

数学という名の自由の翼

第10回 2015年1月

偏差値について その①



かなり昔のことですが、ある脳科学者が、ツイッターで偏差値について過激な発言を行ったことに端を発して、侃々諤々の偏差値論争が起こったことがあります。

でも、論争している人自身が偏差値とは何かをちゃんとわかっていない場合もあります。偏差値を語る前に、偏差値のことを良く知ることが、賢い市民の考え方ですよ。

そこで、今月号と来月号の2回に分けて、偏差値とは何かについて、考えてみようと思います。

1 代表値と平均値

うかれぎつね君とマドンナうさぎさんが、全国規模の2種類の数学の模試を受けました。うかれぎつね君の得点は90点で、マドンナうさぎさんは70点でした。

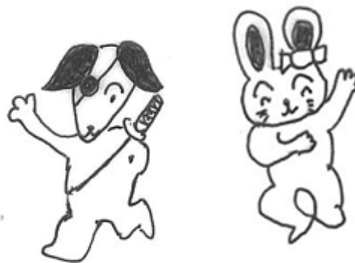
さて、このときどちらが優秀な成績といえるでしょう。うかれぎつね君の得点が高いからといって、単純に彼が優れているといってしまうはいけません。

もし、彼が受けた模試の平均点が95点で、マドンナうさぎさんの受けた模試の平均点が40点だとしたら、マドンナうさぎさんの方が平均より30点も上回っている一方、うかれぎつね君は平均より低い点数だったということになります。

このようにテストの点数を見る場合は、自分の点数だけでなく、平均値からどれだけの隔りがあるかを見ることで、集団の中で、自分の相対的な位置を知ることができるわけです。

ここで、「平均値」のことをその集団を代表する値として「代表値」といいます。代表値は平均値の他に得点分布の中間の値を示す「中央値」(メジアン)

や、最も頻度が多い値を示す「最頻値」(モード)などがあります。



キャラクターは伊藤潤一先生作成

2 散布度

私は、統計の最初の授業で、次のような問題を出して生徒に考えさせています。

【問題1】

コジロー企画と、しもまち商事という2つの会社があります。どちらも社員数は100人で、業務内容は同じなのですが、コジロー企画の社員の平均月給は50万円、一方、しもまち企画の方は30万円です。

ところが、コジロー企画の多くの社員が会社をやめたがっているというのです。なぜだと思いますか？

【問題2】

ガンバミズモーターズと、元気ネコ自動車教習所という、2つの自動車学校があります。

それぞれに通う教習生100人の、運転免許の学科試験の平均得点は、ガンバが60点、元気が70点ということでした。

ところが、人気があるのはガンバミズモーターズの方なのだそうです。なぜだと思いますか。

このような問題を提示し、班で話し合わせます。発表してもらおうと、実にいろいろな意見が出てきてとても楽しい授業になります。

問題1でよく出てくる解答は、「しもまち商事の方が福利厚生がしっかりしている」とか、「コジロー企画はバファハラが横行している」など。

問題2では「ガンバミズモーターズの教官が優しい」などという意見が現れます。

まあ、これらは、いわば答えのない問題ではありますが、私が用意した解答は次の通りです。

【問題1の解答】

実は、コジロー企画の社員の給料を度数分布表で表すと次のようになっていました。

給与	人数	給与×人数
¥ 40,100,000	1	40,100,000
¥ 190,000	40	7,600,000
¥ 100,000	19	1,900,000
¥ 10,000	40	400,000
合計	100	50,000,000
平均		500,000

一方しもまち商事は次のようになっていました。

給与	人数	給与×人数
¥ 500,000	5	2,500,000
¥ 400,000	25	10,000,000
¥ 300,000	38	11,400,000
¥ 200,000	26	5,200,000
¥ 150,000	6	900,000
合計	100	30,000,000
平均		300,000

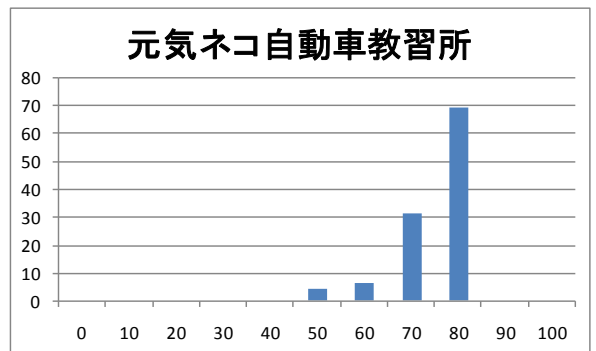
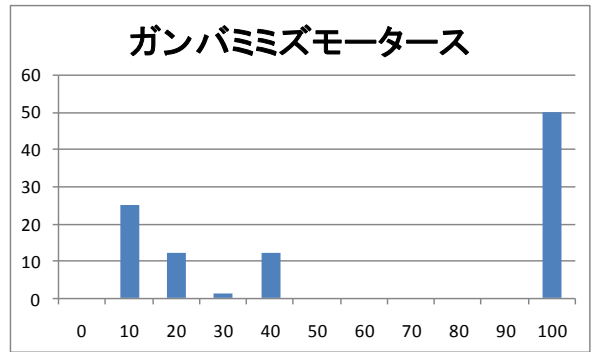
つまり、コジロー企画では、社長？1人が4千万円以上の給料をもらっていて、彼を除いた社員の平均給料は、10万円であることがわかります。これでは、社員は不満タラタラだったのは当然ですね。

