



ユリ

夏、山道を歩いていると、思いがけずユリの花にめぐりあうことがある。静けさのなかで、ひときわあでやかな花である。そんなとき、けだるさも忘れ、立ちどまってしまうものだ。

女性の清らかさをユリにたとえることがある。ユリはしとやかでありながら、はつらつとした花である。大柄な花びら、そして鮮やかな色あいは、夏の紫外線にもみごとに調和する。そのおもかげに、みはてぬ憧れをいただいていたのは、私ばかりではないだろう。

上の写真はテッポウユリだが、ユリといっても品種は多い。なじみ深いものでは、ヒメユリ・ヤマユリをはじめ、シラユリ・スカシユリなど、日本には15種が自生するという。

7月のおもな行事

- 4日～5日 毎月勤労統計調査関東ブロック会議(山梨県)
社会生活指標統計調査会議(東京都)
- 7日 七夕
- 10日～11日 北関東4県所得推計研究会(栃木県)
- 12日～13日 商業統計実務担当者会議(青森県)
- 18日～19日 関東5県統計主管課長会議(千葉県)
- 26日～27日 国勢調査調査区設定ブロック会議(神奈川県)

標本設計 (3)

— 抽出から集計・誤差計算まで —

11. 標本の規模と層別配分

標本学級数は各学年、市部郡部別に各100、計600、生徒数は各標本学級について10人、計6000人としてみました。1つの副標本でみれば、学級数は各学年・市部郡部別に各25、計150、生徒数は各標本学級について10人、計1500人となります。1つの副標本はちょうど全標本の4分の1となっています。

(解説) 標本の規模をどのくらいにするかは大きな問題です。過去に同じ調査をしていれば、増減なり同数なり決めやすいのですが、初めての調査では色々と下調べをしなければなりません。まず類似の調査、参考となる調査の有無を調べます。いつどこで実施した調査でもかまいません。見つかったら、その調査の設計を研究します。すると標本の規模や精度や結果表様式の細分状況がわかりますから、それらを参考にすれば、これから実施しようとする調査に必要な標本の規模は大体見当がつくものです。しかし、参考になる資料が見つからないときは、標本規模をきめるよりどころがありませんから、できれば試験調査(パイロットサーベイ、プリテストなどともいわれる)を行って資料を集めるのが望ましいわけですが、費用や時間の関係でそれもかなわぬとなると、とりあえず、ということで適当と考えられる大きさにきめるよりほかはありません。その代

り、誤差計算をしっかりと行って、次回の改善に備えることとします。

この例では、他に参考になる調査の結果や資料がないとして、冒頭のような決め方をしてみました。したがって、これには特別な根拠はありませんが、一応これで様子を見ようということです。なお、主要な調査項目の平均値の目標標準誤差が示されており、母集団における標準偏差の概数もわかっているときは、標本抽出方法を無層化1段単純任意抽出法であると仮定して、次の式によって標本数を出してみるのも一つの方法です。しかし、実際は層化・

$$(\text{標本数}) = \left(\frac{\text{主要項目の母集団における標準偏差}}{\text{目標標準誤差}} \right)^2$$

2段・系統抽出などを行いますから、計算の根拠はあまりないのです。あくまで参考です。こうして出した標本数に対する実績精度は、層化抽出を行えば良くなり、2段抽出を行えば悪くなるのが普通ですから、この辺のことを頭において、計算された標本数を適当に加減して用いるのがよいのです。筆者の知るところでは、層化2段抽出法においては、層化の効果はあまり高くなく、2段抽出による精度の低下の方が大きく、全体としてはさきの目標標準誤差を上回る誤差となる例が多い様です。したがって、こうして出した標本数は、実際に必要とする標本数より少な目であると考える方が無難だと思います。

表-4 1つの副標本の層別配分数
(上段 学級数)
(下段 生徒数)

学年	市部	郡部	計
1	25	25	50
	250	250	500
2	25	25	50
	250	250	500
3	25	25	50
	250	250	500
計	75	75	150
	750	750	1500

表-5 全標本の層別配分数
(上段 学級数)
(下段 生徒数)

学年	市部	郡部	計
1	100	100	200
	1000	1000	2000
2	100	100	200
	1000	1000	2000
3	100	100	200
	1000	1000	2000
計	300	300	600
	3000	3000	6000

12. 標本抽出の事務

抽出作業に入る前に抽出単位名簿を整備しなければなりません。抽出単位名簿の基礎資料は前号4ページの表-2です。その「学校記入欄」が記入されて県に集められたら、所定の層に分類し、一定の順に配列し、「県記入欄」に所要事項を記入し、標本抽出のお膳立てをします。

