

平均のトリック



中小企業診断士
米倉 博彦

- 国税庁の調査によれば、日本人の平均年収は414万円だそうだ。(平成25年分 民間給与実態統計調査 国税庁長官官房企画課)
- 読者の実感に合うだろうか? 「わたしはそんなにもらってないよ」と思っただろうか?
- 今回は、平均値のトリックとそれに騙されないための注意点について書く。

■通常の平均(算術平均)

皆が通常使っている平均値は、正確には「算術平均」と呼ばれるものだ。データを全て合計してデータの個数で割り算するだけだ。学校でも習うし、直感的にも理解しやすい。ちなみにエクセルではAVERAGE関数を使って計算できる。

上記の平均年収(414万)も、算術平均を用いて算出している。「みんなにそんなもらっているのか・・・」と落ち込む必要はない。なぜなら「外れ値」が大きな影響を及ぼしているからだ。

年収データの場合、最も低い年収はゼロ円である。マイナスになるとすれば、お金を払いながら仕事をすることになり、そんな奇抜な方はあまりいないと思う。

しかし「最も高い年収」というのは存在しない、お金持ちにはきりが無い。国税庁の調査では年収2,500万円以上の方は11万3千人いるようだ。このなかには、2,501万円の方もいれば、数億円の方もいるだろう。

極端な例で考えてみよう。

A氏の年収300万、B氏の年収600万。二名の平均年収は450万だ。この中に、年収3千万のC氏が入るとどうなるか? 平均値はいっきに押し上げられて、平均年収1,300万になってしまう。A氏とB氏は「そんなにもらってないけどなあ」と不思議に思うことだろう。

この場合のC氏が「外れ値」だ。他のデータと大きく違う値が存在すると、算術平均に大きな影響を与える。(図1)

図1の場合、一人だけずば抜けて年収の高い方(8千万円)がいるために、平均値が押し上げられて、多数の方の実感とはずれた値となっている。

■中央値(メジアン)

中央値(メジアン)は、データの総数のちょうど中心に位置するデータが取る数値のことである。

データが51個ある場合は、26個目のデータの値が中央値である。データが偶数、たとえば50個の場合は、25個目と26個目の値の平均が中央値となる。

エクセルではMEDIAN関数で計算できる。図2のデータだと、中央値は500万となる。算術平均と比較して、外れ値の影響を受けにくいためにより実感に近い数値が出る。

図1：算術平均。外れ値により平均が実感とずれる。

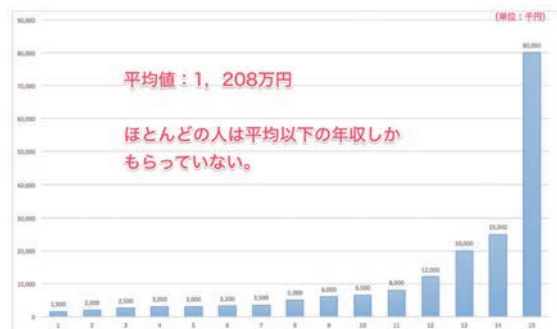
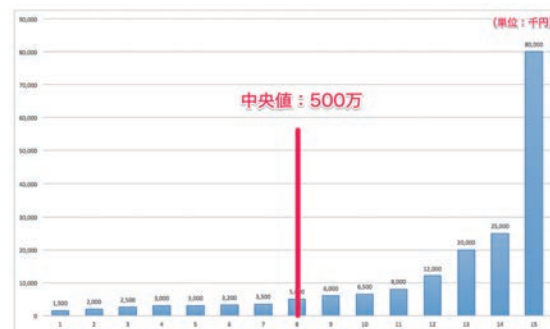


図2：中央値(メジアン)



