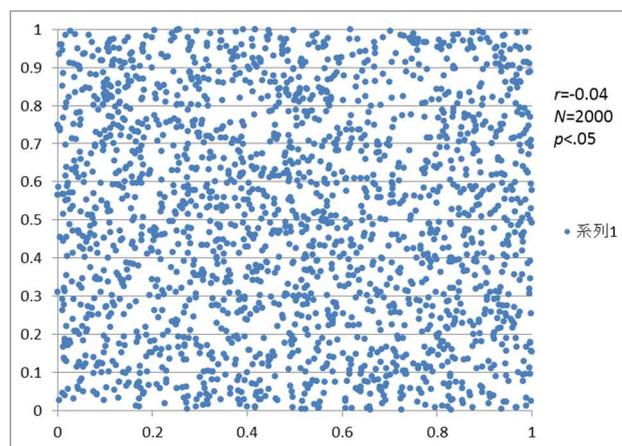


## ●効果の大きさ(効果量)(effect size: ES) : 検定力 : 標本の大きさ : 第 I ・ II 種の誤り ● M.Kishi

### 1 標本の大きさ(サンプルサイズ)( $N$ )が大きいと?

量的研究では、標本の大きさがある程度以上にならないと、得られた結果の誤差が大きくなり、結果から一般化を行うときの信憑性が問題になります。では、大きければ、問題ないのでしょうか？実は、とても大きな標本で統計的検定を行ったとしても、困った問題が生じてくるのです。

たとえば、今、2つの群で  $t$  検定を行ったとします。A 群は、 $N=1000$ 、平均=37.1、 $SD=1.2$  で、B 群は  $N=1100$ 、平均=37.0、 $SD=1.1$  であったとします。両群の間の平均差は0.1です。有意差がありそうですか？結果は5%水準で有意差ありです ( $t(2098)=1.992$ ,  $p<.05$ ) !そして、2つの  $SD$  も有意差があります ( $F(999, 1099)=1.190$ ,  $p<.01$ )。ほとんど差が無いように思えても、実際には有意差がみられました。では



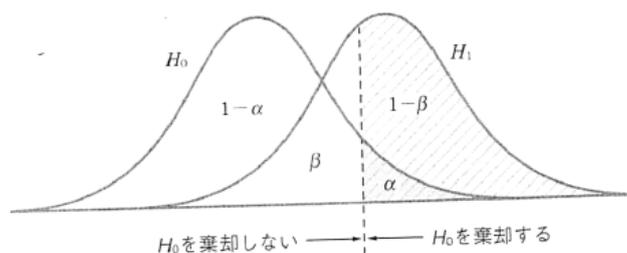
は相関はどうでしょうか？右図を見てください。これは、 $N=2000$  で  $r = -0.044$  となり、5%水準で有意な相関ありとなりますが、散布図を見る限り、単にバラバラとしか言いようがありません。

統計的検定については、検定によってどの程度の差や関係を見いだしているのか(見いださうのか)という視点からも検討する必要があります。そのとき、標本の大きさが大きく関わります。

### 2 第 I 種の誤り・第 II 種の誤り

「帰無仮説を棄却する/しない」は、ある確率、すなわち有意水準と比較し、判断します。しかし、あくまでも確率に基づく判断であって、本当に棄却してよい/悪いのかはわかりません。

① いま、2つの母集団(1年生と2年生:母平均はそれぞれ  $\mu_1$  と  $\mu_2$ ) から得られた標本について、 $t$  値を求め、 $t$  値の分布を右図のように描いたとします。横軸は  $t$  値です。



第 I 種の誤りをおかす確率( $\alpha$ )と第 II 種の誤りをおかす確率( $\beta$ )に対応する領域

② 図内の左側の分布は、帰無仮説  $H_0$  に従った場合の  $t$  の値の分布 ( $H_0: \mu_1 = \mu_2$ , あるいは  $\mu_1 - \mu_2 = 0$ )

