

# 新課程入試における「データの分析」の出題について

## ～センター試験での出題を中心に～

「高校数学・新課程を考える会」事務局長 / 予備校講師

大淵 智勝

### 1. はじめに

平成 27 年度(2015 年度)から新しい学習指導要領(新課程)下での入試が始まった。ただ、2015 年度は「旧課程移行措置」の関係から「新課程と旧課程の共通分野」からの出題が目立った。そのため、新課程特有の分野である数学 I の「データの分析」からの出題はセンター試験や一橋大学などの一部に限られた。しかし、2016 年度入試以降は旧課程移行措置の制約がなくなるため、本格的な「データの分析」の出題が考えられる。

ここでは、2015 年度のセンター試験・数学 IA の問題を中心に、2015 年度の「データの分析」の出題をまとめるとともに、次年度以降の出題の傾向を考えていく。

### 2. センター試験・数学 I/IA での出題と正答率などについて

まず、2015 年度大学入試センター試験の本試験での「数学 I・数学 A」(以下、「数学 IA」とする)と「数学 I」におけるデータの分析の分野の出題の様子は以下ようになっていた。

(数学 IA・第 3 問) 配点 15 点  
 生徒 40 人のハンドボール投げの記録を元にした問題  
 [1](1) 1 回目の記録についてのヒストグラムから第 3 四分位数が含まれる階級を求める。(3 点)  
 (2) 与えられた 6 個の箱ひげ図から、ヒストグラムと矛盾する箱ひげ図を 4 個選ぶ。(4 点)  
 (3) 1 回目と 2 回目での「記録の変化の記述」と「2 回目のデータについての箱ひげ図」との 4 つの組のうち、矛盾するものを 2 組選ぶ。(6 点)  
 [2] 40 人のうち 39 人についての 1 回目と 2 回

目のハンドボール投げの記録についての散布図と各回の平均値、中央値、分散、標準偏差と 1 回目と 2 回目のデータの共分散が与えられているところから、1 回目と 2 回目の相関係数を求める。(2 点)

(数学 I・第 4 問) 配点 20 点

[1] 数学 IA と同じ(計 13 点)

[2](1) 数学 IA の [2] と同じ(2 点)

(2) 2 回目で欠席していた 1 人のデータを含めた 40 人での共分散と相関係数と、元の共分散と相関係数の大小を答える。(5 点)

「データネット 2015」による調査でのこれらの問題の正答率は下表の通りである。数学 IA の他の分野の得点率は、第 1 問(2 次関数)が 56%、第 2 問(集合と命題、図形と計量)が 65%、第 4 問(場合の数)が 65%、第 5 問(整数の性質)が 63%、第 6 問(図形の性質)が 63%(いずれも「データネット 2015」より)であることを考えると、第 3 問の得点率は 78%と高く、それぞれの問題の正答率もそこそこ高いといえる。

問題	問題内容	IA の正答率	I の正答率
[1](1)	ヒストグラムから第 3 四分位数を含む階級を答える	83.4%	60.0%
(2)	ヒストグラムと矛盾する箱ひげ図を 4 つ選ぶ	84.1%	60.0%
(3)	2 回目の分析結果と箱ひげ図の組で矛盾するものを 2 つ選ぶ	78.0%	53.3%
IA[2] I[2](1)	与えられた値から相関係数を答える	58.5%	46.7%
I[2](2)	1 人加えたときの相関係数、共分散の変化を答える		20.0%

表 1 数 IA 第 3 問、数 I 第 4 問の小問別正答率  
 (「データネット 2015」より)

[1]の問題は四分位数の計算方法を知っているだけでは解けず、四分位数や箱ひげ図の意味がわかり、また、2回目とのデータの比較について箱ひげ図からわかることの判断といったことができるかどうか問われていた。しかも(2)と(3)はそれぞれ選択肢を4個と2個答えるが、それぞれ完答をしないと点を与えられないものであった。

「データの分析」を勉強するにあたり、実際にヒストグラムや箱ひげ図を作成したり、その箱ひげ図から「正しいと言えること」と「正しいとは言えないこと」といったことを考えたことのある受験生にとっては容易に答えが出たと思われる。

