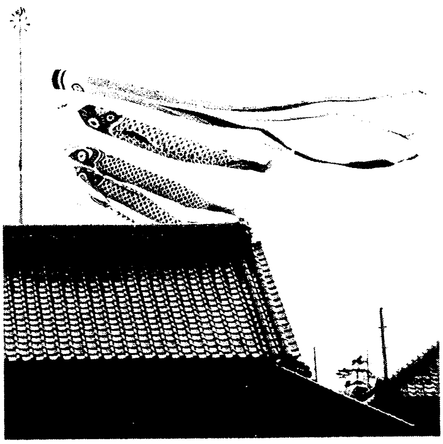
歳時記



端午の節句は，子供の成長を祝う子供の日ということで，祭日になっています。最近，３月３日の上巳（じょうし）の節句も祭日として休みにしようという動きがあります。男の子の節句が休みなのに，女の子の節句が休みでないのはおかしいということらしいですが，女の子を持つ世の親からすれば，そうあって欲しいと思うのは自然かもしれません。

こういうことは，「五月の鯉の吹流し」のように，さっぱりとした気持で，決めて欲しいものです。

今月のおもな行事

○１日	学校基本調査調査日
○８～10日	市町村統計担当者会議（大子町）
○９～10日	関東甲信静統計主管課長会議（東京都）
○10～12日	統計グラフ作成指導者講習会（水戸市・土浦市・結城市）
○15～16日	北関東４県県民所得事務研究会（栃木県）
○18～19日	事業所統計第二次ブロック会議（静岡県）
○23日	住宅統計全国統計主管課長会議（東京都）
○23～24日	統計調査員研修会（群馬県）
○25～26日	事業所統計第二次市町村事務打合せ会（水戸市・ときわ荘）
○31日	商業動態統計調査ブロック会議（埼玉県）

さいじき

●シリーズ「短期統計実務講座」

標本数をどう決めるか（上）　標本調査のために………

全数調査に比べて労力と費用の両面で経済的であり，集計も早くできる便利な標本調査ですが，実際に調査を企画・設計するとなるとなかなか難しいところが有ります。まず，最大の難問が「数」の問題です。のっけから「数」の問題で申し訳ないのですが，お金の問題よりは深刻ではないと思ってあきらめて下さい。この数とは標本数のことです。一体どれだけの数を調べれば，統計として利用できる結果が得られるのか，という問題です。数多く調べればよいのかもしれませんが，それでは標本調査らしくなくなってしまいますし，第一金はかかるは，正確さの程度はわからないでは，それは大変です。できれば標本数というのは，最小の費用でもって必要な精度を達成できるところに置くのが良いのです。この標本数を計算するには，あらかじめ公式が用意されていて，母集団の大きさ，母集団の**標準偏差**または**変動係数**などがわかれば公式を使いながら標本数を計算できるようになっています。この公式は，標本を抽出する方法の違いによってさまざまな公式を利用しなければなりません。標本調査の調査方法には，

① 単純任意抽出法

② 集落抽出法

③ 層別抽出法

④ ２段（多数）抽出法

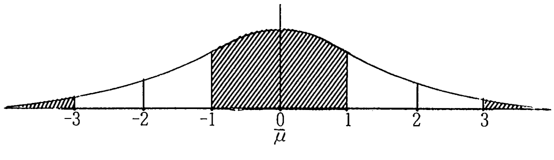
⑤ 確率比例抽出法

などがあります。それぞれの抽出法の説明はここでは省いて，実際の問題を例にとって標本数を計算してみることにします。その前に確率についてどうしても必要欠くべからざる，必要書くべきことを書かなければなりません。

単純任意抽出法によって抽出された標本の標本平均の分布については，次のような性質があります。

「母集団の分布がどのような分布であっても，**標本数** n が十分大きい場合は，標本平均の標本分布は近似的に**正規分布**となる。」（**中心極限定理**と言われる）