それに比べて、しもまち商事は、バランス良く給料が配分されているように見えますね。

【問題2】

それぞれの自動車学校の得点分布をヒストグラムにかくと次のようになりました。

ここで、自動車免許の学科試験は90点以上とらなければ合格とはならないことに注意が必要です。



このヒストグラムを見ると、ガンバは、50名の合格者を出しているけれど、平均点の高い元気ネコの方は、1人も合格者を出していないことが判明しました。

このような問題によって、平均という代表値だけで物事を判断することは危険であることを説明するわけです。

では、また、テストの話に戻しましょう。あるテストを、いずれも人数が40人のクラス、A組とB組で行ったとき、2クラスとも平均点が50点だったとします。このとき2つの集団は同等のレベルの集団といえるのでしょうか。

先ほどの2つの問題と同じように考えると、もしかしたらA組は100点満点の人が20人、0点の人が20人で結果として平均が50点で、B組は40人全員が50点で平均点が50点かもしれませんね。

そうなる2つのクラスはまるで性質の異なる集団といわざるを得ません。つまり、平均点だけではその集団の性質を見ることはできなくて、「得点分布の散らばり具合」という、もう一つの観点も必要に

なります。この「得点の散らばり具合」のことを「散布度」といいます。

テストが終わると平均点を気にする人がよくいますが、平均点と散布度をあわせて考察しないと意味がありません。

そういう意味でも、全国学力調査で、平均点を上回った学校を公表するという発想は、理解に苦しみます。

さて、散布度には、最高点と最低点の差をとった「レンジ」や、最近高校の教科書に登場した、現場であまり評判の良くない「四分位範囲」、平均からの差の絶対値の平均の「平均偏差」、平均からの差の2乗の平均をとった「分散」、そして、その分散のルートをとった「標準偏差」などがありますが、テストの評価を行う場合はほとんど「標準偏差」が用いられます。

「得点分布の散らばり具合」というのはいわば、「平均からの隔たり具合」なので、それぞれの得点の平均からの差をとって、その平均を求めれば良いだろうと思うかもしれませんが、そうすると、平均が50点の場合、60点の人は+10点、40点の人は-10点ということになり、相殺されて0になってしまう散らばりを評価できません。

じゃあ「差の絶対値」をとって平均偏差を考えるのが普通と思う人もいるかもしれませんが、それでもよいのですが、一般に統計資料を分析する場合

「微分積分」という手法が用いられたりするので、絶対値が付いた式はそういう処理に向いていないという欠点があります。

そこで、差の絶対値ではなく、「差の2乗」をつくることでプラスマイナスが相殺されることを防ぎ、解析可能な変数にするのです。また、2乗することで、差を大きくして違いを強調するという意味もあります。

標準偏差は分散のルートをとることで1次のオーダーに置き換えたものです。例えば100点満点のテストで、分散が100といわれてもピンときませんが、標準偏差が10点といわれれば、平均点が

60点の場合、55点だった人は標準の範囲内だなどほっとしたり、80点だった人は、標準からかなり離れた優れた点数だなどと思ったりできるわけですね。

それがどの程度優れているのか、並はずれているのか、そうでもないのか、ということを数量化するための根拠として正規分布の理論があり、微分積分など数学の手法が用いられるわけです。

そして、テストなども含め、一般の統計資料は、しばしば他者や過去との比較が行われなければならないので、「世界標準」の規格が必要です。みんなが「標準偏差」を用いているときに一人だけ「平均偏差」を用いるのはうまくありません。そういうわけで、世界中の殆どすべての統計資料は「平均値」と「標準偏差」を用いているのです。

以上で、偏差値について考える準備ができました。

3 偏差値とは

偏差値は、ある集団において行ったテストの点数を、次の式で置き換えた得点です。

$$\text{偏差値} = \frac{\text{得点} - \text{平均}}{\text{標準偏差}} \times 10 + 50$$

ここで、偏差値を z 、自分の得点を x 、平均を m 、標準偏差を h とおくと上の式は次のように文字式で表すことができます。

$$z = \frac{X - m}{h} \times 10 + 50$$

では、この式の意味は何なのか、ステップを踏んで調べていきましょう

おっと、紙面が尽きてしまいました。

この続きは次号で。