



ブドウ

子供の頃、冒険物語に夢中だった。自分が主人公になりきって、波瀾万丈の活躍をするのである。『オデュッセイア』の児童版など繰返し読んだものだ。オデュッセウスが巨人キュクロプスの洞窟にとじこめられたとき、彼は巨人をブドウ酒で酔わせ、牝羊の腹にかくれて脱出する。なにが気に入ったのか、この場面は今でも覚えていて、ブドウを摘む季節になると、昨日のこつのように想いだす。

オデュッセウスのブドウ酒は悪酔いするが、ブドウ酒には好ましい効用がある。長期にわたって飲めば、脳に栄養がゆきわたり、精神活動を衰えさせないらしい。これとは対照的に、ビールには精神をおだやかにする効用があると言う。もち論、酒も度をこせば「気違い水」にちがいない。

九月ともなれば、夏の終りをまちかねたブドウは、日一日とその輝きをましていく。

9月のおもな行事

- 3～7日 学校保健統計調査審査会(水戸市・日立市・鹿島町・土浦市・下館市)
- 5～6日 農業基本調査3県打合せ(群馬県)
- 10～11日 関東ブロック県単統計調査研究会(長野県)
- 12～13日 統計グラフ県審査(水府荘)
- 14～15日 第6次漁業センサス事後調査ブロック会議(神奈川県)
- 25～26日 特定サービス業実態調査ブロック会議(静岡県)
- 26～27日 関東ブロック統計刊行物研究会(埼玉県)
- 31日 消費動向調査基準日

標 本 設 計 (5)

.....

——想定調査；県政に関する住民意識——

今月は、前号で説明した国勢調査の調査区関係資料を利用して、県政に関する住民意識調査を想定し、その標本設計を行ってみます。本当の標本調査の設計資料を整えることは大変なことですが、1度しておけば後で類似の調査に使えますから、面倒でも得策です。

1. 調査目的

県政に関する諸事項について、県民の意識を把握することとします。

2. 調査の地域的範囲

昭和50年国勢調査調査区の後置番号が5, 6, 7, 9の地域を除く県の全域とします。

3. 調査単位

20歳以上の個人とします。

4. 調査票の流れ

次の系統によることとします。

県 ←→ 市町村 ←→ 調査員 (面接聴取) → 世帯員 (個人)

5. 抽出単位

昭和50年国勢調査の調査区または単位区を第1次抽出単位、その中に住む調査単位(個人)を第2次抽出単位とします。(注. 第2次抽出単位を世帯とし、個人を第3次抽出単位とする方法も考えられ、その方が抽出作業は楽です。しかし精度の面で損失をきたすおそれがありますので、多少面倒でも世帯を抽出単位としない方がよいと思います。また、単位区というのは、前号では紙面の都合上説明できませんでしたが、大きな準世帯からなる調査区(分類符号02に属する調査区)などを、並みの大きさになるようにいくつか分割したものをいいます。抽出単位としては当然、より望ましいものとなっています。以下では、単位区と調査区を一括して調査区と呼ぶこととします。)

6. 集計時の電子計算機使用の可否

調査結果の集計には電子計算機が使用できるものとします。これによってきめの細かい集計計画や誤差計算が可能になります。

7. 標本誤差の計算方針

相互貫入標本方式(副標本方式)により5つの副標本を作って行うこととします。

標本誤差の計算は面倒なものですから、電子計算機が普及していない時代では、なかなか実行できませんでした。しかし現在では普及がかなり進んでいますから、誤差計算をしようと思えば、できるわけですが、現実には誤差計算

に対する理解が必ずしも十分でなく、あまり実行されていないようです。

1つの標本設計に対する標本誤差の算式はいくつかありますが、すべてが現実的であるとは限りません。たとえば、層化2段抽出の場合ですと、1段目の抽出による誤差分と2段目のそれを合わせた形の公式が理論書に出っていますが、実際にはこれにあてはめる数値的材料に乏しく、適用は困難です。標本誤差の現実的な計算には、副標本方式が一番便利のようです。

8. 標本の抽出確率

調査区については、その調査区に付けられているウェイトに比例した確率に従って抽出し、個人については、その調査区のウェイトの逆数に比例した確率に従って抽出することとします。こうすると、全体の抽出確率の種類が少なくなり、集計が容易になります。

