

卒業

卒業の季節、三月は惑いの月である。なにかが一段落したことの安堵。なにかをやりおさせたことの虚脱。親しい友人と別れることの哀切。新しいものに一步ふみだすことへの不安。なにかが開けることへの期待。突然の自由に情念はいりみだれ、心がさわぐ。映画『明日に向けて撃て!』を見たのは、私が高校を卒業して間もない頃だった。

舞台は19世紀末のアメリカ西部、そして南米ボリビア。主人公は、二人組の強盗と彼らの恋人である。映画のなかで、恋人たちが自転車に相乗りし、草原を走っていくシーンがある。セビアを基調にした画面とバック・メロディがかさなり、私は強烈にさわやかな自由を感じた。

やがて、強盗に見切りをつけた彼らは、新しい生活を南米にもとめ、ボリビアへ旅だっていく。しかし、奔放であることに慣れすぎた彼らが辿ったのは、以前のような強盗稼業だった。・・・ボリビア軍の包囲網のなか、彼らが撃ってでるシーンで画面は止まる。そして、明日に向けた銃声だけがこだまする。自由とは何か。

3月のおもな行事

- 1～7日 労働力調査実地検査
- 1～9日 市町村民所得推計指導
- 1～31日 農林業センサス調査票等取りまとめ
- 3～4日 国勢調査実施本部設置打合せ
- 10日 全国統計大会班長会議
物価指数(水戸市2月)速報公表予定
- 11～12日 都道府県民所得推計講習会(経済企画庁)
- 13～14日 都道府県産業速関表研究会(行政管理庁)
- 14日 労働力特別調査打合せ(水戸市・ときわ荘)
- 17日 物価指数(茨城県1月)公表予定
- 24日 統計実務講習会(水戸市・常陽開発センター)
- 25日 学校基本調査・学校保健統計調査電算講習会(東京大学)
- 28～29日 市町村統計主管課長会議(大洗町・かもめ荘)

統計データの見方・表わし方 (5)

—— 分析手段としての比率 ——

1. 対照群と実験群

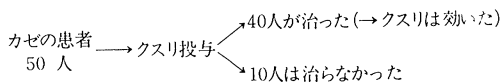
統計データは、グラフに表わすにしろ、数字を読むにしろ、その数字が何を言おうとしているのかを把握しておかなければなりません。そのためにも、統計の分析の手段についても基本的な概念を知っていなければなりません。そこで、統計データの分析に当たっての基本的なポイントを紹介します。

まず、〔例-1〕をみてください。

〔例-1〕 カゼ薬Aの効果をたしかめるために、カゼの患者50人にそのクスリを飲んでもらったところ、2日後には40人がなっていました。この事実にもとづいて「クスリはカゼに対して効果がある」と結論することができますか。

統計的議論で大事なのは、データに基づいて言えることは言う、言えないことは言わないということです。データの読み方が大事なのです。この例題で言っていることを整理すると、図-1のようになります。

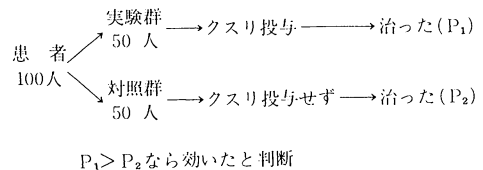
図-1 例題の図解



例題では、このデータを「薬は効いた」と読んだわけですが、そう読んでもよいでしょうか。これは、言い過ぎているのです。考え方のプロセスが大事です。問題点の1つは、カゼの患者50人のなかにどういう人が混じっているかということです。治りかけの人もいるでしょうし、投薬しても治らない人もいるということです。しかし、もう1つの根本的な問題点は、「薬が効いた」という判断の根拠です。こう判断する人は、50人患者がいて薬を与えたら40人($\frac{40}{50}=80\%$)も治ったんだから「薬が効いた」と考えているんだと思います。しかし、それは誤っています。カゼなどは、薬を飲まなくとも放っておいても治ります。薬を飲んだから80%の人が治ったのか、放っておいて何%の人が治ったのか、不

