水泳シーズン

うつとうしい梅雨もあけ，からつと晴れあがつた青空，まつたく気持のよいものである。しかし，連日30°を過す暑さにはまいつてしまう。そこで我々大人達は冷いビールをグツト一ぱい飲みほしたところを想像する。

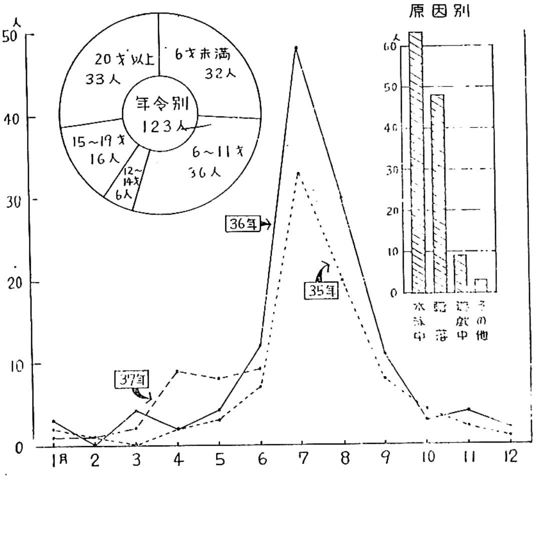
子供達は夏休みの宿題で頭がいつぱいのところえ，流行の「水道方式」とやらでお母さんからもつめ込まれ，それにこの暑さでは，カツパではないが水がこいしくなるのも無理からぬこと，太陽がカンカン照りつける昼どきともなれば，近くの海や川，そしてプールにと泳ぎに行く。頭を冷し体を鍛えるこの水泳も，父兄の健全な指導と監視があつてこそ楽しいものとなるのである。でないと思いもよらぬ事故を起し，とりかえしのつかないことになる。

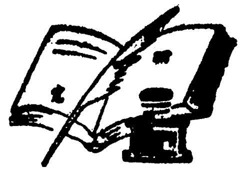
これを物語るように昨年の水の事故による死亡者の統計をみると，123人の犠牲者のうち，56%にあたる68人は11才未満の者であり，これらの尊い人命は父兄の注意ということによつて救うことが出来たのではなかろうか。原因別では水泳中が最も多く63人，次いで転落の48人となつている。どんな場所で起つたかというと，河川が57人，用水堀23人，海21人，湖沼11人などが主なところ，時期的にはグラフにみるように，7，8月に最も多くの犠牲者を出している。

子供達の水遊びは，親達の深い愛情によつて見守つて，悲しい事故が起らないようにしたいものである。

水の事故による死亡者

資料 県警察本





標本調査への手引 (10)

総理府統計局　高　橋　史　朗

第2部　標本選定の技巧（つづき）

9　グループ分けによる調査地域の節約

さきほど第8節（6月号）でみたように，1地域当たりの平均従業員数を許容区間の相対誤差$\eta=5\%$，危険の確率$\lambda=2$でもとめるには，250の調査地域が必要です。ところで，これは東京都を区部とそれ以外とに分けた場合の数ですから，いま，そのまえの層化していない場合の数を第5節（3月号）でみてみると，280となつています。したがつて，調査地域の散布状態を安定させるために層化したときの方が，層化していないときにくらべて，必要とする調査地域の数が小さくなつているわけです。これは非常に好ましい傾向ですが，それがこの場合だけに限られた偶然の結果なのか，それとも層化によつて起る必然の結果なのかをここで確かめてみます。

まず，調査地域の数lを決める式を，層化していない場合と層化した場合についてまとめてしめすと，次のとおりになります。なお，層化していない場合の式で，従業員数の地域間変動係数 CVbx および地域内変動係数 CVwx は，そのままでは，層化した場合の層間変動係数 CVbx および層内変動係数CVwxとそれぞれ混同する恐れがあるので，次のように書き変えました。

CVbx→CV.bx
CVwx→CV.wx

(1)　層化していない場合

$$l \geqq \left(\frac{\lambda}{\eta}\right)^2\left(CV_{.bx}^2+\frac{N-m}{N-1}\;\frac{CV_{.wx}^2}{m}\right)$$

(2)　層化した場合

$$l \geqq \left(\frac{\lambda}{\eta}\right)^2\left(CV_{wbx}^2+\frac{N-m}{N-1}\;\frac{CV_{wwx}^2}{m}\right)$$

