

## 1. 検定力とは

- 統計的検定を行う場合、設定した有意水準よりも  $p$  値が小さくなったときに、分析に使用したデータに「差がある」と判断される。 $p$  値はサンプルサイズが大きくなるほど反比例して小さくなる傾向があり、サンプルサイズが大きい場合データが有意になりやすくなる。サンプルサイズの影響を除外し、検定にかけた群に差があるかどうかを検証する方法として効果量を観察することでサンプル間の実質的な差がわかる。
- 「統計的検定においては、(a) サンプル・サイズ、(b) 有意水準、(c) 効果量、(d) 検定力の 4 つが、検定結果の良し悪しを決定する要素である (水本・竹内, 2010)」
- 有意水準を 5% に設定した場合、5/100 で推定に誤りがあることがある、と言える。この 5 回の検定結果に引っかかってしまった場合、実は有意差がないのに有意差があったということにして結論付けてしまう可能性がある。
- 第 1 種の過誤 (Type I error)  
5% 有意水準を  $\alpha$  としたとき、 $\alpha =$  「実際に差がないのに差がある」と判断してしまう
- 第 2 種の過誤 (Type II error)  
「実際に有意差があるのに有意差がない」と判断してしまうケース。 $\beta$  で表され、 $\beta = 0.20$  が望ましい。これは有意水準を 1% や 0.1% と低く設定した場合に現れることがある。
- 検定力 (power) とは、これらの要素を排除し、正しく有意差を検出できる確率のことで、 $\text{power} = 1 - \beta$  で定義される。
- 通常、検定力は 0.8 に設定され ( $\beta = 0.20$  とするのが望ましいことから、 $1 - 0.2 = 0.8$ )、「0.8 以下の検定力の場合には、第二種の誤りを犯す可能性が高くなる (Cohen, 1992)」とされている。

## 2. 検定力を検出する意義

- 検定力と有意水準はそれぞれ  $\text{power} = 0.80$ ,  $p = 0.05$  と設定されるため、実験計画の段階でサンプルサイズと効果量を設定しなければならない。
- サンプルサイズが小さすぎた場合、検定力が下がってしまう場合があり、逆にサンプルサイズが大きすぎる場合には第 1 種の過誤が発生する可能性が高くなる。
- このことから、実験の前段階で、適切なサンプルサイズを把握していることが重要になる。そこで検定力分析を行い、サンプルサイズを実験前に決めることが重要になる。これを事前分析 (A priori) という。
- 一方で、検定力を実験の後に分析することを事後分析 (post hoc) という。
- 事前分析の場合、先行研究で得られる情報 (有意水準、検定力、効果量) から検定力を計算することができる。効果量を報告していない文献が多いが、効果量は <https://e-learning-service.net/jlta.ac/mod/resource/view.php?id=232> の excel ファイルから計算できる。
- 事後分析の場合は、有意水準、検定力、効果量の値から算出することができる。

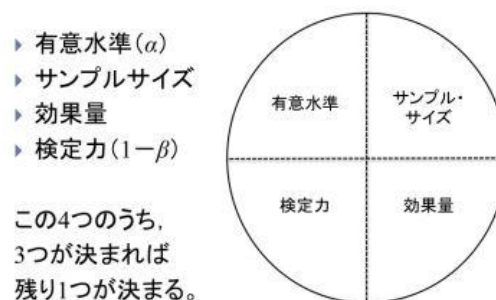


図5 統計的検定における4つの要素

### 3. 検定力分析のツール

- G\*Power 3 (<http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>)
- R pwr パッケージ (<http://cran.r-project.org/web/packages/pwr/index.html>)
- EZR (<http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>)

### 4. 検定力分析のできる統計手法

- $t$ 検定 (対応あり、なし)
- 一元配置分散分析 (対応あり、なし)
- 二元配置分散分析
- 共分散分析
- 多変量分散分析
- カイ 2 乗検定
- ノンパラメトリック検定  
(ノンパラの場合対応するパラメトリック検定の手法を用いればよい。)
- 相関分析
- 回帰分析 (単回帰・重回帰)

