平均と標準偏差、変動係数

＜いみ＞

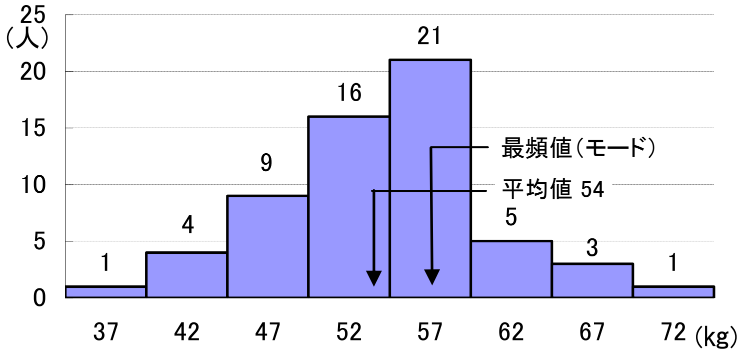
　ある集団についてのデータがどのように分布しているかを表すものとして、その集団の代表値を示す平均値及びそのばらつき具合を示す散布度がある。平均には算術平均が、散布度には標準偏差、変動係数などがある。

＜たとえば＞

表1　度数分布表

体重（階級）kg（中央値）	人数（度数）f
35～39（37）	1
40～44（42）	4
45～49（47）	9
50～54（52）	16
55～59（57）	21
60～64（62）	5
65～69（67）	3
70～74（72）	1
計	60

図1　ヒストグラム



注：平均値はそれぞれの階級の中央値をもとに計算した。

表2　平均と分散・標準偏差の計算

生徒	数学テストの得点（X）	偏差（X－μ）	偏差の2乗（X－μ）^{2}
A	85	15	225
B	82	12	144
C	80	10	100
D	74	4	16
E	72	2	4
F	68	−2	4
G	64	−6	36
H	61	−9	81
I	59	−11	121
J	55	−15	225
合計（10人）	700 ΣX 平均 μ＝70	0 Σ（X－μ）	956 Σ（X－μ）^{2}

平均（μ）（ミュー）

$$\mu = \frac{\Sigma X}{n} = \frac{700}{10} = 70$$

分散（σ^2）（シグマ）

$$\sigma^2 = \frac{\Sigma (X-\mu)^2}{n} = \frac{956}{10} = 95.6$$

標準偏差（σ）

$$\sigma = \sqrt{95.6} \fallingdotseq 9.8$$

注：n はデータの個数である。

表3 身長の平均値･標準偏差･変動係数

（単位：cm、%）

区 分		平均値 ①	標準偏差 ②	変動係数 ②÷①×100
男子	幼稚園	111.6	4.90	4.39
	高校生	170.7	5.86	3.43
女子	幼稚園	110.6	4.91	4.44
	高校生	157.9	5.35	3.39

表４ 偏差値の計算

生徒＼学科	実得点		偏差値	
	数学	国語	数学	国語
A	85	78	65	39
B	82	80	62	50
C	80	78	60	39
D	74	81	54	55
E	72	80	52	50
F	68	83	48	66
G	64	82	44	61
H	61	79	41	45
I	59	82	39	61
J	55	77	35	34
平均（μ）	70	80	50	50
分散（σ^2）	95.6	3.6	―	―
標準偏差（σ）	9.8	1.9	10	10

＜かんどころ＞

１．データの分布

データがどのように分布しているかその実態を把握するには、データをいくつかの階級に区分し、その階級ごとの個数(度数)を表にした度数分布表、あるいは、それを棒グラフにして表わしたヒストグラムが適している(表１、図 1)。

データの分布の中でしばしば現れる代表的な分布として、後半で説明する正規分布がある。正規分布では平均値が最も度数の多い値(最頻値(モード))となるが、平均値とモードが異なるようにデータが分布することもある。

２．平均値

一般に平均値には、算術平均

$$\frac{X_1 + X_2 + \cdot\cdot\cdot + X_n}{n}$$

（X_i（$i = 1, 2, 3 \cdots n$）：各データ、ｎ：データ数）が多く使われている。平均値は通常μ(ミュー)と表示される。

３．標準偏差と分散(表２)

