

## データ分析を役立つものにするために

札幌創成高校 数学科 教諭  
外山 尚生

### 1. はじめに

データ分析は社会や日常生活に近い内容であるために、教え方によっては身近に役立つ面白い分野になり得ると思います。しかしながら現行課程になって今まで2回ほどデータ分析を教えてきましたが「教科書通りではどうもつまらない」というのが正直な感想です。そこで今回、データ分析はなぜつまらないかを分析し、将来に役立つデータ分析の授業案を提案してみたいと思います。

### 2. なぜデータ分析はつまらないのか？

高等学校のデータ分析がつまらない理由はこの分野が「数学的でなく」、「現実的でない」ところに問題点があると思います。ここで2つの問題点をあげてみます。

#### [問題点1] 数学的理論づけが薄く、単なる計算問題になりがちである。

数学Iの範囲では $\Sigma$ を使って計算をすることができず、また正規分布などを用いて細かい数理統計学の議論ができないために、数学的議論をごまかしながら授業を進めていかなければなりません。そのため「こうすれば計算できる」という計算方法を教え、計算するだけの授業になりがちです。考えることをせず、機械的になってしまうので、どうも面白くないのです。(一方で難しいことを考えずに単なる四則演算で答えが出る分野なので喜ぶ生徒もいるのですが…)

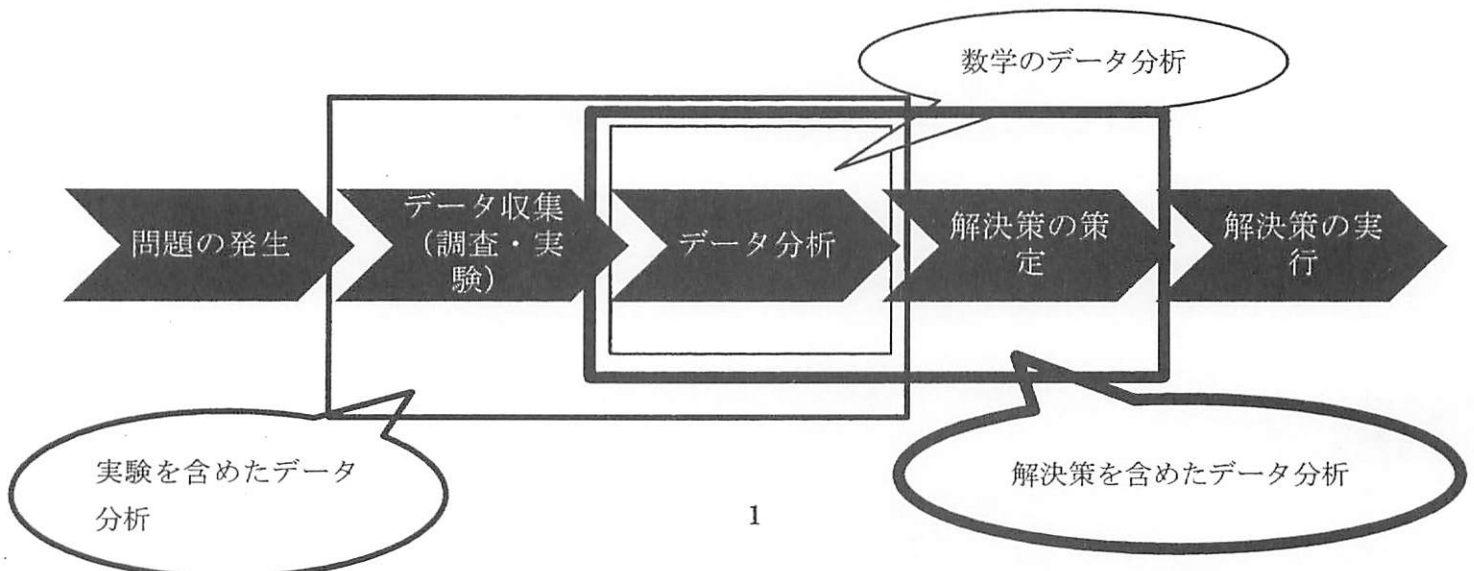
#### [問題点2] データの数が少なく、算出された統計量に統計学的意味がない。

問いによってはたった5つのデータから分散や標準偏差、相関係数などを求めさせる問題があります。手計算をするためには仕方ないことだという理由は納得できるのですが、5つのデータの統計量に何の意味があるのか？はなはだ疑問です。

このようにデータ分析は現実に近い学問分野であるのにも関わらず、手計算をさせるという視点に立つと、数学的にも社会的にも無機質で無意味なものになりがちです。ここにデータ分析がつまらない理由があるのではないのでしょうか？

