『経済セミナー』連載「EBPM の思考法――やってみようランダム化比較試験!」 第5回(2019年12月・2020年1月号)「ランダム化比較試験の実施工程(2)」

# 第5回連載 演習

サンプルサイズの見積もりは Stata や R といった統計解析用のソフトウェアを用いて行 うことができるが、ここでは Optimal Design というソフトウェアを用いて演習を行う。 Optimal Design は the William T. Grant Foundation の支援の下、Stephen Raudenbush ら が開発したサンプルサイズの設計に特化した専用ソフトウェアであり、無償で提供されて いる<sup>1</sup>。Optimal Design は本文の表 1 や表 2 で示したような 2 つのパラメータの関係を平易 な操作で作図することができる。

#### Optimal Design のインストール

Optimal Design は以下の公式サイトからダウンロードすることができる。やや分かりづらいが、図1の赤枠で囲った箇所をクリックすることでダウンロードページが開く。

https://sites.google.com/site/optimaldesignsoftware/home

### 図 1. Optimal Design Software の公式 HP

About	Optimal Design Software
Software	The Optimal Design for Multi-Level and Longitudinal Research is a free stand-alone software
Ags References	statistical power calculations of impact evaluation studies, particularly group-randomized trias. Click here t current version and documentation.
Sitemap	Suggested citation: Baudenbuch S. W. et al. (2011). Ontimal Design Software for Multi-level and Longitudinal Research (Version 3.01) (Software).
	Available from www.wtgrantfoundation.org.
	We also offer a consultation service.
	The work reported here has been generously funded by the William T. Grant Foundation.

ダウンロードページには Optimal Design の詳細なマニュアルも公開されている。本マニ ュアルはソフトウェアの使い方のみならず、検出力分析の非常に分かりやすいテキストに もなっている。特にクラスターランダム化比較試験における検出力分析の説明が詳しい。是 非、読者は本文と併せて本マニュアルを活用し、サンプルサイズ設計の理解を深めて欲しい。

<sup>&</sup>lt;sup>1</sup> 残念ながら Mac には対応していない。公式 HP では、「Unfortunately no, there isn't. To run on a Mac you will need a virtual machine (e.g., virtualbox.org).」とのつれない記載がある。

### Optimal Design の使い方:シンプルなランダム化比較試験の場合

必用サンプルサイズを定めるためには、効果量(介入群と統制群におけるアウトカムの平 均値の差)、アウトカムの分散、有意水準、検出力(Power)の4つの要素を決める必要が あった。本文でも説明した通り、通常は有意水準5%に設定し、介入効果として価値が認め られると判断された標準化効果量(効果量を標準偏差で除した値)を所与とした時、どの程 度のサンプルサイズがあれば80%の検出力を確保できるかを考えていく。

放課後寺子屋事業を例とした本文の表1(下記に再掲する)は、効果量を所与とした時の サンプルサイズと検出力の関係を示していた(この例では、効果量は5点/7.5点=0.67)。 表から、ランダム化比較試験によって放課後寺子屋事業の有効性を検証するためには、介入 群40人、統制群40人からなる計80人弱の生徒が必要なことが読み取れた。この表1に示 されていたサンプルサイズと検出力の関係をOptimal Designを用いて作図してみよう。

サンプルサイズ	10	20	50	80	100	200
(両群合計)						
検出力	15.4%	29.2%	63.7%	83.76%	91.0%	99.7%

表1. サンプルサイズ別の検出力

Step 1. 作成する図の種類を選択する。

脑をダブルクリックすると、図 2 のような Optimal Design の起動画面が表示される。



図 2. Optimal Design の起動画面

ランダム化比較試験といっても、非常にシンプルなものからクラスターランダム化比較 試験のような複雑な構造を持つものまで様々な種類がある。Optimal Design はそうした各 種デザインに応じて設計者が必要とする作図を行うことができる。そこで、まず初めに上 部に並ぶメニューの「Design」から、どのようなデザインにおけるサンプルサイズ設計を 行うのかを選択する。

表1が念頭に置いているデザインは、個々の生徒を対象にランダム割付を行うというものである。そこで、「Design」をクリックし、その後表示される「Person Randomized Trials」、「Single Level Trial」を順次選ぶ<sup>2</sup>。すると最後に、縦軸と横軸に何を取った図を作成するのかを選択することになるので、ここでは「Power vs. Total number of people (n)」をクリックする (図 3)。

