



「データの分析」分野の入試問題の分類と解法 の一考察 入試センターのサンプル問題解説 ～第3問データの分析～

♡ 7

 情報処理学会・学会誌「情報処理」
2021年10月20日 10:21



阿部百合（早稲田大学高等学院）

連載「教科「情報」の入学試験問題って？」第4弾です。前回までの入試問題解説に関してはこちら（<https://note.com/ipsj/n/n81737ef872ec>）をご覧ください。

大学入試センターによる共通テスト『情報』のサンプル問題[1]の大問3を見てみましょう。

（以下、サンプル問題中の記号や文字を**太字**で表現しています）

▼ 目次

問題設定

問1

問2

問3

問4

まとめ

問題設定

実際の状況や、生徒に身近なデータが使われています。今回の問題では実際のサッカーワールドカップのデータ [2] を使って決勝進出チームと予選敗退チームの違いを分析します。問に入る前に2つの分析資料（**表1**、**図1**）が提示されています。データの分析問題で重要、かつ見落としがちな点に生データの数値変換や補完があります。図表にまず目がいきがちですが、資料の前後にある説明文をしっかりと読

み「図表中の数値（図示されている場合は点やグラフ）の意味を理解する」ことが肝要です。数値の意味について書かれている部分を赤線で示しました。グラフの場合は軸やラベルに注目し、単位も必ず見ます。入試問題演習では、数値の意味の部分を自身でチェックするとよいでしょう。

決勝進出チームと予選敗退チームの違いを調べるために、決勝進出の有無は、決勝進出であれば1、予選敗退であれば0とした。また、チームごとに試合数が異なるので、各項目を1試合当たりの数値に変換した。

表1 ある年のサッカーのワールドカップのデータの一部（データシート）

	A	B	C	D	E	F	G	H	I	J	K
1	チーム ID	試合数	総得点	ショートパス 本数	ロングパス 本数	反則 回数	決勝進出 の有無	試合当たりの 得点	1試合当たりの ショートパス本数	1試合当たりの ロングパス本数	1試合当たりの 反則回数
2	T01	3	1	834	328	5	0	0.33	278.00	109.33	1.67
3	T02	5	11	1923	510	12	1	2.20	384.60	102.00	2.40
4	T03	3	1	650	269	11	0	0.33	216.67	89.67	3.67
5	T04	7	12	2257	711	11	1	1.71	322.43	101.57	1.57
6	T05	3	2	741	234	8	0	0.67	247.00	78.00	2.67
7	T06	5	5	1600	555	9	1	1.00	320.00	111.00	1.80

