

雑感 「データの分析」を出題した一橋大

■ 数カリ初年度の今年、一橋大が「データの分析」の問題を出題したことは、もう旧聞に属することかも知れないが…

数列の和に関する [I] との選択問題である。

[II] a, b, c は異なる3つの正の整数とする。次のデータは2つの科目 X と Y の試験を受けた10人の得点をまとめたものである。

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
科目 X の得点	a	c	a	b	b	a	c	c	b	c
科目 Y の得点	a	b	b	b	a	a	b	a	b	a

科目 X の得点の平均値と科目 Y の得点の平均値とは等しいとする。

- (1) 科目 X の得点の分散を s_x^2 、科目 Y の得点の分散を s_y^2 とする。
 $\frac{s_x^2}{s_y^2}$ を求めよ。
- (2) 科目 X の得点と科目 Y の得点の相関係数を、四捨五入して小数第1位まで求めよ。
- (3) 科目 X の得点の中央値が65、科目 Y の得点の標準偏差が11であるとき、 a, b, c の組を求めよ。

問題は難しくはなく、センター試験用にこの範囲を学習してあれば、難なく解けたのではなかろうか。

その意味では、新課程履修者にやや有利な出題だったと言っても良いのかも知れない。

■ 興味・関心は、果たして一橋大が今後もこの範囲の出題を積極的に行う気があるのかなのかだ。一橋大のレベルにふさわしい問題が作成され、今後も出題されていくことを期待したいところだ。

■ このジャンルの出題に関しては、当「雑感」でも93に「データの分析のオリジナル問題」を掲載したが、このような内容の出題も視野に入りたいところだ。

■ 他に想定するのは、数列の和との関連である。

基本的には次のような各データについて、平均、分散を求めさせるのは、定義を知っていれば単に和の計算の問題である。

- (1) データ： $\{1, 2, 3, \dots, n\}$
- (2) データ： $\{1, r, r^2, r^3, \dots, r^{n-1}\}$

もう少し難易度を上げるならば、次のような設定である。

- (3) $k=1, 2, 3, \dots, n$ に対して k と書かれたカードが k 枚ある。カードの枚数の合計を N とするとき、この N 枚のカードに書かれた N 個の値をデータ X とする。

データ $X = \{1, 2, 2, 3, 3, 3, \dots, n, n\}$ について、平均、分散を求めよ。

- (4) (3)で r^{k-1} と書かれたカードが k 枚とする。
- (5) (3)で k と書かれたカードが $n-k+1$ 枚とする。

(3)で言えば、

$$E(X) = \frac{2}{n(n+1)} \sum_{k=1}^{\frac{1}{2}n(n+1)} k \cdot k = \frac{1}{12}(n^2 + n + 1)(n^2 + n + 2),$$

$$\begin{aligned} \sigma^2(X) &= \frac{2}{n(n+1)} \sum_{k=1}^{\frac{1}{2}n(n+1)} k \cdot k^2 - (E(X))^2 \\ &= \frac{1}{96}(n^2 + n + 2)(3n^4 + 6n^3 + n^2 - 2n - 8) \end{aligned}$$

といったところである。