

## 第1章：データ入門

### OpenIntro Statistics 第4版（日本語版）

原著スライド：Mine Çetinkaya-Rundel（OpenIntro）  
CC BY-SA ライセンスのもと使用・翻訳。  
一部の画像はフェアユース（教育目的）に基づき使用。

## 慢性疲労症候群の治療

- **目的**：慢性疲労症候群に対する認知行動療法の効果を評価する。
- **参加者候補**：かかりつけ医や専門医から慢性疲労症候群専門クリニックへ紹介された 142 名の患者。
- **実際の参加者**：紹介された 142 名のうち、実際に研究へ参加したのは 60 名のみ。診断基準を満たさない者、他の健康上の問題を抱える者、参加を拒否した者は除外された。

Deale et. al. *Cognitive behavior therapy for chronic fatigue syndrome: A randomized controlled trial.* The American Journal of Psychiatry 154.3 (1997).

## 研究デザイン

- 患者は治療群と対照群にランダムに割り当てられ、各群 30 名：
  - **治療群**：認知行動療法——協調的・教育的なアプローチで行動面を重視。症状を悪化させることなく活動量を段階的・安全に増やす方法を患者に指導した。
  - **対照群**：リラクゼーション——活動量を増やす指導は行わず、筋弛緩法・視覚化・即時リラクゼーション技法を指導した。

## 結果

下の表は、6か月後の追跡調査における良好な転帰を示した患者の分布を示す。なお、7名（治療群3名・対照群4名）が途中で脱落した。

		良好な転帰		合計
		あり	なし	
群	治療群	19	8	27
	対照群	5	21	26
	合計	24	29	53

- 治療群における良好な転帰の割合：

$$19/27 \approx 0.70 \rightarrow 70\%$$

- 対照群における良好な転帰の割合：

$$5/26 \approx 0.19 \rightarrow 19\%$$

## 結果の解釈

データはグループ間の「真の」差を示しているか？

- コインを 100 回投げることを考えてみよう。表が出る確率は 50%だが、ちょうど 50 回表が出るとは限らない。このようなばらつきは、ほぼあらゆるデータ生成プロセスで生じる。
- 2 グループ間で観測された差 ( $70 - 19 = 51\%$ ) は真の差かもしれないし、自然なばらつきによるものかもしれない。
- 差が非常に大きいため、真の差である可能性が高いと考えられる。
- この差が偶然によるものでないと判断できるかどうかは、統計的手法を用いて判断する必要がある。

## 結果の一般化

この研究の結果は、慢性疲労症候群の患者全体に一般化できるか？

これらの患者は特定の特性を持ち、自ら参加を希望した者たちであるため、慢性疲労症候群の患者全体を代表しているとは言い難い。ただちに全患者に一般化することはできないが、この最初の研究は有望な結果を示している。特定の特性を持つ患者には効果があることが示されており、他の患者にも少なからず効果が期待できる。

## クラスでのアンケート

入門統計学の受講生を対象にアンケートが実施された。以下はアンケートの質問の一部と、それに対応する変数名である：

- gender：あなたの性別は？
- intro\_extra：自分を内向的・外向的、どちらだと思えますか？
- sleep：平均して1日何時間寝ますか？
- bedtime：ふだん何時頃就寝しますか？
- countries：今までに何カ国訪れたことがありますか？
- dread：この授業への「うんざり度」を1～5で表すと？

## データ行列

統計学の受講生から様々な変数についてデータを収集した：

変数

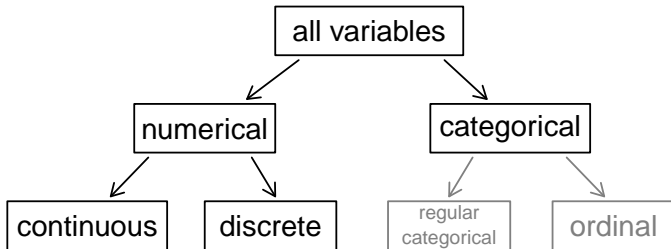


学生	gender	intro_extra	...	dread
1	male	extravert	...	3
2	female	extravert	...	2
3	female	introvert	...	4
4	female	extravert	...	2
⋮	⋮	⋮	⋮	⋮
86	male	extravert	...	3



観測値

# 変数の種類



## 変数の種類 (続き)

	gender	sleep	bedtime	countries	dread
1	male	5	12-2	13	3
2	female	7	10-12	7	2
3	female	5.5	12-2	1	4
4	female	7	12-2		2
5	female	3	12-2	1	3
6	female	3	12-2	9	4