Design Empirically Based MDES Help				
Person Randomized Trials	>	Single Level Trial	>	Power on y-axis
Cluster Randomized Trials with person-level outcomes	>	Multi-site (or blocked) Trials	>	Power vs. Total number of people (n)
Cluster Randomized Trials with cluster-level outcome (measurement of group processes)		Repeated Measures	>	Power vs. Effect size (delta)
Meta Analysis	T I			Power vs. Explained variation by covariate (R2)
	_			MDES on y-axis
				MDES vs. Total number of people (n)
				MDES vs. Power
				MDES vs. Explained variation by covariate (R2)

図 3. デザインと図の選択

すると、以下のような作図のためのサブウィンドウが表示される(図4)。

# 図 4. 描画画面

Optimal Design Plus Emprical Evidence	×
Eile Design Empirically Based MDES Help	
Single Level Trial - Power vs. N	
α δ 1 2 5 X5 5 95	
	nber of subsects ini on x-axis

<sup>&</sup>lt;sup>2</sup> Optimal Design では、アウトカムの測定単位とランダム割付の測定単位が一致するラン ダム化比較試験のことを Person Randomized Trials と呼んでいる。

Step 2. 有意水準、効果量の値を設定する。

次に、赤枠で囲ったアイコンをクリックし、有意水準と標準化効果量の値をそれぞれ入 力する。まず、■をクリックすると図5左のようなウィンドウが開くので、有意水準を入 力する。ここでは有意水準は5%(0.050)とする<sup>3</sup>。すると、サブウィンドウに図が表示さ れるが、一旦気にせずに操作を進めて欲しい。次に、■をクリックして表示されるウィン ドウ上で効果量の値を入力する(図5右)。入力欄が3つあるが、ここでは最上部の一つ のみ入力すればよい<sup>4</sup>。ここでは先の標準化効果量0.67を入力する。これでパラメータの 入力作業は終わりとなる。

図 5. 有意水準と効果量の入力

	Effect size: delta > 0 ×
Cimiliance laudi 0.0 < alaba < 1.0	delta(1) 0.67
Significance leve: 0.0 < alpha < 1.0 ×	delta(2)
alpha 0.050	delta(3)
OK	ОК

Step.3 サンプルサイズと検出力の関係を読み取る。

必要なパラメータの入力が終わると、描画ウィンドウ内に次頁のような図が描かれる (図 6)。この青の曲線がサンプルサイズと検出力の関係を表している。右上に先ほど設定 したパラメータの値(有意水準 $\alpha$ =0.05、標準化効果量 $\delta$ =0.67)が表示されているので、 誤りがないか確認して欲しい。なお、横軸のサンプルサイズ(Total number of subjects) は介入群と統制群に 1:1 で割付けた場合の合計人数を表している。

<sup>&</sup>lt;sup>3</sup> Optimal Design では有意水準のデフォルト値は 5%となっているため、有意水準を 5%と する場合は、この入力作業は省くことができる。

<sup>&</sup>lt;sup>4</sup> 実際のサンプルサイズの見積もりにおいてはアウトカムの分散の値が不確定になるため、いくつかの効果量を想定して必要サンプルサイズの比較を行う必要性に駆られることも少なくない。そのような時は、ここで最大3つの効果量を設定することで、3つのシナリオを同時に図示化することができる。読者は例えば0.2、0.5 といった値を入力して、どのような図が描かれるか試してみて欲しい。



図 6. サンプルサイズと検出力の関係

曲線上にカーソルを移動し左クリックをすると、その座標値が表示される。縦軸上の 0.8 から水平移動して曲線とぶつかる辺りで左クリックをすると必要サンプルサイズとして約 74 人が必要であることが読み取れるであろう(図 7)。Optimal Design は、この座標の読み 取りに若干の癖があり、操作しづらい点が否めない。



図 7. 必要サンプルサイズの読み取り

デフォルトでは横軸のサンプルサイズの最大値が 199 となっているが、標準化効果量が 小さい場合、このサンプルサイズでは曲線が検出力 0.8 に達する領域を表示できない。その 場合は、MMをクリックし適宜横軸の最大値を調整する。値を入力すると、図が更新される。 最後に、本文7節「検出力をあげるための工夫」で触れた分析時における説明変数の追加 について補足しておく。ランダム割付を行っていれば、分析は原則介入群と統制群のアウト カムの単純な平均値比較を行うことで平均介入効果をバイアスなく推定することができる。 しかしランダム割付を行っていたとしても、アウトカム水準に影響を与えうる要素を回帰 分析における説明変数として用いることで介入効果の推定精度を高め、結果として検出力 をあげることができる。つまり、同一の検出力を確保するために必用となるサンプルサイズ を小さくすることが可能となる。では説明変数の影響を考慮して、サンプルサイズを見積も るにはどうしたらよいだろうか?