▲問の前の資料表1（赤線は筆者追加）

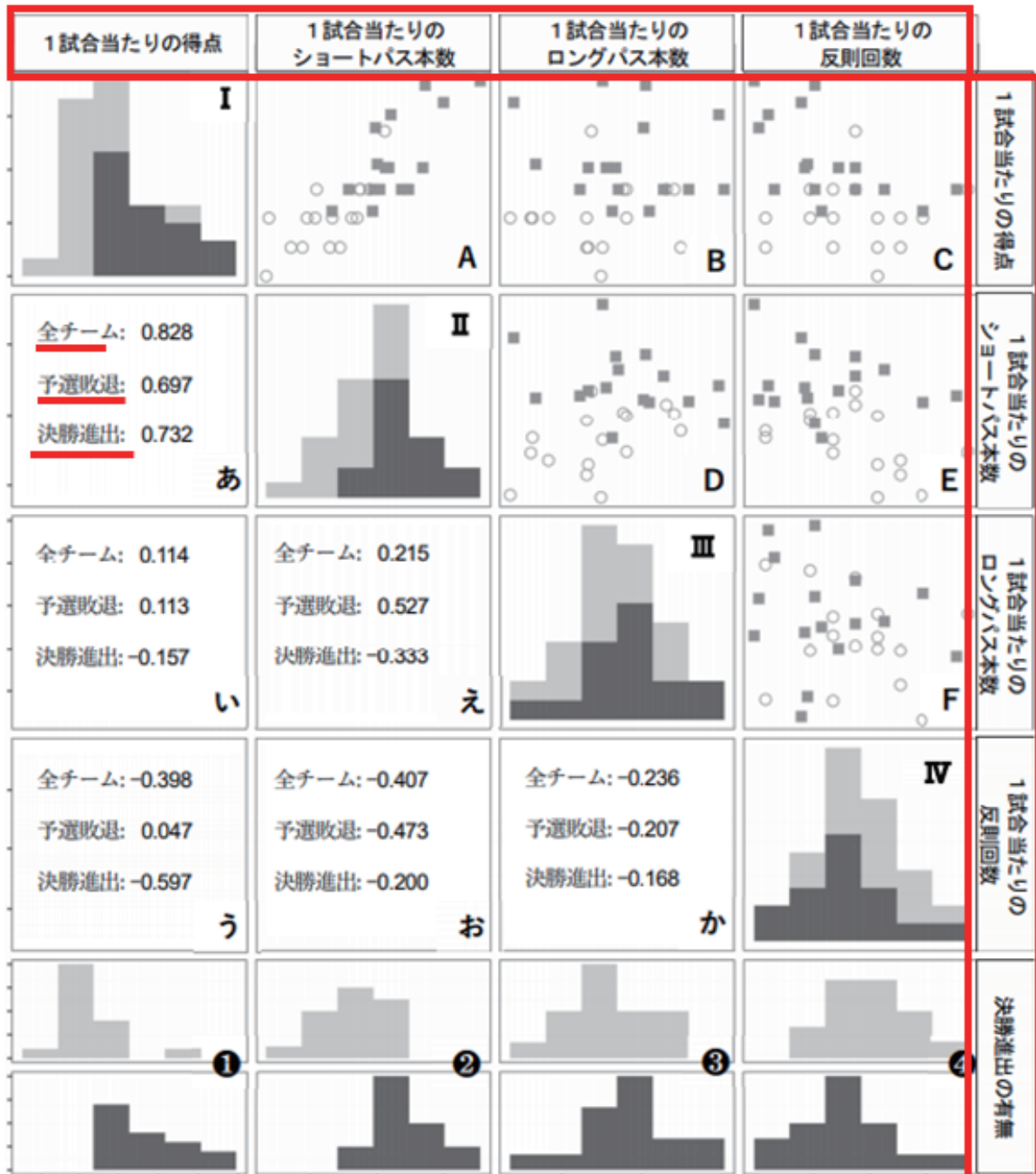


図1 各項目間の関係

▲問の前の資料図1 (赤線は筆者追加)

問1

資料図1を読めるか問う問題です。ここでは散布図行列がプログラムで作れるかは問題ではありません。重要なのは「図を読めること」です。2次元行列の形に面食らった方もいるでしょう。図の下に以下の説明がありますが、かえって読みづら

いかかもしれません。むしろ一度でも図1のような行列（総当たり戦の対戦表など）や散布図行列を扱った経験があれば、図1は読めます。この機会に演習しておくとうれしいです。

図1のⅠ～Ⅳは、それぞれの項目の全参加チームのヒストグラムを決勝進出チームと予選敗退チームとで色分けしたものであり、①～④は決勝進出チームと予選敗退チームに分けて作成したヒストグラムである。あ～かは、それぞれの二つの項目の全参加チームと決勝進出チーム、予選敗退チームのそれぞれに限定した相関係数である。またA～Fは、それぞれの二つの項目の散布図を決勝進出チームと予選敗退チームをマークで区別して描いている。例えば、図1のAは縦軸を「1試合当たりの得点」、横軸を「1試合当たりのショートパス本数」とした散布図であり、それに対応した相関係数はあで表されている。