とします。分布の頂点のところの  $t$  値は、帰無仮説から0になりますが、当然ながら、標本の抽出のしかたによっては  $t$  値が0から離れた値もあり得ます。一方、図内右側は対立仮説  $H_1$  に従った場合の  $t$  値の分布 ( $H_1: \mu_2 > \mu_1$ ) とします。真ん中は  $\mu_2 - \mu_1$  です。そして、ここでは便宜上、片側検定とします。

③ 図内の縦破線をみてください。これは、 $t$  値がこの線より右側の値なら帰無仮説  $H_0$  を棄却して対立仮説  $H_1$  を採択し(珍しいことが起こった)、左側の値なら帰無仮説を棄却しない状態を表しています(別に珍しいとはいえない)。したがって、 $\alpha$  の部分の面積が  $H_0$  における棄却の際の有意確率となります。

④ 表をみてください。これは、検定の結果を研究者がどのように判断したかを横側に、本当はどうだ

ったのかを縦側にした表で、全部で4つの状態が示されています。  $p$  は確率です。

⑤「第Ⅰ種の誤り(type I error)」とは、帰無仮説( $H_0$ )を棄却する、すなわち有意水準  $\alpha$  で有意差や有意な関係ありと判断したが、本当は、有意差がない、あるいは無関係だった場合です。この確率は  $\alpha$  で表されます。「第Ⅱ種の

誤り(type II error)」

とは、有意差や有意な関係は無いと判断し

たが、本当は有意差や有意差や有意な関係があった場合です。この確率を  $\beta$  で表しま

		本当の事態	
		H0は真 本当は差は無かった、あるいは無関係だった	H0は偽 本当は有意差があった、あるいは有意な関係があった
研究者の判断	H0を棄却する 有意水準 $\alpha$ で有意差や有意な関係あり、と判断	第Ⅰ種の誤り $p = \alpha$	正しい棄却 $p = 1 - \beta$
	H0を棄却しない 有意差や有意な関係なし、と判断	正しい採択 $p = 1 - \alpha$	第Ⅱ種の誤り $p = \beta$

す。正しい採択は左下のセルの  $1 - \alpha$  の確率、正しい棄却は右上の  $1 - \beta$  の確率になります。図内の  $\alpha$ 、 $1 - \alpha$ 、 $\beta$ 、 $1 - \beta$  の部分と表内とを対応づけてみてください。

### 3 検定力：検定力を大きくするには

⑥ 検定を実施するにあたっては、有意差や有意な関係があったとき、本当に有意差や関係があったと判断できるのが望ましい、すなわち「正しい棄却： $1 - \beta$ 」をどのように大きくするかが重要になります。この  $1 - \beta$  の大きさを「検定力」あるいは「検出力」といいます。山田・杉澤・村井(2008)によりますと、たとえば、検定力が 0.3 である場合、もし実際に帰無仮説が誤っていたとしても、有意な結果が得られる確率は 0.3 とかなり低くなり、一方、検定力が 0.9 である場合、実際に帰無仮説が誤っていればほぼ確実に (0.9 という高い確率で) 有意な結果が得られるということを示しているのです。別の言い方をすれば、検定力とは、帰無仮説が本当に間違っているときに、きちんと棄却する確率であり、たとえば、母平均が同じでなかった場合、「同じでない」ときちんと指摘できること、と理解してください。

⑦では、どうすれば検定力を大きくできる、すなわち  $1 - \beta$  の値を大きくできるのでしょうか(図参照)。

第一に、有意水準  $\alpha$  の値を大きくすれば(縦の破線を左の方へ移動すれば)  $1 - \beta$  のエリアすなわち検定力は大きくなります。しかし、 $\alpha$  を大きくすると、帰無仮説を棄却する場合が多くなってしまい、第Ⅰ種の誤りをおかす確率がどんどん大きくなるという別の問題が生じます。

第二に、標本の大きさ ( $n$ ) を大きくすることです。標本の大きさが大きくなれば、標準誤差 ( $\sigma / \sqrt{n}$ ) が小さくなります。図では、2つの分布の広がり具合が小さくなるので、 $\mu_1$  と  $\mu_2$  の離れ具合が同じでも、2つの分布の重なり部分が小さくなり、結果として検定力 ( $1 - \beta$ ) が大きくなります。標本の大きさが重要な影響を持っているのです。