- gender : カテゴリ変数
- sleep : 数値変数・連続
- bedtime : カテゴリ変数・順序
- countries : 数値変数・離散
- dread : カテゴリ変数・順序尺度 (数値変数として扱うことも可)

## 練習問題

電話の市外局番はどのタイプの変数か？

- (a) 数値変数・連続
- (b) 数値変数・離散
- (c) カテゴリ変数
- (d) カテゴリ変数・順序

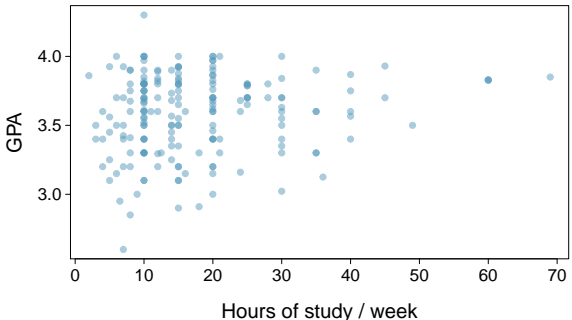
## 練習問題

電話の市外局番はどのタイプの変数か？

- (a) 数値変数・連続
- (b) 数値変数・離散
- (c) カテゴリ変数
- (d) カテゴリ変数・順序

# 変数間の関係

GPA と週の学習時間に関係はあるか？



データ点に何か異常はないか？

GPA>4.0の学生が1名おり、これはおそらくデータエラーである。

## 説明変数と応答変数

- 2つの変数の組において説明変数を特定するには、どちらがもう一方に影響を与えると考えられるかを判断する：

説明変数  $\xrightarrow{\text{影響する可能性}}$  応答変数

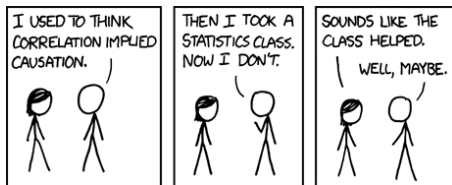
- 変数を説明変数・応答変数と名付けても、2変数間の関係が実際に因果関係であることを保証するわけではない。関連が確認された場合でも同様である。これらのラベルは、どちらがもう一方に影響を与えると予想されるかを追跡するためだけに使う。

## データ収集の2つの主要な方法

- **観察研究**：データが生じる過程に直接介入しない方法でデータを収集する（例：調査）。
  - 変数間に自然に生じる関連の証拠を得ることはできるが、それだけで因果関係を示すことはできない。
- **実験**：研究者が被験者をさまざまな処置にランダムに割り当て、説明変数と応答変数の因果関係を確立する。

## 相関と因果関係

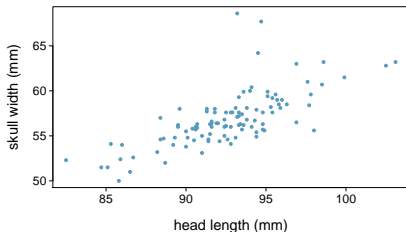
- 2つの変数は何らかの関係を示すとき、それらを**関連した変数**と呼ぶ。
  - 関連した変数は**従属変数**とも呼ばれる。
- 2つの変数が関連していない、すなわち明らかなつながりがない場合、それらは**独立**であると言う。
- 一般に、**相関は因果関係を意味しない**。因果関係はランダム化実験からのみ推論できる。



<http://xkcd.com/552/>

## 練習問題

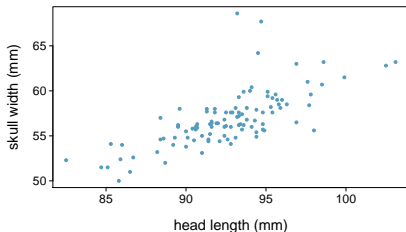
右の散布図に基づき、フクロモモンガの頭の長さ（mm）と頭蓋骨の幅（mm）について正しい記述はどれか？



- (a) 頭の長さ（mm）と頭蓋骨の幅（mm）の関係はなく、独立している。
- (b) 頭の長さ（mm）と頭蓋骨の幅（mm）は正の関連がある。
- (c) 頭蓋骨の幅（mm）と頭の長さ（mm）は負の関連がある。
- (d) 頭が長いほど頭蓋骨が広くなる（因果関係）。
- (e) 頭蓋骨が広いほど頭が長くなる（因果関係）。