明だからです。80%というのは別に薬のせいではないかも知れません。今のこの問題の盲点は、80%の人が薬を飲んで治ったという事実は事実として認めてもよいのです。しかし、薬を投与したことがこの治った原因かどうかの判断に対しては根拠がないということです。統計データを使うときには、データの中に根拠が含まれているかどうかを判断しなければいけません。この例題の答は、図-2のとおりです。

図-2 答の図解



まず50人選んで薬を与え、そのうち何人治ったか調べます。それから逆に、別の50人を選んで薬を与えず、そのうち何人治ったか調べます。そして、薬を与えた人は何%(P_1)治った。薬を与えなかった人は何%(P_2)治った。この P_1 と P_2 を比べて判断するわけです。例題では P_1 しかありませんから、 P_2 と比較しなければ効いたかどうかの判断はできないのです。

なお、統計の専門用語では、実際に薬を飲んでもらう人を実験群、それと比較対照する人を対照群といいます。統計データを議論するときには、どんな場合でも同じような問題がありますから、基本的な用語として知っていてよい言葉だと思います。もう1つ大事なことは、この50人と50人が、一方は治りかけの人が多く含まれていて、他方では治りづらい人が多く含まれているのでは議論できません。ですから、50人・50人の選び方にも気をつけなければいけません。普通は、あわせて100人選んでおいて、クジ引き(あるいはサイコロ)でランダムに50人・50人に分けます。こういう作業をすることによって、同じような人が入るように気をつけます。これは標本調査のときよく言われることですが、それ以外のときでも、データを集める場合に注意しなければならない点です。統計データを対比するときの基本的な原理です。

2. 追跡調査と回顧調査

続いて、〔例一2〕は統計情報の集め方・調査のやり方に関する例題です。

〔例一2〕 年齢40歳台のガン患者100人と同じ年齢層の健康者200人(いずれも男性)について、若いときにタバコを大量にのむ習慣があったか否かを調べたところ次の結果が得られました。

これから「喫煙の習慣のあった人の半数程度は40歳台までにガンのおそれがある」と主張することは正しいでしょうか。

	計	喫煙の習慣	
		あり	なし
ガン患者	100	80	20
健康者	200	80	120

ここにあげた事例は正しくありません。それは、データからこういうことは言えないからです。事例では、表一1のように考えているのではないのでしょうか。

表一1 例題の解釈 (1)

	喫煙の習慣		
	計	あり	なし
患者	100 ^人	80 ^人	20 ^人
健康者	200	80	120
計	300	160	140
患者率	33%	50%	14%

例題では、「喫煙習慣がガンの発生と関係があるかどうか」という問題意識があって、このデータを扱っています。従って、比率の考え方にもとづき、分母に喫煙習慣のあった人、分子にその中でガンにかかった人という比率を作ってみます。同様に、喫煙習慣のなかった人についても比率を作ってみます。すなわち、喫煙習慣ありの人でガンにかかった人の割合(P_1)を出し、それと喫煙習慣なしの人でガンにかかった人の割合(P_2)を出します。そして、2つの比率の対比をすればよいわけです。そうすると、 $P_1 = \frac{80}{160} = 50\%$,

$P_2 = \frac{20}{140} = 14\%$ ですから、喫煙習慣ありの方がガンにかかる率が高い。このことから「喫煙習慣のあった人の半数程度は40歳台までにガンの恐れがある」と言っているわけですが、そこまで言えるでしょうか。比率の値も大きすぎます。どこがおかしいかというと、数字を加算するとき加算できるかどうか考えなければならないということを忘れてのことです。健康者200人とガン患者100人を加算して300人というのは、数学としては加算してもかまいませんが、統計学としては加算しても意味がありません。どこの世界にだって、300人のうちガン患者100人なんてことはありません。統計学的に加算してはいけない数字を加算して分母にした比率を作っているところに問題があるのです。問題があるからこのデータは使えませんし、このデータからそんなことを言うてはいけません。

この例でキチンとした答を出すためには、プラスαの情報が必要です。ガン患者100人、健康者200人というのはサンプルです。どんな方法で、100人、200人が選ばれたかを考えてみるのです。この例題では、何人のうちの100人、何人のうちの200人という抽出率の情報が全く含まれていなかったのだから、答が出せないのです。そこで、〔例一2〕の文章に「40歳台の男性のうち全国の人口の1%はガン患者である」と付け加えれば、データの料理の仕方は変わってきます。そして、このデータは次のように解釈することができます。