### 4 効果の大きさ (効果量: $ES$ )

⑧ 第三は、たとえば最初の図で、 $H_0$  と  $H_1$  の2つの標本平均の差が大きくなる、すなわち分布が離れることで、検定力が大きくなるはずですが。この2つの標本平均の差について、次の式で求めた値を、効果の大きさ(効果量)(effect size :  $ES$ )とといいます(ただし、式は平均値の差の検定の場合です)。

「効果の大きさ( $ES$ ) = (2年生の標本平均 - 1年生の標本平均) / 群内標準偏差の推定値」

式から、2つの平均が離れていればいるほど効果の大きさ  $ES$  の値は大きくなるのです。この、効果量( $ES$ )

とは、山田他(2008,p.358)によると、「例えば相関係数の検定であれば母相関係数の大きさ、平均値差の検定であれば母平均の差の大きさなど、相関がゼロ、平均値差がゼロ、という帰無仮説と、相関がある、平均値差がある、という対立仮説との差異を量的に表現したもの」です。一言でいえば、「はっきり違うかどうか」の程度を表したものと云えます。

なお、効果の大きさは、差の統計量や関係の統計量などの場合で、求め方が異なります。

⑨ このように、有意水準、検定力、サンプル数、効果の大きさ(ES)の4つは、相互に関係があることがわかります。検定力(1-β)の値は、通常0.8程度が必要とされています(Cohen,1992)。有意水準は多くの場合でα=0.05が使われています。すると、重要なのは効果の大きさ(ES)とサンプル数の関係になるのです。すなわち、効果の大きさをどの程度に設定するかによって必要サンプル数が決まり、逆に、あるサンプル数によって実験や調査を実施した場合には、その効果の大きさが得られることになるのです。詳しくは[大久保・岡田,2012][村井 2006], [豊田2009]などを参考にしてください。

⑩ Cohenによると、たとえばt検定の場合、効果の大きさ(ES)の目安は、0.2 → 「小さな効果」、0.5 → 「中くらいの効果」、0.8 → 「大きな効果」がよく使われています。

⑪ 効果の大きさの計算は、豊田(2009)などで、Rを用いた計算例を示しています。また、同書の中で計算式が示されています。大久保・岡田(2012)もわかりやすく説明されています。さらに、表1は、水本・竹内(2008)による、効果量の目安を示したものです。この論文は、効果量を理解する上でお勧めです。

ここに挙げた本、論文はいずれも大変お勧めです。ぜひ詳細に調べてみて

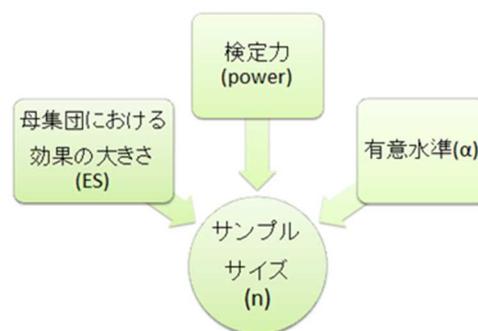


図9-3 サンプルサイズを決める要因 (村井,2006)

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「英語教育研究」31, 57-66.

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
(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	$\eta^2$	.01	.06	.14
		partial $\eta^2$	-	-	-
		$\omega^2$	.01	.09	.25
		$f$	.10	.25	.40
(6) 二元配置分散分析 (Two-way ANOVA)	多重比較	$r$	.10	.30	.50
		$\eta^2$	.01	.06	.14
		partial $\eta^2$	-	-	-
多元配置分散分析* (Multi-way ANOVA) *三元配置以上の分散分析	主効果	$\omega^2$	.01	.09	.25
		$\eta^2$	.01	.06	.14
		partial $\eta^2$	-	-	-
(6) 共分散分析 (ANCOVA)	交互作用	$\omega^2$	.01	.09	.25
		$r$	.10	.30	.50
		partial $\eta^2$	-	-	-
(7) 多変量分散分析 (MANOVA)	多重比較	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 $\eta^2$ (multivariate $R^2$ )	-	-	-
多変量共分散分析 (MANCOVA)	多変量検定	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 ( $\chi^2$ 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。 $\eta^2$ の大きさの目安は文献によっては、 $r$ を2乗した $\eta^2$ 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  $\eta^2$ の効果の大きさの基準は明確なものがない。multivariate  $\eta^2$ と multivariate partial  $\eta^2$ 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。