まず、層の分類は学年(3区分)、市部郡部(2区分)、計6区分とし、各区分内の配列は慣用の順とし、ファイルします。表-6は「市部・3年」の層のファイルについての記入例です。(表-6は説明のためわざわざ作ったもので、実際の記入はファイルの中に行います)

標本抽出の手順は次のとおりです。

..... 総理府統計研修所講師 船津好明

(1) 層内の学級数の合計および生徒数の合計を確かめ、層のファイルの末尾などに記入します。

(2) 表-6の「生徒数の累計」欄に生徒数の、層内を通じた累計を記入します。最終累計値はその層の生徒の総数に一致します。

(3) この例では、層内には360学級あり、この中から最終的に100学級抽出しますが、実際は25学級抽出することを4回くり返します。

(4) (抽出間隔)学級の抽出には生徒数をウェイトとし、生徒を等間隔抽出することによる確率比例抽出法を用います。抽出間隔は、

$$\frac{\text{層内の生徒の総数}}{\text{1つの副標本の抽出学級数}} = \frac{16904}{25} = 676.16$$

とします。

(5) (抽出起番号)抽出起番号は抽出間隔を超えない整数を4個独立に、乱数表から選びます。その結果、抽出順に486, 202, 338, 113となったとしますと、これらをその順に各副標本の抽出起番号とします。(乱数表とその使い方については拙著「調査統計入門」(共立出版)92-97頁に詳しい説明があります。)なお、抽出間隔を小数2位まで取るなら、抽出起番号も小数2位まで取ってよいのですが、ここでは簡単のため、整数としました。

(6) (標本学級の抽出)表-6の「生徒数の累計」欄をみて、これらの抽出起番号を超える最も近い累計値をさがし、それに対応する学級を各副標本の最初の標本学級とします。この例では、「生徒数の累計」欄の515, 224, 368, 132の累計値があたり、従って、第1中の7組, 1組, 4組, 茨城大付中の3組が各副標本の初めの標本学級になります。(表-6の「副標本別標本学級」欄に○印をつけます。)他の標本はこうしてきまった4個の抽出起番号に676.16を順次加えて得られる番号からさきの要領で「生徒数の累計」欄の累計値をさがし、その累計値にあたる学級を選び出すことによって得ることができます。例では、第3中の2組, 第2中の6組, 9組, 4組などとなります。こうして、市部の3年の学級が25学級ずつ4系列, 計100学級抽出されます。これと同じ作

業は市部の1, 2年および郡部の1, 2, 3年の別に行いますから、結局全県で600学級が抽出されることになります。

(7) (生徒の抽出)調査する生徒(第2次抽出単位)の数は、各標本学級において一律に10人としました。一律にすると、第1次抽出単位を学級の大きさに比例した確率によって抽出したこととの関係で、集計が便利になるからです。(一律にしないときは、集計時に学級別の結果に係数をかけて合算するという操作が必要となります。)生徒の抽出は、生徒の席番号による等間隔抽出法とします。抽出間隔は、生徒数の十位の数(48人なら4, 50人なら5, 37人なら3など)とし、表-6の「生徒の抽出間隔」欄に記入していきます。抽出起番号は標本学級ごとに無作為にきめます。乱数表から、その学級の生徒数を超えない整数を拾い出し、表-6の「抽出起番号」欄を埋めていきます。標本となる生徒は、抽出起番号と、それに順次生徒の抽出間隔を加えた番号にあたる生徒とします。抽出間隔を順次加えていった番号が、生徒の数を超えたときは、その番号から生徒数を差引いた番号にあたる生徒を標本とします。そして標本が10人となったところで打ち切ります。たとえば、茨城大付中の3組は生徒数が45人ですから、抽出間隔は4, 抽出起番号は18となったとして、18に4を順次加えていけば22, 26, 30, 34, 38, 42, 46, と46で初めて生徒数45を超えますから、46は46-45=1によって1と書きかえ、以下5, 9, 13, と全部で10人になったところで打ち切ります。これと同様な作業を600個の標本学級について行くと、計6000人の生徒が抽出されることになります。なお、たとえば45人の生徒の中から10人を等間隔抽出するのに、抽出間隔を4とし、抽出起番号を1~4の中の1つの番号として、これに順次4を加えていった番号を標本としても、だいたいよいのですが、末尾の若干名に抽出の機会が与えられないので、標本が偏るおそれがあります。また抽出間隔を5としてもすっきりせず、やはり偏るおそれがあります。上の方法では生徒の抽出される確率はだいたい同じですから、偏りの心配はありません。