正規分布については下のような形を見れば「ああ，あのことか」とおわかりの方も多いでしょう。



確率を考えるにあたって，その基本は正規分布にあると言えます。

母集団の単位の数 N 個から標本として取り出された n についての平均 $\bar{x}$ の分布の状態を度数分布としてヒストグラムに書いてみると，母集団がどのような形でも左の図の形に近いものが出来上がります。そして $\bar{x}$ は n の数をふやしていけばいくほど母集団の平均 μ に近づきます。$|\bar{x}-\mu|$ を**標本誤差**と言います。

左の図で，この分布の平均値 μ が o，標準偏差 σ が１のとき，中心から＋，－の方向に１だけはなれた距離の中に含まれる面積は全体の68.27％になることがわかっています。２離れたときは95.45％です。

Z	α
1	68.27％
2	95.45％
3	99.73％

表にすると右の表になります。このときの α を**信頼水準**といっています。標本設計にあたっては，調査結果の性質あるいは利用方法などの観点からこの信頼水準を決定しておく必要があります。上の表からみれば，信頼水準を68％にとればＺは１，95％にとればＺは２です。どういうことかというと，標本誤差 $|\bar{x}-\mu|$ が，Ｚの値２以下である確率，言い替えれば，Ｚの値２の範囲内にある確率が95％以上，３以下である確率が99％以上だということになります。

平均 $\bar{x}$ の標準誤差については，

$$\sigma(\bar{x})=\sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}}$$

で示されますが，普通は標本数 n というのは母集団の単位の数 N に比べて非常に小さいので，

$$\frac{N-n}{N-1}\fallingdotseq 1$$

となるために，上の式は

$$\sigma(\bar{x})=\sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}}$$

と考えてもかまいません。

それから，信頼水準とともに**許容限界**も決めておかなければ標本数は出て来ません。たとえば，ある市の１世帯の生活費を推定する場合，平均生活費が10万円ぐらいと考え

冨　永　重　己

られる時，その推定を９万９千円から10万１千円の範囲で行いたいと思えば，許容限界は１千円と決めることができます。この許容限界は d で表わされます。

まず，簡単な調査例をとってみることにします。

〔例題１―１〕
ある中学校で生徒の毎月のこづかい額を調査することになった。標本数は何人とればよいか。

さて，調査をしなければならないわけですが，これでは何にもわかりません。ないないずくしです。まず母集団の数 (N) が明らかでなければなりません。N（全校生徒数）は調べたところ800人でした。それから母集団の標準偏差も知りたいものです。これは，過去の経験や，隣町の中学校の結果などから，600円程度であることがわかりました。さて，それから，信頼水準と許容限界を決定しなければなりません。信頼水準を95%にとることに決めて，許容限界を100円以下に押さえることに決めました。なお，信頼水準は95%から99.7%にとるのが普通です。

以上の情報をまとめると，

$N = 800$

母集団の標準偏差 $\sigma = 600$

$95\% \longrightarrow Z \fallingdotseq 2$

$d = 100$

これらを，単純任意抽出の場合の標本数 n を決定する公式，

$$\boxed{n = \frac{Z^2 \cdot \sigma^2}{d^2} \cdot \frac{1}{1 - \dfrac{1}{N} + \dfrac{Z^2 \cdot \sigma^2}{d^2} \cdot \dfrac{1}{N}}}$$

にあてはめて計算します。

$$n = \frac{2^2 \times 600^2}{100^2} \cdot \frac{1}{1 - \dfrac{1}{800} + \dfrac{2^2 \times 600^2}{100^2} \times \dfrac{1}{800}}$$

$$= \frac{1}{\dfrac{799}{800} + \dfrac{144}{800}}$$

$$= 144 \times \frac{4}{\dfrac{943}{800}} = 122$$

122人を調査することになりました。

もっと簡単に計算しようとするときは，上の公式は，

$$\boxed{n = \frac{Z^2 \cdot \sigma^2}{d^2}}$$

でもかまいません。なぜなら，この場合も N に比べて普通は n が非常に小さくなるため

$$\frac{1}{1 - \dfrac{1}{N} + \dfrac{Z^2 \cdot \sigma^2}{d^2} \cdot \dfrac{1}{N}} \fallingdotseq 1$$

