



自動車

つい最近までマスコミによつて、テレビ、電気冷蔵庫、電気洗濯器が三種の神器などともてはやされ、私達消費者もこれに眩惑されこれらを一応とりそろえないと一人前でないような気がして、せつせと月賦で買い込んだものである。しかしこの頃では三種の神器は過去のものとなされ、今や、ステレオ、ルームクーラー、カーという超デラックス消費ブームを起さんと、新聞、ラジオ、テレビなどいろいろの媒体を通してマスコミが、またも私達をおおつております。

それに影響されたわけでもあるまいが、近頃役所でも自家用車で通勤している人がいるとか。

日本にどれくらい自家用車を持つている人がいるかという、経済企画庁が37年2月に、全国の都市世帯4,011を対象に実施した。消費者動向予測調査結果によれば、自家用車の普及率は5.1%で、20世帯に一台の割合となつている、そして50.3%が月賦で購入しており収入階級では80万円以上の層で特に購入が多いとのこと。

私達も早く自家用車を買える身分になりたいものであるが、と同時に交通事故が今日大きな社会問題となつていることを考えなければならぬ。そこでまず自動車に関する統計を眺めておく必要があると思います、その意味で9ページをどうぞ。

統計の楽しみ

総理府統計局長

小田原 登志郎

わが国の統計思想は近年ことに向上したといわれる。その意味は、国民の統計関心が次第に高まって、統計に基づいてものを考え、策を立てる習慣がだんだん身についてきたということであろう。そのお陰で、官公庁の行なう統計調査に対する国民の理解も深まり、いろいろ問題はあるにせよ、統計申告についても大いに協力を得られるようになったことは、まことによろこばしい次第である。

統計を使つてものごとを考えてゆく一法として、身辺日常のことがらについて、自ら統計をとつて考えてみるなど、身にもつき、楽しくもある方法であろう。その意味で、一見些末のような身辺の経験、日常の現象を統計的に観察し、そこによくもの事の法則を見出して行かれた。いわばその道の草分けともいえる寺田寅彦先生のことを思い起さずにはいられない。透徹した科学者であり鋭敏な詩人でもあられた先生の業績を云々する資格はむろん私にはないが、試みに全集のところどころを拾い読みしただけでも、日常当面する何でもなような現象を統計的に明快に考察された興味つきない珠玉の隨筆を、数えきれぬほどたくさん見出すのである。

入学試験の成績について数学と語学の点数を調べて、両者の間に統計的相関があることを見出し、結局「二つの学科がこれを修得するに要する頭脳の働き方の上で本質的に互に共通な因子を持つているような事はないか。」との考えを進められたり（「数学と語学」——昭和4年）、「電車の中で試みに同乗の人々の顔を注意してみわたしてみると、あまり感じのいい愉快な顔はめつたに見当らない……顔全体としての表情が十中八九までともかくも不愉快なものである。」として結局「……からつ風と戦い砂塵に悩まされねばならない事に加えて、この満員電車で割り込んで踏まれ、押され、ののしられなければならないこと……これが毎日日課のように繰り返される間には、自分の顔のしわの一つや二つは増すに相違ない。」と論断されているのは、今日でも面白く十分に急所をついているということができよう（「電車と風呂」——大正9年）

なお、電車については、その混雑の週期的波動ないし律動についての、実測を伴つた、かなり難かしい統計的数学的論証がなされている（「電車の混雑について」——大正11年）。

先生が42才の厄年を迎えられた大正10年には、この厄

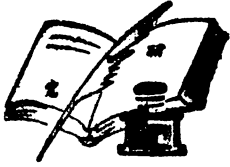
年というものに意味があるのか、ないのか、との疑問を自ら提出し、手近かな資料から各年令における死亡率の曲線を捜し出してみられたが、「曲線の示す範囲では42才における死亡の確率が特別に多くはないという漠然とした結論が得られそうに見える。しかし統計ほど確かなものはないが、また統計ほどうそをつくものもない……厄年の場合でも材料の選み方によつては、あるいは意外な結果に到着することがないものだろうか。」となおも疑問をつづけておられる（厄年とetc.）

