2003 年 5 月 13 日

数学情報ゼミレジュメ (3)　(今野)

散らばりの尺度

値が大きいほど，散らばりの度合いがおおきい！

1. 範囲 — 分布の範囲を示すもの：

$$R = (\text{最大値}) - (\text{最小値}) = \max\{x_1, x_2, \ldots, x_n\} - \min\{x_1, x_2, \ldots, x_n\}$$

<u>データ１</u>：$R_A = R_B = 10,\ R_C = 4$

2. 四分位偏差 — データの第 3 四分位点と第 1 四分位点との差の半分：

$$Q = \frac{1}{2}\{(\text{第 3 四分位点}) - (\text{第 1 四分位点})\}$$

<u>データ１</u>：$Q_A = \frac{1}{2}(7-3) = 2,\ Q_B = \frac{1}{2}(8-2) = 3,\ Q_C = \frac{1}{2}(6-4) = 1$

3. 平均偏差 — 各観測値が (算術) 平均からどれだけ離れているかについての平均を求めたもの：

$$d = \frac{1}{n}\{|x_1 - \bar{x}_n| + |x_2 - \bar{x}_n| + \cdots + |x_n - \bar{x}_n|\}$$

データ１：

$$\begin{aligned} d_A &= \frac{1}{10}\{|0-5| + |3-5| + |3-5| + |5-5| + |5-5| + |5-5| + |5-5| \\ &\quad +|7-5| + |7-5| + |10-5|\} \\ &= 1.8 \\ d_B &= 2.8, \qquad d_C = 0.8 \end{aligned}$$

4. 分散と標準偏差 — 観測値と平均との偏差の 2 乗の平均をデータの分散といい，その (正) の平方根を標準偏差という：

$$\begin{aligned} S^2 &= \frac{1}{n}\{(x_1 - \bar{x}_n)^2 + (x_2 - \bar{x}_n)^2 + \cdots + (x_n - \bar{x}_n)^2\} \\ S &= \sqrt{S^2} \end{aligned}$$

<u>データ１</u>：

$$\begin{aligned} S_A^2 &= \frac{1}{10}\{(0-5)^2 + (3-5)^2 + (3-5)^2 + (5-5)^2 + (5-5)^2 + (5-5)^2 + (5-5)^2 \\ &\quad +(7-5)^2 + (7-5)^2 + (10-5)^2\} \\ &= 6.6 \\ S_B^2 &= 10.8, \qquad S_C^2 = 1.2 \end{aligned}$$

したがって,

$$S_A = 2.569, \qquad S_B = 3.286, \qquad S_C = 1.095$$

分散は

$$S^2 = \frac{1}{n}\{x_1^2 + x_2^2 + \cdots + x_n^2\} - (\bar{x}_n)^2$$

と表現できる.

x の値	0	1	2	3	5	5	7	8	9	10	50
x^2 の値	0	1	4	9	25	25	49	64	64	100	358

よって

$$S_B^2 = \frac{358}{10} - 5^2 = 10.8$$

となる.

5. 変動係数 — 分布の中心が著しく異なる場合 (スケーリングが違う) には,分散をもって分布の散らばり具合の比較をすることができないので,分散を調整したものを用いる.変動係数はその例である:

$$(\text{変動係数}) = \frac{(\text{標準偏差})}{(\text{平均})}$$

標準化得点

データを

$$z_i = ax_i + b, \qquad i = 1, 2, \ldots, n$$

のように1次変換をした場合,その平均,分散,標準偏差は

$$\bar{z}_n = a\bar{x}_n + b, \qquad S_x^2 = a^2 S_z^2, \qquad S_x = |a| S_x$$

のように変わる.特に,$a = S_x^{-1}$, $b = -\bar{x}_n / S_x$ の場合

$$z_i = \frac{x_i - \bar{x}_n}{S_x}$$

を標準化という.すると

$$\bar{z}_n = 0, \qquad S_z = 1$$

となる.さらに,

$$T_i = 10z_i + 50$$

としたものを偏差値得点という.この場合には,平均と分散は

$$\bar{T}_n = 50, \qquad S_T = 10$$

となる.

データ２：仮の平均を用いるといい．仮の平均を 97.0 とする．データは

$$1.0,\ 1.4,\ 0.0,\ 0.9,\ 0.7,\ 2.6,\ 1.8,\ 1.6,\ -0.4,\ -0.5$$

となる．これより，平均は

$$\bar{x}_n = \frac{1}{10}(1.0+1.4+0.0+0.9+0.7+2.6+1.8+1.6-0.4-0.5)+97.0 = 0.91+97.0 = 97.91$$

また，分散の値は仮の平均に影響をうけないので，

$$\begin{aligned} S^2 &= \frac{1}{10}\{(1.0)^2+(1.4)^2+(0.0)^2+(0.9)^2+(0.7)^2+(2.6)^2+(1.8)^2+(1.6)^2+(-0.4)^2 \\ &\quad +(-0.5)^2\} - (0.91)^2 \\ &= 1.723 - 0.8281 = 0.8949 \end{aligned}$$

データ１

A	0	3	3	5	5	5	5	7	7	10
B	0	1	2	3	5	5	7	8	9	10
C	3	4	4	5	5	5	5	6	6	7

データ２

98.0	98.4	97.0	97.9	97.7
99.6	98.8	98.6	96.6	96.5

問題 2

つぎのデータの (a) 平均 (b) メデアン (c) 範囲 (d) 平均偏差 (e) 分散を求めよ．

5.01	14.67	8.60	4.42	4.95	7.24
7.51	12.30	14.59	7.98	11.53	4.10

ヒント

x	x^2	$x-10$	$(x-10)^2$	$\lvert x_i - \bar{x}_n \rvert$
5.01	25.1001	-4.99	24.9001	3.565
14.67	215.2089	4.67	21.8089	6.095
8.60	73.9600	-1.40	1.9600	0.025
4.42	19.5364	-5.58	31.1364	4.155
4.95	24.5025	-5.05	25.5025	3.625
7.24	52.4176	-2.76	7.6176	1.335
7.51	56.4001	-2.49	6.2001	1.065
12.30	151.2900	2.30	5.2900	3.725
14.59	212.8681	4.59	21.0681	6.015
7.98	63.6804	-2.02	4.0804	0.595
11.53	132.9409	1.53	2.3409	2.955
4.10	16.8100	-5.90	34.8100	4.475
102.9	1044.715	-17.1	186.715	37.63