▲問の前の資料図1の下にある説明文

a問題

a問題では文章中に問いが隠れています。赤い線の部分が問いです。

図1を見ると、予選敗退チームにおいてはほとんど相関がないが、決勝進出チームについて負の相関がある項目の組合せは、1試合当たりのアとイである。また、決勝進出チームと予選敗退チームとで、相関係数の符号が逆符号であり、その差が最も大きくなっている関係を表している散布図はウである。したがって、散布図の二つの記号のどちらが決勝進出チームを表しているかが分かった。

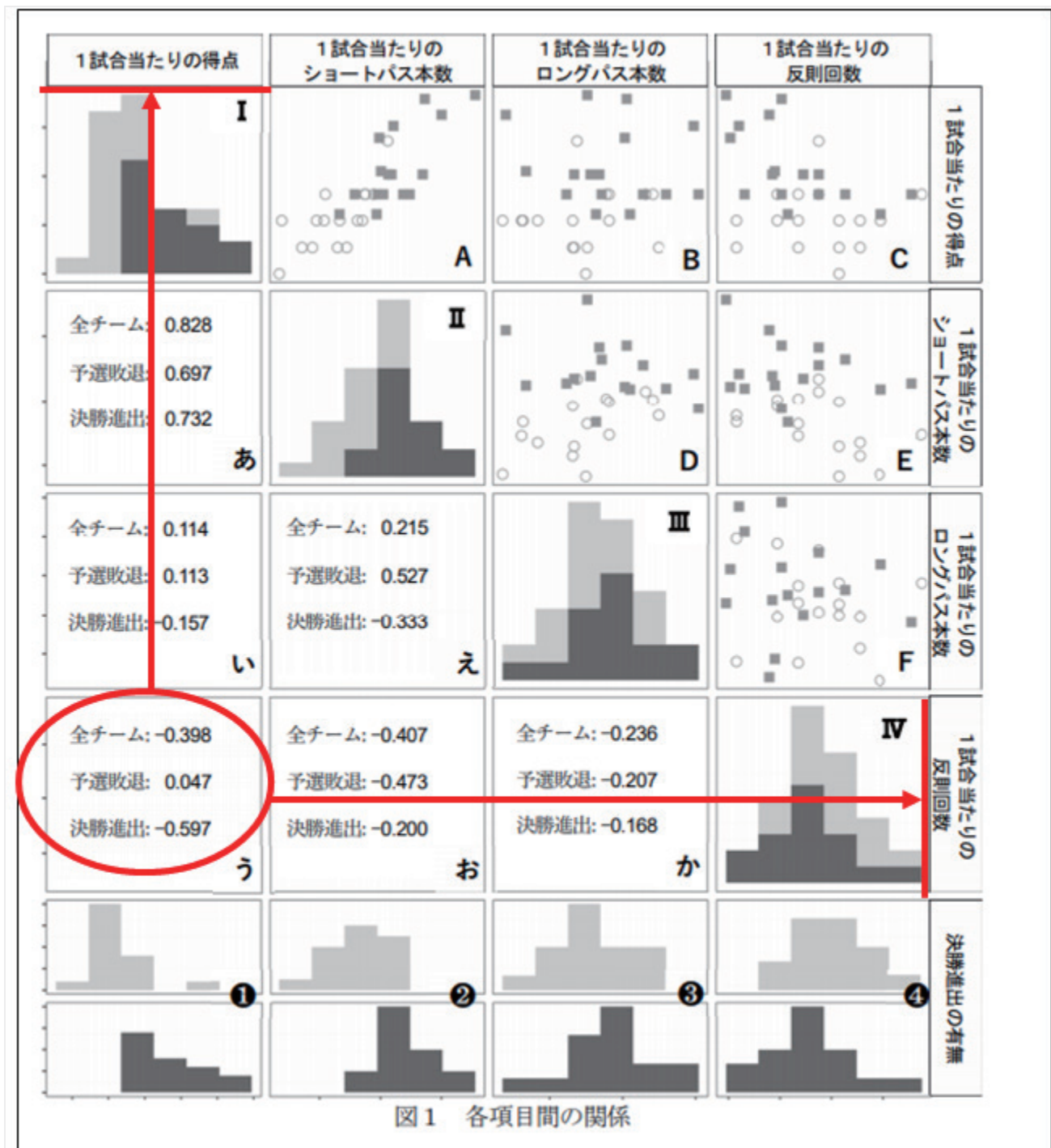
ア・イの解答群

- ① 得点 ② ショートパス本数 ③ ロングパス本数 ④ 反則回数

ウの解答群

- ① A ② B ③ C ④ D ⑤ E ⑥ F

アとイの組合せは、図1のあ～かの予選敗退と決勝進出の数値を比較すれば③得点と③反則回数であるとすぐに解けます。「相関」の資料がI（ヒストグラム）・A（散布図）・あ（相関係数）のどの形式か分かっているかを、問われています。（図-1.1）



