

平均と標準偏差

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^{*}(中心の値)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がよく用いられている。

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

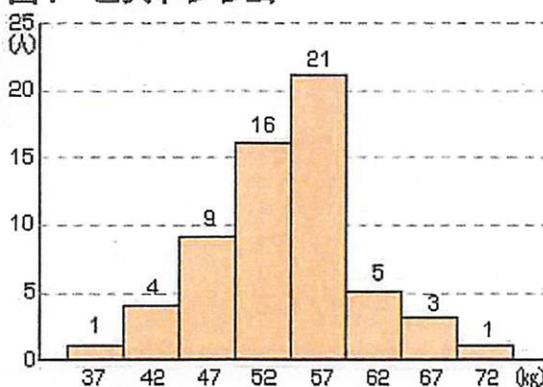


表2 平均と分散・標準偏差の計算

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

平均 (\bar{x})
(μ)
 $\mu = \frac{700}{10} = 70$

分散 (\bar{s}^2)
(σ^2)
 $\sigma^2 = \frac{956}{10} = 95.6$

標準偏差 (σ)
 $\sigma = \sqrt{95.6} \approx 9.8$

表3 あるクラスのテスト結果

生徒	学科	実得点		偏差値	
		数学	国語	数学	国語
A		61	79	41	45
B		74	81	54	55
C		55	77	35	34
D		85	78	65	39
E		68	83	48	66
F		72	80	52	50
G		64	82	44	61
H		80	78	60	39
I		82	80	62	50
J		59	82	39	61
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. 度数分布表・ヒストグラム

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、データをその大きさによりいくつかの階級に区分し、その階級ごとの個数(度数)をカウントして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)。

例えば、年齢別人口や従業者規模別事業所数など多くの統計表は度数分布表の形で作成され、また、年齢別人口をヒストグラムにした人口ピラミッドは人口構造の分析等によく用いられている。

2. 平均値*

一般に平均値には、単純平均

$$\frac{X_1 + X_2 + \dots + X_n}{n} \quad (X_i (i=1, 2, 3 \dots n) : \text{各データ}, n: \text{データ数})$$

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くなってしまうため、データの平均値のほか、データのばらつきの範囲を示す散布度を使用する機会が多い。

データの散布度を示すものとして、データの平均値との差(偏差)の2乗を平均し、これを変数と同じ次元で示すために平方根をとった標準偏差が最も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、平方根をとる前の値(標準偏差の2乗)を分散と

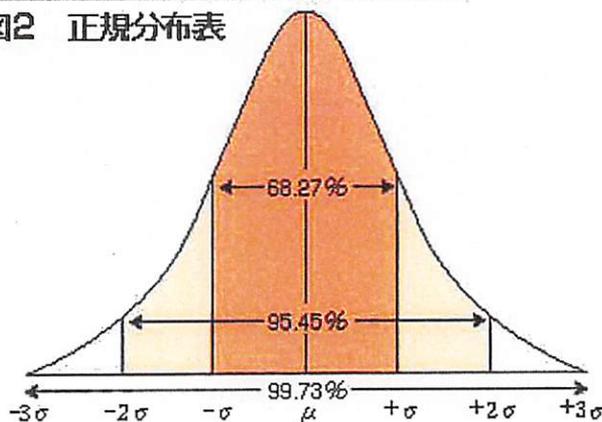
いい σ^2 と表示される。

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。

$\mu \pm \sigma$	0.6827 (約2/3)
$\mu \pm 2\sigma$	0.9545 (約19/20)
$\mu \pm 3\sigma$	0.9973

図2 正規分布表



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。

5. 偏差値(表3)

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。たとえば、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で比較すればそれが可能となる。

共通のものさしとしてよく用いられる偏差値(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倍を加えた(引いた)数値を対応させて、規準化した数値のことで、次式で表される。

$$\text{偏差値}(T) = 50 + 10 \times \frac{\text{素点} X - \text{平均} \mu}{\text{標準偏差} \sigma} = 50 + 10 \times Z \quad (\text{偏差値の平均は50、標準偏差は10である。})$$

偏差値は、平均点からの偏差を示す得点なので、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以下または70以上の人は、全体の4.55%(=100%-95.45%)で

ある。70以上の人だけならば、半分の2.275%となり、100人中2位か3位までの順位と判断される。