平均値だけでは、データがどのように分布しているかが分からないため、平均値のほか、データのばらつきの範囲を示す散布度を使用する場合が多い。

データの散布度を示すものとして、平均値との差(偏差)の２乗を平均した値の平方根である標準偏差がよく用いられている。標準偏差は通常σ(シグマ)で表示される。

また、標準偏差の２乗を分散といいσ^2と表示される。

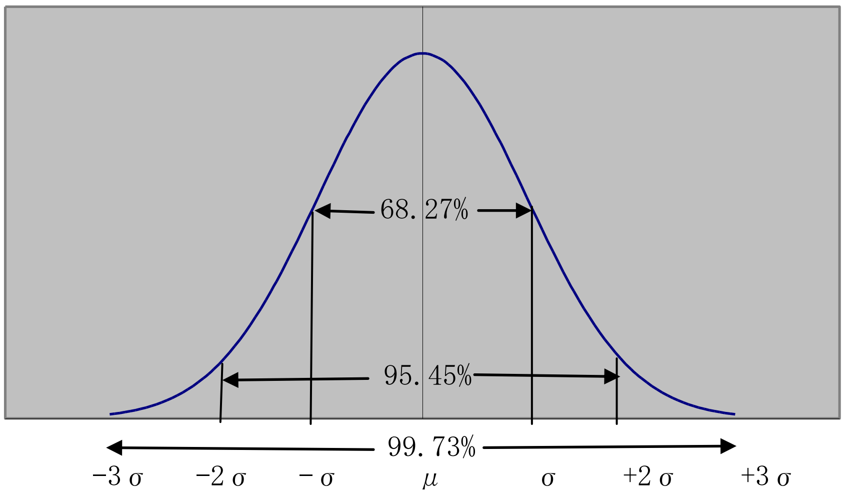
４．標準偏差の意味

標準偏差は、データの分布の広がり幅(ばらつき)をみる一つの尺度である。平均値と標準偏差の値が分かれば、データがどの範囲にどのような割合で散らばっているか(分布)がある程度明らかになる。図２のような平均値μを中心に左右対称の釣り鐘型の分布(正規分布)では、平均値(μ)と標準偏差(σ)及び度数の間に次の関係が成り立っている。

(範囲)	(その中に入るデータの割合)
μ±σ	0.6827（約2/3）
μ±2σ	0.9545（約19/20）
μ±3σ	0.9973

これは平均値±標準偏差の範囲に全データの 68.27%が、±標準偏差の２倍の範囲内に全データの 95.45%が分布するという意味である。

図２　正規分布表



５　変動係数（表３）

標準偏差はデータの分布のばらつきをみる一つの尺度であるが、２つの集団のばらつきの程度を比較する場合は、必ずしも有効な分析であるとは限らない。個々のデータの値が大きい（＝平均値が大きい）集団のほうが、標準偏差が大きくなる可能性が高いからである。

こういう場合、標準偏差をそれぞれの平均値で割ることで、集団の規模を考慮したうえでの比較を行うほうが有効である。標準偏差を平均値で割ったものを変動係数(ＣＶ）という。

幼稚園児と高校生の身長を比較した場合、男女とも、標準偏差は平均値の大きい高校生のほうが大きいが、変動係数は幼稚園児が大きいことから、幼稚園児のほうが高校生に比べ身長にばらつきがあることがわかる。

６．偏差値（表４）

学力テストの場合、素点自体よりも受験者全体の分布の中での個人の位置が問題とされることが多い。例えば、数学と国語のテスト結果では、難易度が異なっているために素点では自分の位置を比較できないが、標準偏差を用いた共通のものさし、偏差値で比較すればそれが可能となる。

偏差値（Ｔ）は、①平均点には 50 を対応させ、②平均から標準偏差のＺ倍だけ上回る（下回る）点数には 50 にＺの 10 倍を加え、規準化した数値のことで、次式で表される。

$$偏差値（\mathrm{T}）= 50 + 10 \times \frac{素点\mathrm{X} - 平均\mu}{標準偏差\sigma} = 50 + 10 \times \mathrm{Z}$$

（偏差値の平均は 50、標準偏差は 10 になるように設定されている）

偏差値を求めることで、異なるテストの得点であっても同一集団内での自分の位置を比較することが可能となる。また、全員の素点が正規分布に従うと考えられるとき、偏差値もまた正規分布に従う。$\mu \pm 2\sigma$ の範囲は（50±２×10 より）30 から 70 であり、偏差値が 30 より小さいまたは 70 より大きい人は、全体の 4.55%（＝100%－95.45%）である。70 より大きい人だけならば、半分の 2.275%となり、100 人中２位か３位までの順位と推測することができる。

表４では、生徒Ｅは実得点では国語のほうが高かったものの偏差値は数学のほうが高いことから、実際には数学のほうが成績がよかったことがわかる。