## 練習問題

右の散布図に基づき、フクロモモンガの頭の長さ  
と頭蓋骨の幅について正しい記述はどれか？



- (a) 頭の長さ  
と頭蓋骨の幅に関係はなく、独立している。
- (b) 頭の長さ  
と頭蓋骨の幅は正の  
関連がある。
- (c) 頭蓋骨の幅  
と頭の長さは負の  
関連がある。
- (d) 頭が長いほど  
頭蓋骨が広くなる  
(因果関係)。
- (e) 頭蓋骨が広いほど  
頭が長くなる  
(因果関係)。

## 母集団と標本

PHYS ED | AUGUST 29, 2012, 12:01 AM | 21 Comments

### Finding Your Ideal Running Form

By GRETCHEN REYNOLDS



David De Lossy/Getty Images

<http://well.blogs.nytimes.com/2012/08/29/>

*finding-your-ideal-running-form*

**標本：**最近ランニンググループに参加した成人女性のグループ  
**結果を一般化できる母集団：**データがランダムにサンプリングされていれば成人女性

**研究の問い：**走ることで、より効率的なランナーになれるか？

**対象母集団：**すべての人

## 逸話的証拠と喫煙研究の黎明期

- 禁煙研究は 1930～40 年代、タバコの喫煙が急増した頃に始まった。タバコの煙に敏感な喫煙者がいる一方、まったく影響を受けないように見える喫煙者もいた。
- 禁煙研究は「私の叔父は一日 3 箱吸っているが至って健康だ」というような逸話的証拠（限られた標本に基づく根拠で、母集団を代表しない可能性がある）に基づく反論に直面した。
- 「喫煙は複雑な人間行動であり、その性質上研究しにくく、個人差によって交絡されている」という結論が出された。
- やがて研究者たちがより大きな標本（喫煙者）を検討できるようになると、喫煙が健康に悪影響を及ぼすという傾向がはっきりと見えるようになった。

Brandt, *The Cigarette Century* (2009), Basic Books.

## 全数調査

- 全員を対象にして、母集団全体を「サンプリング」した方がよいのではないか？
  - これを全数調査という。
- 全数調査には以下のような問題がある：
  - 全員を調査するのは難しい：見つけにくい個人や測定しにくい個人は常に存在する。しかも、そのような見つけにくい人々は、他の人々と異なる特性を持っている可能性がある。
  - 母集団は変化し続ける。全数調査ができたとしても、母集団は常に変化するため、完全な測定は不可能である。
  - 全数調査は標本抽出より複雑になる場合がある。

# Illegal Immigrants Reluctant To Fill Out Census Form

by PETER ODOWD

March 31, 2010 4:00 AM

 from **KJZZ**



**Listen to the Story** 

Morning Edition 3 min 48 sec

[+ Playlist](#)  
[+ Download](#)

There is an effort underway to make sure Hispanics are accurately counted in the 2010 Census. Phoenix has some of the country's "hardest-to-count" districts. Some Latinos, especially illegal residents, fear that participating in the count will expose them to immigration raids or government harassment.

<http://www.npr.org/templates/story/story.php?storyId=125380052>

## 探索的分析から推測へ

- 標本抽出は自然な営みである。
- 料理中の食べ物を味見することを考えてみよう——料理全体の味を知るために、その一部（標本）を味見（検査）する。
- スープを一口味見して「塩気が足りない」と思うのが**探索的分析**である。
- そこから「スープ全体に塩が足りない」と結論づけるのが**推測**である。
- 推測が有効であるためには、味見した一口（標本）が鍋全体（母集団）を**代表している**必要がある。
  - 表面からしか取っていない一口が、塩が底に溜まっているスープを代表しているとは言えない。
  - よくかき混ぜてから味見すれば、一口はスープ全体を代表しやすくなる。

## 標本バイアス

- **無回答バイアス**：ランダムにサンプリングされた人のうちごく一部しか調査に回答しない場合、標本はもはや母集団を代表していない可能性がある。
- **自発的回答バイアス**：その問題について強い意見を持つ人が自ら進んで回答することで標本が構成される場合。このような標本も母集団を代表しない。

### Quick vote

Do you get paid sick days at your job?

- Yes       No
- What job?

**VOTE** or view results

### Quick vote

Do you get paid sick days at your job?