この両式を比較してみますと，どちらの場合の調査地域数lが小さいかは，結局，地域間変動係数 CV.bx と層内地域間変動係数CVwbx および地域内変動係数 CV.wx と層内地域内変動係数CVwwxとのうちで，それぞれ，どちらが小さいかにかかつていることが分かります。

そこで，いま，従業員数の変動係数CVxについてかんがえてみると，これは，地域間変動係数CV.bxおよび地域内変動係数CV.wxあるいは層間変動係数 CVbxおよび層内変動係数CVwxによつて，それぞれ，次のように表わされます。

$$CV_x^2=CV_{.bx}^2+CV_{.wx}^2$$

$$CV_x^2=CV_{bx}^2+CV_{wx}^2$$

そして，このうち，第2式の層内変動係数 CVwxは，層内地域間変動係数CVwbxおよび層内地域内変動係数CVwwxによつて，次のように表わされますから，

$$CV_{wx}^2=CV_{wbx}^2+CV_{wwx}^2$$

したがつて，この関係を第2式に入れると，第2式は次のようになります。

$$CV_x^2=CV_{bx}^2+CV_{wbx}^2+CV_{wwx}^2$$

ところで，この第1式と第2式とのあいだに，第1式の地域内変動係数 CV.wx が第2式の層内地域内変動係数 CVwwx に等しいという関係が成り立つのです。

$$CV_{.wx}=CV_{wwx}\quad あるいは\quad CV_{.wx}^2=CV_{wwx}^2$$

したがつて，このことから，第1式の地域間変動係数CV.bxと第2式の層間変動係数 CVbx および層内地域間変動係数 CVwbx とのあいだに，次の関係が成り立ちます。

$$CV_{.bx}^2=CV_{bx}^2+CV_{wbx}^2$$

そこで，いま，これらの関係式にもとずいて，調査地域数lを決める(1)式および(2)式を比較してみますと，その第2括弧内の第1項は，層化した場合の方が，層化していない場合にくらべて，一般に小さく，また，第2項は層化してもしなくても変わりないことが分かります。したがつて，調査地域数lは，層化した場合の方が，層化していない場合にくらべて，一般に小さくなるわけです。したがつて，さきほどの1地域当たりの平均従業員数の場合の結果は必然であり，偶然ではないことが分かります。

これは，層化が，また別の意味をもつていることを意味します。すなわち，層化は，調査地域の散布状態を安定させて，調査が円滑におこなえるようにするだけではなく，調査地域の数を少なくして，調査が経済におこな

えるようにするわけです。そこで，このことから，調査を円滑におこなうために層化し，それにともなつて，調査地域数の節約をはかるというだけではなく，さらに，調査地域数を節約して，調査を経済におこなうだけのために層化するということがかんがえられてきます。

この後者の立場からおこなう層化は，結局，層化によつて，層間変動係数 CVbx をできるだけ大きくするということになりますが，これには，大きく分けて，２つの方向がかんがえられます。

方向 1 同じ層に属する飲食店の従業員数はできるだけ相等しく，また，異なる層に属する飲食店の従業員数はできるだけ相異なるようにすると，層間変動係数CVbxは大きくなります。したがつて，従業員数が等しい飲食店はできるだけ同じ層に，また，従業員数が異なる飲食店はできるだけ異なる層にまとめるとよいわけですが飲食店は，調査員の活動能力をかんがえて，地域にまとめてありますから，このゆき方はそのままでは採れません。

そこで，これよりは劣りますが，同じ層に属する地域はその１飲食店当たりの平均従業員数ができるだけ相等しく，また，異なる層に属する地域はその１飲食店当たりの平均従業員数ができるだけ異なるようにすると，層間変動係数 CVbxは大きくなりますから，したがつて，１飲食店当たりの平均従業員数の等しい地域はできるだけ同じ層に，また，異なる地域は異なる層にまとめるとよいわけです。しかし，このゆき方も，あまり忠実に守つて層をまとめると，各層に属する地域はあちこちに散在することになりますから，取り扱いが煩雑になつて好ましくありません。そこで，一般には，むしろ，各層は隣接する地域からまとめることにして，そのさい，できるだけ，このゆき方を採り入れてゆくというようにするのがよいとおもいます。

方向 2 層は，細分されて，その数が増えると，層間変動係数CVbxは大きくなります。したがつて，層の数を調査地域の数あるいは調査員の数まで増やし，各層に属する飲食店の数が相等しくなるようにすればよいわけですが，しかし，そこまで細分して，各層に属する飲食店の数を相等しくすることは非常に困難ですから，このゆき方もこのままでは採れません。