### 3. そもそもデータ分析は何のために行われるのか。

データ分析は社会からの要求に基づき取り込まれた分野です。それでは社会の中でデータ分析はどのように行われているのでしょうか？社会の中でデータ分析は下の図のように進められます。



高校数学のデータ分析では「データ分析」のみが行われています。資料をもとに表やグラフを作成し、そのデータに対する統計量（平均値や分散、標準偏差、偏差値、相関係数、中央値など）を調べます。この結果をもとに分析を行い、「わかったこと」をまとめます。

学校や先生によってはテープを10cmに切る、時間を計る、アンケート調査を行うなどの実験（心理学実験）を行い、そこで集まったデータをもとにデータ分析を行う授業を行っているところもあるでしょう。この実験は心理学的にはとても有用な実験であると思われます。しかしながらこの実験を有用なものにするためには、数学ではなく心理学的な理論づけが必要であるため、数学の範囲では十分な考察が行われなまま終わってしまいます。「結果、こうなったね」→「ホントダ！」で終わってしまうのです。

ところでなぜ私たちはデータを分析するのでしょうか？それは何か現実に直面した問題を解決するためです。データ分析のメインは「データを分析する」ことではなく、「解決策を作り、実行する」ことなのです。「データを分析する」ことではなく、「解決策を作る」ことにデータ分析の面白さがあるのではないのでしょうか？生徒に解決策を考えさせることによって、データ分析は社会に役立ち、面白い授業にすることができるのです。

#### 4. 面白く社会に役立つデータ分析

今までの分析をもとに情報の授業を使って「解決策を含めたデータ分析」を行ってみました。

##### <<授業案のイメージ（1テーマ3時間）>>

時間	内容	詳細
1時間目	統計量の算出方法、理論を理解する。	教科書を用いた一般的なデータ分析の授業。 ここである程度の計算練習もさせておく。
2時間目	<個人ワーク> エクセルを用いてデータ分析	課題を設定し、それに対するデータも与える。 エクセルを用いて統計量を計算し、グラフなどを作成する。 個々がデータを分析し、解決策を考える。
3時間目	<グループワーク> グループディスカッション	2時間目に作成したデータをもとにグループディスカッション（1グループ5人程度）を行う。グループで1つの解決策をまとめてもらい、発表する。 最後に教員の立場からデータ分析を行った結果を提示する。

生徒が興味をもち、なおかつ楽しんで考えることができるように次のことに気を付けて題材選びをしました。

- ①題材はできるだけ生徒の日常に関わる題材、または社会の中で役に立つ題材を選ぶ。
- ②データは少なすぎでは現実味がないので、ある程度多くする。データが多くなると計算が煩雑になってしまうのでPCソフトのエクセルを用いて計算をメインにせず、データを分析し、解決策を立てることをメインに授業を行います。
- ③解決策は深いところまで議論できるようにある程度高度なものを選ぶ。「正しい答えを出す」ことを目標とせず、「考え、議論する」ことを目標にします。（解決策はひとつとは限りません。きちんと理由づけできていた答えならすべてが正解となります）

この授業を行うことにより、生徒の資料に対する態度が変わってきました。解決策を考えるためには資料を深く読み込んで、考えなければなりません。自然と生徒はデータを深いところまで読み取り、考えるクセがつくようになりました。中には先生顔負けなハイレベルでユニークな結論が出たグループもあり、生徒にとって楽しく、ためになる授業を行うことができたように思います。

以下、いくつか題材を提示します。これからのデータ分析の授業の参考にいただければ嬉しいです。

5. 面白く社会に役立つデータ分析の題材例

5-1. 度数分布、箱ひげ図、平均、分散、標準偏差に関する題材

[題材1]

数学の実力試験（100点満点）を行った。生徒の数学の実力アップのために、この結果をもとにしてそれぞれのクラスの数学の教科指導方針を考えたい。以下の資料はそれぞれのクラスの平均、分散、標準偏差、度数分布表である。このデータをもとに後の問題に答えなさい。

[問題1]データの分析結果をもとにして、それぞれのクラスの数学の成績を上げるためにはどのような対策を行えばよいですか？具体的に考えて対策案を立てなさい。

[問題2]A組、B組、C組の1クラスを数学特待クラスとしたい。どのクラスを数学特待クラスとすべきか？理由も含めて書きなさい。

[問題3]A組、B組、C組のうち、一番授業をするのが大変だと思われるクラスはどのクラスですか？理由も含めて書きなさい。

	A組	B組	C組
平均	70.0	70.0	57.0
分散	82.10	20.19	20.19
標準偏差	9.06	4.49	4.49

度数分布表

点数	A組	B組	C組
～45	0	0	0
45～50	0	0	1
50～55	2	0	7
55～60	1	1	8
60～65	5	2	4
65～70	4	8	1
70～75	4	8	0
75～80	2	2	0
80～85	1	0	0
85～90	2	0	0