英語やドイツ語の中にも日本語とよく似た音をもつた同義の語に出合うことを発見しそれがすべて偶然の暗合であるともいえぬとされて、一つの仮説をたててこれ統計的に数学的に吟味してゆかれた比較言語学上の一文も面白いものであるが（「比較言語学における統計的方法の可能性について」——昭和3年）。さらには日本独特の俳諧連句の分野において、連句心理の統計的分析を試みられているのには一驚させられる（「連句心理の諸現象」——昭和6年）。

さて、統計の楽しみは、このように、まず身近なことからについて、実測を重ね、資料を整理してみることから初まる。たとえば家計調査のようなものでも、家庭がこれに協力しつつ、同時に、家計簿をつけ、自分の世帯の統計をつくることによつて、他の家計と比べ又各月の状況を系列としてみるができるということになると世帯自身にとつても充分に興味と関心と呼び起すことになるにちがいない。個人企業経済調査なども、とくに中小個人業主の方々には、この調査に答えて、帳簿を整え売上高、必要経費の計算を自ら行つてこれを月々比較することにより、自分で絶えず自家企業の診断を続けてゆくことができるのではなからうか。

科学的管理法の普及に伴つて、自企業工場内の各種の統計をとつてみたいからとて、統計講習を希望し統計局に講師の斡旋を依頼される向きも少くない。その結果経営上の問題点を自ら発見して行くことは興味深く大切なことである。さらに製品の品質管理、市場調査を行うところまで進むことは、むしろ今日の企業の常識となつていくといつてよいであろう。

国民のすべての層において、統計づくりの楽しみを味わう域にまで到達し、真の意味の統計思想が国民的習性になる日もそう遠いことではないと思われる。



標本調査への手引(9)

総理府統計局 高橋史朗

第2部 標本選定の技巧(つづき)

7 なお残る標本選定の制約

2段抽出によつて、標本選定はどうかやら実用化されました。しかし、これはまだどうかやらという段階であつて十分にではありません。十分に実用化されるには、なお工夫によつて取り除かなければならない制約が残つております。この制約とは、一口に言えば、標本の散布状態が不安定であるということですが、具体例による方が説明しやすいので、また、ここでも、東京都で営業している飲食店の平均売上高をもとめるという場合をとりあげてみましょう。

制約3 まえに述べた2段抽出によりますと、1次抽出単位を地域、2次抽出単位を飲食店として、飲食店が選びだされますが、このままですと、調査地域が東京都内のどこから選びだされてくるのか、まったく予想がつかえません。いいかえると、調査地域が、たとえば、東京都内のどの区で、どの市で、どの町で、どの村で、それぞれどれだけ選びだされるのか、まったく予想がつかないわけです。それは、まったく、サイコロを実際に振つたときの目の出方によるわけですが、しかも、もしも、サイコロを振り直して、調査地域を選び直したとしますと、その散布状態は、また、まえとはすつかり変つてしまうはずでです。このように、このままの選び方では調査地域の散布状態が非常に不安定になりますが、このことが、次のように、標本選定の制約になります。

(1) たとえば、この標本調査を実施する本部は千代田区においてあるとしましょう。しかし、この調査を実施するには、本部だけでは東京都が広すぎるので、東京都のうち、区部(23区)は本部が直接に担当することにして、区部以外については、八王子市に支部において、そこが担当することにしたとしましょう。

さて、調査を実施するには、まず、調査票の印刷、調査員の手当、調査員の交通費、統計表の作成などに必要な費用を確保し、また、調査員を雇わなければなりません。ところで、この費用のうち、統計表の作成などは本部がおこない、また、調査票の印刷なども本部が一括しておこなつて、支部と分ければよいでしょうが、調査員の手当、交通費などについては、あらかじめ本部と支部とで分けておいて、それぞれが受け持つている調査員に

支給する方がむしろ便利であるとかんがえられます。しかし、さきのような選び方では、いくらずつを本部と支部とで分けておけばよいのか、この調査を企画したさいに決めておくことができません。なぜなら、本部と支部とでそれぞれ調査員を何人ずつ受け持つかは、区部と区部以外から、それぞれ、どれだけずつ調査地域が選びだされるかにかかつており、それは、さきに見たように、非常に不安定であるからです。したがつて、たとえば、仮に区部と区部以外とでそれぞれ営業している飲食店の数に比例して分けておいたとしても、いざ実施したさいには、その配分を大きく変えなければならない事態もおこりうるわけです。このような事態はできれば避けたいところです。