表一2 例題の解釈 (2)

	喫煙の習慣		
	計	あり	なし
患者	10 ^人	8 ^人	2 ^人
健康者	990	396	594
計	1000	404	596
患者率	1%	2%	0.3%

全国では、患者10人、健康者990人の割合です。そこで、先程の情報にもとづき、ガン患者については喫煙習慣のあり・なしは80%と20%の割合、健康者については40%と60%の割合だから……として表一2のように数字を分けるわけです。その上で喫煙習慣のある人の数を分母にとって、そのうちガンにかかった人の数を分子にとって $\frac{8}{404} = 2\%$ という答が得られ、同様に、喫煙習慣のない人についても

● シリーズ「統計研修から」

$\frac{2}{596} = 0.3\%$ という答が得られます。こういう見方をすれば、データを正しく読んだことになるわけです。ただし、喫煙習慣のある人がない人に比べてガンになる確率が高いというのは事実です。そのことは表一1から読みとれますから、表一1の計算は無意味とはいえません。しかし、比率を計算してガンに対する危険が50%とってはいけないのです。こういうパーセンテージを出すためには、世の中には何%のガン患者がいるのかという別の情報が必要になってきます。この例のように原因と結果との間に時間がかかる場合、データの求め方が問題となります。原因は20歳台に喫煙習慣があったか否かということ、結果は40歳台にガンになったか否かということで、20年の時間をかけて因果関係を調査しなければなりません。したがって、こうした因果関係を議論しようとする、統計データはおいそれとは得られません。とにかく、分析のためには、20歳台の情報と40歳台の20年位の時間間隔のある情報を組み合わせなければなりません。因のデータと果のデータが同じ時点では困ります。時点の違ったデータの組み合わせをしなければならず、データの集め方に1つの大きな問題点があるわけです。

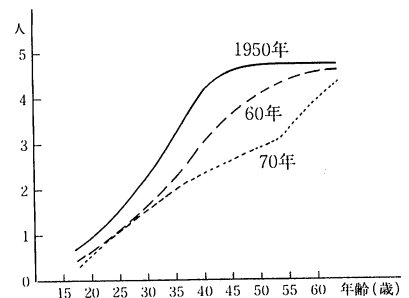
20歳台の人について、タバコをのむ習慣がある人を把握して、その人が40歳台にガンになる時間の経過を追いかけて調査することができれば、分析の段階でのデータの扱い方は楽になります。これを追跡調査と言います。追跡調査ができない場合は、それに代わる方法として、その逆をやっていく調査もあります。それを回顧調査と言います。まず、今現在40歳台の人を選び、ガン患者何人・健康者何人と把握しておきます。それから、そのガン患者・健康者それぞれについて20歳台に喫煙習慣があったかどうかを調べるわけです。回顧調査の場合、例えば、ガンの患者を病院の患者から調べ、健康者は別の調査で調べるというやり方をします。そうして、対比のためには、2つのグループのサンプル数をなるべく同数にする方がよいので、世の中の実態(ここで例では1:99)とは違った割合でサンプルをえらぶこととなります。その意味で、患者の数と健康者の数を加算する前に、細工が必要になってくるのです。回顧調査の場合、患者1人、健康者99人の割合で調査する必要はありません。患者1人を調べても仕方ありません。たとえば、患者100人くらいは調べないと正しい判断ができません

ん。これと対比するので、健康者は999人もいないわけです。その当然の結果として、回顧調査の抽出の割合が違ってきます。そのことを忘れて、両者をウっかり加算してはいけません。この事例に限らず、因果関係を議論するときには、時間差のあるデータが必要であり、それを求めるために回顧調査をすることが多く、こういう問題が起こることも頭においてほしいと思います。

