

■ 数Iの「データの分析」。言い方が悪いが、センター試験専用の内容のような印象で、国公立大でもこれまで一橋大などで出題があるものの、センター試験で出題されている内容と、大きく傾向が変わるものではなかったように思う。

しかし、ここに来て「本気の出題」が、次々登場している。

■ まず、福井大医学部・前期である（「医学部だからね」と言い放っておいて良いものか…）。

変数  $x$  のデータの値を  $x_1, \dots, x_n$ 、変数  $y$  のデータの値を  $y_1, \dots, y_n$  とする。変数  $x$  の標準偏差を  $s_x$ 、変数  $y$  の標準偏差を  $s_y$  とする。また、変数  $x$  と変数  $y$  の相関係数を  $r$  とする。このとき、以下の問いに答えよ。

(1) 変数  $x$  の最大値を  $\max(x)$ 、最小値を  $\min(x)$  とする。このとき、

$$s_x \leq \max(x) - \min(x)$$

が成り立つことを示せ。さらに、等号成立の条件を調べよ。

(2) 変数  $z$  のデータの値を  $z_1 = x_1 - y_1, \dots, z_n = x_n - y_n$  とする。このとき、

$$r = \frac{s_x^2 + s_y^2 - s_z^2}{2s_x s_y}$$

が成り立つことを示せ。ただし、 $s_z$  は変数  $z$  の標準偏差とする。

(3) 次の表は、ある運動部に所属する10名の身長（変数  $x$ 、単位  $\text{cm}$ ）と体重（変数  $y$ 、単位  $\text{kg}$ ）のデータ、および変数  $x$ 、変数  $y$ 、変数  $x - y$  の平均、分散、標準偏差を計算した結果である。ただし、 $y_1 < y_2$  とする。

No.	1	2	3	4	5	6	7	8	9	10	平均	分散	標準偏差
身長 $x$	157	163	178	180	164	161	179	185	165	168	170	83.4	9.13
体重 $y$	$y_1$	$y_2$	63	77	61	63	70	79	62	65	65	64.8	8.05
$x - y$	$157 - y_1$	$163 - y_2$	115	103	103	98	109	106	103	103	105	19.0	4.36

①  $y_1, y_2$  の値をそれぞれ求めよ。

② 変数  $x$  と変数  $y$  の相関係数  $r$  を求めて、このデータの傾向について説明せよ。なお、 $r$  の値は小数第3位を四捨五入して小数第2位まで求めよ。また必要ならば、 $9.13 \times 8.05 \approx 73.5$  を用いてもよい。

出題意図や解答例を公表しているが、それぞれ「分散、標準偏差、相関係数の意味とデータ分析力を問う」、「(1) 略 (2) 略 (3) ①  $y_1 = 50$ ;  $y_2 = 60$  ② 略」と、おざりな印象だ。

実際解こうとすると、(1)から手が止まってしまうのは私だけか、まだ、(2)の設定の方が鉛筆がサクサクと動く。

■ 空欄補充だが、同志社大・文系で回帰方程式の係数決定にかかわる内容の出題がある。2次関数の最小問題とは言え、本格的だ。

$n$  個の2変数データ  $(x_i, y_i)$  ( $i = 1, 2, \dots, n$ ) がある。 $n \geq 2$  とし、 $x_i$  ( $i = 1, 2, \dots, n$ ) は互いに異なるとする、 $x_i$  と  $y_i$  に対して関数  $y = ax + b$  より得られる値  $ax_i + b$  と  $y_i$  との差の2乗和を  $n$  で割った

$$R = \frac{1}{n} \sum_{i=1}^n \{y_i - (ax_i + b)\}^2$$

を最小にする  $a, b$  を求めることにする。 $R$  を  $a$  と  $b$  について展開した展開式の各項の係数を、 $x$  の平均値  $\bar{x}$  と分散  $s_x^2$ 、 $y$  の平均値  $\bar{y}$  と分散  $s_y^2$  および  $x, y$  の共分散  $s_{xy}$  を用いて表すと、

$$R = b^2 + 2\bar{x}ab + (\text{ア})a^2 - 2\bar{y}b - 2(\text{イ})a + \text{ウ}$$

となる。この式の右辺を、まず  $b$  について平方完成し、次に  $a$  について平方完成することにより、 $R$  を最小にする  $a, b$  は  $\bar{x}, \bar{y}, s_x^2, s_{xy}$  を用いて  $a = \text{エ}$ 、 $b = \text{オ}$  と求めることができる。

なお、回帰方程式については、今年(2019)のセンター試験IAの追試験にも出題がある。要注意である。

■ また、神戸大の理系後期では、ベクトルと絡めた出題がある。  
[http://www.office.kobe-u.ac.jp/stdnt-examinavi/admission/admission01/examin/exam/kouki\\_sugaku\\_riki\\_2019.pdf](http://www.office.kobe-u.ac.jp/stdnt-examinavi/admission/admission01/examin/exam/kouki_sugaku_riki_2019.pdf) の最終問題、5番である。

信州大では、数列を絡めた、等差数列をなす  $n$  個のデータの平均、分散を問う出題がある（こういった内容の出題は、私には想定内だが）。

■ このように見てみると、「大学(出題者)は素材を探している」。

素材が、データの分析であれ何であれ、問題になりそうなら、これまでの傾向などお構いなく、様々なものと組合せて問題に仕立てている。

その意味で、「データの分析」を侮ってはいけない。