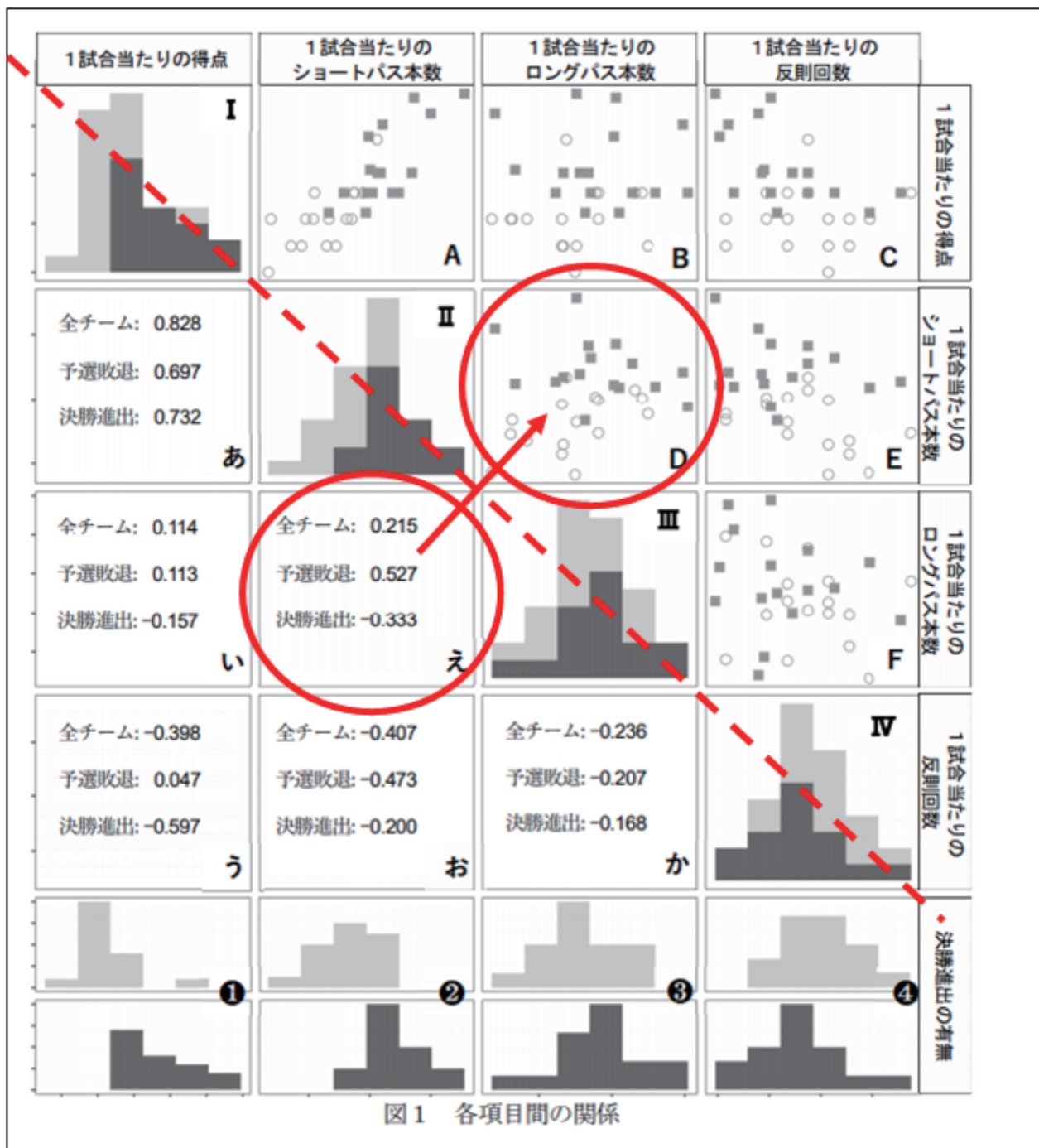
▲図-1.1 問の前の資料図1からの読み取り1（赤線部分）

アとイの答え(順序問わず)

①得点

③反則回数

ウは、問題文に釣られて散布図を見ても解けません。まずはあ～かの相関係数を見て、対応する散布図が答えとなります。総当たり戦の対戦表同様、左上から右下への対角線が対称軸となって、データの組合せが同じです。(図-1.2)



▲図-1.2 問の前の資料図1からの読み取り2（赤線部分）

ウの答え

③ D

b問題

b問題では、選択肢から誤っているものを選びます。

b 図1から読み取れることとして誤っているものを解答群から一つ選べ。

エ

エの解答群

- ① それぞれの散布図の中で、決勝進出チームは黒い四角形（■），予選敗退チームは白い円（○）で表されている。
- ② 全参加チームを対象としてみたとき、最も強い相関がある項目の組合せは1試合あたりの得点と1試合あたりのショートパス本数である。
- ③ 全参加チームについて正の相関がある項目の組合せの中には、決勝進出チーム，予選敗退チームのいずれも負の相関となっているものがある。
- ④ 1試合あたりのショートパス本数の分布を表すグラフ②で、下の段は決勝進出チームのヒストグラムである。

定石では資料図1をもとに各選択肢の正誤を判定しますが、選択肢すべてに先に目を通すと疑わしい選択肢②「全体の相関が正だが、個別の相関が2つとも負」があるため、まずは選択肢②を確認します。資料図1のあ～かの相関係数の符号だけ取り出してみます。

	あ	い	う	え	お	か	選択肢②