また、調査員についてですが、調査の成否がまず調査員の質にかかつていることは明らかですから、十分に慎重に雇わなければなりません。また、調査の内容を十分に理解してもらうように、しつかりと訓練することも必要です。したがつて、よい調査員を雇うには、どうしても、かなりまえから、たとえば、一応の企画がまとまつたあたりから準備を始める必要があるでしょう。ところで、準備をするには、本部と支部でそれぞれ何人の調査員が必要なのか分からないと困りますが、それがまえにも言つたように非常に不安定であるわけです。したがつてここからも、調査地域の散布状態の不安定さが、調査の大きな障害になつていくことが分かります。

(2) 企画のはじめに、調査地域の散布状態が非常に不安定であることは、確かにあまり好ましいことではありませんが、しかし、調査の企画と実施に若干の食い違いのあることは、ある程度やむをえないことでしょうからどうしてもどうにもならないというのであれば、我慢できないことはないかも知れません。しかし、もしも、この調査が長期にわたつて、続けておこなわれてゆくというのであれば、これはどうしても取り除かなければならない障害となつて現われてきます。

調査が長期にわたつて続けておこなわれてゆくということになれば、適当な期間ごとに調査飲食店を新しく選び直して交代してゆくことになりませんが、それも、はじ

めはおなじ調査地域のなかで選び直すとしても、長期にわたれば、調査地域も新しく選び直してゆかなければなりません。その場合、新しく選び直された調査地域が、これまでの調査地域の近くにくるとはもちろん限りませんし、むしろ、区部にあつた調査地域が選び直すと区部以外へゆき、また、逆に、区部以外の調査地域が選び直すと区部にゆくといった可能性も非常に高いわけです。このようなわけですから、したがつて、調査地域が選び直されるごとに、本部と支部とで受け持つている調査地域したがつて調査員の数は相互に変わり、それにつれて調査員の手当、交通費などの配分も変わつてゆくわけです。また、そうなれば、一般に、これまでの調査地域と新しく選び直された調査地域との距離はかなり遠いものになるでしょうから、これまでの調査員がひきつづいて新しく選び直した調査地域を担当してゆくことができなくなり、一方では、これまでの慣れてきた調査員に辞めてもらい、他方では、新しい不慣れな調査員を雇わなければならないとなつてしまいます。そして、それらはすべて調査地域の散布状態が不安定なためにおこつてくるのです。

8 東京都の地域のグループ分け

この調査地域の散布状態が不安定であるという標本選定の制約を取り除くには、次のような工夫がかんがえられております。すなわち、調査地域を東京都内のどこからでも直接に選びだすというのではなく、まず、東京都を区部と区部以外とに分けてかんがえ、選びだす調査地

域の数をあらかじめ区部と区部以外とに分けておき、次いで、区部と区部以外とから、それぞれに割り当てられた調査地域を別々に選びだすのです。(第7図参照)

このように工夫すれば、はじめ企画したさいに、それぞれ区部と区部以外とに予定した調査地域の数があつて変わるということはありません。それは調査地域を新しく選び直しても変わることはありません。したがつて調査地域の散布状態はかなり安定して、区部と区部以外とにたいする調査員の手当、交通費などの配分は常により一定になります。

なお、区部と区部以外とに分けるだけでは、調査地域を新しく選び直す時、区部あるいは区部以外といつてもその面積はかなり広いですから、これまでの調査地域と新しく選び直した調査地域との距離はかなり遠くなり調査員に辞めてもらつたり、雇つたりしなければならぬ危険がかんがえられますが、これはただ説明を簡単にするためにしたこと、実際には、もつと細かく東京都を分ければ、その危険はなくなるわけです。

なお、選びだす調査地域の数をどのように区部と区部以外とに分けるかですが、さしあたり、区部と区部以外とでそれぞれ営業している飲食店の数に比例させることにします。さしあたりというのは少し奇妙な言い方ではないかとおもわれるかも知れませんが、ここは、いろいろと技巧をこらせるところなので、それについては、また、あとで述べることにしたいとおもいます。