Read Related Articles

Yes	████████████████████	63%	20056
No	████████	21%	6816
What job?	████	15%	4885

Total votes: 31757  
This is not a scientific poll

cnn.com, 2012年1月14日

- **便宜的標本**：アクセスしやすい個人が標本に含まれやすくなる。

## 標本バイアスの例：ランドン対FDR

偏った標本が誤った結果をもたらした歴史的な例：



1936年、ランドンはFDR（フランクリン・ルーズベルト）の再選に対抗して共和党の大統領候補を目指した。





## リテラリー・ダイジェスト誌の世論調査——何が間違っていたか？

- この雑誌が調査対象としたのは、
  - 自誌の読者、
  - 自動車の登録者、そして
  - 電話の契約者であった。
- これらのグループは当時（大恐慌の時代であることを思い出してほしい）の全国平均を大きく上回る収入を持っており、共和党を支持する有権者が多く含まれていた。つまり標本は当時のアメリカの**典型的な**有権者を代表していなかった。

## 大きな標本は望ましいが……

- リテラリー・ダイジェスト誌の世論調査は 240 万人という巨大な標本に基づいていたが、標本に**バイアス**があったため、正確な予測はできなかった。
- スープの例えに戻ると：スープがよくかき混ぜられていなければ、スプーンがいくら大きくても味は正しく分からない。よくかき混ぜてあれば、小さなスプーンで十分である。

## 練習問題

ある学区が、最近2件の生徒の重傷事故を受けて、高校生の校内駐車を禁止するか検討している。最初の一步として、保護者に郵送でアンケートを行い、この方針変更に対するかどうかを尋ねた。送付した6,000通のうち、返送されたのは1,200通であった。回答のあった1,200通のうち、960通が方針変更賛成、240通が反対であった。次の記述のうち正しいものはどれか？

- I. 郵便物が保護者に届かなかった可能性がある。
  - II. 学区は方針変更について保護者から強い支持を得ている。
  - III. 高校生の保護者の多数が方針変更反対している可能性がある。
  - IV. すべての保護者に調査票を送ったのだから、調査結果にバイアスが生じる可能性は低い。
- (a) Iのみ      (b) IとII      (c) IとIII      (d) IIIとIV  
IV

## 練習問題

ある学区が、最近2件の生徒の重傷事故を受けて、高校生の校内駐車を禁止するか検討している。最初の一步として、保護者に郵送でアンケートを行い、この方針変更に対するかどうかを尋ねた。送付した6,000通のうち、返送されたのは1,200通であった。回答のあった1,200通のうち、960通が方針変更賛成、240通が反対であった。次の記述のうち正しいものはどれか？

- I. 郵便物が保護者に届かなかった可能性がある。
  - II. 学区は方針変更について保護者から強い支持を得ている。
  - III. 高校生の保護者の多数が方針変更反対している可能性がある。
  - IV. すべての保護者に調査票を送ったのだから、調査結果にバイアスが生じる可能性は低い。
- (a) Iのみ      (b) IとII      (c) IとIII      (d) IIIとIV  
IV

## 観察研究

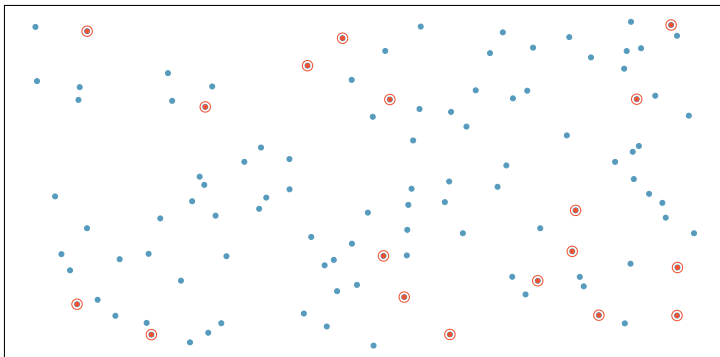
- 研究者はデータが生じる過程に直接介入しない方法でデータを収集する。
- 観察研究の結果は、説明変数と応答変数の間の関連を示すために使えるが、因果関係を確立することは一般的にできない。

## 良い標本の取り方

- ほぼすべての統計的手法は、暗黙の無作為性という概念に基づいている。
- 観察データが母集団からランダムな枠組みで収集されていない場合、これらの統計的手法——推定値やその誤差——は信頼できない。
- 最もよく使われる無作為抽出法は、**単純無作為抽出**、**層化抽出**、**クラスター抽出**である。