●シリーズ「短期統計実務講座」*****

表-6 標本抽出作業表

(市部・3年)

学校記入欄			県 記 入 欄							
学校名	学級	生徒数	生徒数の累計	第1次抽出←→第2次抽出				抽生徒の 出 間 隔	出 生 徒 番 号 抽	標本となる生徒の番号
				副標本	別標本	学級	学級			
茨城大付	1	42	42							
	2	45	87							
	3	45	132				○	4	18	18, 22, 26, 30, 34, 38, 42, 1, 5, 9
	4	44	176							
第 1	1	48	224		○			4	39	39, 43, 46, 2, 6, 10, 14, 18, 22, 26
	2	50	274							
	3	47	321							
	4	47	368			○		4	3	3, 7, 11, 15, 19, 23, 27, 31, 35, 39
	5	48	416							
	6	49	465							
	7	50	515	○				5	22	22, 27, 32, 37, 42, 47, 2, 7, 12, 17
	8	48	563							
第 2	1	47	658							
	2	47	705							
	3	49	754							
	4	48	802				○	4	45	45, 1, 5, 9, 13, 17, 21, 25, 29, 33
	5	47	849							
	6	48	897		○			4	8	8, 12, 16, 20, 24, 28, 32, 36, 40, 44
	7	47	944							
	8	47	991							
	9	47	1038			○		4	12	12, 16, 20, 24, 28, 32, 36, 40, 44, 1
	10	49	1087							
第 3	1	48	1135							
	2	46	1181	○				4	19	19, 23, 27, 31, 35, 39, 43, 1, 5, 9
	3	47	1228							
七 重	1	35	16868							
	2	36	16904							
計	360 学級	16904 人	-	25 学級	25 学級	25 学級	25 学級	-	-	1000人

(学級数と生徒数は仮想)

13. 実地調査

標本となる学級と生徒の番号がきまったら、調査の連絡網に従って調査票等の用品と共に学校に送り、調査を依頼し返送を待ちます。学校では生徒を通じて、保護者に記入を依頼します。

記入済の調査票が県に集まったら、学級ごとに、全部の調査票に、指定されている副標本番号を印します（ゴム印が便利）。

14. 集計

結果数字の集計は副標本の別、層の別に行い、全県の結果はそれらを統合して出します。そして、標本誤差は副標本別の結果を利用して、結果数字と同時に算出します。よって同じ項目の数字は少くとも4（副標本）×6（層）+1（総計）+1（標本誤差）=26通り計算されることになります。このほかに中間計が必要なら更に若干通り増えます。たとえば、市部、3年の第1副標本についての集計は次のようにして行います。1つの集計項目をX（例、その生徒の1週間の家庭における学習時間数）とすると、市郡・学年・副標本の区分が同一な標本250人について単純平均すると、その区分の平均値となります。（式-1参照）

式-1 市郡別学年別副標本別調査項目の平均値の算式の構造

$$Y = \frac{1}{25} \cdot \frac{1}{N} \sum_{i=1}^{25} \frac{1}{N_i} \sum_{j=1}^{n_i} X_{ij} = \frac{1}{250} \sum_{i=1}^{25} \sum_{j=1}^{10} X_{ij}$$

（平均レベルへの縮小項）
 標本の学合、級算、抽出率、抽出数、標本学級内の全生徒（保護者）数、標本学級内の合算、標本生徒（保護者）数=10、単純平均、調査項目の値、標本生徒（保護者）番号

1つの副標本についての全県の数字は、その副標本の市郡・学年の区分ごとの生徒数をウェイトとした加重平均によって求めます。（式-2参照）

式-2 全県全学年全標本における調査項目の平均値の算式（添字は市郡・学年の別を表わす）

$$Z = \frac{1}{104116} \sum N_{ij} Y_{ij} = \frac{1}{104116} (17037 Y_{市・1} + 16990 Y_{市・2} + 16904 Y_{市・3} + 17742 Y_{郡・1} + 17728 Y_{郡・2} + 17715 Y_{郡・3})$$

全県全学年全標本についての数字は、4つの副標本ごとの数字を単純平均して求めます。（式-3参照）

式-3 全県全学年全標本における調査項目の平均値の算式（添字は副標本の別を表わす）

$$\bar{Z} = \frac{1}{4} (Z_1 + Z_2 + Z_3 + Z_4)$$

15. 標本誤差の計算

4つの副標本は、いずれも母集団を代表するように独立に抽出されていますから、式-2によって算出される副標本別の4つの推定値はみな母集団の同一の値に対する推定値となります。それらは一般に多少異った値をとりますが、それはこれらが標本誤差を持っているためです。

