

⑨ 動態統計に伴うサムアリンクの一例

遠藤 健児

内田 良男

建設省の建築着工統計に於ける調査事項のうちに建物の延べ面積と工事費予定額とが含まれているが、これは着工届に記載されるものであるから実際に竣工した建物について得られるべき値とは相当に喰い違ふものである。

殊に工事費については税金に対する考慮から他の所得などに於けるのと同じ理由に依つて、過小に申告され勝ちである。

従つてこれらの数値から評価される坪当り工事費は着しく信頼し難いものである。

これが五千円以下となつて現れる様な場合も少なくない。そこで着工届に依つて登録された新建築物について、主として一棟毎の延べ面積と総工事費とに關して補正的な調査が必要とされるに到つた。

建築着工統計は都道府縣の建築監督主事の下に提出される建築物着工届に依る一つの動態統計であつて、集計は一ヶ月毎に行われてその結果は建設省住宅局の建築着工統計月報に発表される。

しかしながら補正調整は竣工した建築物について行われるべきものであるから種々の制約をうける。従つてこの場合の標本調査に當つては特に次の点を考慮しなければならない。

(1) 着工届が提出されてから実際にその建物が出來上る迄には数ヶ月を要し、しかも竣工届の様なものも提出されない

(2) 補正調査の調査は建築監督主事、其の他極く少数の専門技術者の査定に依るものであるから同時に多数の棟について調査することは難かしい。

(3) ある期間に於る着工建築物に対する補正調査の結果は、その期間から余りに隔つてはいない時期に発表しなければならない。

この結果補正調査は次々に提出される着工届のうちからその一部を抽出し、抽出された分についてそれが出来上るべき頃合を見計らつて査定に出向く方法をとらねばならない。

叔てこの様な場合の抽出計画を行うに當つては次の二点が問題となる。

(4) 標本抽出計画を行う際には、調査の対象となつている期間中に着工されるべき建築物の確定的な棟数は未知である。即ち対象となる母集団の大きさを確認することが出来ない。

(5) 従つて指定された大きさの任意標本を抽出することは不可能であつて、抽出比を予め指定出来るだけである。逆に云へば実際に得られるべき標本の大きさは計画時には確定出来ない。

(6) 標本の抽出は系統的抽出法に依らねばならない。

この様な條件の下で、先づ全都市に関する補正調査のための標本抽出法と母数推定法を次の様に考案した。普通に行われる調査の場合の同様な考慮から全都市を適当に層別し、市又は區を第一抽出單位とし、着工された棟を第二抽出單位とする副次抽出法をとる。

今或る一つの層には M 個の市（又は區）が含まれるものとし、この第 j 層目の單位について対象となる一定期間中に着工されるべき棟数を N_j 、その推定値を N_j' 、この層に関する合計を夫々 N 、 N' とする。予め指定された抽出比を r とし、比例割当法に依つて標本を各々の層に配分するものとすれば、この單位に関する抽出比 r_j 並にこれが抽出された時に実際におめられる標本数（棟数） n_j は唯一つの第一抽出單位を抽出する時には次の様に定められる。

$$N'r = n', \quad n'/N'_j = r'_j, \quad N'_j r'_j = n'_j$$

一方ある測定 Y (例えば一棟毎の工事費) の第 j 単位に於る平均を \bar{Y}_j , これが含まれる層に於る平均 \bar{Y} は夫々

$$\bar{Y}_j = \frac{1}{N_j} \sum_{k=1}^{N_j} Y_{jk}, \quad \bar{Y} = \frac{1}{N} \sum_{j=1}^M N_j \bar{Y}_j = \frac{1}{\sum N_j} \sum_j \sum_k Y_{jk}$$

と書き表わされる。更に第 j 単位に抽出される確率を P_j , それ抽出された時、得られた標本に関する測定値を

$$y_{j1}, y_{j2}, \dots, y_{jn_j}$$

とし、この平均値を $\bar{y}_j = \frac{1}{n_j} S(y_{jk})$ としよう。

此処で予め M 個の単位に対応する M 個の常数の system k_1, k_2, \dots, k_M を適当に定めておいて

$$\bar{y} = k_j \bar{y}_j = \frac{k_j}{n_j} S(y_{jk})$$

を以て母数 \bar{Y} の推定値とする。この統計量 \bar{y} の分布については、一応標本の棟が單純に任意抽出されたものとするれば次のことが容易に示される。

$$(i) \quad E(\bar{y}) = \sum_j P_j \bar{Y}_j k_j \quad (= \hat{\bar{Y}} \text{ とおく})$$

$$(ii) \quad D^2(\bar{y}) = \sum_j P_j (k_j \bar{Y}_j - \hat{\bar{Y}})^2 + \sum_j P_j k_j^2 \left(\frac{1}{n_j} - \frac{1}{N_j} \right) \sigma_j^2$$

$$\text{但し} \quad \sigma_j^2 = \frac{1}{N_j - 1} \sum_{k=1}^{N_j} (Y_{jk} - \bar{Y}_j)^2$$

