

統計の読み方

～もっともらしさに騙されないために～

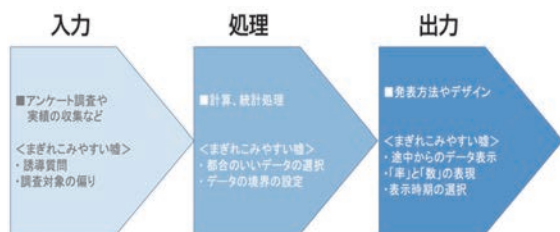


中小企業診断士
米倉 博彦

- 「世の中には3種類の嘘がある：嘘、大嘘、そして統計だ」。イギリスの政治家、ベンジャミン・ディズレーリの言葉だと言われている（異説もある）。
- ビッグデータ（大規模情報分析）の時代が到来しても、統計の嘘に騙されないための基本は何ら変わっていない。

■入力、処理、出力。各段階で「嘘」が入り込む

統計データを作るためには、大きくわけて「入力」「処理」「出力」の3つの段階がある。それぞれの段階で嘘が入り込む余地がある。具体的に見てみよう。



〈図表1：統計データ作成の3つの段階〉

1. 入力（誘導質問、調査結果の偏り）

データを集める段階。アンケート調査が代表的で、我が国で最も大きなものは国勢調査だろう（昨年実施された）。

この段階でまぎれこむ嘘は「誘導質問」がもっとも多い。

たとえば、あるテーマに対して、ポジティブな説明を行った後でそのテーマの是非について質問する。

回答者は、直前に与えられたポジティブな説明に引きずられて、肯定的な回答をしがちになる。

逆にネガティブな情報を与えれば、回答者は当該テーマに対して否定的な回答をすることが多くなる。

そのテーマに対して明確に賛成・反対の意見を持っている回答者は設問の影響を受けない。しかし、賛成でも反対でもない、このテーマに対して明確な意見のない層は、このような設問に影響を受けやすい。

結果として、アンケート結果は出題者の意図した方向に歪んでしまう。

また、調査対象がそもそも偏っているケースも

ある。日中平日の電話調査であれば、いくら無作為に電話をかけたとしても、その時間に自宅に居る人達（主婦、老人がほとんどだろう）からしか調査ができない。逆に、ネットを使ったアンケートでは、若者に偏る傾向がある。

2. 処理（都合のいいデータの選択、データの境界の恣意的な設定）

つぎは、集めたデータを処理する段階だ。

この段階では、都合のいいデータを選択し、都合の悪いものは排除するという誘惑が発生する。

統計データを集める場合、作業者は何らかの「仮説」を持っている場合がほとんどだ。その仮説を証明するために統計データを作るのだが、仮説に反するデータが出ると、どうしても「これは例外」と思いたくなり、実際にそのデータを排除してしまう。

また、データの境界線を恣意的に設定することで、結論に影響を及ぼすことができる。たとえば年代別のデータを分析するとしよう。

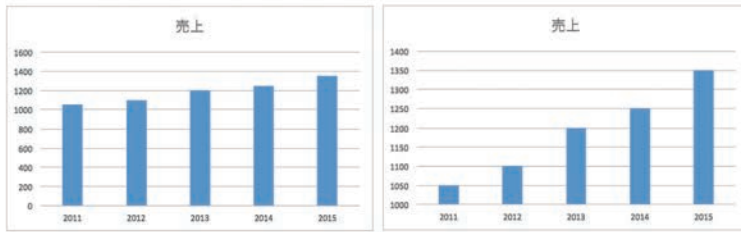
10歳刻みでやるのか、5歳刻みなのか、25歳～34歳（アラサー）といった区切りなのか・・・境界線をどうするかで、データの見え方が変わってくるかもしれない。

3. 出力（途中からのデータ表示、「率」と「数」の表現、表示時期の選択）

最後に、分析（統計処理）の終わったデータを出力する段階だ。

みんなに理解してもらえるよう、グラフや表などに加工した上で、ネットや紙媒体で公開することになるが、この段階でも「嘘」が入り込む。

ひとつは、「途中からのデータ表示」だ。図表2はその一例だが、グラフの目盛りをゼロからはじ



〈図表2：目盛りのスタート地点を調整する〉

めるか、途中からはじめるかでグラフのイメージは大きく異なる。

左のグラフだと「売上は微増」と見える。右だと「売上は大きく増加」と解釈されそうではないだろうか。

次は「率」と「数」の表現についてだ。例えば売上が昨年の100億円から110億円に上昇したとしよう。率で言えば10%アップ、数(金額)で言えば、10億円の上昇だ。同じように、売上が昨年の10万円から11万円に増えたとしよう。率は10%アップ、数(金額)で言えば、1万円のアップだ。あなたなら、率と数どちらで表現するだろう？

表示時期、表示期間を変えることで、統計の解釈を歪めることもできる。

2008年に、いわゆるリーマンショックが起きた。特に経済関係の統計データであれば、2008年以前を含めるかそうでないかで、グラフのイメージは大きく変わる。

見てきたように、統計データを作る際には複数の段階で「嘘」が入り込む余地がある。注意深く検証すれば、作成者の嘘(もしくは恣意的な解釈)を見破ることが可能だ。ただし、いくつかの項目については検証ができない(例えば、質問票が公開されないアンケート結果など)ので、その部分に関しては作成者の「良心」を信じるしかない現状だ。

■統計の嘘を見破る5つの鍵

ダレル・ハフという学者が1954年に書いた「統計でウソをつく方法」という本がある(ブルーバックス刊)。最終章のタイトルは「統計の嘘を見破る5つの鍵」。現代でも通用すると思うので、下記で説明する。

1. 誰がそう言っているのか？

統計データの出所はどこか、このデータによってメリットを得る組織・団体はないか。大学教授や社長、政治家などのいわゆる「権威」を利用していないか。

2. どういう方法でわかったのか？

前述の「入力」の段階。調査方法は適正か。アンケート内容に特定の回答を誘導するような記述はないか。

3. 足りないデータはないか？

やはり前述の「入力」の段階。調査対象者に偏りはないか、意図に合わないデータを排除していないか。

4. 言っていることが違ってやしないか？(問題のすり替えに注意)

「統計によれば、ガンになる人が年々増加している、これは加工食品の影響だ！自然に帰れ！」という意見があったとする。

本当にそうだろうか？単に高齢化の影響で、いままではガンになる前に死亡していた人々が長生きした結果ガンにかかるようになっただけでは？

少しずれるが、「数字は達成していないが、社員はがんばった」という主張も問題のすり替えである。社員が頑張ったのに、数字を達成していないということは、頑張る方向が間違っていたのかもしれない。

5. 意味があるのか？(どこかおかしくないか)

「コンサル実績多数！」→本当に多数なら、数字を示すのでは？

「5年間で1,000件のコンサルティング」→1年で200社の指導ができるのか？ちょっとアドバイスただけでは？セミナーの参加者もカウントしてないか？

「ここ数年の上昇傾向から、来年も上昇します！」→ここ数年上がったから、来年も上がるとは限らない。上がっている原因をきちんと説明しなければいけない。

なんとなくそれらしいけれど、実は意味のないデータや表現を使って自説を主張する人に気をつけよう。

■まとめ

統計をうまく操作すれば簡単に「相手をこちらの思った方向に動かす」ことができる。

統計で、調査で、アンケートで、このような結果が出ている！という主張を鵜呑みにするのではなく、今回の内容をもとにちょっとだけ立ち止まって、数字を検証してみよう。

それだけで、もっともらしい統計に騙されることはなくなるはずだ。