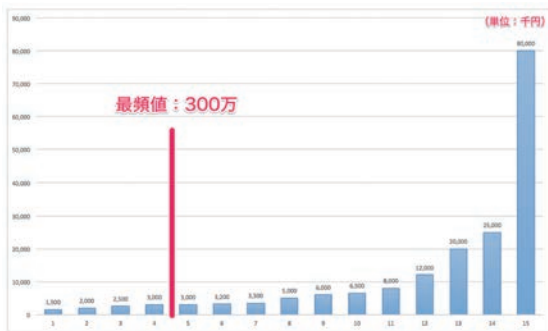


■最頻値 (モード)

最頻値 (モード) は、データ全体の中でもっとも数の多い数値のことである。エクセルではMODE関数で計算できる。図3で見ると、ここでの最頻値は300万となる。

最頻値に位置する人が最も多い、つまり、多数派なわけである。こちらも算術平均と比較すると、中央値とは違う意味で実感に近い数値が出てくる。

図3：最頻値 (モード)



■算術平均が有効なケース

算術平均が有効なのは、データが正規分布 (釣り鐘型のグラフ) に従っている場合だ。例えば人間の身長などは多少の幅はあれども、100センチとか3メートルといった極端な身長を持つ人はほぼ存在しない。こういう場合は、算術平均、中央値、最頻値すべてがだいたい同じくらいの数値になるため、算術平均のみ計算すれば事足りるだろう。

■元データを疑う

ここまで3種類の平均について見てきた。しかし、どの方法を使ったとしても、そもそものデータの解釈がおかしければ正しい分析はできない。

年収なら、調査する団体によって世帯あたりだったり一人あたりだったりそもそもの基準が違っている場合もある。(国税庁の調査は一人あたりのもの)

世帯あたりの調査であれば、共働きの家庭が増えている現在では、一人一人の年収が下がっていても世帯でみれば増加しているケースもあるだろう。

調査対象がネットアンケートなら若い人のデータが多くなるし、平日昼間の電話調査であれば、どうしても高齢者や主婦の回答が多くなる。

アンケートの質問内容も問題だ。なかには、特定の回答に誘導するような質問を書いているケー

スもあるので注意が必要である。

「かなり良い」「良い」「普通」「悪い」「かなり悪い」の5段階で回答する場合と、「普通」を抜いた4段階で回答する場合でも数値が変わる可能性がある。5段階の場合、真ん中の選択肢である「普通」を選ぶ人が多くなるからだ。

平均値を算出するために使ったデータ (生データという) をどうやって取得したのかを把握しなければ、そこから出てくる結論はあまり意味のないものになってしまう。

■最後に

算術平均は便利な概念だ。ただし、前述のように算術平均はデータのばらつきが大きすぎると実感とズレる場合がある。中央値 (メジアン) や最頻値 (モード) も適宜計算することで、そういったズレをなくすることができる。

また、平均値を取るということは、逆に言えば個々のデータの特徴を消してしまうことになる。平均値はあくまでデータを大づかみする際の参考としてとらえ、できるだけ棒グラフや折れ線グラフなど、より生データに近いもので分析するよう心掛けた方が良い。

ここまで書いておいてなんだが、性格や企業経営など多数の項目が複雑に絡み合ったものに関しては、平均はほとんど意味はなさない。「平均的な性格」や「平均的な会社」は統計上は存在するが、現実には存在しないのだ。

よって、どんな性格の女性にも通じる魔法の口説き文句や、どんな会社の売上も数倍になる魔法のテクニックは、存在しない。あると主張するのであればその方は詐欺師か、論理的な考え方ができない方かのいずれかだろう。

絶対の手法が存在しないということは、個別の事情に合わせたオーダーメイドの対応が必要になるということだ。

だからこそ、筆者のような経営コンサルタントの存在意義があるのだろう。いまのところは。

参考：平成25年分 民間給与実態統計調査 (国税庁 長官官房企画課)
<https://www.nta.go.jp/kohyo/tokei/kokuzeicho/minkan2014/minkan.htm>