

1 3. 統計調査とは

(2011年10月20日授業)

2011/10/20

公的統計・統計機関の行動原則

- 国際連合統計委員会 UNSC
 - 公的統計の基本原則 (Fundamental Principles of Official Statistics)
- 欧州委員会統計総局 eurostat
 - 欧州統計行動準則 (European Statistics Code of Practice)
- 総務省統計局・政策統括官(統計基準担当)
 - 使命と行動指針

2011/10/20

公的統計の基本原則 (国連統計委員会: UN Statistical Commission)

1. Relevance, impartiality and equal access (妥当性、公平性および入手の平等性)
2. Professional standards and ethics (専門家の規範と倫理)
3. Accountability and transparency (説明責任と透明性)
4. Prevention of misuse (誤用の防止)
5. Sources of official statistics (公的統計の出所・情報源)
6. Confidentiality (機密保持)
7. Legislation (法律の制定)
8. National coordination (国内における調整)
9. Use of international standards (国際標準の使用)
10. International cooperation (国際協力)

2011/10/20

欧州統計行動準則

1. Professional Independence (専門的独立性)
2. Mandate for Data Collection (データ収集権)
3. Adequacy of Resources (適切な資源の使用)
4. Quality Commitment (質に対する責務)
5. Statistical Confidentiality (統計上の機密保持)
6. Impartiality and Objectivity (公平性および客観性)
7. Sound Methodology (堅実な方法論)
8. Appropriate Statistical Procedures (統計手法の妥当性)
9. Non-Excessive Burden on Respondents (調査対象への過重負担の禁止)
10. Cost Effectiveness (コストパフォーマンス)
11. Relevance (妥当性)
12. Accuracy and Reliability (正確性および信頼性)
13. Timeliness and Punctuality (適時性および定時制)
14. Coherence and comparability (一貫性および比較可能性)
15. Accessibility and Clarity (アクセスのし易さ、明確性)

2011/10/20

統計局・政策統括官(統計基準担当) の使命と行動指針

- 社会に役立つ、客観的で正確な統計を追求します
- 多様な利用者に利用しやすい形で統計を提供します
- 調査の対象となる方の秘密の保護に万全を期します
- 調査の対象となる方の負担に常に心を配ります
- 高度の専門性を備え、内外の統計の発展に尽くします
- 政府横断的な調整を通じて統計の体系的な整備に努めます

2011/10/20

(公的)統計の分類

作成方法による分類

- 第一義統計
 - 統計作成目的の調査を行い、その結果得られた統計 (例) 国勢調査、労働力調査(完全失業率)
- 第二義統計(業務統計)
 - 行政上の目的で収集された情報を集計して得られた統計 (例) 通関統計、住民基本台帳人口移動報告、職安統計(有効求人倍率)
 - 人口動態調査は行政上収集された戸籍情報の集計であり、第二義統計に分類されるが、市区町村を対象とした統計「調査」である点では、第一義統計の性格も有する。

2011/10/20

(公的)統計の分類(続き)

加工の有無による分類

- 一次統計
 - 調査の結果、直接得られるもの
- 二次統計(加工統計)
 - 一次統計を加工して得られるもの、解析的性格をもつ
(例)国民経済計算、消費者物価指数

2011/10/20

第一義統計の種類

- センサス(全数調査) census 記述
 - 母集団(関心対象)の全ての個体を調査するもの。「悉皆調査」ともいう。
 - 得られる分布の特性値は定数。
 - 国勢調査(総務省)
 - 経済センサス-基礎調査(総務省)
- 標本調査 sampling survey 推測
 - 母集団から無作為抽出(random sampling)によって得た標本を調査するもの。
 - 得られる標本分布の特性値を、母集団分布の特性値の推定量として用いる。⇒ **確率変数**
 - 国民生活基礎調査(厚生労働省)

2011/10/20

無作為抽出の方法

- 母集団リスト(=サンプリング台帳)から確率的に抽出
 - 母集団の中の各個体が等確率で抽出される
 - 単純ランダムサンプリング
 - 系統抽出法(等間隔抽出法)
 - 多段抽出法
 - 層別抽出法
- 等の方法がある

2011/10/20

単純ランダムサンプリング

simple random sampling

- サンプリング台帳から乱数(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

2011/10/20

単純ランダムサンプリング(続き)

- 標本数と精度の関係
(例)視聴率 関東地区(調査世帯600)
 - ドラマ「江・姫たちの戦国」第40回
NHK総合 2011/10/16(日) 20:00~45 15.2%
二項分布より、視聴率 p の推定量の分散は $p(1-p)/n$
 $p=0.152, n=600$ として計算すると、 0.000214826 。
標準偏差は、 0.0147 。したがって、95%信頼区間は $\pm 2.9\%$ 。
 - 視聴率 18.1% と 視聴率 15.2%
の響きの違いは大きい!!
- (視聴率データはビデオリサーチ社の公表にもとづくものである)

初回の視聴率は
21.7% !!

