雑感　偏差値に関する誤解と盲信

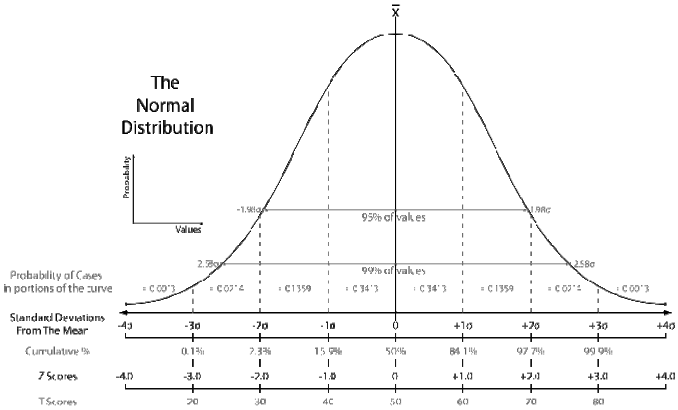
■　いわゆる進学校に勤務していると，生徒に受験させる「模試」の結果が明らかになり，昨年度との比較や，他校との比較等といった情報まで出回る．そういったデータ処理は，ほとんどが「偏差値」で行われている．

■　偏差値 Z は言うまでもなかろうが，得点 X と平均値 m，標準偏差 σ に対して， $Z=\dfrac{X-m}{\sigma}\times 10+50$ で定義される．

得点 X を見ただけでは分からないが，平均値や標準偏差によって「標準化」（「正規化」？）された偏差値 Z を見れば，平均よりどの程度良いのか，全体のどの位置にいるのかが分かるし，他の試験との比較もできるというふれ込みの「値」である．

偏差値が世の中に出てから 60 年ほど経つらしいが，誤解や盲信も多いのが実情だ．

■　次に載せたのは，例の Wikipedia にある図であり，



「偏差値の利用価値が高いのは、母集団の数値の分布が正規分布に近い状態の時である。分布のピークが 2 箇所ある場合など、正規分布と大きく異なる場合には適切な指標となりえない場合がある。

分布が正規分布に近い場合は、40 から 60 の間に約 68.3%、30 から 70 の間に約 95.4%、20 から 80 の間に約 99.73%、10 から 90 の間に約 99.9937%、0 から 100 の間に約 99.999953%が含まれる事が知られている。つまり、
偏差値 60 以上（あるいは 40 以下）は、全体の 15.866%。
偏差値 70 以上（あるいは 30 以下）は、全体の 2.275%。
偏差値 80 以上（あるいは 20 以下）は、全体の 0.13499%。
偏差値 90 以上（あるいは 10 以下）は、全体の 0.00315%。
偏差値 100 以上（あるいは 0 以下）は、全体の 0.00002%。

例えば、全受験生が 100 万人いた学力試験で偏差値を求めると、偏差値 80 以上となる者は、ほぼ 1350 人となる。

平均値から大きく離れた場合は 0 から 100 の間に収まらないが、その割合は非常に低く、約 0.000047%、つまり約 200 万分の 1 しかない。偏差値の上限値、下限値は元となる標本の分布によって決まるものであり、いかなる実数をもとりうる。」
との解説がある．

■　ここで重要なのは，「分布が正規分布に近い場合」という「条件」である．実際の試験では（模試を含めて），そんなに正規分布近似できるようなヒストグラムにならないというのは経験上の真実である．

したがって，正規分布近似できないような分布では，偏差値は余り役に立たないことがある．

■　実際，先日見た模試（記述である）の数学の度数分布が，昨年度は偏差値 75 以上まで分布があるのに対して，偏差値 67 程度で頭打ちになっていた．

平均点を見ると，200 点満点で，昨年は 80 点台，今年は 120 点ほどあった．これから予想できるのは，問題が易しく（もしかすると満点続出で），満点 200 点の偏差値が 67 程度になったのであろうということである．

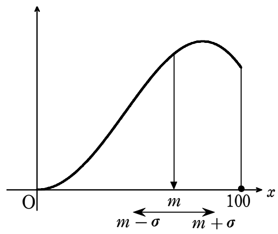
このような試験だと，前回の偏差値が 70 の生徒が，今回満点を取って偏差値 67 になり，「偏差値が下がってしまった（出来が悪かった）」ように見えてしまう．

実際は，200 点以上の点数がないため，それ以上の「学力（？）」を判定できない試験だったということである．

偏差値だけを見ていると，こういった間違った判断をすることになる．

■　偏差値の分布は「正規分布」だと信じている人も少なくないが，1 次式での変換だから，「山」の形は変わらない．

例えば， $y=\dfrac{1}{4162.12}x\sin\dfrac{x}{40}$ $(0\leqq x\leqq 100)$ を確率密度関数とする得点分布 $\left(\displaystyle\int_0^{100} y\,dx=1\right.$ となるように係数調整 $\left.\right)$ に対して，



平均 $m=\displaystyle\int_0^{100} xy\,dx=67.62$，

分散 $\sigma^2=\displaystyle\int_0^{100} x^2y\,dx-m^2$

$=5001.28-67.6151^2=429.5$ から，

標準偏差 $\sigma=20.72$ である．

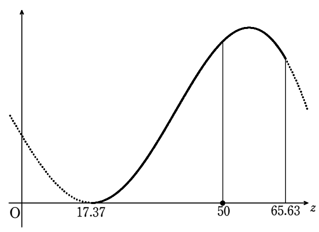
度数分布グラフは右のようになり， $m+2\sigma$ は 100 点を超えている．

偏差値 z は $z=\dfrac{x-67.62}{2.072}+50$ で定義される．

$x=2.072z-36.00$ であるから， z の分布は

$y=\dfrac{1}{4162.12}x\sin\dfrac{x}{40}$ $(0\leqq x\leqq 100)$ にこれを代入して，おおよそ

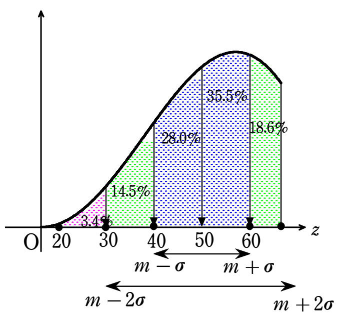
$y=(0.00050z-0.0086)\sin(0.052z-0.90)$　$(17.37\leqq z\leqq 65.63)$ となる．



このグラフは次の通りであり， x の分布の「山」と形は同じで，適切な縦横方向の拡大・縮小をし，平行移動して重ね合わせることができる．

満点の 100 点でも偏差値は 65.63 しかない．

この分布では，右図のような割合で分布しており，正規分布の割合とはだいぶ異なる．



■　多くの場合，得点分布は正規分布ではないから，「偏差値からおおよその順位が分かる」というのは迷信に過ぎないと思っているし，偏差値比較をしたから，正しい比較ができているというのも間違いであると思っている．

■　なお，偏差値は 100 を超える値もあり，模試などで散見される．偏差値が負になることもあり，40 年の教員生活の中で一度だけ目撃がある．ほとんどの生徒が 70～100 点といった数学の試験で，30 点を取った 1 人の生徒の偏差値である．

2014 年 6 月 9 日