ここで、二三の用語について説明しておきましょう。まず、区部と区部以外とはそれぞれ層といい、また、東京都を区部と区部以外とに分けることを層化するというです。また、層化をもちいる標本選定を一般に層化抽出といいます。ここでは、2段抽出もあわせてもちいていますから、したがつて、ここでもちいている標本選定は層化2段抽出ということになるわけです。普通、層化2段抽出というときの「層化」は、1次抽出単位の層化を意味します。

さて、各層すなわち区部と区部以外とでの地域および飲食店の選定の方法は、まえとおなじように、それぞれ比例確率重複抽出および等確率非重複抽出によることにし、また、これもまえとおなじように、この6月の1地域当たりの平均従業員数および1従業員当たりの平均売上高をもとめることにしてみましよう。すると、これらの推定値の推定式と推定子は、それぞれ、次のようになります。

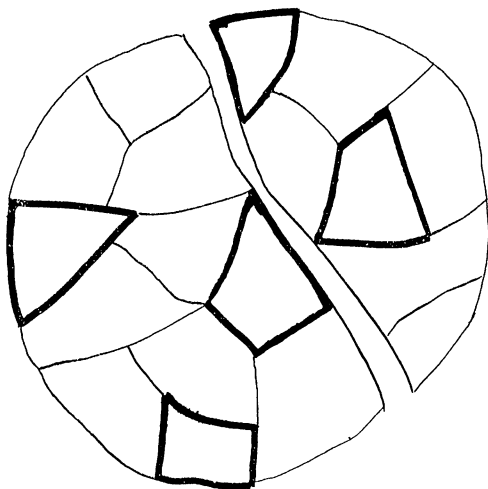
(1) 1地域当たりの平均従業員数の場合

まず、推定式は、まえと等しく、次のとおりになります。

$$\text{従業員数の算術平均} \times \frac{\text{飲食店総数}}{\text{地域総数}}$$

したがつて、このように層化をもちいても、推定式は

第 7 図



◻ 調査する地域

◻ 調査しない地域

とくに煩雑にはならないわけです。

なお、調査する6月における、それぞれの地域の飲食店の数があらかじめ分かっているということは、よほどの幸運に恵まれなにかぎり、期待できないことなので、地域の選定は、やむをえず、過去の資料によつて、「不

完全」におこなうしかありませんが、それを、それぞれの調査地域の調査飲食店数の修正によつて「完全」に切り替えますと、その調査飲食店数および推定式の修正はまえとおなじように、次のとおりになります。

$$\text{ある調査地域の調査飲食店数} = \text{その調査地域の「設計」調査飲食店数} \times \frac{\text{その調査地域の飲食店数}}{\text{その調査地域の「設計」飲食店数}}$$

$$\text{従業員数の算術平均} \times \frac{\text{調査飲食店数}}{\text{「設計」調査飲食店数}} \times \frac{\text{「設計」飲食店総数}}{\text{地域総数}}$$

次に、推定子ですが、いま、選びだす地域の数を ℓ として、推定子を \bar{U}_ℓ と表わしてみると、その期待値 $E(\bar{U}_\ell)$ はもとめようとする1地域当たりの平均従業員数に等しく、したがつて、第1部第10節(9月号)の条件1を満たしております。また、変動係数 $CV(\bar{U}_\ell)$ は、次のように

$$CV(\bar{U}_\ell) = \frac{1}{\sqrt{\ell}} \sqrt{CV_{wbx}^2 + \frac{N-m}{N-1} \frac{CV_{wvx}^2}{m}}$$

と表わされますから、これを条件2に入れて、 ℓ について解いてみると

$$\ell \geq \left(\frac{\lambda}{\eta}\right)^2 \left(CV_{wbx}^2 + \frac{N-m}{N-1} \frac{CV_{wvx}^2}{m} \right)$$

となります。

ここで、 CV_{wbx} は従業員数の層内地域間変動係数、 CV_{wvx} は層内地域内変動係数、 N はそれぞれの地域の飲食店数の最大、最小間のある数、また、 m はそれぞれの調査地域の調査飲食店数を表わしています。