9. 調査区の層化

分類符号別の調査区一覧表を基本とし、これを応用して層を作ることとします。こうして作る層がこの調査に対する最善のものになるかどうかわかりませんが、住民の生活上の種々の立場も反映しながら調査区を分類したものですから、有望であるといえましょう。

茨城県の場合の基本表は表-1のようになります。分類符号別の調査区数やウェイト数を見ると、少な過ぎて1つの層を設けにくいものもあるので、なるべく類似のもの同志でくり、表の太線のように仕切って組かえ、層とすることとしました。

(注. この分類符号による区分のほか、住民意識に関係の深そうな既知の特性があれば、なるべくそれを層化の指標に取り入れるのもよいと思います。たとえば、調査区の近辺の市街化の程度は、都市対策に関する住民の意識にかなり関係があると考えられますから、調査区をその前置番号によって人口集中地区、準人口集中地区、その他の地区の3種に分けることは意味があります。しかし、この3種の調査区一覧表あるいはこの3種と前記の分類符号をクロスした調査区一覧表は既製のものがありませんから、必要なら新たに作らなければなりません。本県の場合では、約1万2千個の調査区を再点検して、3種の別に調査区数やウェイト数などを集計する作業となり、短時間では困難でしょう。そのためここでは取り上げませんが、時間をかけてもやっておけば将来の標本設計に役立ちます。)

表-1 層別基本数一覧表

層	分類符号	調査区数 (単位区数)	ウェイト数	1副標本あたりの抽出 調査区数 (単位区数)	抽出調査区 (単位区)数 の合計
1	01	27	27	0	0
2	02	208	889	1	5
	0251	(28)	106		
	0252	(49)	201		
	0253	(38)	103		
	0254	(0)	—		
	0255	(6)	20		
	0256	(11)	331		
	0257	(1)	7		
	0258	(5)	20		
	0259	(17)	58		
	0260	(0)	—		
0261	(12)	43			
3	計	303	885	1	5
	03	77	77		
	05	34	115		
	06	158	583		
	07	8	26		
	08	26	84		
4	09	413	1571	2	10
5	計	1434	4423	5	25
	04	200	200		
	10	1147	3963		
	11	87	260		
6	12	504	1537	2	10
7	13	3881	13647	16	80
8	14	1259	5358	6	30
9	15	2274	9171	10	50
10	計	1557	6123	7	35
	16	1473	5786		
	17	84	337		
計		11860	43631	50	250

昭和50年国勢調査調査区関係資料利用の手引シリーズNo.2第4表、第6表にもとずいて作成。分類符号03を第3層に含め、04を第5層に含めたのは、集計の算式の簡素化のための便宜であり、本来は避けたいところ。

10. 標本の規模

確たる算出基礎はありませんが、1副標本について50調査区、各調査区について10~15人の個人を抽出してみることにします。すると全標本で250調査区、約3000人となります。3000人規模の世論調査はよく行われております。精度が十分であるかどうかは、標本の設計方法や結果表の細かさなどによって一概にいえませんが、あまり細かな集計をするのでなければ、まず利用できるものと思われま。全国レベルの調査でも3000人規模は並みの規模です。とにかく、調査結果の精度は標本から計算してみることにします。

11. 抽出調査区数の層別配分

各層に割りあてるべき抽出調査区数は、集計の便宜のため層のウェイト数に比例的にきめることにしました。表-1の「1副標本あたりの調査区数」欄が各層への配分基本数です。

12. 調査区の抽出単位名簿

層内の調査区の配列順は、分類符号、市町村、調査区番号の順とし、これを抽出単位名簿とします。この配列順の精度への効果はどの程度かわかりませんが、管理の上で便利です。各調査区にはウェイトが示されていますから、調査区ごとに配列順に従った累計値を計算し記入します。累計値の最後はその層のウェイトの総数に一致します。(表-2参照)

13. 調査区の抽出

調査区の抽出は調査区のウェイトを等確率系統抽出することによって行います。抽出間隔は

$$\frac{\text{層内のウェイト数}}{1 \text{ 副標本の抽出調査区数}}$$

〔例〕第4層の場合

$$\frac{1571}{2} = 786$$

とします。そして抽出間隔を越えない数を乱数表等によって独立に5個(整数でよい)選び、それぞれを5つの副標本の抽出起番号とします。(乱数表の使用法は拙著『調査統計入門』(共立出版)に詳しく説明してあります。)