### 5. 事前分析 (対応あり $t$ 検定の場合)

#### ■ 事前分析の例

短期海外留学の効果を TOEIC の得点の伸びを見ることで推定しようとした。統計的に正しい分析を行うために検定力分析を行う。

- この場合、サンプルサイズを求めたいので、(1)有意確率、(2)効果量、(3)検定力を入力して算出する。
  - (1) 有意確率は単純に 5%有意水準か 1%有意水準かを入力する。
  - (2) 効果量については、先行研究の効果量を用いるか、期待される効果量 (小 :  $d = .20$ , 中 :  $d = .50$ , 大 :  $d = .80$ ) を代入する。不明瞭な場合は中程度の効果量と仮定する。
  - (3) 検定力は Cohen (1992)で推奨される  $\text{power} = 0.80$  とし、この値を代入する。

### 5.1. G\*Power 3.1.7 を利用した実習（事前検定）

(1) G\*Power 3 を使用し、同じ条件で分析を行う。

事前分析のため、①効果量、②有意水準、③検定力を想定する。このうち②、③は基本的に  $\alpha = 0.05$ ,  $\beta = 0.80$  として考えれば良いため、効果量を先行研究などから推定する。

(2) それぞれ適当な分析方法（今回は対応あり t 検定のため、Test family では「t.test」、Statistical test では「Means: Difference between two dependent means (matched pairs)」、Type of power analysis では「A priori: Compute required sample size – given  $\alpha$ , power, and effect size」）を選択する。

(4) Input parameters の選択画面では、Tail(s)を「two」に、Power (1 –  $\beta$  err prob)を 0.80 に修正する（デフォルトでは、Power が 0.95 になっているため注意）

(5) 検定力がわからない場合は、左側の「Determine =>」をクリックし、「From group parameters」にチェックを入れ、各群の平均値と標準偏差、群間の相関を入力するか、「From differences」に平均値の差と標準偏差の差を入力し、下部の「Calculate and transfer to main window」をクリックする。すると、Input Parameters の効果量の部分に効果量が自動で入力される。そのうえで右下の「Calculate」をクリックすると分析が開始される。今回は、平均値の差を 50、標準偏差の差を 75 と仮定して分析を行う。

(6) 今回は平均と標準偏差の差から効果量を推定したが、この結果、効果量は 0.667 となった。

図 7. G\* Power 3 の出力結果（事前検定）

(7) 分析の結果、Total sample size が 20 のとき、検定力が 0.807 となり、十分な検定力を持つサンプルサイズは 20 以上であることがわかる。なお、画面上部の「Protocol of power analysis」をクリックすると、テキストファイルで分析の結果が提示される。

### 5.2. G\*Power 3.1.7 を利用した実習（事後検定）

- 事後検定の場合、その時点で(1)サンプルサイズ、(2)有意水準、(3)効果量の 3 要素がわかっているため、これらからそのデータの検定力が水準に達しているかどうかの分析を行う。
- TOEIC スコアの伸びを比較するために t 検定を行った。その結果、 $t(49) = -3.495$ ,  $p = .001$ ,  $d = 0.37$  であった。この結果をもとに、事後検定を行う。
- その結果、検定力=0.929 と高かったことがわかった。

### 5.3. EZR を使用した実習（事前検定）

- EZR (<http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>) を利用して事前分析を行う。  
このソフトウェアは R をカスタマイズしたもので、パッケージ化されているため、CRAN からインストール可能。R パッケージインストーラーからもインストール可能だが、複数のパッケージを手動でインストールする必要があるため、上記 URL からダウンロードすることを推奨する。
- (1) EZR では入力画面で「2 群間の平均値の差」と「2 群共通の標準偏差 (SD)」の入力が求められるので、これらの情報があれば良い。
- (2) これらの情報が準備できたら、図 8 の手順に従い、「統計解析」→「必要サンプルサイズの計算」→「対応のある群の平均値の比較のためのサンプルサイズの計算」とたどり、図 9 の入力画面に「2 群間の平均値の差」と「2 群共通の標準偏差 (SD)」の情報を入力する。どちらの情報も、t 検定を行った上

で結果に出力される情報なので、そこから M と SD を入力すれば良い。なお、「 $\alpha$ エラー」と「検出力（検定力）」はそれぞれ  $\alpha = 0.05$ , power = 0.80 と設定する。