[題材2]

あなたはハンバーガーショップAバーガーの店長です。ある日Aバーガーに次のようなクレームがありました。

「AバーガーのポテトはBバーガーのポテトに比べて長さが短いような気がします。

Aバーガーの方が短いポテトが多いので、Aバーガーで買うポテトは損です。

AバーガーでもBバーガーくらい長いポテトを提供してください。」

そこであなたはAバーガーとBバーガーのポテトの長さを調査しました。

下の表は調査の結果のAバーガーとBバーガーのポテトの長さの平均、分散、標準偏差です。

このデータをもとに、このクレームに対する返答を考えてください。

	Aバーガー	Bバーガー
平均	4.57	4.61
分散	0.68	2.58
標準偏差	0.83	1.62

## 5-2. 回帰分析に関する題材

### [題材1]

以下のデータはA高校の数学の定期試験の結果と模擬試験の結果を表にまとめたものである。

この結果をもとに後の問いに答えなさい。

[問題1] A高校の数学の定期試験の結果は生徒の実力を適正にはかっているといえるでしょうか？

[問題2] 定期試験の結果および、模擬試験の結果をもとにして習熟度別グレード編成を考えたい。グレードはA、B、Cの3段階とし、Aは基礎、Bは標準、Cは応用的な内容を授業で行うものとする。グレード編成の基準を決め、個々人のグレードを決めなさい。ただし、グレード編成にあたって、人数配分は自由に行ってもいいが、ある程度人数にバランスをもたせること。

※この資料に関しては実際の過去の生徒の5回分の試験結果データを①A試験[基礎問題=定期試験]とB試験[応用問題=模擬試験]に分け、その平均を100点満点になるようにとる。②さらにランダムで点数を少しずつ変える③できた点数をランダムに並び替えることによって、現実のデータを個人特定ができないようにした。

### [題材2]

2013年度の〇〇総合商社の決算書である。〇〇総合商社はA支店、B支店、C支店の3つの支店を持ち、それぞれの支店の支店長は限られた販売予算の中でそれぞれが効果的であると考えられる施策を「重点施策」として販売促進の対策を行っている。この決算書は2013年度の月ごとの重点施策の支出額を「施策費」とし、それぞれの支店の売上額を「売上」として表したものである。

[問題1] 施策費と売り上げにはどのような関係がありますか？

[問題2] A支店、B支店、C支店の売上について比較したとき、一番問題を抱えていると考えられる支店はどれか。また、なぜその支店が問題であるといえるか。理由を2つ書け。

[問題3] A支店、B支店、C支店の施策費について比較したとき、一番施策費を上手に利用していると考えられる支店はどれか。また、なぜその支店がそういえるか。理由を書け。

[問題4] 〇〇総合商社では2014年に新たにD支店を開店させ、営業を行う予定である。D支店の施策をこの資料をもとにして考えたい。D支店の特徴は次のようである。次の問いに答えよ。

- ・D支店を利用する客層はA支店、B支店、C支店の客層とそれほど変わりはないと予想される。
- ・D支店の施策費は月4～5百万円を考えている。
- ・新規開店の店であるため、宣伝によるリスクは最小限にしたい。また、宣伝によって顧客を集め、できるだけ多くの売り上げを上げたいと考えている。

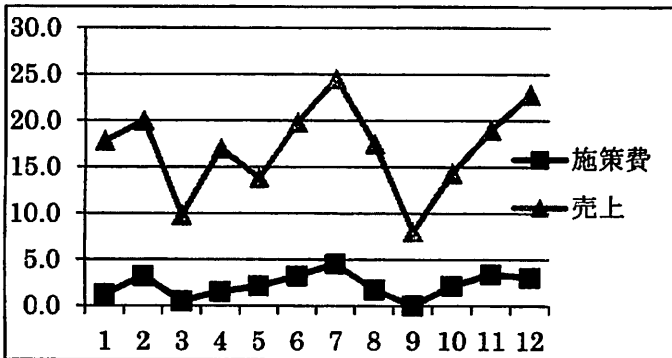
(1) D支店の重点施策として「TVコマーシャル」「割引券」「店頭イベント」のどれをあげればよいか。一つ選べ。また、なぜそれがよいと考えるか理由を述べよ。

(2) D支店では2014年に一度バーゲンセールを行う予定である。このときには施策を重点的に行うとともに、売り上げアップを狙っている。何月にバーゲンセールを行うのがよいか。また、その月が適切である理由を答えよ。