〔例〕第4層の場合

544, 576, 310, 109, 768

抽出調査区は、抽出起番号とそれに抽出間隔を順次、所要の回数だけ加えた番号のウェイトに対応する調査区とします。

●シリーズ「短期統計実務講座」*****

表-2 第1次(調査区)抽出単位名簿

第4層

調査区	ウェイト	ウェイト の累計	副 標 本					備 考
			1	2	3	4	5	
水戸市								
0037	4	4						
0089	2	6						
0373	3	9						
0375	3	12						
0414	4	16						
0419	3	19						
0428	3	22						
0430	4	26						
0441	2	28						
0474	4	32						
0479	3	35						
0514	5	40						
0613	4	44						
0726	3	47						
0759	3	50						
0760	6	56						
0769	6	62						
……	……	……						
他市町村につづける	他市町村につづける	累計をとる						
		全県通して						
		1571						
計	1571	—						

第4層は分類符号09「換算世帯数が16以上で、15歳以上就業者中に占める建設業および製造業の自営業主数の割合が1割以上の調査区」からなり、茨城県全体で413調査区、1571ウェイトあります。(昭和50年国勢調査結果による標本調査基礎資料、表1-1「分類符号別通し番号による調査区一覧表」により作成)

(例) 第4層の場合

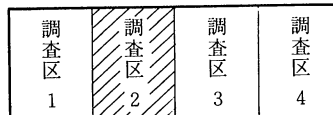
副標本別抽出ウェイト番号

副標本番号	1	2	3	4	5
抽出ウェイト	544	576	310	109	768
番号	1330	1362	1096	895	1554

このようにして、すべての層から所定の数の調査区を抽出したら、各調査区の所在地を確かめ、調査の始まる少し前に調査区要図の写を持って調査区内の踏査をします。そして、調査区内のすべての世帯を訪問して調査対象となる人のリストを作り、第2次抽出のための抽出単位名簿とします。(注。調査区の所在地の名称(〇〇番地など)だけをたよ

りに、そこに住む人々のリストを作ることは困難です。役場にある住民基本台帳によるとしても、調査区の範囲に住む人だけをえり分けるのは不可能ですから、調査区要図の写を持たないときは、調査区の地理的範囲を厳密に考えないで、便宜上調査区の所在地として表わされている番地などに住む人々を住民基本台帳から抽出する、という方法で済ませることもあります。そのときは、調査区の範囲と、実際に調査する範囲がずれることがあります。(筆者の経験では、このずれはよく起こります。)ずれが大きくなると先に定めた調査区の抽出確率や調査区内の個人の抽出確率の意味が薄くなってしまいます。)

(例) 同一番地内に4つの調査区がある場合



(調査区2が抽出されても、実際に調査されるのは調査区1~4となる。)

調査区要図の写を持って調査区内のすべての世帯を訪問し、調査対象となるべき人を識別して名簿に記載する作業には一苦労がありますが、これをなしてこそ計画通りの調査となるものです。名簿の様式は表-3のようにするのが適当でしょう。

14. 個人の抽出

個人の抽出は、調査区内を踏査して作った名簿(表-3)にもとづいて、等間隔系統抽出によって行います。抽出間隔はウェイトの3倍としてみました。何倍でもかまいませんが3倍とすると次のような利点があります。すなわち、ウェイトの1調査区あたりの平均が約4、20歳以上人口の平均が約150人弱であることから、ウェイトを3倍しますと平均約12となり、これを抽出間隔とすると多くの調査区

表-3 第2次(個人)抽出単位名簿

(市町村)

調査区番号	所在地	層	ウェイト	抽出間隔	修正倍率	副標本番号
番号	氏名	住所	男女	年齢	備考	
1						
2						
3						
4						
5						

と10~15人ぐらゐの個人が抽出されることとなつて、本番の調査の作業量が適当になるというわけです。もちろん、1調査区の抽出人数がもっと多くてもよければ、ウェイトの2倍とするのもよし、ウェイトをそのまま抽出間隔としてもかまいません。

