

10 章 データ分析力

組 番 氏名

今回の授業の位置づけ

○探究スパイラル(課題設定→情報収集→分析と整理→まとめ・表現)の過程にあつて、課題に対して立てた仮説の検証に向けて、収集したデータの真偽や妥当性を判断する際に駆使するのが「データ分析力」である。

ねらい

○「収集したデータは正しいのか?」「正しいとしても裏付けは確かなのか?」「サンプルの採り方は客観的かつ公平か?」「数値同士の関係の有無は?」「数値が対応しているように見えるが見せかけではないか?」これらの問いをぶつけて検証する、いわゆる「クリティカル・シンキング」(批判的な捉え方)の視点を身につける。

準備

- クラスを6つのグループに分けておく。各グループに進行役を1名決めておく。
- 数学Ⅰの教科書、相関係数の計算のための電卓(もしくはスマホの電卓機能)を用意する。

【以下の1と2を、各自読む 15分】

※先生はタイムキーパー役をお願いします

1 「統計を使った嘘」の3つのパターン

「本製品は利用者の90%の満足度をいただいております」という宣伝文句を信用する?しない?

利用客から返送されたアンケートハガキによる数値だったならばどうだろうか。アンケートハガキを返送すること自体がそもそもこの製品に好意的であるとも考えられるのではないだろうか。

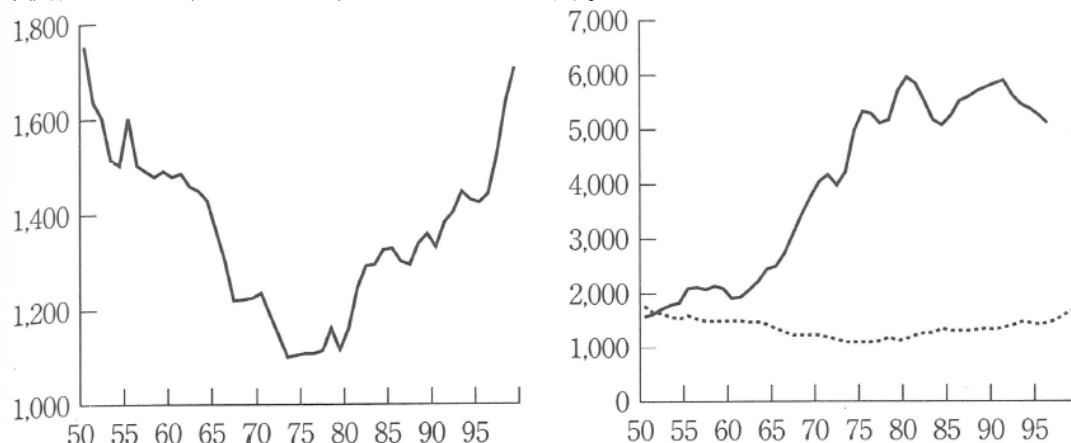
あるいは、店頭でその製品を買おうとした客をねらって満足度を聞いたのかもしれない。その製品に不満ならば、そもそも買おうとしないのではないだろうか。

人数でなくパーセント表示したのも疑わしい。少ない母集団から全体を推測することには無理がある。

(1) 見せ方による嘘

よくあるのが、グラフの描き方によって、目盛りの取り方によって嘘をつくものです。

下の図は、日米の犯罪率を比較したグラフです。左が日本の犯罪率で、右がアメリカの犯罪率を表しています(どちらも、横軸は1950年から2000年までをとっています)。



(『大学生のためのリサーチリテラシー入門』 p 179 のより)

このグラフをぱっと見ると、「アメリカ(右)は、1960 年代に数値が上昇し、1980 年代以降は横ばいで推移している。それに対して、日本(左)は、1980 年代まで下がってきて、その後急激に値が上昇し続けているのだな」と思ってしまうでしょう。しかし、このグラフには嘘があります。

