

散布度

有田光一郎
a2p21004

データのばらつきを測る尺度

- 例

表A	データ	算術平均	最大値	最小値	範囲	平均偏差
A群	10 20 30 40 50	30	50	10	40	12
B群	27 28 30 32 33	30	33	27	6	2
C群	10 30 30 30 50	30	50	10	40	8

絶対的散布度

- 度数分布の集中の程度および幅を表すのに変量の絶対値を用いて表す方法。

変量相互間の差にもとづくもの

- 四分位範囲
- 四分位偏差

四分位範囲

- 第一分位数をQ1、第三分位数をQ3とすると四分位範囲QRは $QR=Q3-Q1$ で求められる。
- 例 $QR=162-155=7$ となる。

昇順データ	149	152	155	157	158	161	162	164	167
大きさの順位	1	2	3	4	5	6	7	8	9
	E1		Q1		Me=Q2		Q3		E2

四分位偏差

- 四分位偏差QDは $QD=(Q3-Q1)/2$ で求められる。
- 例 $QD=7/2=3.5$ となる。

昇順データ	149	152	155	157	158	161	162	164	167
大きさの順位	1	2	3	4	5	6	7	8	9
	E1		Q1		Me=Q2		Q3		E2

算術平均に対する偏差に基づくもの (その1)

- 平均偏差
- 個々のデータ X_i 算術平均 M との差の絶対値の総和をデータ数で割った値を平均偏差という平均偏差 D は $D = \sum (X_i - M)$ で求められる。

算術平均に対する偏差に基づくもの (その2)

- a群
- $Da = ((10 - 30) + (20 - 30) + (30 - 30) + (40 - 30) + (50 - 30)) / 5 = 60 / 5 = 12$ となる。
- 同様にB群とC群を求めると $DB = 2$ 、 $DC = 8$ となりデータのばらつきが大きい順に平均偏差も「DA, DC, DB」という順番になった。

表A	データ	算術平均	最大値	最小値	範囲	平均偏差
A群	10 20 30 40 50	30	50	10	40	12
B群	27 28 30 32 33	30	33	27	6	2
C群	10 30 30 30 50	30	50	10	40	8

算術平均に対する偏差に基づくもの (その3)

- 分散
- 個々のデータを X_i ,算術平均を M としたとき、 $(X_i - M)$ を偏差といい、偏差の平方和をデータ数 N で割った値の事をいう。
- 分散 V は $V = \sum (X_i - M)^2 / N$ で求められる。

算術平均に対する偏差に基づくもの (その4)

- a群
- $V2 = (10 - 30)^2 + (20 - 30)^2 + (30 - 30)^2 + (40 - 30)^2 + (50 - 30)^2 / 5 = 1000 / 5 = 200$ となる。
- 同様に $Vb = 5.2$ $Vc = 160$ となる。

表A	データ	算術平均	最大値	最小値	範囲	平均偏差
A群	10 20 30 40 50	30	50	10	40	12
B群	27 28 30 32 33	30	33	27	6	2
C群	10 30 30 30 50	30	50	10	40	8

算術平均に対する偏差に基づくもの (その5)

- 標準偏差
- 分散は散布度として重要だが、その単位は生データの単位の平方になっているので、生データとの比較には不便である。比較をするには、ばらつきの度合いを平均と組み合わせて使用するために分散の平方根をとって、平均と同じ単位にする方法である。
- 標準偏差 $= \sqrt{\text{分散}}$

算術平均に対する偏差に基づくもの (その6)

表A	データ	算術平均	範囲	平均偏差	分散	標準偏差
A群	10 20 30 40 50	30	40	12	200	14.142
B群	27 28 30 32 33	30	6	2	5.2	2.28
C群	10 30 30 30 50	30	40	8	160	12.65

相对的散布度

- 四分位偏差係数
- $VQ = Q/Me = ((Q3 - Q1)/2)/Me$
- 平均偏差係数
- $VD = D/M$ or $VD = D/Me$
- 標準偏差係数
- $VS = S/M$
- 變異係数
- $CV = Vs \times 100$