[2]の相関係数を求める問題は正答率が低い、細かく見ると正解が⑦ 0.95 であるのに対して、間違いである⑥ 0.91 と⑧ 0.99 にマークをしたのがそれぞれ 9.7% と 6.4% であった。これらの誤答がすべて「定義がわかっていたのに計算をミスした」というものによるものとし、これを正答とすれば、正答率は 74.6% と [1] の問題の正答率に近づく。

またこの [2] の問題は、数学 I での (2) まですでに 1 つの問題としたかったのだと思うが、数学 IA で数学 I での (2) が問われなかったのは、全体で 60 分という時間制約による作問側の判断だとと思われる。

今年は 1 年目であるが故に、大学入試センター側としてもモニターチェックなどをかけていても、多少、難易度については手探りだった部分があるのだろうと思われる。今回の実施結果によって出てきた正答率などを鑑みたと、もう少し問題の難易度を上げていくことは可能だと判断する可能性はあり、数学 I の [2](2) や試作問題での「相関と因果」といったものを、数学 IA の中でも今後、出題していくことは十分に考えられる。

### 3. 一橋大学での出題

一橋大学の前期日程の 2 次試験で、選択問題の 1 つとして「データの分析」の問題が出題された。旧課程下では医学部の入試で、当時は数学 C にあった統計の分野が出題されていることが

あったが、新課程では「データの分析」で扱うレベルの統計の知識は最低限もっている生徒に入学してほしいという文系の学部も多いと思われ、これからは文系の数学で出題されるケースが増えてくるのではないと思われる。

一橋大学で出題された「データの分析」の問題は以下の通りである。

[ 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。

(2015 年度 一橋大学 前期)

文系の入試数学の中でも一橋大学の問題は難易度が高く、苦慮する受験生が多い。しかし、この問題は最後の (3) が整数問題っぽくはなるが、「データの分析」の学習がしっかりと出来ていれば計算も比較的容易で取っつきやすいと思われる。なお、この略解は以下の通りである。

- (1) 科目 X, Y の平均が等しいので、

$$3a + 3b + 4c = 5a + 5b$$

$$\therefore 2c = a + b$$

よって、平均は

$$\frac{5(a+b)}{10} = c$$

このことから、それぞれの分散は

$$s_X^2 = \frac{6}{10} \left( \frac{a-b}{2} \right)^2, \quad s_Y^2 = \frac{10}{10} \left( \frac{a-b}{2} \right)^2$$

よって、 $\frac{s_X^2}{s_Y^2} = \frac{3}{5}$

(2) 共分散  $s_{XY}$  は

$$s_{XY} = \frac{2}{10} \left( \frac{a-b}{2} \right)^2$$

よって、相関係数  $r$  は

$$r = \frac{2}{\sqrt{6 \times 10}} = \frac{1}{\sqrt{15}} = \frac{\sqrt{15}}{15}$$

$$= \frac{3.872 \dots}{15} = 0.258 \dots$$

より、 $r = 0.3$

(3)  $a < b$  としても  $a > b$  としても、 $c$  は  $a$  と  $b$  の間の値であるので、科目 X での中央値は  $c$  より

$$c = \frac{a+b}{2} = 65$$

また、教科 Y の標準偏差が 11 であるので、

$$\sqrt{\left( \frac{a-b}{2} \right)^2} = 11 \quad |a-b| = 22$$

これらより、

$$(a, b, c) = (76, 54, 65), (54, 76, 65)$$

#### 4. 「統計検定」の活用

新課程において「データの分析」が必修科目である数学 I の必須分野として入ってくるのと同じ頃に始まった「統計検定」であるが、この 3 級が数学 I の「データの分析」の分野を含んだ出題となっている。

まず、統計検定では、統計に関しての以下の 3 つの力を見る問題を出題している。

- (1) 基本的な用語や概念の定義を問う問題 (統計リテラシー)
- (2) 用語の基礎的な解釈や 2 つ以上の用語や概念の関連性を問う問題 (統計的推論)
- (3) 具体的な文脈に基づいて統計の活用を問う問題 (統計的思考)