1つ目の嘘は図 7-2 の左の図において「日本の V 字型の左側は粗暴犯や凶悪犯が多かったのに対し、反対に右側は窃盗が圧倒的に多い」という**数字の性質の違い**です。

そして、もう1つは**目盛りの違い**です。縦軸の目盛りが左と右のグラフで異なっているのです。右の図で点線で描かれているのは、左のグラフを右の目盛りに合わせて書き直したものです。点線を見ると、つまり、同じ目盛りで比較すると、日本は 1950 年代以降、1,000 と 2,000 の間を安定して推移していることがわかります。これが見せ方による嘘の例です。

(2) データ選択の嘘

もともとの条件が異なるものを比較しようとする、おかしな結果が導かれる場合があります。これを利用して、嘘をつくことが可能です。

たとえば、年間の自動車事故による死亡者数と、飛行機事故による死亡者数を比較したところ、自動車事故による死亡者数の方が多かったとします。この結果から、「飛行機の方が自動車よりも安全な乗り物だ」と判断してよいでしょうか。それはまずいですよね。

なぜなら、自動車は毎日乗る人も多いでしょうし、利用される機会＝事故に遭う可能性のある機会が、そもそも、飛行機のそれよりもはるかに多いからです。そうした前提条件が異なるものを単純に比較してはいけません。

さらに、こんな例も考えられます。都会にある A 市と田舎にある B 市で、中学生の子どもがいる家庭のうち、何%が子どもを塾へ通わせているかを調べました。すると A 市では 68%が、B 市では 44%が子どもを塾に通わせているという結果になりました。この結果を見て、「都会のほうが教育熱心な親が多く、子どもを塾へ通わせる家庭が多いのだ」と考えたとしたらどうでしょうか。それは正しいといえるでしょうか。

A 市と B 市ではそもそも塾の数が違うかもしれません。都会の A 市には塾がたくさんあって、田舎の B 市には塾そのものの数が少ないということも考えられます。また、年収が多い人のほうが子どもを塾に通わせる割合が高いということもあるかもしれません。同じ年収で比較したら塾へ通わせる割合は変わらないのに、都会である A 市の方が高収入の人が多いために、A 市と B 市で塾へ通わせる割合に差が生じているという可能性もあります。

(3) データ収集の嘘

どのようにデータを集めるかによって、まったく違った結果が得られることがあります。

高校生の化粧使用頻度についてアンケート調査を行う場合に、渋谷駅周辺の高校生にアンケートを実施した場合と、地方の農村部でアンケートを実施した場合とでは、まったく違った結果になるでしょう。たとえば、前者だけのデータでは「日本の女子高校生は化粧ばかりしている」という結論が得られるかもしれないし、後者だけのアンケートでは「日本の女子高校生は今なお質素で清楚である」という結論が得られるかもしれません。いずれの場合も、全国の高校生を代表する集団とは見なせないですね。

あるいは、「大学生の授業関心度」を調べるための質問紙調査を、大学の 1 限で実施したとします。その調査結果を一般的な大学生の授業関心度と見なすのは問題がありそうです。というのは、1限の授業にきちんと出て

くる学生は、もともと授業への関心が高い可能性があるからです。1限の授業の受講生だけを調べたとしたら、本来の調査対象全体をうまく反映していないことになりかねません。

ここで取り上げた2つの例は、いずれも関心のある調査対象全体の一部からデータを収集して、そのデータを分析することで、調査対象全体についての様相を明らかにしようとするものです。こうした調査方法は、標本調査（サンプル調査）といわれるものです。関心のある調査対象全体を母集団といいます。そして、母集団の一部であり、実際に調査を行った集団を標本（あるいはサンプル）といいます。