3. クロスセクション分析とコーホート分析

2と関連して、〔例一3〕をみてください。

〔例一3〕 次の図は、各年次のセンサスによって求めた「既婚女子1人当たり子供の数」を母親の年齢別に示したものである。この図示法に問題はないか。

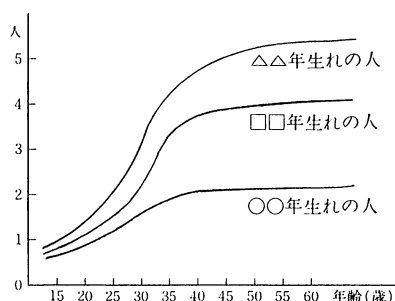


図表の問題みたいですが、本当のねらいは、統計データの見方ないしは調査の仕方・分析の基本にかかわる問題です。この図は、30歳の母親は平均子供何人くらい、50歳の母親は平均子供何人くらいというデータです。正確には、「産んだ子供の数」です。(産んだ子供の数は国勢調査では得られませんから、センサスデータを使って推計したと読んでください。)たとえば、1950年に30歳の母親は2.5人子供を産んだというように読むわけです。こういう情報を、1950年にセンサスの結果を使ってまとめ、子供の数をタテ軸に、母親の年齢をヨコ軸にとって、それらの関係を図示しているわけです。同じことを1960年について、1970年について示してあるのですが、60年、70年の場合何となくおかしくなっています。1970年の数字をグラフから読んでみます。まず、25歳台の母親をみると子供の数は0.9人、30歳

になると結婚してかなりたっていますから1.5人、・・・ところが60歳になると急にふえて4人になります。もとの数字自体は間違っていない。それにしても、これほど急にふえるのでしょうか。人間の心理として、線にそってグラフを読みたくるわけですが、そうするとおかしなことになるわけです。50歳を過ぎてから子供の数が急にふえるなど考えられません。そういう意味で、この図は人を誤解させます。

1970年のデータということ、1970年の調査によって得たデータということ。そのことは知っておくべきですが、データを見る観点では、70年調査という枠のなかで見たいけないのです。70年の調査の対象という意味では同じであっても、質的に異なったいろいろな人が混じっています。例えば、戦前生れと戦後生れという具合です。この例題では「70年にいる人と50年にいる人を比べて子供の数が多い少ない」をみるのではなく、「戦前生れと戦後生れでは子供の数が違う」のをみるのがポイントです。言い換えれば、「戦前は子供5人が普通だったが、戦後は2人になった」ということを表わしたいのがこのデータなのです。そこで、仮りに1925年生れを境目してみます。1970年に1925年以前生れの人、年齢でいいますと45歳以上ですから、年齢45歳のところに区切りを入れれば、そこから後は1925年以前生れになります。同様に、1960年センサス、1950年センサスについても区切りを入れてみます。そうして、これらの区切りを線で結んでみれば、1925年生れの人について、50年、60年、70年の状態を、経過を追ってみていくことができるわけです。

図一三 母親の生年別子供数



図一三は、このように線を結びなおしたものです。〔例一三〕の表わし方では、生れた年次を無視して、ある時点に現存

している人についてみることになります。これをクロスセクションでみると言います。これに対して、図一三の見方は、「同じ年に生れた」という条件で集団区分をし、その集団についてみるわけです。これをコーホートでみると言います。

調査をするときは、1950年センサス、1960年センサス、1970年センサスと一定の時点でおさえますが、データを見る時は必ずしも一定の時点でおさえるのではなく、同じ年に生れた人のその後の経過をみるという見方をする場合が多いのです。だから、統計データの見方としてコーホートは、たいへん重要な概念です。日本語では「同時出生集団」と呼んでいます。コーホートでみるというテクニックを知っていると、いろいろな分析ができます。特に用いられるのは人口統計の場合ですが、それ以外の場面でもコーホートの考え方は重要です。例えば、近頃国民の社会観・生活観が変わってきたと言われていています。その変化が、戦前生れの人の特長なのか、戦後生れの人の特長なのか、それとも総ての人を含めて社会観が変わってきたのか、それを判断したいという問題意識があるとき、コーホートによる分析をしなければなりません。そうした意味では、コーホートについて分析するのか、一定の時点で分析するのかということは、大変重要な分析のテクニックなのです。

編集子より； このシリーズは、上田先生が昭和54年3月に総理府統計研修所で講義されたものを収録・編集したものです。

昭和55年国勢調査の計画概要

本年10月1日には、全国いっせいに国勢調査が実施される。特に今回の調査は、10年ごとのいわゆる大規模調査であること、国際連合の提唱に応じて世界各国が調査を行う1980年世界人口・住宅センサスの一環でもあることが特徴である。