抽出起番号は、抽出間隔の範囲内で乱数表等によって1つの整数を選ぶこととします。そして、抽出起番号と、それに抽出間隔を順次加えた値を名簿の最後まで求めていき、その番号にあたる個人を標本とします。(表-3の番号を○でかこむのが便利です。)抽出作業は、名簿の最後まで行ったら終りとし、先頭にはもどりません。

こういう方法で抽出しますと、標本となる個人の数は、多くの場合1調査区あたり10~15人ぐらゐになりますが、調査区によってはこの範囲から大きくはみ出すこともあります。たとえば国勢調査の時点での換算世帯数が50ぐらゐの調査区はウェイトが4となっていますが、その後この調査区の世帯数が100に増えたのに、ウェイト4として抽出しますと30人ぐらゐ抽出されるかもしれません。これを15人ぐらゐにするには抽出間隔を標準の倍に、つまり抽出確率を標準の半分にする必要があります。表-3の「修正倍率」はそのための欄です。

15. 集計および誤差計算

標本に対する本調査が終れば、調査票を集めて集計をはじめますが、調査票ごとに副標本番号、層番号、ウェイト、修正倍率が識別されていなければなりません。集計の算式は次のようになります。

式-1 1つの副標本における比率の算式

$$P = \frac{1}{N} \sum_i \frac{1}{m_i} \sum_j \frac{1}{w_j} \cdot 3w_{ij} \sum_k X_{ijk}$$

- X: 調査項目
- k: 標本となった個人の番号
- j: 標本調査区番号
- i: 層番号
- w: ウェイト
- W: ウェイトの合計
- m: 標本調査区数
- N: 県内20歳以上人口

- 調査区内合計の推定値
- 層内合計の推定値
- 総平均の推定値
- 総合計の推定値

この式は繁雑ですが、 w_{ij} が消えること、 m_i/W_i をほぼ一定としたこと、 \bar{n} を1調査区あたりの標本(個人)数とすれば $N \approx 3W\bar{n}$ となることを利用して整理すると結局

$$P \doteq \frac{1}{n} \sum_i \sum_j \sum_k X_{ijk}$$

(ただし n は1副標本の標本(個人)総数)となり、単純平均の形となります。なお、抽出間隔を修正した調査区を集計するに際しては、その修正倍率を r_j とし、上式において

$$\sum_k X_{ijk} \text{の代りに } r_j \sum_k X_{ijk}$$

を用いることとなります。修正倍率を用いる調査区の数が少ければ、 n は変えなくても支障はありません。

これによって同じ項目について5つの比率が計算されますが、これらを用いて全標本による比率は次のようにして求められます。

式-2 全標本による比率の算式

$$\bar{P} = \frac{1}{5} \sum_{s=1}^5 P_s$$

\bar{P} に対する標準誤差は、各副標本の P_s を用いて、次の式によって算出されます。

式-3 全標本による比率の標準誤差

$$\sqrt{\frac{1}{5(5-1)} \sum_{s=1}^5 (P_s - \bar{P})^2} = \sqrt{\frac{1}{20} \sum_{s=1}^5 (P_s - \bar{P})^2}$$

この式で計算される標準誤差は、 \bar{P} で割ると相対標準誤差(変動係数)になります。いずれにしても、たくさんの項目についての誤差を計算すると、標本の大きさや比率の大きさに応じた誤差の傾向を知ることができます。一般には標本数が大きくなると、誤差は小さくなる傾向があります。比率の大きさと関係では、標準誤差は比率が0と1に近いほど小さく、相対標準誤差は比率が1に近いほど小さくなるという性質があります。

また、式-3によって計算した誤差の値を、無層化1段単純任意抽出法による誤差 ($\sqrt{\frac{P(1-P)}{5n}}$) とくらべてみるのは、実際の標本設計の効果を検討するのに参考になります。経験によれば、層化2段抽出法は無層化1段単純任意抽出法よりは精度が悪いようです。層化のメリットと2段化のデメリットが重ってなおデメリットが残ったものを考えられます。

上記の抽出方法は、一般に層化2段抽出法と呼ばれるものです。おおざっぱに書きましたが、実際にはもっと細かい配慮と工夫をしようところが多々あります。ここでは基本的な考え方にとどめました。