ここでの2例は、どちらも、偏った標本（サンプル）を分析しても、その結果は、母集団の様子を正確に表すものにはならないということを伝えてくれるものです。

『大学生のためのリサーチリテラシー入門』（山田剛史・林創著 ミネルヴァ書房 2011）より抜粋

2 相関関係と因果関係

夏の暑い日に外へ出たところ、ウチワや扇子であおいでいる人を多く見かけたでしょう。

その時、「ウチワや扇子であおいでいる人たちがいるから、暑くなる」と判断する人はいないだろう。暑くて汗が出るから、ウチワや扇子であおいで涼んでいる（気化熱を利用している）のは言うまでもない。

では、クルマの保有率が減少し、地価が高騰した地域があったとしたらどうか？さらに、その地域で商店街が急に衰退してしまったとしたら？

クルマと地価に直接の関係は無いはずだし、一般的には、商店街が衰退すれば地価が下落するだろうと思われるがちだが、その地域に利便性の高い鉄道が敷設されて駅ビルができたという事実を聞けば、納得するのではあるまいか。クルマ、地価、商店街という要素の変化に、駅ビルという第四の変数が介在したということである。

こうしたさまざまな関係性についての説明を、谷岡一郎著『データはウソをつく』から以下に引用する。

〈「**相関関係 (correlation)**」と「**因果関係 (causality)**」とは、よく混同される概念ですが両者は同じではありません。まず両者の定義を見てもらいます。

☆**相関関係**…変数の一方□が変化するとき、もう一方の変数□の変化がランダムでない関係（または同じことだけど、変数□が変化するとき、変数□の変化がランダムでない関係）。（記号で□↔□のように、両方向の矢印を使用する）

☆**因果関係**…変数の一方□の変化によって、もう一方の変数□の変化を引き起こす関係。つまり、変数□が原因で、変数□が結果。（記号で□→□のように、一方向の矢印を使用する）

相関関係と因果関係の違いをわかりやすく説明すると、因果関係の変数□、□は、**時間的に□が□に先行**しますが、相関関係ではどちらが先かわかりません。□が□に先行するかもしれないし、その逆かもしれないし、はたまた同時かもしれないのです。

「**(統計学上有意な)相関関係が存在する**」ことは、**因果関係の前提**です。つまり有意な相関関係は、因果関係があることの必要条件となります。逆に言えば、相関関係がある、というだけでは、因果関係についてはまだ何も言えません。

さて、変数□イと□ロの間に今、統計的に有意な相関関係があるものとします。つまり、「□イ ↔ □ロ」が成立しています。この時、□イと□ロの間には、次のような9つの可能性があります。

可能性1 調査方法の問題(サンプル数、母集団など)

可能性2 □イが□ロの直接原因 □イ → □ロ

可能性3 □ロが□イの直接原因 □イ ← □ロ

可能性4 □イと□ロが相互に影響 □イ ↔ □ロ

可能性5 □イが□ロの間接原因 □イ → ■ → □ロ ■ → □ロ □イ → ■ ■ ↔ □ロ

可能性6 □イも□ロも「第三の変数」の結果 ■ → □イ ■ → □ロ

可能性7 上記1～5の組み合わせ

可能性8 循環因果(計測不能)

可能性9 単なる偶然＝見せかけの相関)

『データはウソをつく』(谷岡一郎著 ちくまプリマー新書2007)より抜粋

【課題1】：(20分) ※先生はタイムキーパー役をお願いします

ねらい

ワーク3 の司会進行をお願いします

○因果関係について理解する。

内容

○「見せかけの因果関係」を各自考えてみる。

ワーク1 「見せかけの因果関係」(一見すると「原因とその直接的な結果」のように受け取れるが、厳密には「原因→結果」になっていない関係)を、その解説も含めて各自1つ考える (5分)

解答例：「天気が晴れる → アイスクリームの売り上げが増加する」

解説 : 気温が上がるとアイスクリームが売れるのであって、晴れていても気温が低ければ売れない。

(補足) 22～23度になるとアイスクリームの売り上げが増加するが、30度以上になると売り上げが鈍り、かき氷が売れてくる。

ワーク2 グループ内でそれぞれが考えた「見せかけの因果関係」を発表する。そして、グループの最優秀作を話し合って決める。 (7分)

