

# 疫学 (Epidemiology)

## 第4回

標本抽出法  
誤差やバイアスの制御

中澤 港 (内線 1453)

<minato@yvu.jp>

<http://phi.yvu.jp/epidemiology/>

---

---

# 第4回講義概要

- これまでも多少触れたが，標本抽出法について概説し，標本抽出を含むデザインによって誤差やバイアスをどのように減らすことができるのかを説明する。
  - 日本疫学会編「疫学」では概ね第17章と第10章に相当する。
- 
-

# 標本抽出法とは？

- 疫学調査で得られた結果を適用したい集団（介入研究でいう参照集団にあたる。統計的に考えるならば母集団）の全数を調査する悉皆調査は、費用や時間などの制約、あるいはその必要がないなどの理由で実施されないことが多い。
  - その代わりに、集団全体を代表する適当なサイズの標本をうまく選んでやればよいことになる。
  - 如何にうまく集団全体を代表するような標本を選ぶかという目的で考案されたさまざまな方法を総称して標本抽出法と呼ぶ。
- 
-

# なぜ標本抽出するのか？

- 悉皆調査には欠点がある
  - 集団には多様性がある
  - 代表性をもった部分を標本として抽出できれば、その部分を調べることによって全体の性質を推測できる(統計的推測の考え方)。
  - 記述疫学では母集団が確定していないと無意味。分析疫学や介入研究では、必ずしもそうではない。
- 
-

# サンプルの種類

- area sample
- cluster sample
- grab sample
- probability sample ( 広義の random sample)



# 単純無作為抽出法

- まず母集団の全員をリストし連番を割り振る。
- 乱数表, さいころ, コンピュータなどを使ってランダムな番号を必要な個数選ぶ。例えば,
  - 全員に  $(0,1)$  の一様乱数を与える
  - 小さい順に並べ替える
  - 小さい方から必要なところまで対象とする
- 統計ソフトを使うと簡単。N 人からなる母集団から p 人を抽出するとき, R なら `rank(runif(N))` の出力結果の左から p 個の番号に当たる人を標本とすればよい。

# 層別抽出法

- 年齢別，性別，職業別など，既知の階層毎に単純無作為抽出する
  - 層によって調査指標が異なることが既知の場合は単純無作為抽出より代表性がいい
  - 層ごとの集計ができる
  - サンプルング以前に，階層の情報がわかっていなければならない(が，予備的にその集団について階層を調べたりすると，それ自体が本調査に影響するかもしれない)
  - 階層の出現頻度が事前にはわからない
  - 時間と金がかかる
  - 総サンプル数が決まっている場合，階層毎のサンプル数が減ってしまう
- 
-

# 集束抽出法

- 集落抽出法ともいう。Cluster sampling
  - 多段抽出の1つで、最終段階では全数を標本とする。最終段階の1つ前で選ばれる集団を単位として抽出する方法と考えられる
  - 途上国の調査ではよく使われる。例えば、複数の村を含む州の調査などで、村をランダムに選んで、選ばれた村は全数調べる
  - 比較的安上がりで同意を得やすく短期間で調査できる場合が多い
- 
-



# 確率比例抽出法

- Probability Proportionate Sampling (PPS)
  - 母集団が不均質なとき、均質と考えられるブロックに分け、各ブロックの人口に比例した確率でいくつかのブロックが選ばれた後、各ブロックからは同数のサンプルを抽出する方法
  - 逆にブロックサイズによらず等確率でいくつかのブロックが選ばれた後、各ブロックからそのサイズに比例した数のサンプルを抽出する方法を副次抽出法という。
- 
-

# 標本抽出法の選択

- 母集団が小さいときは、単純無作為抽出か層別抽出
  - 母集団が大きいときは、資金が豊富にあれば層別抽出
  - 母集団が大きく資金が乏しいとかアクセスが悪いときは多段抽出
- 
-

# 標本サイズの計算

- 表 17.2 を参照
- 目的によって計算式が異なる
- 測定値に関する予測（先行研究などから）が得られ、有意水準と検出力を決めれば計算できる

# 誤差

- 誤差 = 真値との差
  - 標本誤差 = 標本抽出の偶然変動に伴う母集団の真値との差 (標本サイズが大きいほど小さい)
  - 非標本誤差 = 標本誤差以外の誤差 (例えば不適切な標本抽出による誤差)
  - ランダムな誤差 (偶然誤差)
    - ランダムな誤差が小さい = 精度 (precision) が高い
  - 系統的な誤差 = バイアス
    - 研究デザイン, データ収集, 分析, レビュー, 出版など, 研究のさまざまな段階で起こりうる。
    - 系統的な誤差が小さい = 正確さ (accuracy) が高い
- 
-

# バイアス

- 選択バイアス (selection bias)
    - 観察する集団が母集団を正しく代表していないときに起こる偏り
  - 情報バイアス (information bias)
    - 観察するときに得られる情報が正しくないために起こる偏り
  - 交絡バイアス (confounding bias)
    - 分析疫学で起こる特殊な偏り。要因と疾病の両方と関連する交絡因子の存在によって起こる
- 
-

# バイアスの制御

- 無作為化 (randomization) : ランダムに群分けをすることで介入以外の条件を確率的に均質化
  - マッチング (matching) : 交絡因子の条件が似るように対照群を選ぶ
  - 層化 (stratification) : 交絡因子のカテゴリ別に解析することで交絡因子の影響を除く
  - 標準化 (standardization) : 基準集団を決めて交絡因子のカテゴリ別割合を調整することで交絡因子の影響を取り除く
- 
-