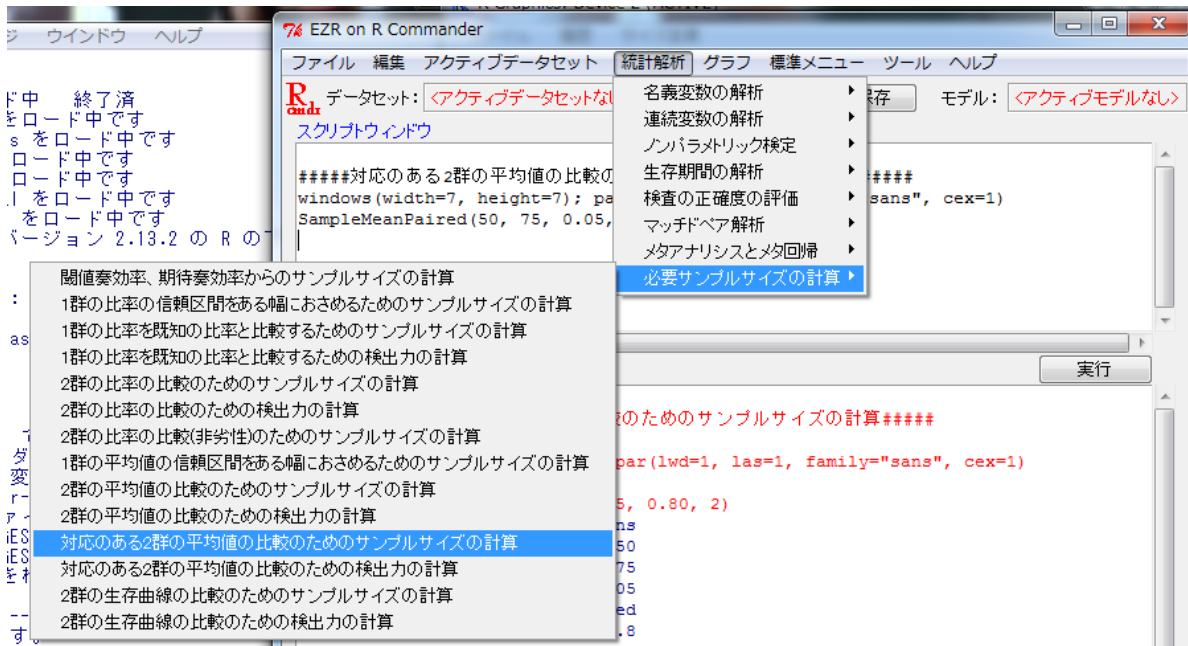


図 2. 事前検定の選択画面

(3) G\*Power3 の際と同様の数値を入力し、分析を行う。

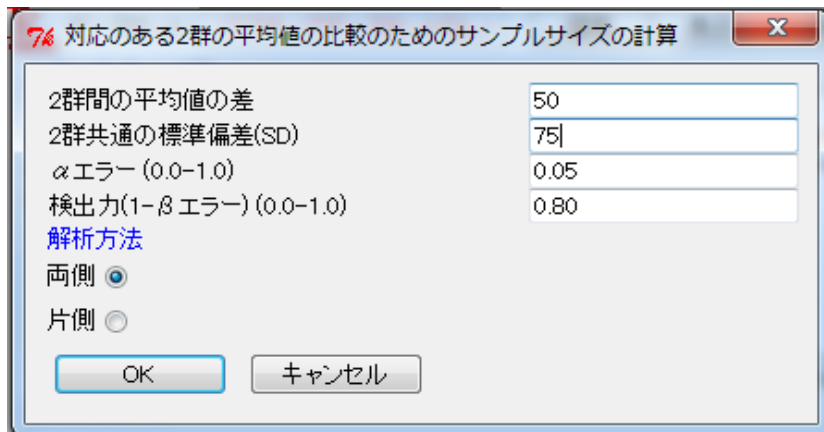


図 3. 入力画面

(4) これで OK を押すと、以下のように出力される。

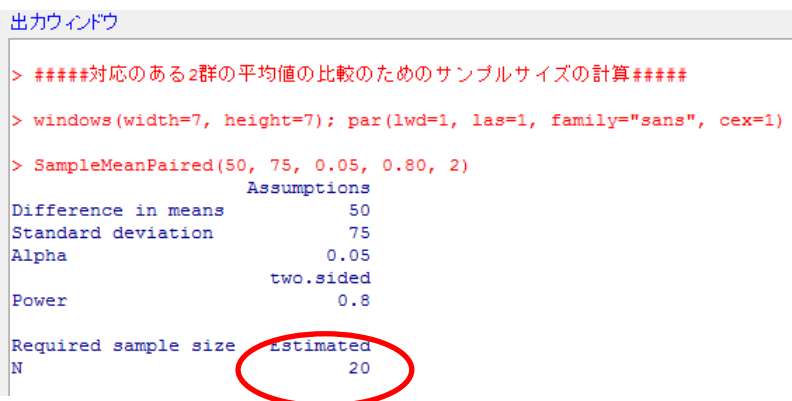


図 4. 出力画面

赤丸で示された部分に推定値が現れる。この場合、36以上のサンプルがあれば、検定力があるとみなすことができる。