全チーム	$\begin{bmatrix} + \\ + \\ + \end{bmatrix}$	$\begin{bmatrix} + \\ + \\ - \end{bmatrix}$	$\begin{bmatrix} - \\ + \\ - \end{bmatrix}$	$\begin{bmatrix} + \\ + \\ - \end{bmatrix}$	$\begin{bmatrix} - \\ - \\ - \end{bmatrix}$	$\begin{bmatrix} - \\ - \\ - \end{bmatrix}$	$\begin{bmatrix} + \\ - \\ - \end{bmatrix}$
予選敗退							
決勝進出							

全チームが正で予選敗退と決勝進出両方が負のものではなく、他の選択肢を確認せずとも答えは②です。

エの答え

②

問2

鈴木さんは、この結果からショートパス 100 本につき、1 試合当たりの得点増加数を決勝進出チームと予選敗退チームで比べた場合、0. **オカ** 点の差があり、ショートパスの数に対する得点の増加量は決勝進出チームの方が大きいと考えた。

また、1 試合当たりのショートパスが 320 本るとき、回帰直線から予測できる得点の差は、決勝進出チームと予選敗退チームで、小数第 3 位を四捨五入して計算すると、0.0 **キ** 点の差があることが分かった。鈴木さんは、グラフからは傾きに大きな差が見られないこの二つの回帰直線について、実際に計算してみると差を見つけられることが実感できた。

さらに、ある決勝進出チームは、1 試合当たりのショートパス本数が 384.2 本で、1 試合当たりの得点が 2.20 点であったが、実際の 1 試合当たりの得点と回帰直線による予測値との差は、小数第 3 位を四捨五入した値で 0. **クケ** 点であった。

