

## intro - PSS 分析の概要 【評価版】

PSS (power and sample-size) 分析は統計学的な研究を企画する上で大変重要な役割を担っています。それは研究目的を達成できる確度を高める上で最適な研究資源の配分について情報をもたらします。

1. PSS 分析
2. 仮説検定
3. PSS 分析の構成要素
  - 3.1 研究デザイン
  - 3.2 統計的手法
  - 3.3 有意水準
  - 3.4 検定力
  - 3.5 臨床的に意味のある差と効果量
  - 3.6 標本サイズ
  - 3.7 片側検定、両側検定
  - 3.8 ドロップアウト
4. 生存時間データ
5. 感度解析
6. PSS 分析の用例

## 1. PSS 分析

PSS (power and sample-size) 分析は統計学的な研究を企画する上でキーとなる要素です。研究目的を達成できる確度を高めるためには研究資源をどのように配分するのが最適であるかについて、PSS 分析は情報をもたらします。

- 与えられた研究目的を達成するためにはどれだけ多くの被験者が必要か？  
被験者の数が少なすぎた場合には重要な効果を検出できる確率が低いものとなります。逆に被験者の数が多すぎた場合には時間と資源の浪費を招きます。
- 与えられた資源のもとでは研究目的が達成できる確率はどの程度か？
- 与えられた資源のもとで検出できる最小の効果はどの程度か？

PSS 分析はこれらすべての問いに答えるための情報をもたらします。

PSS 分析と言った場合、後向きのもので（すでに実行された研究について分析するもの）もありますが、Stata では前向きの分析（これから行おうとする研究についての分析）機能を問題とします。

仮説検定のような統計的推論手法が研究目的を評価する上で用いられます。[PSS] マニュアルにおいては、関心の対象となる研究目標を評価する上で仮説検定を用いるタイプの研究を分析対象とします。用いられる手法には平均値や分散、比率、相関等に関する 1 標本検定や 2 標本検定が含まれます。これらの手法の詳細については [PSS] power をご参照ください。

PSS 分析の要素技術について議論を始めるに先立ち、仮説検定の基本事項について復習しておくことにします。

## 2. 仮説検定

評価版では割愛しています。

## 3. PSS 分析の構成要素

PSS 分析の目標とするところは、選択された統計的手法が関心の対象である特定の効果を検出できるだけの検定力を持つよう、研究の企画立案を支援することです。例えば PSS 分析は、ある選ばれた統計的検定があらかじめ規定された有意水準において指定された大きさの効果を検出できるだけの検定力を有するためにはどれだけの標本サイズが必要かを決定するためによく用いられます。ここで言う“効果を検出する”という表現は一般的に収集されたデータが対立仮説を支持することを意味します。例えば効果を検出するということは

- 2 群間の平均値が異なったものである
- 疾病の確率とリスク因子の間には関連性がある
- 2 つの測定値間には何らかの相関がある

といったことに該当するわけです。

PSS 分析のゴールはいくつかの方法によって達成できます。

- 指定された有意水準、検定力、効果量、他のパラメータから直接標本サイズを算出する
- 指定された有意水準を前提にした上で、ある範囲の標本サイズや効果量に対する検定力を評価する
- 特定の標本サイズの場合に妥当な検定力を伴って検出し得る効果の大きさを評価する
- 検定力や標本サイズに対する要件の他のパラメータに対する感度を評価する

PSS 分析を構成する主な構成要素には次のようなものがあります。

- 研究デザイン
- 統計的手法
- 有意水準  $\alpha$
- 検定力  $1 - \beta$
- 関心の対象である効果の大きさ、あるいは臨床的に意味のある差の大きさ（しばしば効果量 (effect