データ・ベース・システムについて……

茨城県では、総務部電子計算課、企画部企画調整課、企画部統計課の3課が共同して、昭和53年度に「茨城県データ・ベース・システム」を開発し、54年度から利用を開始しています。このシステムの概要と問題点等について、以下に述べてみたいと思います。

1. データ・ベース・システムとは

データ・ベース・システム（以下DBSと略す）とは、行政を進める上で必要な情報をコンピュータの中に蓄積し、必要に応じてこれを取り出して利用しようとするものです。

ところで「行政上必要な情報」といっても、いろいろなものが考えられます。例えば、県内の道路施設に関する情報を蓄えておいて道路整備に役立てることも考えられますし、職員の人事管理なども考えられます。

今回茨城県で開発されたDBSは「予測と計画のためのDBS」です。即ち、各種の行政計画を策定する際に必要なデータ（人口、所得など）を中心に蓄積してあります。また、利用の方法も計画策定上のデータ分析に便利ように考えられています。しかし計画策定以外の目的に利用しても一向に差し支えありません。

DBSを使ってどんなことができるのかということを表-1にまとめてあります。この中で基本的な機能は数表作成です。利用者の指示に従って、蓄積データの中から必要なデータを取り出し、必要に応じて加工もして編集し、用紙にプリントします。

表-1 データ・ベース・システムの機能

数表作成	編集・加工は自由自在。
グラフ作成	横棒グラフ、横折れ線グラフの2種類。数表の一部をグラフ化する。
地図作成	茨城県の市町村別濃淡地図をコンピュータが描き出す。
統計解析	

① 時系列分析	過去の時系列データをもとに将来予測を行なう。予測手法は17種類。
② 重回帰分析	1つの現象を他のいくつかの現象によって説明する。
③ 重相関分析	いくつかの現象間の因果関係を分析。

2. データ・ベース・システムの内容

(1) 蓄積データについて

昭和54年4月1日現在で蓄積されているデータは102項目、約53万件になっています。主な項目を表-2に示します。これらのデータは原則として過去10年分のものが時系列的に蓄積されています。

毎年発生する新しいデータについては、年数回にわけて追加蓄積するようになっていきます。また、利用者のニーズを調査して、新しい項目を追加することも考えられています。

一方では、データの細かさについても、常に検討される必要があります。例えば、現在市町村別の人口は年齢5歳階級別に蓄積されていますが、これを各歳別にする必要があるかどうかというようなことです。

しかしながら、利用者の要望に従ってどんなデータでも蓄積してしまうことには問題があります。システムの管理運営やコンピュータの容量などを考え合わせると、利用頻度の高いものを優先させざるを得ません。この意味で、蓄積データについては定期的な見直しが必要となります。

(2) 地域区分について

データは原則として県内市町村別に蓄積されていますが、重要なものについては都道府県別にも蓄積されています。

また、これらの基本区分を組み合わせた地域のくくりについても、各種のものが利用可能になっています。

(3) 加工機能について

四則演算の他に特定演算と呼ばれる加工機能があります。

これは、構成比（百分率）や伸び率のように、利用頻度の多い加工を簡単にできるようにしたものであり、使い勝手が非常によくなっています。

また、人口の多い市町村から順番に番号をつけたり、対数関数や指数関数も使用できるようになっています。

(4) 編集機能について

数表を作成すると、用紙にプリントされてきますが、行間隔を適当にあけたり、1 ページ内の行数や列数を指定したりすることができます。これによって見やすい表を作ることができます。

3. 利用効果について

D B Sを利用して効果が上がる場合と、さほどでない場合とがあります。ここでは効果が上がると思われる例を2、3示してみます。

(1) データの時系列変化を知りたいとき

一般に統計書は1年分が1冊になっていますので、10年間の変化を調べるときは、10冊の統計書を用意して各年の該当する場所を拾い集めなければなりません。この点D B Sは10年分のデータが一括して蓄積してありますので非常に便利です。