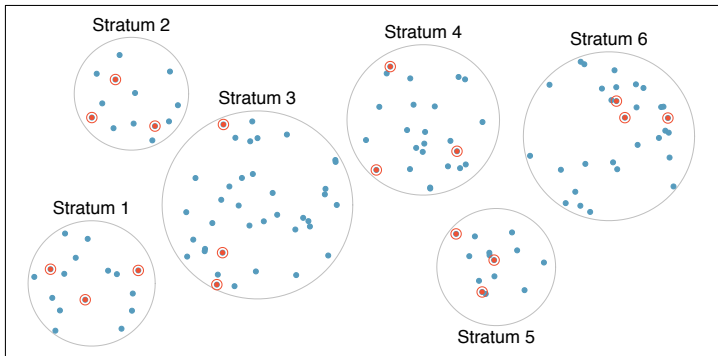
## 単純無作為抽出

母集団から、選ばれた点同士に関係性がないようにケースをランダムに選択する。



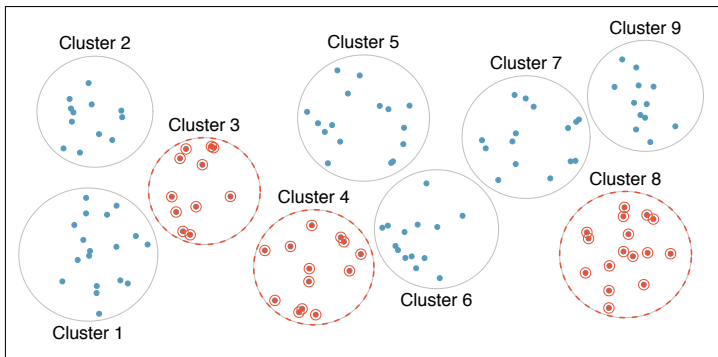
## 層化抽出

層は似た観測値からなる。各層からそれぞれ単純無作為抽出を行う。



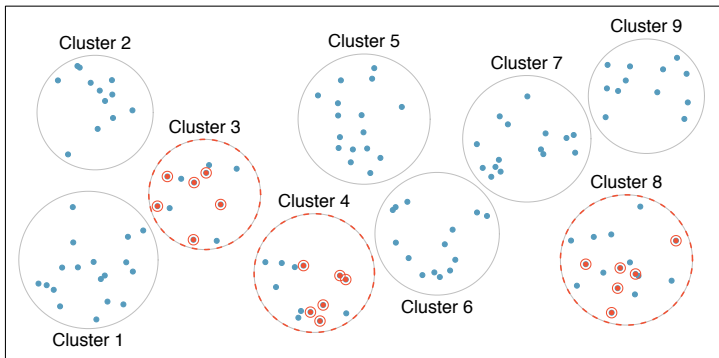
## クラスター抽出

クラスターは通常、均質な観測値からなるわけではない。クラスターを単純無作為抽出し、そのクラスター内のすべての観測値を調査する。経済的な理由から選ばれることが多い。



## 多段抽出

クラスターは通常、均質な観測値からなるわけではない。クラスターを単純無作為抽出し、さらに抽出されたクラスターから観測値を単純無作為抽出する。



## 練習問題

ある市議会が、市の郊外地区で世帯調査を実施するよう要請した。この地区は多くの異なるユニークな地域（大型住宅が多い地域、アパートだけの地域など）に分かれている。次のどの方法が最も効果的でないか？

- (a) 単純無作為抽出
- (b) クラスター抽出
- (c) 層化抽出
- (d) ブロックサンプリング

## 練習問題

ある市議会が、市の郊外地区で世帯調査を実施するよう要請した。この地区は多くの異なるユニークな地域（大型住宅が多い地域、アパートだけの地域など）に分かれている。次のどの方法が最も効果的でないか？

- (a) 単純無作為抽出
- (b) クラスター抽出
- (c) 層化抽出
- (d) ブロックサンプリング

## 実験計画の原則

1. **統制**：直接研究している変数以外の変数の（潜在的な）影響を統制する。
2. **無作為化**：可能な限り、被験者を処置にランダムに割り当て、母集団からランダムにサンプリングする。
3. **反復**：研究内では十分に大きな標本を収集することで反復する。あるいは研究全体を繰り返す。
4. **ブロッキング**：応答変数に影響を与えることが分かっている、または疑われる変数がある場合、まずその変数に基づいて被験者を**ブロック**にグループ化し、各ブロック内でケースを処置群にランダムに割り当てる。

