データの分析【変量の変換】　p.180～182

1 学習内容の説明 ⇒ 2 問題演習 ⇒ 3 振り返り（確認テスト・相互採点・リフレクションの記入）

【内容目標】変量が変化したときに平均値や分散、標準偏差に起きる変化を理解しよう

□変量の変換

データの各値に一斉に同じ数を加えたり，一斉に同じ数を掛けたとき，平均値，分散，標準偏差がどのように変化するかを考えてみよう。

変量 x についてのデータが，n 個の値 x_1，x_2，……，x_n であるとし，x のデータの平均値を $\overline{x}$，分散を $s_x{}^2$，標準偏差を s_x とする。

a，b を定数として，**x を a 倍され b だけ増加させてつくる**式 $z=ax+b$ で新たな変量 z を作る。このとき，z のデータは次の n 個の値である。

$$z_1=ax_1+b,\ z_2=ax_2+b,\ \cdots\cdots,\ z_n=ax_n+b$$

変量 z のデータの平均値 $\overline{z}$ は

$$\begin{aligned}\overline{z}&=\frac{1}{n}(z_1+z_2+\cdots\cdots+z_n)\\&=\frac{1}{n}\{(ax_1+b)+(ax_2+b)+\cdots\cdots+(ax_n+b)\}\\&=\frac{1}{n}\{a(x_1+x_2+\cdots\cdots+x_n)+nb\}\\&=a\cdot\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)+b\end{aligned}$$

散布図を考えると
イメージしやすいかも…

よって　$\overline{z}=a\overline{x}+b$ となり、**z の平均値も a 倍され b だけ増加する。**

また，$z_k-\overline{z}=ax_k+b-(a\overline{x}+b)=a(x_k-\overline{x})$ であることから，変量 z のデータの分散 $s_z{}^2$ と標準偏差 s_z については

$$\begin{aligned}s_z{}^2&=\frac{1}{n}\{(z_1-\overline{z})^2+(z_2-\overline{z})^2+\cdots\cdots+(z_n-\overline{z})^2\}\\&=\frac{1}{n}\{a^2(x_1-\overline{x})^2+a^2(x_2-\overline{x})^2+\cdots\cdots+a^2(x_n-\overline{x})^2\}\\&=a^2\cdot\frac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}\end{aligned}$$

$$\therefore\quad s_z{}^2=a^2s_x{}^2$$

$$s_z=\sqrt{a^2\cdot s_x{}^2}=|a|\cdot|s_x|=|a|\cdot s_x$$

したがって，**z の分散は a^2 倍**、**z の標準偏差は $|a|$ 倍**される

すべての変数を実数倍すると散らばりに影響が出るが、すべての変数に足したり引いたりしても散らばりに影響は出ない

一般に，次のことが成り立つ。

変量の変換

a，b は定数とする。変量 x のデータから $z=ax+b$ によって新しい変量 z のデータが得られるとき，x，z のデータの平均値を $\overline{x}$，$\overline{z}$，分散を $s_x{}^2$，$s_z{}^2$，標準偏差を s_x，s_z とすると

平均値　$\overline{z}=a\overline{x}+b$，　**分散**　$s_z{}^2=a^2s_x{}^2$，　**標準偏差**　$s_z=|a|s_x$

分散は偏差を２乗しているのだから係数の２乗が現れる
（ $+b$ は散らばりに影響しないので消えてしまう）

データの分析【変量の変換】　p.180～182

さらに 変量 y についてのデータが，n 個の値 y_1，y_2，……，y_n であるとし，y のデータの平均値を $\overline{y}$，分散を $s_y{}^2$，標準偏差を s_y， $X=ax+b$ ， $Y=cy+d$ したときの共分散を s_{XY} とすると

$$s_{XY}=\frac{1}{n}\{(X_1-\overline{X})(Y_1-\overline{Y})+(X_2-\overline{X})(Y_2-\overline{Y})+\cdots\cdots+(X_n-\overline{X})(Y_n-\overline{Y})\}$$

$$=\frac{1}{n}\Big[\{(ax_1+b)-(a\overline{x}+b)\}\{(cy_1+d)-(c\overline{y}+d)\}$$