- (5) EZR でこの出力を行うと、同時に図 4 が現れる。これは検定力とサンプルサイズの関係を示した図である。

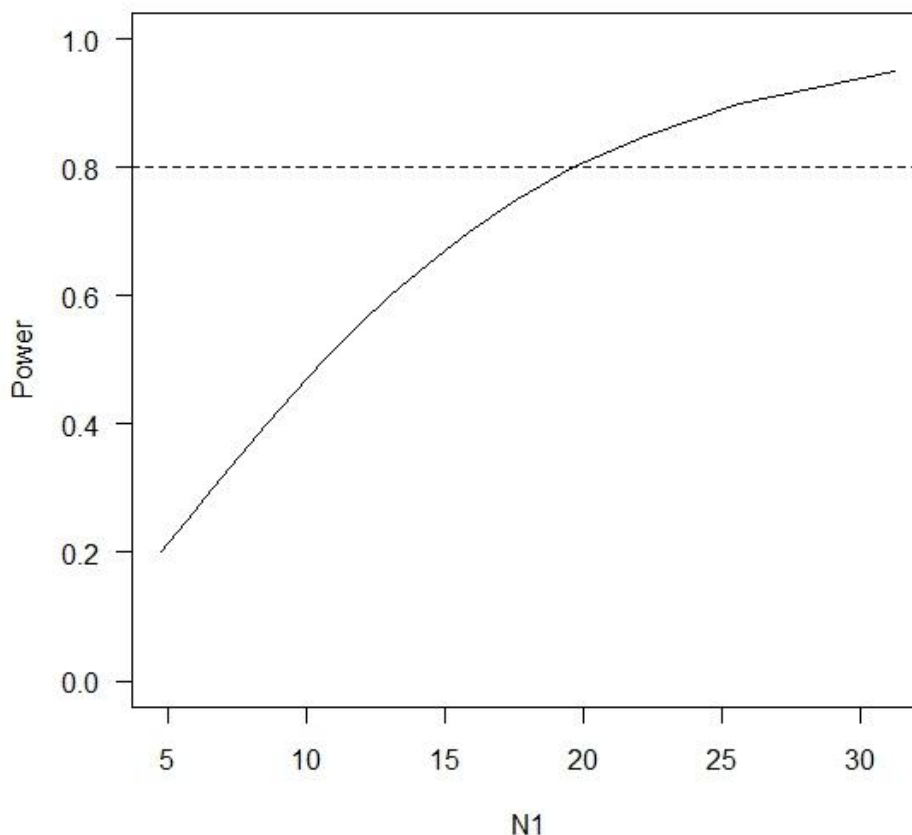


図 5. サンプルサイズと検定力の関係

## 6. EZR を使用した実習（事後検定）

- 事後検定の場合、その時点で(1)サンプルサイズ、(2)有意水準、(3)効果量の 3 要素がわかっているので、これらからそのデータの検定力が水準に達しているかどうかの分析を行う。
- TOEIC スコアの伸びを比較するために t 検定を行った。その結果、 $t(49) = -3.495$ ,  $p = .001$ ,  $d = 0.37$  であった。この結果をもとに、事後検定を行う。