そこで，各層に属する飲食店の数がその層の調査地域の数に比例するという関係は崩れないようにしておいてできるだけ層の数を増やすというのがよいということになります。

これまで，算術平均の１例である１地域当たりの平均従業員数について述べてきましたが，こんどは，２個の算術平均の比率の１例である **１従業員当たりの平均売上高**について説明することにします。

まず，調査地域の数 ℓ を決める式を，層化していない場合と層化した場合とについてまとめてしめすと，次のようになまります。なお，層化していない場合の式で，従業員数の地域間変動係数CVbxと地域内変動係数 CVwx，売上高の地域間変動係数CVbyと地域内変動係数 CVwy，それに従業員数と売上高の地域間相関係数 ρbxy と地域内相関係数ρwxyは，そのままでは，層化した場合の，従業員数の層間変動係数 CVbx と層内変動係数CVwx，売上高の層間変動係数 CVbyと層内変動係数 CVwy，それに従業員数と売上高の層間相関係数 ρbxy と層内相関係数ρwxyに，それぞれ，混同する恐れがあるので，次のように書き変えました。

CVbx→CV.bx　CVby→CV.by　ρbxy→ρ.bxy

CVwx→CV.wx　CVwy→CV.wy　ρwxy→ρ.wxy

(1) 層化していない場合

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2 \left[\left(CV_{.bx}^2 + \frac{N_x - m}{N_x - 1}\,\frac{CV_{.wx}^2}{m}\right) + \left(CV_{.by}^2 + \frac{N_y - m}{N_y - 1}\,\frac{CV_{.wy}^2}{m}\right) - 2\left(\rho_{.bxy} CV_{.bx} CV_{.by} + \frac{N_{xy} - m}{N_{xy} - 1}\,\frac{\rho_{.wxy} CV_{.wx} CV_{.wy}}{m}\right)\right]$$

(2) 層化した場合

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2 \left[\left(CV_{wbx}^2 + \frac{N_x - m}{N_x - 1}\,\frac{CV_{wwx}^2}{m}\right) + \left(CV_{wby}^2 + \frac{N_y - m}{N_y - 1}\,\frac{CV_{wwy}^2}{m}\right) - 2\left(\rho_{wbxy} CV_{wbx} CV_{wby} + \frac{N_{xy} - m}{N_{xy} - 1}\,\frac{\rho_{wwxy} CV_{wwx} CV_{wwy}}{m}\right)\right]$$

そこで，いま，この両式を比較してみますと，次のような関係式

$$CV_{.bx}^2 = CV_{bx}^2 + CV_{wbx}^2 \quad CV_{.wx}^2 = CV_{wwx}^2 \qquad CV_{.by}^2 = CV_{by}^2 + CV_{wby}^2 \quad CV_{.wy}^2 = CV_{wwy}^2$$

$$\rho_{.bxy}\, CV_{.bx}\, CV_{.by} = \rho_{bxy}\, CV_{bx}\, CV_{by} + \rho_{wbxy}\, CV_{wbx}\, CV_{wby}$$

$$\rho_{.wxy}\, CV_{.wx} CV_{.wy} = \rho_{wwxy}\, CV_{wwx}\, CV_{wwy}$$

が成り立ちますから，したがつて，調査地域数 ℓ は，一般に，層化した場合の方が，層化していない場合にくらべて，小さいことが分かります。

そこで，この場合も，調査を円滑におこなうために層

化して，その副産物として，調査地域の節約をはかるというだけではなく，もつと直接に，すなわち，調査を経済におこなうだけのために層化するということがかんがえられます。この後者の立場に立つ層化には，さきの1地域当たりの平均従業員数の場合とおなじように，２つの方向がかんがえられますが，しかし，それは，さきの場合の従業員数を，この場合には

$$\frac{\text{従業員数}}{\text{平均従業員数}} - \frac{\text{売上高}}{\text{平均売上高}}$$

とするだけで，あとはまつたく変わりありませんから，その説明は，ここでは，省略したいと思います。

ところで，これまでの説明から明らかなように，方向1で述べた層のまとめ方は，もとめようとする統計によつて変わつてきます。したがつて，いくつかの統計を同時にもとめようとする場合には，方向1で述べた層のまとめ方は，各統計のあいだで必ずしも一致せず，むしろ，矛盾する方が多いわけです。そこで，そのような場合には，もとめようとする統計のなかで，とくに重要な統計だけに，できるだけ満足がゆくように層をまとめるということになります。