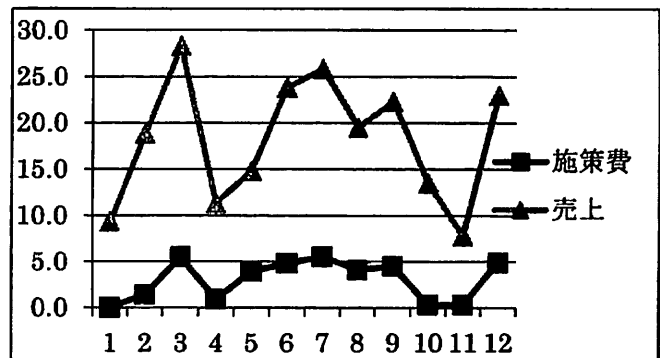
※資料は次ページ

支店 施策 月	A支店 TVコマーシャル		B支店 割引券		C支店 店頭イベント		A+B+C	
	施策費(x)	売上(y・百万)	施策費(x)	売上(y・百万)	施策費(x)	売上(y・百万)	施策費(x)	売上(y・百万)
1	1.2	17.8	0	9.3	1.2	4.1	2.4	31.2
2	3.2	20.0	1.4	18.8	3.3	18.9	7.9	57.7
3	0.5	9.8	5.5	28.3	0.3	28.8	6.3	66.9
4	1.5	17.0	0.9	11.3	3.1	26.7	5.5	55
5	2.1	13.8	3.9	14.8	0.3	7.1	6.3	35.7
6	3.2	19.8	4.8	23.8	0.4	14.7	8.4	58.3
7	4.5	24.5	5.5	25.8	1.3	24.9	11.3	75.2
8	1.7	17.5	4.1	19.5	0.3	18.2	6.1	55.2
9	0.0	8.0	4.5	22.3	0	2.4	4.5	32.7
10	2.1	14.3	0.3	13.5	2.8	18.2	5.2	46
11	3.4	19.0	0.3	7.8	2	9.2	5.7	36
12	3.0	22.8	4.9	23	0.7	5.1	8.6	50.9
合計	26.4	204.3	36.1	218.2	15.7	178.3	78.2	600.8
平均	2.2	17.025	3.00833	18.1833	1.30833	14.8583	6.51667	50.066667
標準偏差	1.25366	4.67246	2.12464	6.46566	1.14924	8.81528	2.18702	13.450052
相関係数	0.877		0.904		0.364		0.765	

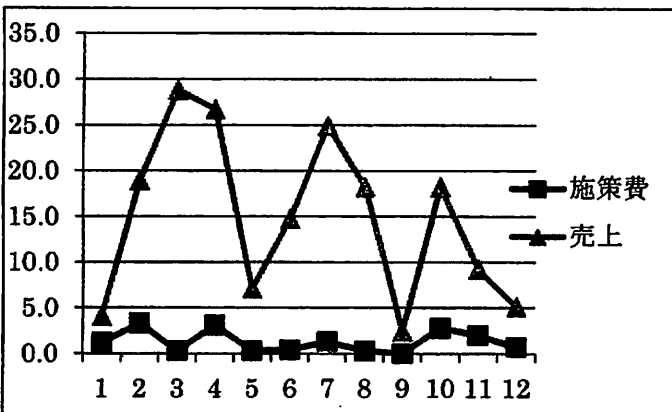
A支店



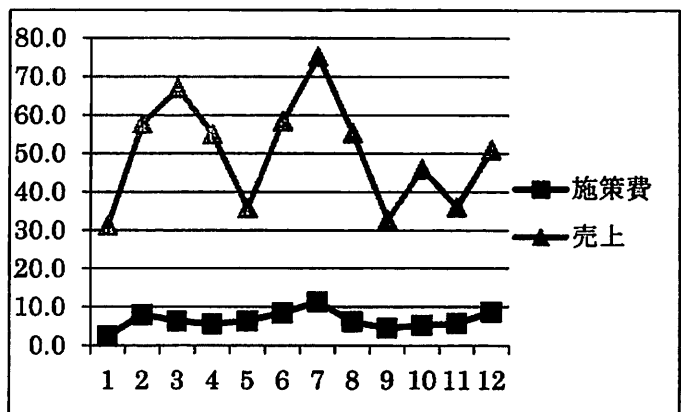
B支店



C支店



合計



## 6. 参考文献

- ・「ファーストブック 統計学がわかる」 向後千春・富永敦子 技術評論社
- ・「ファーストブック 統計学がわかる[回帰分析・因子分析編]」 向後千春・富永敦子 技術評論社
- ・「「それ、根拠あるの?」と言わせないデータ・統計分析ができる本」 柏木吉基 日本実業出版社