問2では回帰直線の理解、計算力を確認しています。資料**図2**にある回帰直線の式に値を代入し計算します。このあたりは情報Iの授業で扱われるので点が取れる問題となるはずです。問1のa問題同様、問2も問題文の中に問いが含まれる形式です。

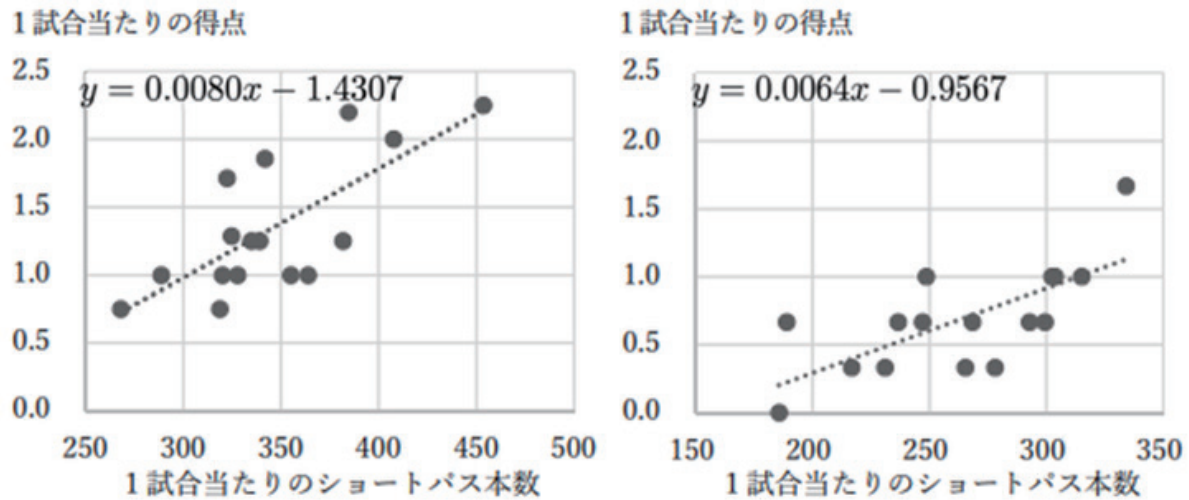


図2 決勝進出チーム(左)と予選敗退チーム(右)の
1試合あたりの得点とショートパス本数の回帰直線

▲サンプル問題の問2の資料図2

オカ：資料図2の決勝進出チーム（左）と予選敗退チーム（右）のショートパス100本につき，1試合あたりの得点増加数の差を聞いています。

単純に x に100を代入して計算しますが「増加数」を聞かれているので，直線の傾き（つまり x の前の数）を100倍して比較すれば十分です。

よって， $0.8 - 0.64 = 0.16$ ，答えは0.16です。

オカの答え

16

キ：問題文の誘導通り，1試合あたりのショートパスが320本のときを計算します。

$x = 320$ を代入して計算すると，決勝進出チーム1.1293，予選敗退チーム1.0913よって差は0.0380で小数第3位を四捨五入し答えは0.04です。

キの答え

4

クケ：決勝進出チームの実際の得点と回帰直線による予測得点の誤差の確認です。図中の式の y が予測得点であることを理解していれば、計算は単純です。図2の左の式 $y = 0.0080x - 1.4307$ に $x = 384.2$ を代入した値と、実際の得点2.20の差を求め、小数第3位を四捨五入すると答えは0.56点です。

クケの答え

56

問3

鈴木さんは、この分析シートから **コ** と **サ** について正しいことを確認した。

コ ・ **サ** の解答群

- ① 1試合当たりのロングパス本数のデータの散らばりを四分位範囲の視点で見ると、決勝進出チームよりも予選敗退チームの方が小さい。
- ② 1試合当たりのショートパス本数は、決勝進出チームと予選敗退チームともに中央値より平均値の方が小さい。
- ③ 1試合当たりの反則回数の標準偏差を比べると、決勝進出チームの方が予選敗退チームよりも散らばりが大きい。
- ④ 1試合当たりの反則回数の予選敗退チームの第1四分位数は、決勝進出チームの中央値より小さい。