なお、推定子の標準偏差 $\sigma(\bar{U}_\ell)$ はあまり利用されませんし、また、これからはかなり式が増えて煩雑になってくるので、省略することにしました。

そこで、いま、1地域当たりの平均従業員数を、たとえ、許容区間の相対誤差 $\eta = 5\%$ 、危険の確率 $\lambda = 2$ ともとめることにしてみましょう。

まず、過去の資料から

$$CV_{wbx} \approx 0.23$$

$$CV_{wvx} \approx 1.09$$

ぐらいであり、また、それぞれの地域の飲食店数も、過去の資料から、最大で23、最小で11ぐらいであるので、もつとも安全をとつて

$$CV\left(\frac{\bar{U}_\ell}{U_\ell}\right) \approx \frac{1}{\sqrt{\ell}} \sqrt{\left(CV_{wbx}^2 + \frac{N_x - m}{N_x - 1} \frac{CV_{wvx}^2}{m} \right) + \left(CV_{wby}^2 + \frac{N_y - m}{N_y - 1} \frac{CV_{wvy}^2}{m} \right) - 2 \left(\rho_{wbxy} CV_{wbx} CV_{wby} + \frac{N_x - m}{N_x - 1} \frac{\rho_{wvx} CV_{wvx} CV_{wvy}}{m} \right)}$$

と表わされますから、これを条件2に入れて、 ℓ について解いてみると、

$$\ell \geq \left(\frac{\lambda}{\eta}\right)^2 \left[\left(CV_{wbx}^2 + \frac{N_x - m}{N_x - 1} \frac{CV_{wvx}^2}{m} \right) + \left(CV_{wby}^2 + \frac{N_y - m}{N_y - 1} \frac{CV_{wvy}^2}{m} \right) - 2 \left(\rho_{wbxy} CV_{wbx} CV_{wby} + \frac{N_x - m}{N_x - 1} \frac{\rho_{wvx} CV_{wvx} CV_{wvy}}{m} \right) \right]$$

となります。

ここで、 CV_{wbx} は従業員数の層内地域間変動係数、 CV_{wvx} は層内地域内変動係数、 CV_{wby} は売上高の層内地域間変動係数、 CV_{wvy} は層内地域内変動係数、 ρ_{wbxy} は従業員数と売上高との層内地域間相関係数、

$$N = 23$$

とし、次に、調査員の活動能力をかながえて、

$$m = 8$$

と決めますと、選びだす地域の数 ℓ は

$$\ell \geq \left(\frac{\lambda}{\eta}\right)^2 \left(CV_{wbx}^2 + \frac{N-m}{N-1} \frac{CV_{wvx}^2}{m} \right) = 246. \dots$$

となります。したがつて、有効桁数をかながえて、2桁までとると、選びだす地域の数は250となります。

なお、これを区部と区部以外とに分けるには、それぞれで営業している飲食店の数に比例させればよいわけですから。

(2) 1従業員当たりの平均売上高の場合

まず、推定式は、まえと等しく、次のとおりになります。

$$\frac{\text{売上高の算術平均}}{\text{従業員数の算術平均}}$$

なお、地域の選定は、やはり「不完全」になり、それを、それぞれの調査地域の調査飲食店数の修正によつて「完全」に切り替えることにはなりますが、まえとおなじように、修正になつても、推定式に変化はありません。

次に、推定子ですが、いま、選びだす地域を数を ℓ として、推定子を $\frac{\bar{V}_\ell}{U_\ell}$ と表わしてみると、その期待値

$E\left(\frac{\bar{V}_\ell}{U_\ell}\right)$ はもとめようとする1従業員当たりの平均売上高に近似的に等しく、したがつて、第1部第10節(9月号)の条件1を近似的に満たしております。また、変動

係数 $CV\left(\frac{\bar{V}_\ell}{U_\ell}\right)$ は、近似的に、次のように、

ρ_{wvx} は層内地域内相関係数、 N_x 、 N_y 、 N_{xy} (注)はそれぞれの地域の飲食店数を表わしています。

(注)それぞれの地域の、従業員数と売上高との相関係数は、すべて、同符号と仮定します。