可能性がある (「これを検出力が低い」と言う)。
 - この場合、「差がない」と判断するのは危険である。
- (2) データ数が多いときは、実際よりも差 (関係) が誇張されてしまう。
 - 統計的に有意であっても、それが意味のあるものとは限らない。例えば相関の場合、データ数が 1,000 あれば $r = .062$ というとても弱い相関でも有意性が検出される (表1、図3を参照)。

2. p 値（有意性）以外の指標

1.5. で挙げた問題を回避するには、データ数に左右されない指標もあわせて検討する必要がある。

2.1. 相関の場合

相関の場合にはピアソンの r そのもの（下の式を参照）、もしくは r^2 （説明率）：

$$r = \frac{\sum_{i=1}^N X_i Y_i}{N} - \overline{XY}}{SD_X SD_Y}$$

（ r の数値の意味については図 3 を参照）

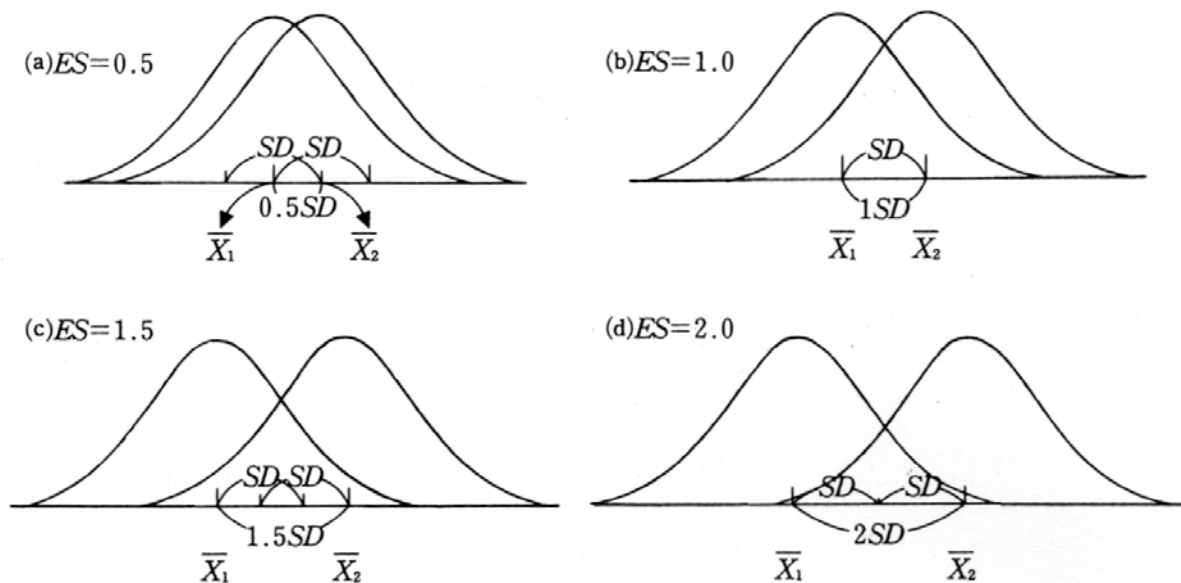
2.2. t 検定の場合

t 検定の場合には効果量（effect size: ES）。

Cohen's d （代表的な ES 指標）の算出方法：

$$d = \frac{|\bar{X}_1 - \bar{X}_2|}{SD_{pooled}}$$

← ● 2 条件の平均の差
← ● 2 条件の標準偏差の（各条件のデータ数の違いを考慮した）平均値



注：2つの条件の標準偏差が等しく、かつ両条件とも分布が正規分布に従っていると仮定

図5. ES (= Cohen's d) が 0.5、1.0、1.5、2.0 である2つの条件の分布の重なり程度（吉田, 1998, p. 75）

表3. ES (= d) に対応した分布の重なり（重ならない部分の割合で表示）

ES (d)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
non-overlap (%)	0.0	7.7	14.7	21.3	27.4	33.0	38.2	43.0	47.4	51.6

1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
55.4	58.9	62.2	65.3	68.1	70.7	73.1	75.4	77.4	79.4	81.1

2.3. その他の検定の場合

Cohen (1988) を参照。

3. 補足事項

3.1. p 値の問題点の認知度

この発表で議論してきた p 値のみに頼った分析、解釈の問題点は心理学や応用言語学でも理解が進み、例えば APA (American Psychological Association) のガイドラインでは、p 値と共に効果量も論文に掲載することが奨励されている。研究誌 *Language Learning* では、投稿要領に効果量が必須であると明記されている。

3.2. メタ分析

先行研究の分析をしていると、似たような実験や調査を行っているのに有意差が検出された研究と検出されない研究が混在していることがある。それぞれの研究で被験者数が異なるためにそれぞれの研究結果を直接比較することが難しいので、被験者数の影響を受けない効果量を計算して比較・分析を試みるのがメタ分析。

3.3. ソフトウェア

G*Power (検定力算出プログラム)

[<http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>]:

データ数、効果量、検定力といった数値の換算をしてくれるソフト（フリーウェア）。例えば、先行研究やパイロット実験の結果からある効果量が期待できるとわかっている場合、その効果量で有意性を検出するにはどの程度の被験者数が必要か、といった計算をしてくれる。

4. 参考文献

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』京都: 北大路書房.

マニュアル的な how to 本が多い中、本書は統計を理解する上で重要なコンセプトを丁寧に解説していておすすめ。

森敏昭・吉田寿夫 (編著) (1990). 『心理学のためのデータ解析テクニカルブック』京都: 北大路書房.

吉田 (1998) でカバーしていないより専門的な説明はこちらで。

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

効果量や検定力分析を勉強するのに必携。ただし、ページ数も多く内容も簡単ではないので、独学は難しいかもしれません。