表2 1試合当たりのデータに関する基本的な統計量（分析シート）

	A	B	C	D	E	F	G	H	I
1		決勝進出チーム				予選敗退チーム			
2	統計量	1試合当たりの得点	1試合当たりのショートパス本数	1試合当たりのロングパス本数	1試合当たりの反則回数	1試合当たりの得点	1試合当たりのショートパス本数	1試合当たりのロングパス本数	1試合当たりの反則回数
3	合計	21.56	5532.21	1564.19	41.30	11.00	4213.33	1474.33	48.00
4	最小値	0.75	268.00	74.40	1.50	0.00	185.67	73.67	1.67
5	第1四分位数	1.00	321.82	92.25	2.10	0.33	235.25	87.67	2.58
6	第2四分位数	1.25	336.88	96.02	2.40	0.67	266.83	91.67	3.00
7	第3四分位数	1.75	368.33	103.50	3.00	1.00	300.08	98.00	3.42
8	最大値	2.25	453.50	118.40	4.50	1.67	334.00	109.33	4.67
9	分散	0.23	1926.74	137.79	0.67	0.15	1824.08	106.61	0.61
10	標準偏差	0.48	43.89	11.74	0.82	0.38	42.71	10.33	0.78
11	平均値	1.35	345.76	97.76	2.58	0.69	263.33	92.15	3.00

▲サンプル問題の問3の表2

表2の考察問題です。2つ選ぶ必要があるため定石通り選択肢を1つずつ吟味します。中高の数学で学んだ統計を理解していれば難しくないでしょう。

問3の答え

②と③

問4

複数の資料、情報から必要な情報を見つけ、答えを導く問題です。クロス集計表が読めることが大前提です。クロス集計表は数学でも条件付き確率で同様の表を使っており、見たことがあるでしょう。

シ：

鈴木さんは、作成した図1と表2の両方から、**シ**ことに気づき、決勝進出の有無と1試合当たりの反則回数の関係に着目した。そこで、全参加チームにおける1試合当たりの反則回数の第1四分位数(Q1)未満のもの、第3四分位数(Q3)を超えるもの、Q1以上Q3以下の範囲のものの三つに分け、それと決勝進出の有無で、次の表3のクロス集計表に全参加チームを分類した。ただし、※の箇所は値を隠してある。

選択肢の吟味は先頭からする必要はありません。確認しやすい選択肢から見ます。選択肢③は、どの資料を見るか書いてあり、しかも読み取るものはヒストグラムと視覚的です。今回はちょうど見やすい選択肢が正解でした。

シの答え

③

ス：

表3 決勝進出の有無と1試合当たりの反則回数に基づくクロス集計表

	1試合当たりの反則回数			計
	Q1 未満	Q1 以上 Q3 以下	Q3 を超える	
決勝進出チーム	※	※	※	16
予選敗退チーム	2	※	ス	16
全参加チーム	8	※	7	32

▲サンプル問題の問4の表3

スの値は表3の中の数字から求められないので、続く問題文を見ます。

この表から、決勝進出チームと予選敗退チームの傾向が異なることに気づいた鈴木さんは、割合に着目してみようと考えた。決勝進出チームのうち1試合当たりの反則回数が全参加チームにおける第3四分位数を超えるチームの割合は約19%であった。また、1試合当たりの反則回数がその第1四分位数より小さいチームの中で決勝進出したチームの割合は **セソ** %であった。

「決勝進出チームのうち1試合当たりの反則回数が全参加チームにおける第3四分位数を超えるチームの割合は19%」とあります。決勝進出チームの内訳は表の1段目です。決勝進出チームの全数は1段目右端の16, 「決勝進出チームの第3四分位数を超えるチーム数」は表の1段目右から2列目スの上の※です。つまり（**ス**の上の※の数字） $\div 16 \times 100 = 19$ ということです。この方程式を解くと**ス**の上の※の数は3と分かります。よって表の**Q3**を超える列に注目して $7 - 3 = 4$ です。