また，このようにして，２つの方向から，層間変動係数CVbが大きくなるように工夫しても，この工夫を，調査地域数ℓの決定に反映させることは，一般に，非常に困難です。それは，ここで述べたような各種の変動係数および相関係数の資料が，通常は，なかなか手に入らないからです。したがつて，ここに述べた２つの方向からの工夫は，本来なら，危険の確率を一定として，ある統計を与えられた幅の許容区間でもとめるのに必要な調査地域の数をできるだけ少なくすることに利用すべきですが，しかし，調査地域数ℓを決めるのに必要な各種の変動係数および相関係数が手に入らない最悪の場合には，逆に，その統計を与えられた数の調査地域でもとめるのに伴う許容区間の幅をできるだけせばめるのに利用するということになりましよう。

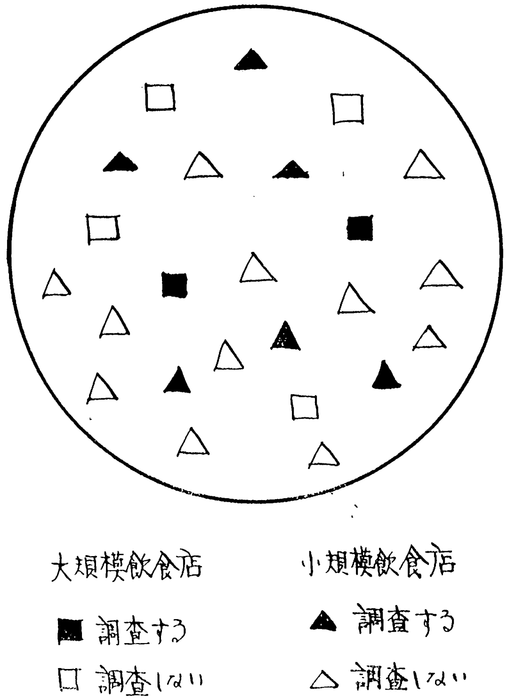
10　調査地域の飲食店のグループ分け

これまで，層化といつても，1次抽出単位すなわち地域の層化だけをかんがえてきましたが，層化はなにも1次抽出単位にかぎつたものではなく，２次抽出単位すなわち飲食店の層化もかんがえられます。ただし，地域の層化が，調査地域の散布状態の安定を主に，調査地域の節約を従におこなわれるのとは異なつて，飲食店の層化は，さしあたつて，もつぱら調査地域したがつて調査飲食店の節約のためにかんがえられます。

そこで，いま，調査地域の飲食店を，たとえば，営業規模によつて，大規模と小規模の２グループに分け，各グループからは，さしあたり，それぞれに属する飲食店の数に比例して，調査飲食店を選びだすことにしてみましよう。なお，ここでも，さしあたり，さしあたりとくり返して，奇妙に感ぜられるかも知れませんが，ここは非常に含みのあるところで，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。

第 8 図

調査地域



さて，そこで，1地域当たりの平均従業員数と1従業員当たりの平均売上高の，推定式と推定子をしめせば，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，次のとおり，変化ありません。なお修正も，まえどおりで，変化はありません。

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

次に，推定子ですが，まえどおりU_ℓと表わしますとその期待値E($\bar{U}_\ell$)は，まえどおり，第1部第10節（9月号の条件1を満たしており，また，変動係数CV(U_ℓ)は，次のようになります。

$$CV(\bar{U}_\ell) = \frac{1}{\sqrt{\ell}}\sqrt{CV^2_{wax} + \frac{N-m}{N-k}\,\frac{CV^2_{wiwx}}{m}}$$

ここで，CVwbxは従業員数の層内地域間変動係数，CVwwwxは層内地域内規模内変動係数，Nはそれぞれの地域の飲食店数の最大，最小間のある数，kはそれぞれの地域の飲食店の大規模，小規模の割合の逆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ます。

したがつて，条件2に入れて，ℓについて解くと

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2\left(CV_{wbx}^2+\frac{N-m}{N-k}\ \frac{CV_{wwwx}^2}{m}\right)$$