になるからです。これで計算すると

$$n = \frac{2^2 \times 600^2}{100^2} = 144$$

144人調査すればよいことになります。そうすれば標本誤差は100円以下であり，95%の信頼水準にある結果が得られるはずです。

〔例題１―２〕
上記の中学校で122人を単純任意抽出してそのこづかいを調査したところ，平均は1,300円，標準偏差は580円であった。全校生徒の平均は信頼水準95%でどの範囲にあるといえるか。（平均の推定）

わかっていることは，

$n = 122,\ \bar{x} = 1{,}300,\ s = 580$

標準誤差の推定には，標準誤差 $\sigma(\bar{x})$ を求める公式を使います。

$$\boxed{\sigma(\bar{x}) = \sqrt{\frac{N-n}{N}} \cdot \frac{s}{\sqrt{n}}}$$

ここでもやはり，$\sqrt{\dfrac{N-n}{n}} \fallingdotseq 1$ と考えれば，

$$\sigma(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{580}{\sqrt{122}} = 53$$

信頼水準が95%ということより，$Z = 2\sigma(\bar{x})$ であるから，母集団の平均 μ は，

$$\bar{x} - 2\sigma(\bar{x}) < \mu < \bar{x} + 2\sigma(\bar{x})$$

$$1{,}300 - 2 \times 53 < \mu < 1{,}300 + 2 \times 53$$

$$1{,}194 < \mu < 1{,}406$$

全校生徒の平均は，1,194円から1,406円の間にあることがわかります。

〔例題2―1〕

ある学校の同窓会で会則を改訂しようとして，会員に郵送でその賛否を問うことにした。予算を節約するため結果に若干の誤差があっても仕方ないとして，一部の会員を選定して手紙を出し賛否の比率を推定することにした。95%の信頼水準で，標本誤差を2%に押さえるためには何人の会員に手紙を出せばよいか。なお，会員総数は32,000人である。

わかっている条件は，

$N = 32{,}000$

$\alpha = 95\% \longrightarrow Z \fallingdotseq 2$

$d = 2\%$

この例は比率の問題になっています。この場合，母集団の調査単位のうちある属性を持つ抽出単位数（ここでは賛成の会員）をN_1，そうでない方（ここでは反対の会員）をN_2とすると，

$N_2 = N - N_1$

比率 $p = \dfrac{N_1}{N}\ (q = 1 - p)$

この場合の公式は前の平均を求める公式と同じ方法で，

$$n = \frac{Z^2 pq}{d^2} \cdot \frac{1}{1 - \frac{1}{N} + \frac{Z^2 pq}{d^2} \cdot \frac{1}{N}}$$

を利用します。ここで，pとqが不明ですが，ここでは，$p \times q$が最大となるように$p = q = \frac{1}{2}$としておきます。公式に数字をあてはめると，

$$n = \frac{2^2 \times \frac{1}{2} \times \frac{1}{2}}{0.02^2} \times \frac{1}{1 - \frac{1}{32{,}000} + \frac{2^2 \times \frac{1}{2} \times \frac{1}{2}}{0.02^2} \cdot \frac{1}{32{,}000}}$$

$$= 2{,}500 \times \frac{1}{\frac{31{,}999}{32{,}000} + \frac{2{,}500}{32{,}000}} = 2{,}500 \times \frac{1}{\frac{34{,}499}{32{,}000}}$$

$$= 2{,}500 \times \frac{32{,}000}{34{,}499} = 2{,}319$$

この式でも，$\dfrac{Z^2 \cdot pq}{d^2}$を計算してそれが$N$に比べて十分小さければそれでかまいません。つまり2,500人に手紙を出せばよいことになります。