- size)  $\delta$  と表現される )
- 標本サイズ  $N$

### 3.1 研究デザイン

評価版では割愛しています。

### 3.2 統計的手法

評価版では割愛しています。

### 3.3 有意水準

評価版では割愛しています。

### 3.4 検定力

評価版では割愛しています。

### 3.5 臨床的に意味のある差と効果量

評価版では割愛しています。

### 3.6 標本サイズ

評価版では割愛しています。

### 3.7 片側検定、両側検定

評価版では割愛しています。

### 3.8 ドロップアウト

評価版では割愛しています。

## 4. 生存時間データ

英文マニュアルのセクション“*Survival data*”をご参照ください。

## 5. 感度解析

評価版では割愛しています。

## 6. PSS 分析の用例

ここでは SAT (Scholastic Assessment Test) の数学の点数に関する分析について考えます。今、新たな教育プログラムがある年度の米国平均である 514 点と比べてそれを 20 点高める効果があると言えるかどうかを検定したいものとします。研究者たちは点数の標準偏差が国平均の 117 を越すことはないと思っています。また、点数の差の検定には 1 標本  $t$  検定が使用される予定です。調査の実行に先立ち研究者たちは、有意水準 5% の両側検定を前提にして、検定力 90% で予想される差を検出するために必要な標本サイズを推定しようとしています。この目的のためには `power onemean` コマンドが使用できます。仕様については [PSS] `power onemean` をご参照ください。

- Statistics ▷ Power and sample size と操作
- PSS ダイアログ: Population parameter: Means: One sample
  - ▷ Test comparing one mean to a reference value
- `power onemean` ダイアログ: Main タブ: 図 1 参照

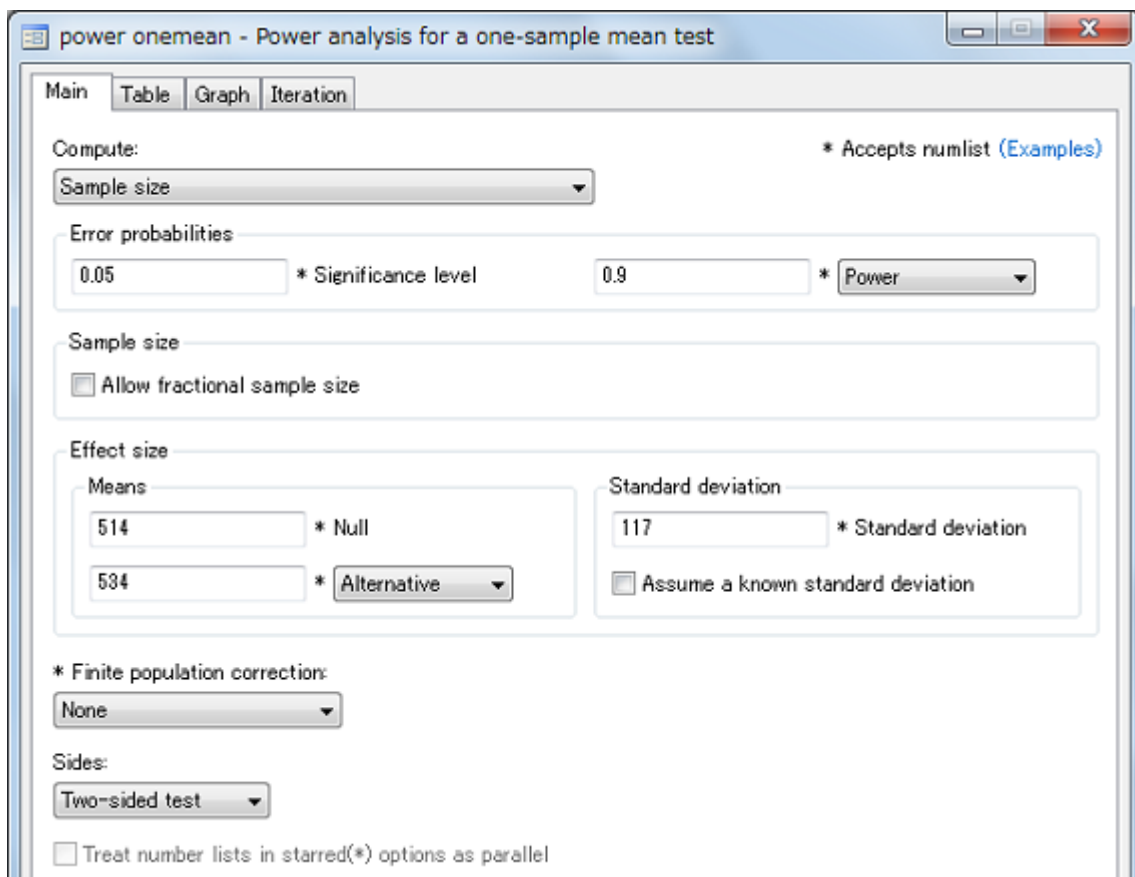


図 1 `power onemean` ダイアログ - Main タブ - 標本サイズ

```

. power onemean 514 534, power(0.9) sd(117)

Performing iteration ...

Estimated sample size for a one-sample mean test
t test
Ho: m = m0 versus Ha: m != m0

Study parameters:

      alpha =    0.0500
      power =    0.9000
      delta =    0.1709
      m0 =    514.0000
      ma =    534.0000
      sd =    117.0000

Estimated sample size:

      N =        362