- (1) 「統計解析」 → 「必要サンプルサイズの計算」 → 「対応のある 2 群の平均値の比較のための検出力の計算」を選択（図 6）

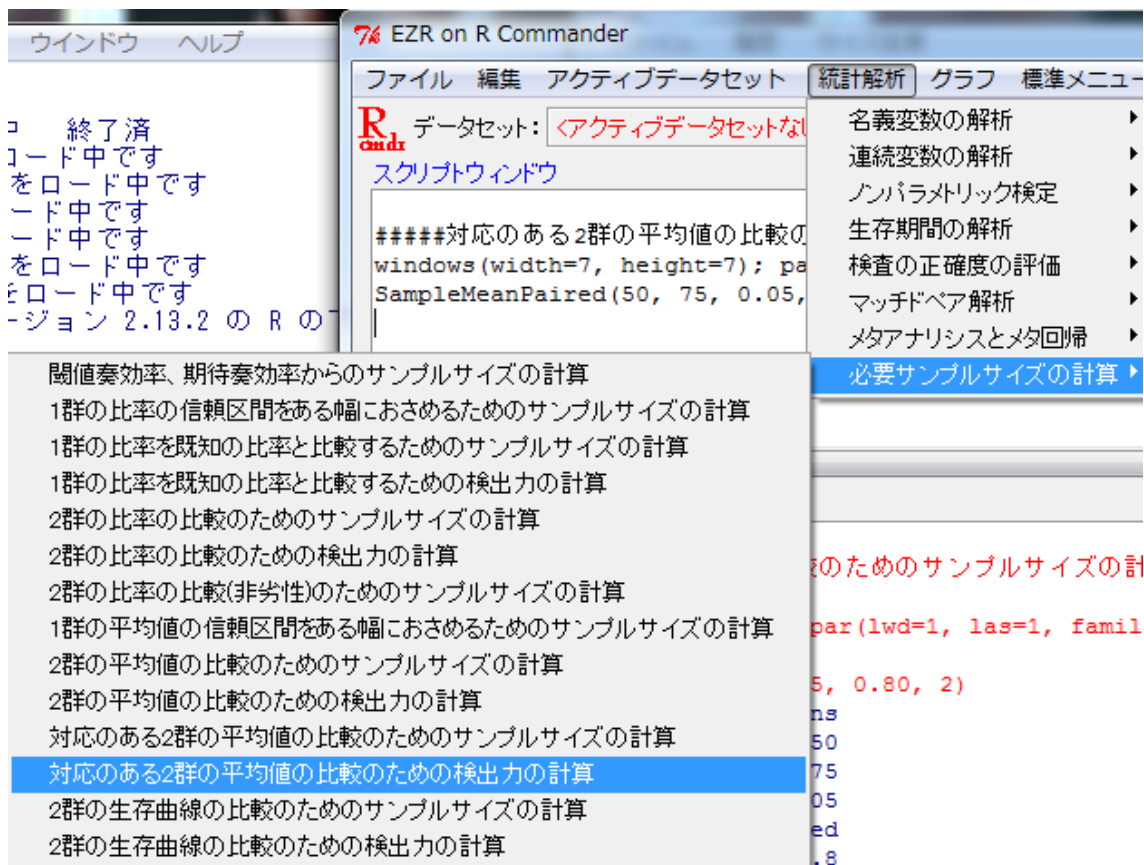


図 6. 事後検定の選択画面

(2) t 検定で得られた結果をそれぞれ図 6 のように当てはめる。この際、サンプルサイズ =  $df + 1$  であることに注意。

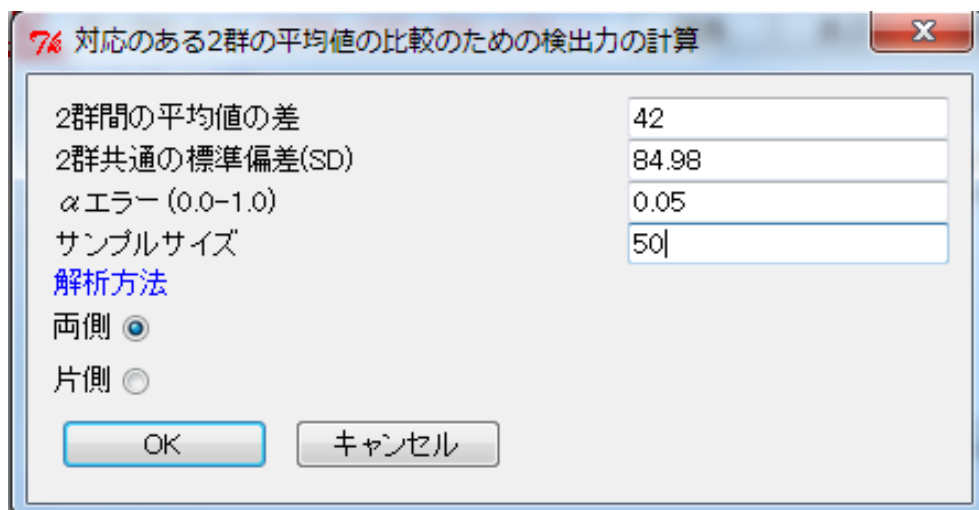


図 7. 事後検定の入力画面

(3) すると、図 8 にあるように結果が算出される。

```

> #####対応のある2群の平均値の比較のための検出力の計算#####
> windows(width=7, height=7); par(lwd=1, las=1, family="sans", cex=1)
> PowerMeanPaired(42, 84.98, 0.05, 50, 2)
      Assumptions
Difference in means      42