1つの副標本による推定値の分散は、4個の副標本の推定値を使って、いわゆる不偏分散の形で計算されます。（式-4参照）

式-4 1つの副標本による推定値の分散

$$\frac{1}{k-1} \sum_{i=1}^k (Z_i - \bar{Z})^2 = \frac{1}{3} \sum_{i=1}^4 (Z_i - \bar{Z})^2$$

↑
副標本の数
k = 4

就任にあたって

統計課長 鷺見 丈

このたび6月1日付の異動で統計課長を命ぜられました鷺見(すみ)でございます。

いまさら私が申し上げるまでもなく、わが国の統計は戦後30数年の歴史を経て著しく改善整備され、今日に至っております。

現在では、数多くの統計が国・地方公共団体の行政分野で、それぞれの施策の企画立案、効果の測定等の基礎資料として重要な役割を果たしております。また、民間企業の経営方針の策定、資金の決定、学校教育における教材、家庭における生活設計に至るまで広範な分野で利用されております。

さて、最近の国や地方公共団体さらには民間企業においても、オイルショック以来低迷をつづける経済環境の悪化により、財政的・経営的に窮地に陥り、減量経営に転換せざるをえなくなっております。財源の伸びみやみの中で高度経済成長期にのびきったままの多様な行政需要に対応していくためには、正に厳しい施策の選択が余儀なくされております。

このような社会経済の変化に対応して、統計に対する需要はますます増大の傾向にあり、その内容も一層多様化することが考えられます。統計調査もその種類の増加、内容の複雑かつ高度化が要求されてくるものと思われれます。

統計調査の種類の増加、内容の複雑高度化の進行は、調査の実施面においていくつかの難しい、かつ深刻な事態を招来することになりはしないかと考えております。すなわち、調査客体である個人・法人などの負担が増大することとなります。

現在でも、調査拒否あるいは回収期限までに記入してもらえない、また、統計調査員が中々みつからないなど、いわゆる統計調査環境の悪化ということが言われております。統計調査員の確保とその処遇改善あるいは調査対象者に対する協力依頼の考え方など、調査実施上の基礎的な問題点について、今ひとつ円滑な運用ができるよう改善すべきことがあるようです。

現在の統計調査が、市町村とりわけ第一線統計調査員の並々ならぬ努力と奉仕の精神により調査の実を上げている現状を思うとき、十分配慮しなければならない問題であり

ます。

また、一方では、統計調査に対する県民の意識改革を大々的に啓発していく必要があります。調査の実施上の改善も重要ですが、調査の円滑化を図るためには、県民の統計調査に対する重要性の認識と協力意識がなければなりません。それがなければ、正確な信頼できるデータの収集も困難であります。したがって、あらゆる広報媒体の協力と活用によって、広く県民各層の理解と協力をうるよう更に一層のPRに努めなければならないと考えます。

統計需要の増大に対処するためには、統計調査の重複を除き、調査内容などの簡素化・合理化を図るとか、調査技術の開発による報告者負担の軽減を図るなど調整・改善を検討しなければならないと考えます。

各種統計調査については、統計部門の所掌する統計調査(殆んど国の委託調査)の間で、また、他の業務部門で実施する調査との間において、重複しているようなものがあるはしないかどうか。これは統計調査結果の情報の集中化・一元化と併せて今後検討されなければならない問題であります。

幸い本県では、いよいよ本年3月から電算組織によるデータ・ベース・システムの運用が開始されました。これが今後ますます充実されて、総合的な情報の一元化とその多目的利用について検討を加え整備されるならば、多大な効果と機能を発揮することが期待できると思われれます。

最後に、統計調査結果の公表について、もっと意を配らなければなりません。多数の県民の協力によってでき上がった調査結果であり、行政施策の基礎資料として役立てることはもとより、広く県民の利用に供することを怠ってはなりません。県の行政資料室は正にそのために設置されているものであり、目でみる統計情報として整備充実を図ってまいりたい。

本年度は、10年毎に行われる1980年世界農林業センサスをはじめ30の統計調査が行われます。県民の皆様をはじめ関係各位の御協力をお願い申し上げます。

何分にも着任早々で、思いついたまま申し述べましたが、よりよい統計づくりに全力を尽くす所存でありますので、各位の御指導と御協力をお願い申し上げます。