統計検定の 3 級は高校数学の「データの分析」に対応しており、中学数学の「資料の活用」に相当する 4 級の分野に加えて、以下の分野が出題範囲となっている。

- ・標本調査 (母集団, 標本, 全数調査, 無作為抽出, 標本の大きさ, 乱数)
- ・データの散らばりの指標 (四分位数, 四分位範

囲 (四分位偏差), 標準偏差, 分散)

- ・データの散らばりのグラフ表現 (箱ひげ図)
- ・2 変数の相関 (相関, 散布図 (相関図), 相関係数)
- ・確率 (独立な試行, 条件付き確率)

この統計検定において、2015 年度センター試験での「データの分析」の分野の出題に似た問題が出題されているが、その例を見ていく。

センター試験・数学 IA の第 3 問の中で最も正答率の低かった [2] の相関係数を聞いている問題はつぎのようになっている。

[2] ある高校 2 年生 40 人のクラスで一人 2 回ずつハンドボール投げの飛距離のデータを取ることにした。次の図 2 は、1 回目のデータを横軸に、2 回目のデータを縦軸にとった散布図である。なお、一人の生徒が欠席したため、39 人のデータとなっている。

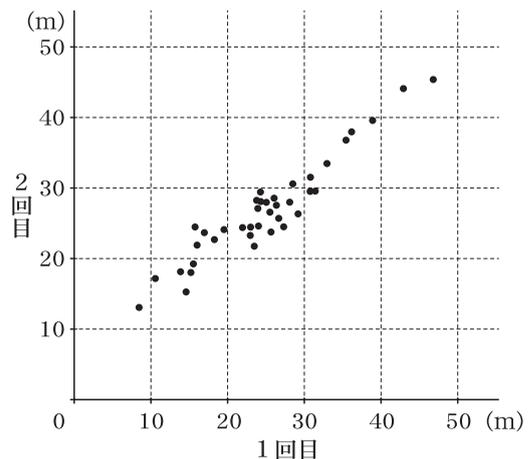


図 2

	平均値	中央値	分散	標準偏差
1 回目のデータ	24.70	24.30	67.40	8.21
2 回目のデータ	26.90	26.40	48.72	6.98

1 回目のデータと 2 回目のデータの共分散	54.30
------------------------	-------

(共分散とは 1 回目のデータの偏差と 2 回目のデータの偏差の積の平均である)

次の「ク」に当てはまるものを、下の①～⑨のうちから一つ選べ。

1 回目のデータと 2 回目のデータの相関係数に最も近い値は、「ク」である。

- ① 0.67   ② 0.71   ③ 0.75   ④ 0.79  
 ⑤ 0.83   ⑥ 0.87   ⑦ 0.91   ⑧ 0.95  
 ⑨ 0.99   ⑩ 1.03

( 答え ⑦ )

(2015 年度大学入試センター試験 本試験)

これに対し、2013 年 11 月実施の統計検定 3 級の問 17[1] は次のような問題である。

問 17 次の表は、ある車の速度と、ブレーキを踏んだときの停止距離の関係を表している。なお、1 マイルは約 1.6km, 1 フィートは約 0.3m である。(表略:  $x$  がスピード(マイル毎時),  $y$  が停止距離(フィート)である表)

[1] スピードと停止距離の間の相関係数として最も適切なものを、次の①～⑤のうちから一つ選べ。ただし、必要に応じて次の数値を用いてよい。

$x$ : 平均値 35.0 中央値 35.0 標準偏差 22.91

$y$ : 平均値 154.4 中央値 122.5

標準偏差 133.31  $x$  と  $y$  の共分散 2978.13

- ① 0.98   ② 0.65   ③ 0.05   ④ - 0.65  
 ⑤ - 0.97

( 答え ① )

(統計検定 3 級 2013 年 11 月試験)

いずれも、相関係数が共分散と標準偏差からどのように計算されるかを知っている上で、多少、桁数が多い計算であっても処理することが出来れば解ける問題となっている。したがって、統計検定のこの問題を解いたことがある生徒にとって、センター試験の第 3 問 [2] の問題は有利であったのではないだろうか。