昭和55年国勢調査の計画は、第4次試験調査の検討結果をふまえて固めていくことになるが、現段階における計画の概要は次のとおり。

1. 調査の目的

国勢調査は、国内の人口の分布とその構成を明らかにし、国や都道府県、市区町村の各種行政施策その他の基礎資料を得るために行われる。

なお、法令上の主な利用としては、都道府県議会や市町村議会の議員定数の決定、市の設置要件、地方交付税の算定基準、過疎地域の指定等があげられる。

2. 調査の時期

昭和55年10月1日午前零時現在により行われる。

3. 調査の範囲

(一) 調査の地域

国勢調査は、我が国の領土の全域について行うが、現在我が国の行政権の及んでいない地域は除かれる見込みである。

(二) 調査の対象

国内に常住するすべての人を対象とし、これらの人をその人がふだん住んでいる場所で、世帯ごとにまとめて調査する。

外国人も調査するが、外国軍隊の軍人や軍属とその家族及び外国の外交団や領事団(随員やその家族を含む)は、調査の対象から除外する。

今回の調査では、昭和50年に比べて全国の人口は約570万人増加し、1億1760万人位に、世帯数は約410万世帯増加し、3630万世帯位になる見込みである。

4. 調査の事項

調査事項の選定に当たっては、過去における調査の経緯、各省庁からの要望、学識経験者等の意見、社会情勢の変化等を考慮し、さらに試験調査における検討をふまえた22項目を予定しているが、最終的には、第4次試験調査の結果及び統計審議会の答申をまけて確定することになる。

〔世帯員について〕

- (1) 氏名
- (2) 男女の別
- (3) 世帯主との続き柄
- (4) 配偶者の有無
- (5) 出生の年月
- (6) 国籍
- (7) 現住居に入居した時期

- (8) 前住地
 - (9) 教育
 - (10) 就業状態
 - (11) 従業地又は通学地
 - (12) 従業地又は通学地までの利用交通手段
 - (13) 勤めか自営かの別
 - (14) 所属の事業所の名称及び事業の種類
 - (15) 本人の仕事の種類
- 〔世帯について〕

- (16) 世帯の種類
- (17) 世帯人員
- (18) 家計の収入の種類
- (19) 住居の種類
- (20) 居住室数
- (21) 居住室の畳数
- (22) 住宅の建て方

なお、調査事項は、最近の大規模調査であった昭和45年国勢調査では23項目、簡易調査であった昭和50年国勢調査では16項目であったが、今回は、前回の大規模調査に比べ、結婚年数及び出生児数の2項目を削除し、住宅の建て方を追加したものとなっている。

5. 調査の方法

(一) 調査の系統

調査は、総理府統計局を主管官庁として、内閣総理大臣一都道府県知事一市町村長一指導員一調査員の系統を通じて行うこととしている。

(二) 国勢調査指導員及び国勢調査員

ア 国勢調査指導員及び国勢調査員は、市町村長の推選に基づき内閣総理大臣が任命することとしている。

イ 国勢調査指導員は、市町村長の指揮監督の下に、国勢調査員の指導及び調査書類の審査を行う。

ウ 国勢調査員は、市町村長の指揮監督及び国勢調査指導員の指導の下に、調査票の配布及び取集、世帯名簿及び調査区要図の作成その他これらに附帯する事務を行う。

(三) 申告者

世帯主又は世帯の代表者は、調査票に記入し、又は調査員の質問に答えることにより、その世帯の世帯員について、各調査事項を申告する。

(四) 調査票

調査票は、4名連記(表裏両面記入)のマークシート形式の世帯票とする。なお、不在世帯については、マークシート形式でない調査票を用いることとしている。

6. 集計区分と結果の公表

集計は次の区分により総理府統計局で行うこととしている。なお、今回の集計に用いる光学式読取装置は、両面同時読取可能な最新鋭機を用いることとしている。

(一) 要計表による人口及び世帯数の集計

従来から行ってきた集計で要計表を用いて全国、都道府県、市町村別の人口及び世帯数を集計し、その結果は昭和55年12月末日までに公表することとしている。

また、今回は、公表と同時に官報に公示することを予定している。

なお、前回行った「世帯名簿による人口」の集計は、今回は行わない予定である。

(二) 抽出速報集計

調査票の中から抽出した百分の一の世帯の調査票を用いて集計し、全国、都道府県、人口50万以上の市の段階までの速報結果を集計し、昭和56年3月末日までに公表する予定としている。