(2) 大量の加工を伴うとき

特定の地域についてデータを集計したり、加工量が多い場合はコンピュータにやらせるべきです。人間がやると必ず間違いです。

(3) 統計的な分析をする場合

最小二乗法など統計解析は一般に計算量が多く、コンピュータに最適の仕事です。

4. 利用対象と手続き

現在のところ、D B Sの利用対象は本庁及び出先機関の県職員（教育庁、県警本部を含む）に限っています。しかし、将来は一般に開放することも検討しています。

さて利用希望者が踏む手続きは以下のようになっています。まず、「利用申込書」と「リクエスト用紙」に必要事項を記入します。これらの用紙は電子計算課と企画調整課に置いてあります。また記入に当っては『利用の手引き』と『コード表』を参照して下さい。これらは各課に配付してあります。

2枚の書類は係長決裁を済ませて電子計算課に申込みます。電子計算課ではD B Sの操作をして、申込のあった翌日までに原課に連絡があります。

5. 課題と展望

D B Sに関する当面の課題としては、P Rの問題と管理運営体制の問題があります。

P Rについては、企画調整課が事務局となって「茨城県データ・ベース・システム研究会」を設置し、各部の代表の職員等にD B Sの利用法に関する説明会を開いています。が、今後は更に広い層にD B Sを理解してもらう努力が必要です。

管理運営体制については、現在のところ開発3課の共同体制をとっていますが、特にデータの保守管理に関しては問題点が多く、1日も早くD B S担当の組織の設置が望まれています。

将来の展望としては、データ蓄積方法の改善とオンライン化の検討があげられます。

現在、データを蓄積する場合、統計書をもとにカードをパンチして蓄積する方法がとられていますが、今後は集計が機能化されている統計調査業務については、集計結果テープからD B Sへデータを蓄積する方法へ転換することが検討されています。

また、データが欲しい時に次の日まで待たされることを解消するために、データを端末機のブラウン管上に表示させるオンライン化についても検討が進められています。

◆統計の窓◆

表-2 蓄積データの種類

1. 資	源	—土	地	—行政区画面積等	3,629件												
2. 人	口	—総	人	口	—夜間人口, 昼間人口	37,434											
				—常住人口	32,544												
		—就	業	人	口	—常住地就業人口, 従業地就業人口	31,611										
		—世	帯	—普通世帯数, 普通世帯人員等	5,859												
				—常住人口世帯数	1,023												
		—人	口	動	態	—出生数, 死亡数	6,012										
				—転出入人口	7,440												
3. 産	業	—第	1	次	産	業	—農業粗生産額, 生産農業所得, 農家数, 農家人口等	30,906									
		—第	2	次	産	業	—製造品出荷額等, 製造業事業所数, 工場敷地面積等	84,562									
		—第	3	次	産	業	—年間商品販売額, 商店数等	23,860									
						—道路延長等	8,235										
		—事	業	所	—事業所数, 従業者数	4,836											
4. 経	済	—総	支	出	—名目総支出, 実質総支出	4,456											
		—分	配	所	得	—所得の分配	3,747										
		—純	生	産	等	—純生産, 法人所得等	5,490										
5. 福	祉	・環	境	—人	口	動	態	—出生数, 死亡数	38,711								
				—医	療	—病院数, 医師数等	15,345										
				—住	宅	—住宅数等	7,488										
6. 財	政	—国	財	政	—一般会計歳出入決算額	22											
				—特	別	—	22										
				—そ	の	他	—	22									
		—県	財	政	—普通会計歳出入決算額	8,670											
				—一	般	会	計	歳	出	入	決	算	額	—	345		
				—税	収	入	済	額	—	196							
		—市	町	村	財	政	—普通会計歳出入決算額	8,928									
				—税	収	入	済	額	—	7,586							
		—公	共	投	資	—行政投資実績額(国)	3,168										
				—普	通	建	設	事	業	費	(県)	—	3,455		
				—普	通	建	設	事	業	費	(市	町	村)	—	2,992
7. 教	育	・文	化	—幼	稚	園	—幼稚園数, 幼稚園児数等	29,219									
		—小	学	校	—公立学校数, 児童数等	27,807											
		—中	学	校	—公立学校数, 生徒数等	24,273											
		—高	等	学	校	—学校数, 教員数, 全日制生徒数	47,807										
8. 安	全	—犯	罪	—犯罪認知件数, 犯罪検挙数	1,061												
		—交	通	事	故	—事故発生件数, 死亡者数等	2,232										
		—火	災	—火災発生件数, 死亡者数等	6,696												
					蓄積データ合計	527,689件											

(注) 蓄積データ合計は、合計値、小計値も含む