

統計調査について

統計調査のむずかしさ

調査によるデータの収集は、統計的処理によるデータ解析の出発点であり、ここが正確でないと後の処理は全く無意味になってしまいます。しかし、不注意な調査は案外いろいろなところにあります。また、わざと不当な調査を行い、都合の良い結論を導く「我田引水」的な調査例も、残念ながらよくあります。

世の中に「アンケート」があふれているように、統計調査というのは簡単にできるように思われているふしがあります。しかし、例えば「質問数が多すぎて、だんだんいい加減に回答するようになってしまう」アンケートには、日常よく出会うものです。

統計調査は、その方法についてだけでひとつの本が書かれるほど、慎重にあつかわなければならないむずかしいものです。そこで、この本の最初では、「悪い調査」の例をいくつか見ていくことで、調査のありかたについて考えてみましょう。この章にあげた例は、いずれも「悪い例」として有名なものです。

なお、国が行う国勢調査などの統計調査については、調査が「ズル」によって歪められることのないように、統計法という法律によって、きびしい規制・罰則があります。2013年には、ある町で、「市に昇格するために、2010年の国勢調査の結果を改ざんして人口を水増しした」という疑いで当時の副町長が「統計法違反」で逮捕され、有罪となるという事件がありました。

良い質問とは

良い質問とは何かを考えるため、次の「悪い質問」の例を見て、何が悪いのかを考えてみてください。

1. もしあなたが車を運転していて、がまんできないくらい眠くなったら、あなたはどうしますか。
2. あなたは、世の中で一番大切なのはお金だと思いますか。
3. あなたは、「学歴よりも実力が大切だ」と思いますか。
4. あなたは、「宇宙開発は、安全保障上重要なので、進めるべきだ」という意見に賛成ですか。
5. 当店のお客様に対する対応はいかがでしたでしょうか、次の中からお選びください： (1) 不満
(2) おおむね満足

心理的効果を考える

人間の判断は、直観や印象といった心理的効果に大きく左右されます。その影響の例を見てみましょう。

1. 宴会コースのメニューをいくつか設定するとします。店主は、7000円のコースに人気が出てほしいとひそかに考えています。下のどちらにすれば、その意図が実現されやすいでしょうか。
 - (a) 宴会メニューは、6000円・7000円・8000円・9000円の4通りです。
 - (b) 宴会メニューは、4000円・5000円・6000円・7000円の4通りです。

2. 大学で、学生が講義を評価するアンケートを行います。下の2つの尋ね方で、回答に違いが出るでしょうか。

(a) あなたはこの講義に満足しましたか。次の数字に○をつけてください： 1 (非常に不満)
2 3 4 5 (非常に満足)

(b) あなたはこの講義に満足しましたか。次の数字に○をつけてください： 1 (非常に不満)
2 3 4 (非常に満足)

誰を調査しているのか

調査をするときは、当然「調べたい集団」を想定しています。ところが、調査した対象が、想定している集団と違ってしまふことがあります。また、集団の一部のデータを取り出す「標本調査」では、「無作為抽出」をする必要があります。無作為抽出については第8章で取り上げますが、簡単にいえば「調べたい集団からまんべんなく標本をとりだす」という意味です。しかし、この場合も、集団のうち偏った一部分だけを取り出してしまふことがあります。例を見てみましょう。

1. 同窓会の案内を出しました。返事に現在の年収を書いてもらい、同窓生の平均年収を求めました。この結果は信用できるでしょうか？

2. 日本家屋の伝統的規格では、鴨居の高さは五尺七寸（約173cm）または五尺八寸（約176cm）となっています。戦後、日本人の体格が大きく向上したところに、「最近の若い人は背が高い人が増えて、鴨居で頭を打つ人が多い」といわれ、規格を変えなければいけないという声があがりました。

しかし、当時の20歳前後の人について、統計データを用いて調べてみると、176cm以上の人は5%前後しかいませんでした。なぜ、鴨居で頭を打つ人がたくさんいるような気がしたのでしょうか？

3. 戦時中の英国での有名な例です。英国軍では、爆撃から帰還したパイロットに調査を行い、「敵攻撃機からの攻撃方向で一番多かったのは後ろ上方からである」という結果が得られました。この結果は信用できるでしょうか？

統計資料の読み方—数字にだまされないように

人は、数字で表現された情報はつい盲目的に信用してしまいがちです。しかし、その数字が表している意味をよく知っていないと、適用の限界や結果の意味を誤解し、あやまった結論を信じてしまうことにもなりかねません。統計的手法について勉強する前に、「数字で表された表現」を正しく読む方法を、例を通して見てみましょう。

数字から受ける印象

数字自体には、その値以上の意味はありませんが、人間は自分の常識や感覚と比較することによって、勝手な印象をもつものです。逆に、どのような印象を与えるかを考えて、数字を公表するという手が使われることもあります。

次のような表現について、どのような問題があるかを考えてみましょう。

1. この宝くじには、一等1億円が100本入っています。
2. (阪神高速道路の入り口の表示) 環状線までの所要時間：70分
「分」という単位には「短い時間」という印象があります。「70分」と書いてあるほうが「1時間10分」と書いてあるよりも短く感じます。
3. (a) 平成22年国勢調査によると、日本の人口は12,805,7352人、65歳以上の人口は2,924,5685人である。
(b) 平成22年国勢調査によると、日本の人口は約1億2805万人、65歳以上の人口は約2925万人である。

何を表現する数字なのか

日常見かける数字は、現実の何かを数字で表したものです。ですから、その「現実の何か」を誤ると、全く意味のない数字になってしまいます。

次のような表現について、どのような問題があるかを考えてみましょう。

1. いくつかの食料品販売店を調査したところ、全体の3%の商品について、内容量が表示されている量に足りない、という不正が発見された。したがって、消費者は平均して3%損をしていることになる。
2. 1時間あたりに発生する交通事故件数は、朝と夕方が多い。朝や夕方の時間帯には、何か生理的に事故を起こしやすい原因があるのだろうか。
3. 日本での実際のデータによると、たばこを1日に20本吸う人は、1本も吸わない人に比べて10倍肺がんで死亡しやすい¹。したがって、たばこを吸うことは大変危険である。

統計量を正しく理解しているか

統計量とは、調査したデータに対して何かの計算をして、データを要約したものです。わたしたちに一番なじみのある統計量といえば、「平均」でしょう。平均の計算のしかたはよく知っていると思いますが、「なじみがある」ことを過信すると誤解を生みます。

次のような表現について、どのような問題があるかを考えてみましょう。

1. 気象用語でいう「雲量」とは、快晴を0、本曇りを10として空のうち雲の占める割合を表現する。平均は5程度である。したがって、一年間のうち多少雲がある曇りの日が一番多い。
2. 「人生50年」という言葉があったように、100年前の日本人の平均寿命は50歳程度であった。当時の女性は5人6人の子供を生むことは当たり前であったので、子育てが終わった後の人生はたった数年しかなかった。
3. 「500系のぞみ号」は、運行開始当時、最高速度300km/hの「世界一速い列車」であった。

¹愛知県がんセンター中央病院のウェブサイト「がんの知識」より

今日の演習

1936年の米国大統領選挙の時の有名な話です。ある雑誌社は世論調査によって民主党／共和党のどちらの候補が当選するかを予想することにしました。そこで、自社の雑誌購読者と電話所有者・自動車所有者の1000万人を対象にアンケートを行いました。その結果、共和党の候補が当選すると予想しましたが、見事に外れてしまいました。さて、調査にどのような問題があったのでしょうか？