

# 統計解析応用研究 無作為抽出（藪の 1.3 節）

丸山 祐造 Yuzo Maruyama

神戸大学 大学院経営学研究科

## データの収集（無作為抽出） I

- ▶ できるだけバイアス（偏り）がない形で、母集団を代表するデータを取り出す必要
- ▶ バイアス：母集団を代表しないデータを取り出してしまうこと
- ▶ バイアスのあるデータの分析  $\Rightarrow$  バイアスのある結果

## データの収集（無作為抽出） II

### ▶ 例

- ▶ 大学生の意識調査をするため，友人にだけ聞き取り調査
- ▶ あなたの友人はあなたに似た人が多いはず（類は友を呼ぶ）
  - ↑ 同性が多いかも，性格の偏りがあるかも
- ▶ 「友人たち」は母集団である大学生を代表していない可能性
- ▶ 無作為抽出：バイアスの発生を回避（母集団を構成するどの個体もデータとして選ばれる確率が同じになる抽出法）

## データの収集（無作為抽出）

- ▶ 0 から 9 までの数字が書かれた 10 面体のサイコロによる
- ▶ 例：1000 人の学生から 1 人の無作為抽出
  - ▶ 全員に 000 から 999 までの番号付与
  - ▶ 番号の付け方はどのような順番でも OK
  - ▶ 10 面体のサイコロを 3 回振る
  - ▶ 3,7,4 であれば 374 番の学生
  - ▶ 0,9,1 であれば 91 番の学生
- ▶ 無作為抽出の難しさ：個人情報保護法によって，民間企業は有権者の連絡先を自由に知ることは不可能
  - ↑ 現実的な対応の一つ：日本経済新聞社の RDD 方式（random digit dialing）

## データの収集（失敗例） I

- ▶ 1936 年米国大統領選
- ▶ リテラシー・ダイジェスト誌の勝利者予想
  - ▶ 聞き取り調査：電話や自動車の保有者等から約 237 万人
  - ▶ 調査結果：共和党候補ランドン氏が圧倒的な優勢
- ▶ 実際の選挙：民主党候補ルーズベルト氏の勝利
- ▶ 誤った調査結果の原因の背景
  - ▶ 共和党：競争政策を重視，富裕層からの支持が多い
  - ▶ 民主党：弱者保護を重視，貧困層からの支持が多い
  - ▶ 電話や自動車は富裕層が保有するもの

## データの収集（失敗例） II

- ▶ 電話や自動車の保有者への調査  
＝ 富裕層に対する調査  
↑ 共和党支持者に共和党を支持しているかを聞いた  
に過ぎない
- ▶ ギャラップ社：無作為抽出した約 3000 人に対する  
調査  
↑ ルーズベルトの勝利を正しく予測

## データの収集（質問の仕方）Ⅰ

- ▶ 実施するうえで、無作為抽出以外に注意すべきこと
- ▶ 面接員の誘導的な質問「郵政民営化に賛成か反対か」  
or 「市場原理主義によって格差が拡大しています。  
あなたは郵政民営化に賛成ですか」
- ▶ 質問の設定の仕方によって回答が変わる場合  
↑ ダン・アリエリー『予想どおりに不合理』早川書  
房 2008 年

## データの収集（質問の仕方） II

### A 出版社の雑誌 B の購読料の料金設定の選択肢

- ▶ 1 or 2 or 3    3 を選ぶ人が多い
- ▶ 1 or 3        3 の人が 1 に移る

1. ネットでの雑誌閲覧 5000 円
2. 雑誌の購読 1 万円
3. 雑誌の購読 + ネットでの閲覧 1 万円

なぜ人々は選択肢が変わると行動を変えるか ⇐ 相対性を意識した選択を行う性質



## データの収集（質問の仕方） III

- ▶ 答えにくい質問を聞かれた場合，回答者は嘘をつく可能性
    - ▶ 先生（もしくは上司）が未成年のとき飲酒したことがありますか
    - ▶ 男女差別は許容されますか
- たとえ「はい」という答えを持っていたとしても，違法性や反道徳性を気にする．率直に「はい」と答えることは難しい
- ▶ 回答のランダム化：このような質問に対して正直に答えてもらうための方法（詳細は9ページ例1）