(三) 全数1次集計

今回の調査では、全数による基本的な集計を2回に分けて行う計画としている。第1次集計は、産業及び職業分類の格付けを行う前の段階で光学式読取装置にかけて、基本的な事項について市町村段階までの結果の集計を行う。この集計は、都道府県ごとに行い、集計の完了したのから逐次公表し、最終都道府県の結果は、昭和56年10月末日までに公表することを目標としている。

なお、この集計の段階で、都道府県及び市町村別の人口及び世帯数の最終確定結果を数回に分けて官報に公示することを予定している。

(四) 全数2次集計

1次集計の終わった調査票について、産業及び職業分類の格付けを行い、再度機械にかけて産業別、職業別等就業状態に関する事項を含む市町村段階までの結果を集計し、前記(三)の結果公表後、集計の完了したのから逐次公表する。最終都道府県の結果は、昭和58年2月末日までに公表することを目標としている。

なお、基本的な事項の一部については、全数1次集計、全数2次集計の段階で、調査区別に集計を行い、小地域統計への利用に供することとしている。

(五) 抽出詳細集計

20%抽出調査票を用いて、多重クロス表及び産業、職業に関する詳細な結果を、原則として都道府県段階まで集計し、昭和59年1月末日までに公表する予定としている。

なお、今回の集計上の主な特徴としては、①前回行った20%抽出速報集計をやめ、全数1次集計の早期化をはかる。②小地域別統計(調査区別、国勢統計区別等)の充実をはかる。③特定世帯(例えば、母子世帯、高齢者世帯、夫婦共働き世帯など)に関する統計の特掲を行うことなどがあげられる。

現段階における計画のあらましは以上のとおりであるが、昭和55年国勢調査実施本部の設置により、計画を実行に移していくことになる。

(「統計ニュース」創刊号から)

■ 10月1日を調査期日とした事情

国勢調査の調査期日は、第1回の大正9年以来、一貫して10月1日となっている。この期日を選んだ事情は、「大正9年国勢調査記述編」(報告書)に掲載されているが、その内容を要約すると次のようになる。

「調査の時期いかんは調査目的の成否、調査の難易等に至大の関係を有するものにして」と前置きしながら、

「まず、年末・年始は従来常に本籍人口又は現住人口の調査時期であるから、比較上便宜であり、かつ、年齢計算も容易で極めて好都合ではあるが、諸取引の決算、年賀の風習などがあり、しかも一般に冬期は山陰・北陸・東山・東北・北海道に積雪が深く、実査の時期として不適當である。

次に夏期は炎熱が甚だしく実査の時期としてはこれまた不便であり、春又は秋に求めざるを得ない。しかし、春は旅行、遊山する者が多く、人口分布の常態を失するものがある。

以上のことから、比較的人口分布が常態に在り、人々の職業的活動が盛んで、しかも全人口の大半を占める農業就業者にとって全くの農繁期ではなく、ちょうど年の4分の3を経過した10月1日をもって調査期日とした。」

と述べ、さらに、

「第1回国勢調査は(常住人口ではなく)現住人口^(注1)を調査することとしたが、現住人口の欠点を極力避けるため、調査の時期を人口移動が比較的少ないと認められる10月1日に定めた。」
としている。

一方、この10月1日は、4月1日に始まる会計年度の中央日であるから、その調査結果は年度平均値として、行政上の利用に便利であるという見方もある。

注1) 本籍人口は戸籍簿による人口、現住人口は本籍人口を基礎に寄留届などによる出入りを差増した推計人口で、ともに内閣統計局が算出公表していたものであり、国勢調査以前の明治・大正期における全国地域別人口を示す唯一の資料であった。

注2) 国勢調査では、大正9年の調査から昭和22年の臨時国勢調査までは、その人が10月1日午前零時にいた場所(現在地)で調査してきたが、昭和25年からはふだん住んでいる場所(常住地)で調査する方式に切り替えている。

(「国勢調査ニュース」No.2から)