$$+\{(ax_2+b)-(a\overline{x}+b)\}\{(cy_2+d)-(c\overline{y}+d)\}+\cdots\cdots+\{(ax_n+b)-(a\overline{x}+b)\}\{(cy_n+d)-(c\overline{y}+d)\}\Big]$$

$$=\frac{1}{n}\{(ax_1-a\overline{x})(cy_1-c\overline{y})+(ax_2-a\overline{x})(cy_2-c\overline{y})+\cdots\cdots+(ax_n-a\overline{x})(cy_n-c\overline{y})\}$$

$$=\frac{1}{n}\{ac(x_1-\overline{x})(y_1-\overline{y})+ac(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+ac(x_n-\overline{x})(y_n-\overline{y})\}$$

$$=ac\cdot\frac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$$

$$=ac\cdot s_{xy}$$

また　$r'=\dfrac{s_{XY}}{s_X\cdot s_Y}=\dfrac{ac\cdot s_{xy}}{|a|s_x\cdot|c|s_y}=\dfrac{ac}{|ac|}\cdot\dfrac{s_{xy}}{s_x\cdot s_y}=\dfrac{ac}{|ac|}\cdot r$　なので

変量の変換

共分散　$s_{XY}=ac\cdot s_{xy}$　共分散は偏差の積だから係数の掛け算が現れる

相関係数　$r'=\dfrac{ac}{|ac|}\cdot r$　相関係数は正負のみ散らばり具合は変わらない

例１）　変量 x のデータの平均値 $\overline{x}$ が 37，分散 $s_x{}^2$ が 25 であるとする。

このとき，$y=2x+10$ によって得られる新しい変量 y のデータについて

$\overline{y}=2\times37+10=84$,

$s_y{}^2=2^2\times25=100$

$s_y=|2|s_x=2\cdot5=10$

参考　s_y については，

$\sqrt{分散}$ で求めてもよいので

$s_y=\sqrt{s_y{}^2}=\sqrt{100}=10$

方針			
	平均値	$\overline{X}=a\cdot\overline{x}+b$	**式そのまま**
	分散	$s_X{}^2=a^2\cdot s_x{}^2$	**係数の２乗倍**
	標準偏差	$s_X=\|a\|\cdot s_x$	**係数の絶対値倍**
	共分散	$s_{XY}=ac\cdot s_{xy}$	**係数の積倍**
	相関係数	$r'=\dfrac{ac}{\|ac\|}r$	**正負のみ**

変量の変換によって，平均値や分散を求める計算が簡単になることがある。

例２）　5 人の身長 x のデータ　176，170，167，179，168　の平均値 $\overline{x}$ と分散 $s_x{}^2$ を求めてみよう。x の単位は cm である。$x_0=170$ として，新しい変量 u を $u=x-x_0$ で定める。変量 u のデータ，変量 u^2 のデータの値は，それぞれ次の表のようになる。

u	6	0	-3	9	-2	計 10
u^2	36	0	9	81	4	計 130

$\overline{u}=\dfrac{1}{5}\times10=2$，$\overline{u^2}=\dfrac{1}{5}\times130=26$ である。

よって，u のデータの分散 $s_u{}^2$ は　　$s_u{}^2=\overline{u^2}-(\overline{u})^2=26-2^2=22$

$x=x_0+u$ より，$\overline{x}=x_0+\overline{u}$，$s_x{}^2=s_u{}^2$ であるから　$\overline{x}=x_0+\overline{u}=170+2=172$ (cm)

$s_x{}^2=s_u{}^2=22$　　終

例 2 では，$x_0=170$ として $x-x_0$ のデータを考えることにより，変量 x のデータの平均値や分散を求めている。この x_0 を **仮平均** という。

一般には c を正の定数として $u=\dfrac{x-x_0}{c}$ と変換した変量を考えることが多い。ここで $x_0=\overline{x}$ ，$c=s_x$ とすると，変量 $u=\dfrac{x-\overline{x}}{s_x}$ のデータの平均値は $\overline{u}=0$ ，標準偏差は 0 となる。このときの u を x の**標準化**（**標準測度**）という。これを利用したのが偏差値。