〔例題2―2〕

上の調査により実際に抽出した2,500人の会員に手紙を出したところ65%の賛成があった。全会員の賛成の比率は何%と推定できるでしょうか。（比率の推定）

単純任意抽出で比率計算の場合の標準誤差の推定の公式

$$\sigma(\hat{p}) = \sqrt{\frac{N-n}{N} \cdot \frac{\hat{p}\,\hat{q}}{n-1}}$$

を使いますが，$N \gg n$なので$\dfrac{N-n}{N}$を無視して$\dfrac{\hat{p}\,\hat{q}}{n-1}$だけで計算します。$\hat{p}$は65%から$\hat{p} = 0.65$,

標準誤差 $\sigma(\hat{p}) = \sqrt{\dfrac{\hat{p}\,\hat{q}}{n-1}} = \sqrt{\dfrac{0.65 \times 0.35}{2{,}500 - 1}} = 0.0095$

従って賛成の比率pの95%の信頼区間は，95%信頼区間ですから$Z = 2$をとり，$2\sigma(\hat{p})$,

$\hat{p} - 2\sigma(\hat{p}) < p < \hat{p} + 2\sigma(\hat{p})$

$0.65 - 2 \times (0.0095) < p < 0.65 + 2 \times (0.0095)$

$0.631 < p < 0.669$

よって，95%の信頼水準で，全会員の賛否の比率は63.1%から66.9%の間にあるということがわかるわけです。

〔例題3―1〕

ある市でクーラーを持っている世帯数を推定するため市内に住む全部の世帯から標本世帯を単純任意抽出することにした。信頼水準95%で標本誤差が500世帯を越えないような結果を得ようとする場合には標本世帯をどのくらいとればよいでしょうか。

上の調査を行うためにはまず，ある市の全世帯数がわからなければなりません。いま，それを30,000世帯とします。公式は

$$n \fallingdotseq \frac{Z^2 N^2 pq}{d^2}$$

を利用します。pqというのは母集団の分散ですが，それがわかりませんので，約10%ぐらいの世帯がクーラーを入れていることが予想されたとします。そうすると，

$\sigma^2 = pq \fallingdotseq 0.1 \times (1 - 0.1) \fallingdotseq 0.1 \times 0.9 = 0.09$

また，信頼水準が95%より$Z \fallingdotseq 2$，許容限界を500世帯とすることから$d = 500$にすればよいことがわかります。

これらの値を公式にあてはめれば

$$n \fallingdotseq \frac{Z^2 N^2 pq}{d^2} = \frac{2^2 \cdot 30,000^2 \times 0.09}{500^2} = \frac{1,296,000,000}{250,000} = 1,296$$

30,000世帯の中から1,296世帯を調査することになります。

〔例題3－2〕
実際に1,296世帯について調べたところ，8％の世帯にクーラーがはいっていました。この普及率の標本誤差はどのぐらいでしょうか。(比率の推定)

調査の結果，$\hat{p} = 0.08$がわかったわけです。この場合も〔例題2－2〕と同じ公式を使い，

$$\text{標準誤差}\ \sigma(\hat{p}) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n-1}} = \sqrt{\frac{0.08 \times 0.92}{1,296-1}} = 0.0075$$

$$\hat{p} - 2\sigma(\hat{p}) < p < \hat{p} + 2\sigma(\hat{p})$$

$$0.08 - 0.015 < p < 0.08 + 0.015$$

$$0.065 < p < 0.095$$

95％の確率でこの市のクーラーの普及率pは6.5％から9.5％の間にあることが推測されます。

以上，例題を解きながら単純任意抽出の場合の標本数の決め方や，標本から得られた結果をもとにして誤差の推定の方法をやってみました。それらをまとめますと，標本設計にあたっては次の手続きが必要です。

1．標本誤差（d）または標本誤差率（e）の程度（目標精度）を定める。
2．母集団の標準偏差（σ）または変動係数（c）を推定する。
3．信頼水準（α）を決定する。それに伴なって，（Z）が決定される。
4．$n = \dfrac{Z^2 \cdot \sigma^2}{d^2}$　　$n = \dfrac{Z^2 \cdot c^2}{e^2}$　　（nは標本数）
5．調査終了後，達成精度を推計する。
　$\alpha^2 \approx S^2$