ください。

## 5 効果量の計算例

次に、具体的な効果量の計算方法について解説します。解説は、豊田(2009)、大久保・岡田(2012)を参照しました。詳細はこれらの本で学習してください。

### (1) 相関分析の効果量(ES)

相関分析(相関係数:  $r$ )の効果量は、 $r$ の絶対値がそのまま効果量となります。Cohen(1988)の基準にしたがうと、小さな効果量=0.1、中程度の効果量=0.3、大きな効果量=0.5 という目安です。

### (2) 単回帰分析の効果量(ES)

単回帰分析では、回帰の分散分析で用いられる  $F$  の式(右側) 
$$F = \frac{n-p-1}{p} \times \frac{R^2}{1-R^2}$$
 について、式内の右側の  $R^2/(1-R^2)$  の部分が効果量(ES)になります。 $R^2$  は、決定係数であり、 $r$  の2乗です。 $x$  と  $y$  の2つの変数の  $r$  を求め、それをもとに計算することができます。効果量の目安は表1を参照してください。

### (3) $t$ 検定の効果量(ES)

#### 1) 対応なしの $t$ 検定の場合

$t$  検定の効果量を求めるとき、標本分散(分散の分母が  $N$ )と不偏分散(分散の分母が  $N-1$ )によって名称が異なります。前者をCohenの  $d$ 、後者をHedgesの  $g$  といいます。式は、
$$ES = \frac{\text{mean}_1 - \text{mean}_2}{S}$$

です。 $\text{mean}_1$  と  $\text{mean}_2$  は2群の平均値です。 $S$  は、Cohenの  $d$  では、
$$S = \sqrt{\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2}}$$
 (ただし、 $S_1, S_2$

は分母が  $N$  の標本分散)、Hedgesの  $g$  では、
$$S = \sqrt{\frac{(N_1 - 1) S_1^2 + (N_2 - 1) S_2^2}{N_1 + N_2 - 2}}$$
 (ただし、 $S_1, S_2$  は分母が  $N-1$  の不偏分散)

です。 $N_1, N_2$  は2群の標本の大きさです。なお、 $t$  値を計算してある場合、たとえば、Hedgesの  $g$  なら

$$g = |t| \times \sqrt{\frac{N_1 + N_2}{N_1 \times N_2}}$$
 で計算できます。たとえば、1頁の例で計算しますと、 $g = |1.992| \times$

$\text{sqrt}((1000+1100)/(1000 \times 1100))$  で、結果は 0.087 となります。3頁の表1と照合しますと、とても小さい値であることがわかります。

#### 2) 対応あり $t$ 検定の場合

式は、
$$ES = |t| \times \sqrt{\frac{1}{N}}$$
 で、 $N$  は標本の大きさです(対応ありですから2群の  $N$  は同じ)。

### (4) $\chi^2$ 検定(独立性の検定)の効果量(ES)

まず、 $2 \times 2$  のクロス集計表から効果量は、四分点相関係数( $\phi$  係数: ファイ)になります。 $2 \times 2$  以上のカテゴリー数の場合は、Cramer(クラメール)の  $V$  です。計算は、他書を参照してください。

### (5) $\chi^2$ 検定(比率の差の検定・適合度検定)の効果量(ES)

豊田(2009,p.91)で紹介されているのは、標本効果量として、
$$\sqrt{\frac{\chi^2}{\text{総度数}}}$$
 を示しています。

### (6) 1 要因分散分析の効果量(ES)

要因が実験参加者間要因の場合、効果量 ( $\eta^2$ : イータ 2 乗) =  $SS_A / SS_T$  がよく使われています。 $SS_A$ は要因の平方和、 $SS_T$ は全平方和です。いずれも分散分析表に示されていますので、その値を代入。