となります。

ここで，飲食店を層化していない場合の調査地域数ℓの式

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

とくらべてみましよう。すると，層化した場合と層化していない場合とで，どちらの調査地域数ℓが小さいかは，

$$\frac{N-m}{N-k}\ \frac{CV_{wwwx}^2}{m} \quad と$$

$$\frac{N-m}{N-1}\ \frac{CV_{wwx}^2}{m}\quad でどちらが小さいか$$

によつて決まることが分かります。この比較はかなり複雑なので，ここまでで説明をやめたいとおもいますが，ただ，うまく層化すれば，前者は後者よりも，一般に，かなり小さくすることができる，すなわち，層化した場合の方が，層化していない場合にくらべて，調査地域数ℓを一般にかなり小さくすることができるということを言い加えておきます。

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

まず，過去の資料から，

$CVwbx \fallingdotseq 0.23$

$CVwwwx \fallingdotseq 0.71$

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，

$N = 23$

とし，さらに，それぞれの地域の飲食店の大規模，小規模の割合の逆数も，過去の資料から，最大で4.2，最小で1.3ぐらいであるので，もつとも安全をとつて，

$k = 4.2$

とし，次に，調査員の活動能力をかんがえて，

$m = 8$

と決めると，選びだす地域の数ℓは，

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2\left(CV_{wbx}^2+\frac{N-m}{N-k}\ \frac{CV^2wwwx}{m}\right)$$

$=16.5\cdots\cdots$

となります。したがつて，有効桁数をかんがえて，2桁までとると，選びだす地域の数は170となります。

この調査地域数170は，飲食店を層化していない場合の250にくらべて，かなり小さくなつております。

(2) 1従業員当たりの平均売上高の場合

まず，推定式は，次のとおり，変化ありません。なお修正しても，まえどおり，変化しません。

$$\frac{売上高の算術平均}{従業員数の算術平均}$$

次に，推定子ですが，まえどおり$\dfrac{\overline{V}\ell}{\overline{U}\ell}$と表わしますと

その期待値$E\left(\dfrac{\overline{V}\ell}{\overline{U}\ell}\right)$は，まえどおり，第1部第10節（9月号）の条件1を近似的に満たしており，また，変動係数$CV\left(\dfrac{\overline{V}\ell}{\overline{U}\ell}\right)$は，近似的に，次のようになります。

$$CV\left(\frac{\overline{V}_\ell}{\overline{U}_\ell}\right) \fallingdotseq \frac{1}{\sqrt{\ell}}\sqrt{\left(CV_{wbx}^2+\frac{N_x-m}{N_x-k_x}\ \frac{CV_{wwwx}^2}{m}\right)+\left(CV_{wby}^2+\frac{N_y-m}{N_y-k_y}\ \frac{CV_{wwwy}^2}{m}\right)-2\left(\rho_{wbxy}CV_{wbx}CV_{wby}+\frac{N_{xy}-m}{N_{xy}-k_{xy}}\ \frac{\rho_{wwwxy}CV_{wwwx}CV_{wwwy}}{m}\right)}$$

ここで，CVwbxは従業員数の層内地域間変動係数，CVwwwxは層内地域内規模内変動係数，CVwbyは売上高の層内地域間変動係数，CVwwwyは層内地域内規模内変動係数，ρwbxyは層内地域内規模内相関係数，ρwbxyは従業員数と売上高との層内地域内相関係数ρwwwxyは層内地域内規模内相関係数，Nx，Ny，Nxy(注)はそれぞれの地域の飲食店数の最大，最小間のある数，kx，ky，kxyはそれぞれの地域の飲食店の大規模，小規模の割合の逆数の最大，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。

したがつて，条件2に入れてℓについて解くと，

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2\left[\left(CV_{wbx}^2+\frac{N_x-m}{N_x-k_x}\ \frac{CV_{wwwx}^2}{m}\right)+\left(CV_{wby}^2+\frac{N_y-m}{N_y-k_y}\ \frac{CV_{wwwy}^2}{m}\right)-2\left(\rho_{wbxy}CV_{wbx}CV_{wby}+\frac{N_{xy}-m}{N_{xy}-k_{xy}}\ \frac{\rho_{wwwxy}CV_{wwwx}CV_{wwwy}}{m}\right)\right]$$

となります。

層化していない場合との比較は，やはり，複雑になりますから，説明は省略したいとおもいます。ただ，やはり，うまく層化すれば，調査地域数ℓを一般にかなり小さくすることができるということを言い加えておきます。

なお，このように飲食店を層化しても，この標本選定は，やはり，**層化2段抽出**といいます。いいかえると，2次抽出単位の層化は，一般に，表立つて言い表わされないのです。