さて，標本を単純任意抽出法によって選ぶ方法をとるときにいくつかの欠点があります。それはまず第1に，この方法をとるためには母集団の全部の調査対象をつかまなければならないことです。たとえば，全国の世帯の名簿を作るとすればその仕事の量は膨大で，しかも時間がかかります。第2に，標本数が決まって実査の段階にはいったとき，調査員を使わなければならない調査だとすると客体が散在していることになるので大変困難なことになります。実際には調査ができない場合もあるでしょう。第3に，標本設計をする前に，母集団についてさまざまな情報がわかっている場合，それらを無視した標本設計になってしまうことです。たとえば，ある市の高校生の平均身長を推定する場合，学年ごとの平均身長の格差や女子生徒と男子生徒の間の格差や学校ごとの格差，生徒数などがわかっていればまた別な抽出方法をとることによって標本数を少なくできるし，正確性も高めることができる場合があります。

以上で単純任意抽出法の説明を終わり，次回には別な抽出法をいくつか見てみることにします。

（県統計課・消費統計係）

■統計の窓「お知らせ」

工場も店舗も会社も学校も

―事業所統計調査が行われます―

今年の6月15日には，全国いっせいに事業所統計調査が行われます。

事業所統計調査は，全国のすべての事業所を漏れなく調べる調査で，世帯を漏れなく調べる国勢調査と並ぶ最も基本的な統計調査として，昭和22年に第1回目，翌23年に第2回目が行われ，以後3年ごとに実施されて今回で第12回目を迎えました。

1　調査のねらい

事業所統計調査は，我が国における事業所及び企業の産業，規模等の基本的構造並びに業態，所在する場所等の特性のほか，サービス業事業所の営業状況を全国及び地域別に明らかにすることにより，各種政策の基礎資料とするとともに，事業所名簿を作成し，事業所を対象とする各種統計調査の実施のための資料を提供することを目的としています。

2　調査の期日

調査は，昭和53年6月15日現在で行います。

3　調査の法的根拠

事業所統計調査は，国の行う重要な統計調査として，指定統計第2号に指定され，統計法に基づいて公布された，事業所統計調査規則に従って実施されます。

4　調査の対象

事業所とは，物の生産又はサービスの提供が，営利・非営利を問わず業として行われている一定の場所をいいます。つまり，タイトルの“工場も店舗も会社も学校も”のように，工場，店舗，会社，学校のほか，官公庁，病院，国鉄や私鉄の駅，旅館などから神社，教会，寺院にいたるまで事業を行っている場所が事業所です。事業所統計調査は，これらのすべてを調査の対象とします。

事業所は漏れなく――構内事業所も

(今回調査の特徴①)

事業所統計調査は，我が国の事業所を漏れなく調査する悉皆調査です。従って，今回の調査では，従来に増して対象の完全は握に努めることとしています。このため，そのひとつとして大きな会社や工場の構内にある関連下請事業所やサービス会社，官公庁内の売店や食堂，大学内の生協等の構内事業所も，漏れなく調査することを重点としています。

5　調査の種類及び調査事項

調査は，甲調査，乙調査及び丙調査に分けて行います。

○甲調査は，民営の事業所について，次の事項を調査します。

〔全事業所について〕

①事業所の名称，②所在地，③経営組織，④本所又は支所の別，⑤開設時期，⑥事業の種類，⑦業態，⑧従業者数，⑨所在する場所，⑩形態

主産業と従産業　　**(今回調査の特徴②)**

近年，事業の多角化をはかる企業がふえていますが，これを産業構造面からみると，それらの企業は，二つ又はそれ以上の産業に属していることになります。今回調査では，この点に着目し，事業所について主産業と従産業とに分けてその関連を明らかにし，産業構造を多面的には握します。