要因が実験参加者内要因の場合、効果量 ( $\eta^2_p$ : 偏イータ 2 乗) =  $SS_A / (SS_A + SS_E)$  です。 $SS_E$  は誤差の平方和です。添字の  $p$  は *partial*(偏)の略です。

### (7) 2 要因分散分析の効果量 (ES)

2 要因の場合も、基本的には 1 要因の場合と同様の式で計算することができます。ただし、母効果量、交互作用の効果量、一般化効果量などにつきましては、大久保・岡田(2012)や豊田(2009)で解説されています。

## 6 事前の検定力分析・事後の検定力分析(豊田, 2009)

### (1) 内容

図 9-3 より、効果量(ES)、検定力( $1-\beta$ )、サンプルサイズ(標本の大きさ) ( $n$ )、有意水準( $\alpha$ )の 4 つは相互に関係していることがわかります。ところで、研究を実施するプロセスで、データ収集に先立ち、効果量・検定力・有意水準を設定して必要な標本の大きさ( $n$ )を決定したい場合と、実施した研究について振り返るため、標本の大きさ、標本 (から得た) 効果量、有意水準をもとに、研究の検定力 (標本検定力) を知りたい場合とがあります。前者を事前の検定力分析・後者を事後の検定力分析といいます(豊田,2009 より)。

### (2) 事前の検定力分析 : R を使って計算する

たとえば、今、読解力のテストを作成し、2つの読解指導法の間でテスト得点の平均値の大小を比較したいと思います。比較は対応のない  $t$  検定です。効果量を大程度の 0.8 に、検定力は Cohen の基準にしたがって 0.8 に、有意水準( $\alpha$ )は 5%(0.05)に設定したとして、必要な標本の大きさはどのくらいでしょうか？

R で計算してみますと (標本数が等しい  $t$  検定)

```
> library(pwr)
> power.t.test(power=.8,delta=.8,sig.level=.05,type='two.sample')
```

Two-sample t test power calculation

```
      n = 25.52463  →  n の数
  delta = 0.8      →  効果量 (Cohen の d)
      sd = 1       →  SD を 1 としている
sig.level = 0.05  →  有意水準
      power = 0.8  →  検定力(1-β)
alternative = two.sided →  対応無し   対応ありなら paired   を入力
```

NOTE: n is number in \*each\* group

となります。

### (3) 事後の検定力分析 : R を使って計算する

事後の検定力分析の例では、実施した研究について、研究の検定力を確認したい場合などに使われます。いま、ある作文力テストについて、標本数を各群 20 人、標本（から得た）効果量は 0.91、 $\alpha$ （有意水準）は 0.05 であったとします。この研究の検定力（標本検定力）はどのくらいだったのでしょうか。

計算しますと、

```
> power.t.test(n=20,delta=.91,sig.level=.05,type='two.sample')
```

#### Two-sample t test power calculation

```

n = 20
delta = 0.91
sd = 1
sig.level = 0.05
power = 0.8007492    → ★
alternative = two.sided

```

NOTE: n is number in \*each\* group

となり、検定力、つまり 5%水準で有意差を得る確率 =0.8001（約 80%）でした。

（文献）

- ・Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- ・南風原朝和 (2002). 心理統計学の基礎 有斐閣アルマ
- ・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
- ・村井潤一郎 (2006). サンプルサイズに関する一考察 吉田寿夫(編著) *心理学研究法のあたらしい私たち* 第5章 Pp.114-141 誠信書房
- ・大久保街亜・岡田謙介 (2012). 伝えるための心理統計：効果量・信頼区間・検定力 勁草書房
- ・豊田秀樹(編著)(2009).検定力分析入門 ―Rで学ぶ最新データ解析― 東京図書
- ・山田剛史・杉澤武俊・村井潤一郎 (2008). Rによるやさしい統計学 オーム社