母集団と標本

母集団と標本

全数調査と標本調査

- 全数調査、または悉皆（しっかい）調査
  - 全体について調べる（例：国勢調査）
- 標本調査
  - 全体から選び出した一部分について調べる（例：選挙予測）
  - 時間的・経済的に現実的な調査

母集団と標本

- 母集団（population）
  - 標本調査のもとになる集団
  - 母集団の大きさを、 N であらわす
- 標本（sample）
  - 母集団から抽出（サンプリング）された一部分
  - 標本の大きさ（サンプルサイズ：標本に含まれる個数）を、 n であらわす

標本の抽出

- 推測統計学（inductive statistics）
  - 標本にもとづいて母集団の特性値（母数：平均、分散などのパラメータ）を推定・予測する
  - 2つ以上の母集団の特性値を比較・検討する
- 標本が「母集団の精巧なミニチュア」になるように、偏りなく標本を抽出しないといけない

無作為抽出法（random sampling）

- 母集団から無作為に標本を抽出する
- 乱数表（無規則に数字を羅列した表）等をもとに、データ（個体）を標本として選ぶ
- どの標本も全く等しい確率で選ばれる（主観や作為などが入る余地がない）
- 母集団が大きい場合は実施が困難（例えば数万人の名簿を作るのは容易ではない）

系統抽出法（systematic sampling）

- 最初の標本のみ乱数表などで選ぶ
- あとの標本は一定間隔（抽出間隔；sampling interval）で抽出する
- 実際には抽出間隔が扱いやすい数でない場合が多い
- 無作為抽出法と同じで、母集団が大きい場合は実施が困難

多段抽出法（multi-stage sampling）

- 何段階かのステップを経て標本を抽出する（例：2段階抽出法）
- 何段階かのステップで標本にする対象を絞り込み、最後に無作為抽出法で標本を選ぶ
  - 4段階の例：全国→都道府県→市町村→病院・診療所→看護師（あとは無作為抽出）
- 調査の手間ひまが少なく実践的だが、標本の偏り（バイアス）に注意しないといけない

層別抽出法（stratified sampling）

- 母集団を同じ特徴（年代、職業、都道府県など）で層に分ける（層別化）
- 層の構成比に応じて適切な数だけ、層ごとに無作為抽出法で標本を選ぶ
- 層の違いによって偏りがある状況が事前に予測される場合に適している（地域と政党の支持率、年代と好きな芸能人）
    - 母集団の特徴を少なくとも1つは事前に知っている必要がある