業　　態　　**(今回調査の特徴③)**

我が国経済の特殊性として，産業の二重構造問題がしばしば論議されています。そこで，事業所の生産の態様，系列関係，生産活動分野の分担等を調査し，従業者数や資本金などの規模による指標のほか，新たに業態面からも産業の二重構造等の内部構造を，より深くとらえます。

〔会社の本所事業所について〕

①資本金額，②支所の数，③会社全体の常雇数，④会社全体の主な事業の種類

事業所の集積を立体的にとらえる

(今回調査の特徴④)

事業所又は企業の集積の実態をより明らかにする

ため，事業所がどのような場所にあるか，例えば，駅ビルか，高層ビルか，地下街かなどを調査し，従来の地域的分布，いわば平面的なは握に加え，立体的にもは握します。

○乙調査は，旅館，理髪店，映画館，駐車場業等一部のサービス業の事業所について，次の事項を追加して調査します。

①建物・土地の面積，②雇用者の給与，③最近一年間の売上高

サービス業では建物面積など（今回調査の特徴⑤）

サービス業については，従業者のほかにサービス活動をあらわす指標，例えば，旅館の客室数，理髪店の理髪台数，映画館の客席数などの設備の大小をあらわす共通の尺度として，面積を調査します。

○丙調査は，国，地方公共団体及び公共企業体の事業所を対象として，次の事項を調査します。

①事業所の名称，②所在地，③事業の種類，④従業者数

6　調査の方法

調査は，総理府統計局を主管官庁として，都道府県―市区町村―調査員の系統を通じて行われます。

実地の調査活動は，都道府県知事から任命された調査員が当たります。

○調査員が，6月15日から25日頃までの間に各事業所を訪問し，調査票の記入を依頼します。

○調査票は，市区町村→都道府県を通じて総理府統計局に集められ，コンピューターでいろいろな統計表を作ります。

○申告されたことがらは，統計を作るためだけに用いられ，その他の目的，たとえば徴税の資料などに使うことは決してありません。また，調査員や調査関係者が調査で知りえたことがらを，他に漏らすことは，決してありません。

7　集計事項

次の事項につき，全国，都道府県別，市区町村別，地域圏別及び基本調査区別に集計します。

(1)　全事業所について

産業別，従業者規模別，経営組織別，事業所の形態別，事業所の所在する場所別，業態別及び開設時期別の事業所数及び従業者数

(2)　サービス業の事業所について

産業別，従業者規模別，売上高階級別，経営組織別及び面積階級別の事業所数及び従業者数

(3)　会社について（企業単位の集計）

産業(企業)別，従業者(常雇)規模別，資本金階級別及び支所数階級別の企業数

等詳細な多重クロス集計を行います。

8　公表の時期

集計結果は，要計表（都道府県が集計したもの）による全国及び都道府県別事業所数を53年10月に概数として公表し，また，全事業所のうち，従業者が30人以上の事業所について，全国の産業別従業者規模別事業所数及び従業者数を54年1月に速報として公表した後，詳細な結果報告書は，集計完了次第逐次刊行する予定です。

9　結果の利用

調査の結果は，我が国の産業や規模等の基本構造を，言い換えれば産業の見取り図を示すものとして，国はもちろん，都道府県，市町村での各種の行政・経済施策策定の基礎資料として多方面に利用されます。例えば，①国，都道府県における経済計画，地域開発計画及び都市計画，②社会保障・雇用施策，交通・通信施策の立案，③国民所得，県民所得，市町村民所得，産業連関表等の推計資料等に利用されるとともに，④事業所又は企業を対象とする標本調査のフレームとして，国や都道府県のほか，民間においても利用されます。

このように，多くの分野にわたり，かつ，各方面に活用される統計は，他に類をみないといえましょう。

くわしくは，下記までお問い合せ下さい。
茨城県企画部統計課　県勢統計係（内線 426）
〒 310　水戸市三の丸1－5－38