

## 第10回 6月25日の講義内容

- § 3-5. サンプルングの方法
  - 全数調査と標本調査
  - サンプルングの種類
  - 無作為抽出の方法
    - 単純ランダムサンプルング
    - 系統抽出法
    - 多段抽出法
    - 層化抽出法

6/25/03

---

---

---

---

---

---

---

---

## 全数調査と標本調査

- 統計調査
  - 調査対象 = 母集団
  - 全数調査 (悉皆調査)  
母集団全体を調べる (記述)
  - 標本調査  
母集団の部分集合を調べて、母集団の性質をあてる (推測)

6/25/03

---

---

---

---

---

---

---

---

## サンプルングの種類

- 標本のとり方 推定の精度を左右  
(標本の代表性)
- 有意抽出 (非確率的)
  - 典型調査: 典型的な集団を選ぶ
  - 割当調査: 属性によって調査数を割当
- 無作為抽出 (確率的)
  - ランダムサンプルング

6/25/03

---

---

---

---

---

---

---

---

## 無作為抽出の方法

- ✦ まず、母集団リスト = サンプル台帳が必要
  - サンプル台帳として使用されるもの
    - 「選挙人名簿」
    - 電話帳
    - 各種名簿
- ✦ サンプル台帳から確率抽出

6/25/03

---

---

---

---


---

---

---

---

## 単純ランダムサンプリング

- ✦ サンプル台帳から乱数(一様乱数)によって抽出する方法
- ✦ 利点: 各個体が等確率で抽出
- ✦ 欠点: 母集団の属性構成と抽出後の標本の属性構成が異なることがある。  

- ✦ 事前確率と事後構成比の違いに注意

6/25/03

---

---

---

---

---

---

---

---

## 単純ランダムサンプリング(2)

- ✦ 標本数と精度の関係
  - (例) 視聴率 関東地区(調査世帯600)
  - キリンカップサッカー2003 日本×パラグアイ  
テレビ朝日 '03/6/11(水) 19:15 - 123 min  
視聴率 21.6 %  
二項分布より、視聴率 $p$ の推定量の分散は $p(1-p)/n$   
 $p=0.216$ ,  $n=600$ として計算すると、 $0.00028224$ 。  
標準偏差は、 $0.0168$ 。したがって、95%の信頼区間は  
 $\pm 3.36\%$ 。

6/25/03

---

---

---

---

---

---

---

---

## 系統抽出法

- ✦ サンプル台帳から等間隔で抽出する方法。等間隔抽出法ともいう。
- ✦ まず、母集団の総数 $N$ を抽出する標本数 $n$ で割り、 $k=N/n$ 個体からなる $n$ 個のサブグループに分ける。
- ✦ 次に第1サブグループから乱数によって標本を1つ抽出する。以降サンプル台帳から $k$ 番おきに抽出する。
- ✦ (例) 1,000人の母集団から50人の標本を抽出するケース
  - $k=1000/50=20$ ,
  - サンプル台帳の1~20番から乱数で $r$ 番目の個体を抽出。
  - 以降、台帳から $r+20 \times j$ 番目( $j=1, \dots, n-1$ )の個体を抽出。

6/25/03

---

---

---

---

---

---

---

---

## 多段抽出法 multi-stage sampling

- ✦ 母集団から直接ランダムサンプリングするのではなく、母集団を互いに排反する集団に分割して、まず集団をサンプリングする。次に抽出された集団からランダムサンプリングするという抽出法(二段抽出)。
- ✦ 抽出された第一次集団をさらに互いに排反する小集団に分割し、その小集団をランダムサンプリングする、ということをしていくのが、多段抽出法である。

6/25/03

---

---

---

---

---

---

---

---

## 多段抽出法(2)

- ✦ (例) 近畿圏から100人抽出
  - 第一段: 大阪、京都、兵庫、滋賀、和歌山、奈良の6府県から一つの府県を抽出
  - 第二段: 第一段で抽出された府県から市町村を抽出
  - 第三段: 第二段で抽出された市町村からランダムサンプリングで100人抽出

6/25/03

---

---

---

---

---

---

---

---

### 多段抽出法(3)

- ✦ 抽出の原則: 末端が等確率で抽出されること
- ✦ 確率比例抽出
  - 第一段階の抽出で、個体数に応じて抽出確率を調整し、第二段階の抽出でランダムサンプリングする方法
- ✦ 等確率抽出: 第一段で複数集団の抽出を行う場合
  - 第一段階の抽出では、ランダムサンプリングで抽出。第二段階の抽出で個体数に応じた標本数を抽出。

6/25/03

---

---

---

---

---

---

---

---

### 多段抽出法(4)

- ✦ 数値例
  - A学部 在籍 300名
  - B学部 " 200名
  - C学部 " 150名
  - D学部 " 400名
  - E学部 " 150名
  - 5学部から2学部を選び、100人抽出

6/25/03

---

---

---

---

---

---

---

---

### 多段抽出法(5)

#### ✦ 数値例: 確率比例抽出

	抽出確率	第一段	第二段	最終
■ A学部 300名		1/2	50/300	100/1200
■ B学部 200名		1/3	50/200	100/1200
■ C学部 150名		1/4	50/150	100/1200
■ D学部 400名		2/3	50/400	100/1200
■ E学部 150名		1/4	50/150	100/1200

6/25/03

---

---

---

---

---

---

---

---

## 多段抽出法(6)

### ◆ 数値例: 等確率抽出

割当数 第一段 × 第二段 = 最終

- A学部 300名 2/5
- B学部 200名  $100 \times (200/600)$  2/5 33/200 100/1500
- C学部 150名 2/5
- D学部 400名  $100 \times (400/600)$  2/5 67/400 100/1500
- E学部 150名 2/5

◆ 第一段でどこが抽出されるかによって、最終抽出確率が変化。

6/25/03

---

---

---

---

---

---

---

---

## 層化抽出法 stratified sampling

◆ 事前と事後を一致させる工夫の一つ。

◆ (例) A社社員から50人抽出

- A社社員の学歴構成: 高卒70%、短大卒以上30%
- 母集団(A社社員)からランダム標本を抽出するときの確率分布(事前) 高卒0.7、短大卒以上0.3
- 抽出された標本の学歴構成(事後)  
高卒0.5、短大卒以上0.5 ということもありえる。



標本が母集団を代表していないおそれがある。

6/25/03

---

---

---

---

---

---

---

---

## 層化抽出(2)

◆ 事前と事後の層構成比を一致させるように、標本の割当数を層毎に決め、層毎にランダムサンプリングする。

= 層化抽出法

◆ この例では、「学歴」が層(stratum)に対応。

- 高卒層への標本割当数:  $50 \times 0.7 = 35$ 人
- 短大卒以上への標本割当数:  $50 \times 0.3 = 15$ 人

6/25/03

---

---

---

---

---

---

---

---

### 層化抽出(3)

◆ 層化二段抽出:

- 層化抽出と多段抽出の組み合わせ
- 世論調査などでよく用いられる

大都市 → 地域ブロック or 都道府県 → 市町村 → ランダムサンプリング

10万人以上の都市 → //

その他 → //

6/25/03

---

---

---

---

---

---

---

---