他にも以下のように、センター試験の問題と類似した問題が統計検定にある。

・ IA[1](2) のような箱ひげ図とヒストグラムの対応の問題

→ 2014 年 11 月実施の問 13

・ I[2](2) のような新しいデータが加わったときの相関係数の変化などについての問題

→ 2012 年 11 月実施の問 18

さらに、新課程でのセンター試験に際し、大学入試センターが 2013 年 11 月に試作問題を公表したが、これにはデータの変数変換をした後の各値の変化や、「相関と因果は異なる」といったことを聞いている問題がある。具体的には相関係数の性質についての以下のような問題である。

(3) 相関係数の一般的な性質に関する次の [A] から [C] の説明について、 ス  ということがいえる。 ス  に当てはまるものを、次の①～④のうちから一つ選べ。

[A] 相関係数  $r$  は、常に  $-1 \leq r \leq 1$  であり、すべてのデータが 1 つの曲線上に存在するときには、いつでも  $r=1$  または  $r=-1$  である。

[B] もとのデータを定数倍しても、相関係数の値は変わらないが、もとのデータに定数を加えると相関係数の値は変わる。

[C] 2 つの変量間の相関係数の値が高い場合には、これらの 2 つの変量には因果関係があるといえる。

- ① [A] だけが正しい   ② [B] だけが正しい  
 ③ [C] だけが正しい   ④ [A] だけが間違っている  
 ⑤ ①～③のどれでもない

(大学入試センター試験 試作問題)

これに対して統計検定では「相関があるからといって因果があるとは言えない」ということについて、2013 年 11 月実施の問 16[2] に次のような問題がある。

問 16 次の図は、都道府県別の軽自動車保有車両数と軽自動車を除く登録自動車保有車両数の散布図に、縦軸に登録自動車保有車両数のヒストグラム、横軸に軽自動車保有車両数のヒストグラムをそれぞれ付け加えた図である。

( 図略。図には正の相関がみえる散布図がある )

[2] この結果の解釈として適切でないものを、次の①～⑤のうちから一つ選べ。

- ① 軽自動車保有車両数が多い都道府県は、登録自動車保有車両数も多い傾向がみられる。  
 ② 各都道府県の人口が大きく異なるため、今回

の散布図では人口を踏まえた変数の調整に注意を払うべきである。

③ 散布図の中央上部に、他の観測値とはずれた観測値が2つあるが、これらを除くと相関係数は除く前に比べて1に近づく。

④ 登録自動車保有車両数が多い原因は、軽自動車保有車両数が多いことである。

⑤ 観測値は右上がりの直線に近い形で分布しており、直線の傾向線を描き入れることにより、軽自動車保有車両数から登録自動車保有車両数を予測できる。

(答え ④)

(統計検定3級 2013年11月試験)

他にもデータを変数変換した前後での各値の変化については、2014年11月実施の間22[3]では相関係数の変化、2013年11月実施の間17[2]では共分散、相関係数、標準偏差の変化について聞いている問題がある。

センター試験の「過去問」としては2015年度のものしかないので、「データの分析」の入試のための勉強では、その不足を補うのに、2014年度以前の数学IIBで出題されたものを用いることが多いが、新課程の数学IAの「データの分析」に対応させた問題ということであれば、統計検定3級の問題を用いた演習をした方が有意義ではないだろうか考える。

## 5. 「データの分析」の授業での配慮

センター試験での出題の仕方を見てもわかるとおり、統計の「用語」についてのことだけではなく、それぞれの値の意味、グラフの見方やそこから考えられること、複数のグラフの関係を理解できることが求められている。したがって、「データの分析」の授業の際には、総務省統計局や気象庁から出ている実際のデータを用いたり、演習においては統計検定の問題などを積極的に活用することが重要であろうと思われる。