```

この結果によると必要とされる標本サイズは 362 であることがわかります。

評価版では割愛しています。

graph オプションを使うと結果をグラフ化することができます。ここでは標本サイズと差の点数を変化させたときの検定力の値をプロットしてみます。対立仮説に基づく平均点としては 535 から 550 の範囲内の 4 点を (刻みは 5 点) 標本サイズとしては 200 から 300 の範囲の 11 点 (刻みは 10) を指定する形でグラフ化を行ってみます。

- power onemean ダイアログ: Main タブ: Compute: Power  
Sample size: 200(10)300  
Means: Null: 514  
Alternative: 535(5)550  
Standard deviation: 117
- Graph タブ: Graph the results: ✓

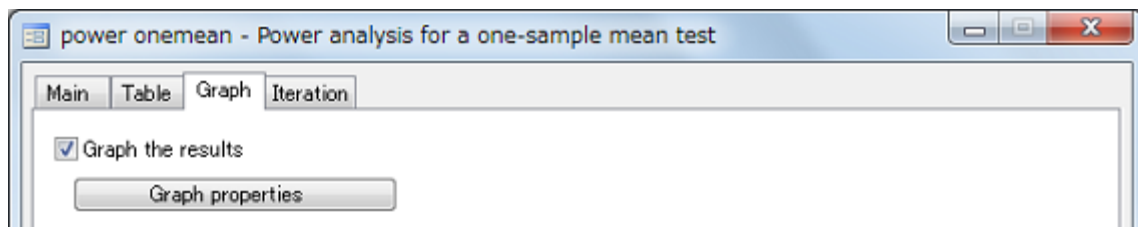
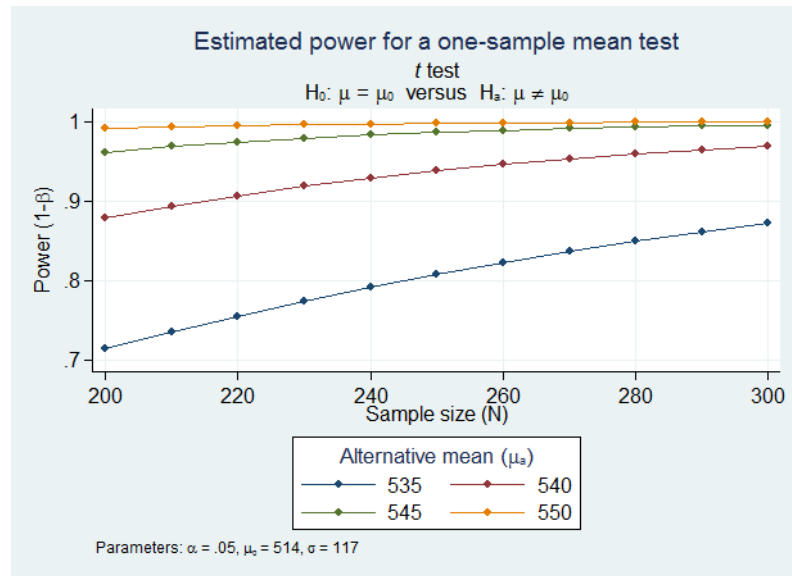


図 4 power onemean ダイアログ - Graph タブ

```

. power onemean 514 (535(5)550), n(200(10)300) sd(117) graph

```



デフォルト設定の場合、 $x$  軸には標本サイズ  $N$  が取られた形でグラフが作成されます。このため対立仮説に基づく平均  $\mu_a$  のそれぞれの指定値ごとに別個の曲線がプロットされています。検定力は標本サイズの増加と共に、あるいは  $\mu_a$  の増加と共に大きな値となることがわかります。例えば  $N = 220$ ,  $\mu_a = 535$  のときの検定力の値は概ね 75% であること、 $\mu_a = 550$  のときには検定力は概ね 1 であることが読み取れます。 $\mu_a = 535$  であっても  $N = 300$  であれば検定力は 87% となります。これらの情報に基づき、 $\mu_a$  と  $N$  の適正な組合せを決定して行くことになります。

評価版では割愛しています。

■