スの答え

4

セソ：「第1四分位数より小さいチームの中で」と問題文にあるので左端の「**Q1未満**」の列を見ます。左端の列は上から※, **2**, **8**です。よって※は $8 - 2 = 6$ と分かるので、その割合は $6 \div 8 \times 100 = 75$, 75%です。

セソの答え

75

まとめ

ここまで「設定に沿って解くタイプの問題」を見てきましたが、慶應義塾大学や明治大学では「設定や分析の不適切な点を考えさせる問題」「調査・分析の修正を

記述させる問題」も出されています[3].

「設定に沿って解くタイプの問題」は、分析データや問題設定にやや不自然さを感じるがあっても割り切って筋書きに乗って解きましょう。ただし、ただ誘導に乗っていればいいのではありません。情報分野を目指す場合や、より難易度の高い大学を目指す場合、「設定や分析の不適切な点を考えさせる問題」、「調査・分析の修正を記述させる問題」に対応できる必要があります。今回の問題でも、予選敗退チームは当然ながら決勝進出チームより試合数が少ない（つまりチームによって試合数が大きく異なる）可能性がある、データサイズが十分でない可能性がある、といったことがあります。常に問題の設定や、「扱われているデータに不適切な点はないか」「どのようにしたら適切か」などの視点から注意して眺める癖も付けたいものです。問題を見るポイントは、

①大問全体の流れ

②データ、図表の軸、単位、ラベル、前後の問題文といった細かいところ
を見ることです。

情報の入試問題演習では、

①長い問題文や複数の資料を読むことに慣れること

②試験時間に焦らずじっくりと問題と向き合うこと
が必要です。

今回扱った問題は試験時間を考慮していない[4]とのことですが、もともとセンター試験『数学I』『情報関係基礎』でのデータの活用分野は問題文や図表が多いです。文章中に問いを含む形式もあり、問題文を読んで聞かれていることを理解する速さも必要であり、ある程度解けるようになってきたら

③制限時間を決めて解く練習

をするとよいでしょう。

また、問題を見て知らない形式の図表で見方が分からない場合や分からない単語が出てきたときは、いったん問題から離れて分からない部分を調べ理解することが基本となります。

参考文献

- 1) 大学入試センター，共通テスト『情報』サンプル問題，https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7ikou.html
- 2) 科学の道工具箱，FIFAワールドカップ2006データの集計，<https://rika-net.com/contents/cp0530/contents/04-14-01.html>
- 3) 情報入試研究会 資料，2018年度明治大学情報コミュニケーション学部問IV他，http://jnsg.jp/?page_id=108
- 4) 全国高等学校情報教育研究会第14回大会(大阪オンライン)基調講演「大学入学共通テスト 新科目「情報」～サンプル問題等とそのねらい～」(2021-8-11)，<https://www.zenkojoken.jp/14osaka/2021053913/>，FIT2021公開シンポジウム大学共通テスト「情報」が目指すもの(2021-8-26)，https://www.ipsj.or.jp/event/fit/fit2021/FIT2021_program/data/html/event/pdf/eventB2_347.pdf

(2021年10月11日受付)

(2021年10月20日note公開)

■阿部百合（正会員）

都内私立高校の情報の授業を担当している。本会CE研委員。

IPSJMooc (<https://sites.google.com/view/ipsjmooc/>) 第4章作成に携わった。

情報処理学会ジュニア会員へのお誘い

小中高校生，高専生本科～専攻科1年，大学学部1～3年生の皆さんは，情報処理学

会に無料で入会できます。会員になると有料記事の閲覧，情報処理を学べるさまざまなイベントにお得に参加できる等のメリットがあります。ぜひ，入会をご検討ください。入会は[こちら](#)から！