ただ、積極的に実際のデータを活用する授業を展開すると、「データをいじること」に終始し、「活

動あって学びなし」という事態に陥る恐れがある。したがって、データをいじるだけではなく、やはり「用語」の定義や意味も理解させていく必要がある。具体的には、代表値である「平均値、中央値、最頻値」のそれぞれの定義とそれによる特徴、中央値を中心としたデータの要約(五数要約)のための四分位数、平均値を中心としたデータの要約に必要な標準偏差、2変量データの相関を見るための共分散とその正規化である相関係数、という具合に、値の名称だけではなく「それがいったい何なのか」を明確にすることで、生徒のこれらの値に対する認識は高くなるのではないだろうか。

例えば、代表値については、平均値、中央値、最頻値といったものをばらばらに話すのではなく、これらがひとまとまりのデータに対する「代表値」であることを認識させ、同じデータでもどの代表値を使うかによって「主張」を変えることができるといったことを例に挙げ、それぞれの定義からその値の意味するところを明確に理解させていくといった授業の展開がポイントとなるのではないだろうか。

さらには、数学Iの[2](2)のように、1つ新しいデータが加わったり、あるいは、1つのデータが変更になったりといったときに、平均、標準偏差、相関係数といったものがどのように変わるのか、というあたりもこれからのセンター試験での出題が増えてくるのではないだろうか。これについては、各値の定義に基づくとどうなるかということと、一方で、各値の特徴からどのように変化すると考えられるかということの両方を丁寧に見ていく授業もまた、大きなポイントになるのではないだろうか。

## 6. 最後に

予備校の冬期講習でセンター試験対策の講座を担当した際、データの分析についてまとめ直した授業をしたところ、アンケートに「データの分析の話をしてくれて助かった」という旨のコメントが多々見られた。高校によっては授業時間の都合

からこの分野を積極的に授業で扱っていない学校もあり、そのためにこのようなコメントが書かれたようである。「入試で出ないから」扱わないというのも問題だが、「入試で出る」のに扱いを軽くしているのはもっと問題なように思う。「数学なんてやっていて意味があるの？」という生徒が多い中、「データの分析」は数学を実際の生活

で使っていくことを強調でき、「数学的活動」を通して「事象を数学的に考察」させるという学習指導要領の数学の「目標」を達成できる分野であると考えられる。そういった点でも、受験だから、というだけでなく、積極的にデータの分析の分野を授業で扱っていくべきではないかと考える。(統計検定<sup>®</sup>は統計検定センターの登録商標です。)

## 実教出版が提案する新たな参考書

### 例題から学ぶ数学 例題編

Co.4311	数学Ⅰ+A	例題編	A5判	224p.	定価(本体476円+税)
Co.4313	数学Ⅱ+B	例題編	A5判	312p.	定価(本体570円+税)
Co.310019	数学Ⅲ	例題編	A5判	228p.	定価(本体570円+税)

- ▼教科書の例から大学入試レベルまでの例題を1冊にまとめました。
- ▼各例題に添えた星印(★)の数で難易度を表しました。
- ▼典型的な解法を網羅しているため、参考書と同様に活用可能です。



### 例題から学ぶ数学 演習編

Co.4312	数学Ⅰ+A	演習編	A5判	128p.	定価(本体429円+税)
Co.310006		演習編解答	A5判	360p.	定価(本体286円+税)
Co.4314	数学Ⅱ+B	演習編	A5判	164p.	定価(本体480円+税)
Co.310018		演習編解答	A5判	524p.	定価(本体300円+税)
Co.310020	数学Ⅲ	演習編	A5判	112p.	定価(本体480円+税)
Co.310021		演習編解答	A5判	368p.	定価(本体300円+税)

- ▼各項目2ページ見開きで、「基本問題→標準問題→応用問題」の構成です。
- ▼教科書の練習問題から大学入試レベルまで、多数の問題を扱っています。
- ▼関連する例題編の問題番号が示されているので、随時例題編を参照して解き方が確認できます。



通巻第71号

2015年9月20日印刷

2015年9月25日発行

©編修・発行

実教出版株式会社

代表者 戸塚雄弐

定価(本体200円+税)

発行所 〒102-8377 東京都千代田区五番町5

TEL. 03-3238-7777

<http://www.jikkyo.co.jp/>