従つて任意に定められた P_j ($j=1, 2, \dots, M; \sum P_j = 1$) に対し、 \bar{y} を求める際には N_j や N は既知となるから

$$P_j k_j = N_j / N, \text{ 即ち } k_j = N_j / (N P_j)$$

とすれば \bar{y} は \bar{Y} に対する不偏推定値となる。

故に第一段階の問題は P_j を如何に定めるかと云うことであつ

て、この P_j に依つて (ii) で與えられる \bar{y} の誤差が變つてくる。如て P_j のとり方によつて大きく影響を受けるのは (ii) の右辺の第一項たる第一次單位間の変動を表わす部分であると考えられるから、差し詰めこの項を出来るだけ小さくする様に P_j を定めればよいわけである。このためには、極端な場合についてみれば

$$k_j = N_j / (N P_j) \text{ であるから}$$

$$(iii) N_j \bar{Y}_j / (N P_j) - \bar{Y} = 0 \text{ 即ち } P_j = (N_j \bar{Y}_j) / N \bar{Y} = \sum_k Y_{jk} / \sum_j \sum_k Y_{jk}$$

とすれば $\sum P_j (k_j \bar{Y}_j - \bar{Y})^2 = 0$ となる。即ち $\sum_k Y_{jk}$ 、換言すれば各單位に於る Y の總計（この例では總工率費）に比例する確率で抽出すればよいわけである。しかしながらこの量は未知量であるからこの方法をそのまま実行することは出来ないのであるが、これに近い量に比例して第一次單位を抽出すればそれだけ精度のよいことが期待される。例へば抽出比を決めるために用いた着工予定棟数 N_j' 、或はその時点に得られた *data* から推定される総工率費の予定額などを用いることが出来る。これに応じて係数 k_j が修正されるからそのために偏倚が入ることはない。

今 P_j として N_j' / N' ($N' = \sum N_j'$) をとつたものとし、
 $N_j' / N \approx N_j \bar{Y}_j / (N \bar{Y})$ となつてゐるものとするれば、推定値 \bar{y} の分散は (ii) 及び (iii) から大約

$$\begin{aligned} v^2(\bar{y}) &\approx \sum \left(\frac{N_j}{N} \right)^2 \frac{N}{N_j} \frac{\bar{Y}}{\bar{Y}_j} \left(\frac{1}{n_j} - \frac{1}{N_j} \right) \sigma_j^2 \\ &= \frac{\bar{Y}}{N} \sum \frac{N_j}{\bar{Y}_j} \left(\frac{1}{n_j} - \frac{1}{N_j} \right) \sigma_j^2 = \frac{1}{N} \sum \frac{\bar{Y}}{\bar{Y}_j} N_j \left(\frac{1}{n_j} \right) \sigma_j^2 \\ &\approx \frac{1}{n'} \left(\frac{1}{N} \sum N_j \sigma_j^2 \right) = \sigma_w^2 / n' \quad (n' \approx n_j, \bar{Y} \approx \bar{Y} \text{ とする}) \end{aligned}$$

となることが期待出来る。

一つの層から實際に得らるべき標本の大小 n_j は N' と r とに依つて予定された数 $n' = N' r$ とは一般に異つてくる。即ち

$$n_j = N_j r_j = n' \cdot N_j / N_j'$$

従つて、例えば $P_j = N_j' / N'$ と定めるときには

$$E(n_j) = n' \sum \frac{N_j'}{N'} \frac{N_j}{N_j} = n' \cdot N_j / N'$$

となる。故に N_j と N_j' とが余りに異なる様をとくときは、比例割当が実現されないことは先づよいとしても、ある層に於る標本数が予定より極端に少くなること有り得る。従つて標本の推定値は適当なものをとる必要がある。

このためには、例文は一ヶ月の間に着工された標本について既得の *data* を利用して十分に分析を行わねばならない。

昭和25年4月から全年10月迄の *data* に依つて丸山文行君が検討した処に依れば、この置は非常に変動が多く正確な予測は困難である様に思われる。

一般に r_j を定めるために用いるべき N_j' にはこの様な差が避けがたいものであるから、標本数が予定より特に少い層については改めて追加的に標本を抽出することにすればよいであろう。

又時に依つては $r_j > 1$ となること有り得るから N_j' の余りに小さい市は他と合併して一つの單位にまとめる様にならなければならない。以上は主として一つの層に於ける考察であるが、全体の平均 \bar{Y} の推定値としては普通に行われる様に各層での推定値から求められる層の大きさ(既知となる)に依る荷重平均をとればよい。

この調査は未だ行われていないのであるが、実際には遙かに容易な条件下で行ふことに要旨が変更されたので上述の企画は実行に移されないことになつた。従つて抽出計画のために必要な既得の *data* の分析も一部を手がけただけで一応打切りとなつたのである。

しかしながら始めに述べた様な條件は、この種の調査に際しては比較的一般的性を持つものと思われるので、その場合の抽出法、推定法についての一つの考察を要約して報告した。

(この研究は文部省科学研究費に依る研究の一部である。)

A sampling design undertaken when the population can not yet be certified.

By K. Endō and Y. Uchida

A sampling method and some unbiased estimators are considered under the following conditions:

(1) Sampling design must be made before all the notifications will be reported.

(2) The size of population can not be certified, but certain estimates of which may be available.