Standard deviation      84.98
Alpha                   0.05
      two.sided
Sample size              50

      Estimated
Power                   0.929

```

図 8. 事後検定の分析結果

この結果から、検定力は 0.929 で高い検定力があるとみなすことができる。

- (3) また、同時に出力される図から、このサンプルの場合、80 弱のサンプルサイズがあれば、検定力=0.80 に近づいたということがわかる (図 9)。

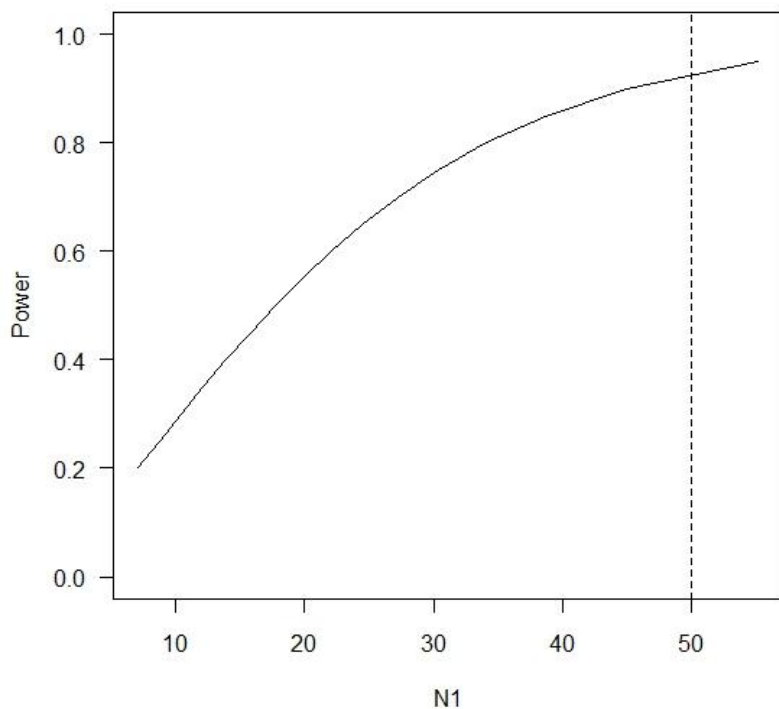


図 9. サンプルサイズと検定力の関係 (事後検定)

- 次回以降の予定
  - 他の分析手法の検定力の求め方について

【気になるひとこと】

水本・竹内(2010)

通常、どの分野のジャーナルでも、有意な結果 ( $p < .05$ ) が出た研究論文を掲載する傾向があるため、 $p$  値が統計的検定においてもっとも重要な指標であると考えてしまう。