先の Step2 では有意水準と標準化効果量の 2 要素のみを指定したが、この時、更に図を クリックすることで説明変数の説明力を入力することができる。本文中でも述べたが、この ような共変量の最も有力な候補はアウトカム変数のベースライン値である。もしこうした 変数がサンプルサイズ設計時に利用可能な場合は、図を指定するとよいだろう。しかし、説 明力がどの程度かを正確に見積もることは容易な作業ではない。そこで、説明変数の影響を 加味する際は、脚注 4 でも触れたように、いくつかの値を設定し説明力によってどの程度 必要サンプルサイズが変動するか確認することで慎重な判断を行って欲しい(図 8)。





以上で、Optimal Design の基本操作は終了となる。上記の Step1 では「Power vs. Total number of people (n)」を選択したが、ここで「MDES vs. Total number of people (n)」を 選択すれば、本文表 2 のようなサンプルサイズとその下で検出可能な最小効果量の関係を 描くこともできる。図 9 は縦軸を標準化効果量、横軸をサンプルサイズとしたものである。 作図にあたっては、先とは違い有意水準と検出力を設定することになる。 **ふ**の代わりに**ご**と いうアイコンが表示されるので、ここから 80%といった検出力を入力して欲しい。なお、 本文の表 2 ではアウトカムであるテストの得点をそのまま用いていたが、この図では標準 化効果量が縦軸に取られていることに注意すること。(標準化)効果量が大きくなるほど、 必要サンプルサイズが小さくなっていくことが見て取れる。



図 9. 効果量とサンプルサイズ

これ以外にも、Optimal Design を用いることで2軸に様々なパラメータをとった作図ができる。本文でも説明したが、目的に応じて適切な作図を行いサンプルサイズの検討を行って欲しい。

演習1:シンプルなランダム化比較試験におけるサンプルサイズの見積もり

Optimal Design を用いて、最もシンプルなランダム化比較試験におけるサンプルサイズの見積もりを行ってみよう。

発展途上国の製造業分野における中小零細企業は、帳簿すらつけておらず、作業効率も悪いことが少なくない。サブサハラアフリカに位置する A 国では、こうした経営状況を改善するために、企業経営者に対してカイゼン研修を提供し利益向上をはかる事業を全国規模で展開することを検討している。事業開始前に十分な数の中小零細企業に対して現状把握のための予備調査を行った結果、ここ数年は毎年、年間の総売上額が平均 25,000 ドル、標準偏差が 10,000 ドルであることが分かっている。

A 国政府は、この事業によって中小零細企業の年間総売上額が 10%向上することを期待 している。本事業が全国展開に値するかを判断するために、パイロット事業を実施し、ラン ダム化比較試験によって有効性を検証する。

### 演習課題1

以下のステップに沿いながら、必要サンプルサイズを見積もってみよう。なお、各ステ ップでは解答に到る思考過程や仮定も明示すること。

- 介入実施後の介入群の平均アウトカム(年間総売上額)は幾らと想定した らよいだろうか?
- ② もし介入がなかったとしたら平均アウトカムは幾らと想定したらよいだろうか?
- ③ ①と②から標準化効果量の値はいくつになるか?
- ④ OD を用いて、必要サンプルサイズを見積もってみよう。

#### 演習課題2

10%向上としていた期待効果を 20%向上に倍増させた場合、必要サンプルサイズはいくつになるだろうか(何倍になるか)?

## 演習課題3

予備調査の標準偏差の値に誤りがあることが発覚した。正しい値は 10,000 ドルではな く 13,000 ドルである。この時、必要サンプルサイズはいくつになるか(期待効果は 10% 向上とする)?

### 演習課題4

13,000 ドルと推定された標準偏差は±5000 ドルの誤差を見込んだ方がよいことが分かった。この場合、サンプルサイズの見積もりはどのように行うことが適切だろうか?