2011/10/20

系統抽出法 systematic sampling

- サンプリング台帳から等間隔で抽出する方法。等間隔抽出法 equi-interval sampling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、サンプリング台帳から 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
 - $k=1000/50=20$ 。
 - サンプリング台帳の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。
- 視聴率調査(ビデオリサーチ社)が採用
c.f. 視聴率調査について

2011/10/20

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 two-stage sampling)。
- 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、ということが続けていくのが、多段抽出法である。

2011/10/20

多段抽出法(2)

- 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
probability proportionate to size sampling (PPS)
 - 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 equal probability sampling
 - 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。

2011/10/20

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。

2011/10/20

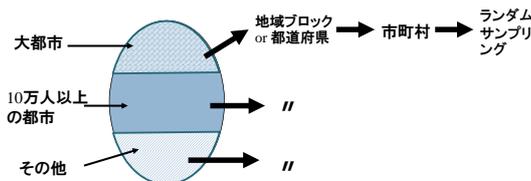
層化抽出法(2)

- 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

2011/10/20

層化抽出法(3)

- 層化二段抽出 stratified two-stage sampling
 - 層化抽出と多段抽出の組み合わせ
 - 公的統計や世論調査などでよく用いられる



2011/10/20

【注意】層別抽出時の母集団平均の推定について

- 層別抽出を行った場合、層ごとの事後抽出率(=抽出確率×回収率)が同一でないとき、抽出(後)標本の単なる算術平均を母集団平均の推定量に用いては **いけない**。
 - このようなケースでは、抽出率調整を行って母集団平均の推定をする必要がある。
 - 抽出率調整が不要な場合は、
 - 抽出後標本における層比率と母集団における層比率が同一のとき
 - 層ごとの事前抽出率が同一にコントロールしてあるときは、層ごとの回収率が同一のとき
- の二ケースのみ

2011/10/20

公的統計調査による標本抽出の例 (平成15年住宅・土地統計調査)

19

層化2段抽出法を採用

- 第1次抽出単位:平成12年国勢調査調査区
- 第2次抽出単位:抽出された標本調査区を基本とする調査単位区内の住戸
 - 標本調査区の抽出に当たっては、都道府県、市部、郡部、市、14大都市の区、大都市圏、都市圏及び人口規模が一定以上の町村のそれぞれについて十分な結果精度が得られるよう、地域ごとに異なる抽出率を適用。

2011/10/20

公的統計調査による標本抽出の例 (平成15年住宅・土地統計調査)(続き)

20

□ 調査単位区の抽出方法

- (1)国勢調査調査区の層化
平成12年国勢調査の結果に基づき、主として「換算世帯数」、「住宅の所有の関係別割合」及び「65歳以上親族のいる一般世帯数割合」により、24層に層別。
- (2)標本調査区の抽出
市、区及び人口1万5千以上の町村については各市区町村ごとに、人口1万5千未満の町村については都道府県ごとに、それぞれ定められた抽出率により系統抽出。抽出率は、人口規模等により変化。
- (3)調査票乙を配布する調査区の指定
十分な結果精度が得られるよう、神奈川県及び福岡県については県庁所在市、政令指定都市とそれ以外の地域ごとに、神奈川県及び福岡県以外の都道府県については県庁所在市とそれ以外の地域ごとに、抽出した標本調査区の中からそれぞれ定められた抽出率により系統抽出。
- (4)調査単位区 of 抽出
抽出された標本調査区を調査単位区に設定。ただし、70戸以上の標本調査区については一つの単位区がほぼ50戸前後になるよう二つ以上に分割し、その中から一つの単位区を無作為に抽出して調査単位区とした。

2011/10/20

調査の実際

21

- 住居が調査区に当たると、調査区内の全ての該当世帯に対し「調査のお知らせ」が配布
- 調査対象世帯になると「調査のお願い」が配布
 - 調査員が訪問(不在の場合は「連絡票」が投函)
 - 調査票と「記入の仕方」が配布
 - 後日、調査員が回収
- 家計調査など調査対象に負担のかかる調査以外、基本的に謝礼はない

2011/10/20