## ブロッキングについて詳しく



- エナジーゲルが走力を向上させるかを調べる実験を設計したい：
  - 処置群：エナジーゲルあり
  - 対照群：エナジーゲルなし
- エナジーゲルがプロとアマチュアのアスリートに異なる影響を与えると疑われるため、プロかどうかでブロッキングを行う：
  - 標本をプロとアマチュアに分ける
  - プロのアスリートを処置群と対照群にランダムに割り当てる
  - アマチュアのアスリートを処置群と対照群にランダムに割り当てる
  - これにより、処置群と対照群にプロ・アマチュアが均等に代表される

なぜこれが重要か？他にブロッキングすべき変数はあるか？

## 練習問題

照明レベルと騒音レベルが学生の試験成績に与える影響を調べる研究を設計する。研究者は照明と騒音が男女で異なる効果を持つ可能性があると考え、両性が各グループに均等に含まれるようにしたいと考えている。次のうち正しいものはどれか？

- (a) 説明変数が3つ（照明・騒音・性別）、応答変数が1つ（試験成績）である。
- (b) 説明変数が2つ（照明と騒音）、ブロッキング変数が1つ（性別）、応答変数が1つ（試験成績）である。
- (c) 説明変数が1つ（性別）、応答変数が3つ（照明・騒音・試験成績）である。
- (d) ブロッキング変数が2つ（照明と騒音）、説明変数が1つ（性別）、応答変数が1つ（試験成績）である。

## 練習問題

照明レベルと騒音レベルが学生の試験成績に与える影響を調べる研究を設計する。研究者は照明と騒音が男女で異なる効果を持つ可能性があると考え、両性が各グループに均等に含まれるようにしたいと考えている。次のうち正しいものはどれか？

- (a) 説明変数が3つ（照明・騒音・性別）、応答変数が1つ（試験成績）である。
- (b) 説明変数が2つ（照明と騒音）、ブロッキング変数が1つ（性別）、応答変数が1つ（試験成績）である。
- (c) 説明変数が1つ（性別）、応答変数が3つ（照明・騒音・試験成績）である。
- (d) ブロッキング変数が2つ（照明と騒音）、説明変数が1つ（性別）、応答変数が1つ（試験成績）である。

## ブロッキング変数と説明変数の違い

- 要因 (factor) とは、実験単位に課することができる条件のことである。
- ブロッキング変数とは、実験単位がもともと持っている特性であり、その影響を統制したいものである。
- ブロッキングは層化に似ているが、サンプリングの際ではなく実験設定においてランダムに割り当てる際に用いられる。

## 実験計画の用語（続き）

- **プラセボ**：偽の処置。医学研究では対照群に用いられることが多い。
- **プラセボ効果**：特別な処置を受けていると信じるだけで、実験単位が改善を示す現象。
- **盲検化（ブラインディング）**：実験単位が自分が対照群か処置群かを知らない状態にすること。
- **二重盲検**：実験単位と、患者に関わる研究者の両方が、誰が対照群で誰が処置群かを知らない状態にすること。

## 練習問題

観察研究と実験の主な違いは何か？

- (a) 実験は研究室で行われるが、観察研究は必ずしも研究室で行う必要はない。
- (b) 観察研究では過去に起きたことしか調べられない。
- (c) ほとんどの実験では無作為割り当てを用いるが、観察研究では用いない。
- (d) 観察研究は、その結果から因果推論ができないため、まったく役に立たない。

## 練習問題

観察研究と実験の主な違いは何か？

- (a) 実験は研究室で行われるが、観察研究は必ずしも研究室で行う必要はない。
- (b) 観察研究では過去に起きたことしか調べられない。
- (c) ほとんどの実験では無作為割り当てを用いるが、観察研究では用いない。
- (d) 観察研究は、その結果から因果推論ができないため、まったく役に立たない。

## 無作為割り当てと無作為抽出

<i>ideal experiment</i>	Random assignment	No random assignment	<i>most observational studies</i>
Random sampling	Causal conclusion, generalized to the whole population.	No causal conclusion, correlation statement generalized to the whole population.	Generalizability
No random sampling	Causal conclusion, only for the sample.	No causal conclusion, correlation statement only for the sample.	No generalizability
<i>most experiments</i>	Causation	Correlation	<i>bad observational studies</i>