ワーク3 各グループの最優秀作を、それぞれ発表する。そして、挙手によってクラスの最優秀作を決める。 (8分)

自分の考えた「見せかけの因果関係」	
-------------------	--

グループの最優秀作	
-----------	--

クラスの最優秀作	
----------	--

【課題2】：（15分）

※先生はタイムキーパー役をお願いします

各自10分間考え、残り5分で先生から解答例を読み上げてください。

ねらい

- 因果関係と相関関係について理解する。

内容

- 実際によく使われる論法に対して、クリティカル・シンキング（批判的思考力）を働かせる。

ワーク1 クリティカル・シンキング問題 （15分）

文部科学省は、「全国学力・学習状況調査」という学力テストの結果を用いて、子どもの学力と家庭環境にどのような関係がみられるかを分析しています。その分析によると、「親の年収や学歴が低くても学力が高い児童の特徴は、家庭で読書をしていること」だとされています。

この結果を受けて、多くのメディアは「子どもに読書をさせることが学力向上に重要だ」と報道をしています。

↓

この報道の間違ひの可能性を指摘してください。

『学力の経済学』（中室牧子 2015 p 20～23）より改題

報道の間違ひの可能性について、自分の考えを下欄に記入する。（10分）

解答例を聴き、要点をメモする。

（5分）

【前半終了】

【課題3】：（40分）

※先生はタイムキーパー役をお願いします

時間終了近くになったら解答例を教室掲示してください

ねらい

- 2種類の数値から相関関係の有無を確認する。

内容

- グループで協力しながら散布図を作り、相関係数を求める。また、関係の可能性について考える。

準備

- 数研出版『数学Ⅰ』教科書 第5章「データの分析」のうち「データの相関」の項 参照
- 電卓があると便利（英語、数学それぞれの平均点から、分散、標準偏差、共分散を計算する）

ワーク 1 A君からO君までの15名の英語と数学のテスト結果は以下のとおりである。グループで協力して、散布図（相関図）をつくり、相関係数を求め、相関関係の有無を答えなさい。ただし、小数第3位を四捨五入して小数第2位まで求めよ。

『課題研究メソッド』（岡本尚也著 啓林館 2016）より一部抜粋

生徒	英語	数学
A	75	68
B	40	50
C	66	62
D	64	52
E	47	56
F	50	55
G	48	51
H	46	44

生徒	英語	数学
I	55	59
J	36	40
K	53	68
L	69	69
M	52	75
N	43	35
O	75	88

ワーク 2

上記の「英語と数学のテスト結果」の関係は、「英語ができるので数学ができる」といった因果関係ではない。では、どのような関係として考えられるだろうか。

4 ページの可能性1～9を参考にしながら、その関係についてグループで話し合って図式化してみよう。

担当の先生へ

解答例です。太字部分を中心にかいつまんでお示ください

【課題2】

ワーク 1

第一に、読書と学力の間に因果関係があるようにとらえているのが誤りです。「因果関係」と「相関関係」、どちらも2つの出来事の関係を示すときに使われる言葉ですが、決定的に違う点があります。

因果関係は「Aという原因によってBという結果が生じた」ことを意味しています。しかし、相関関係は単に、「AとBが同時に起こっている」ことを意味しているにすぎません。相関関係は2つの出来事のうちどちらが「原因」で、どちらが「結果」であるかを明らかにするものではないのです。「相関関係」があるということは、必ずしも「因果関係」があることを意味しません。

つまり、**読書をしているから子どもの学力が高い（因果関係）**のではなく、**学力の高い子どもが読書をしているにすぎない（相関関係）**可能性があるのです。

「読書をする」ことが原因で、「学力が高くなる」という結果がもたらされていることがはっきりしないのに、本を買い与えたり、読み聞かせをしたりしたら、お金や時間の無駄遣いになってしまうかもしれません。

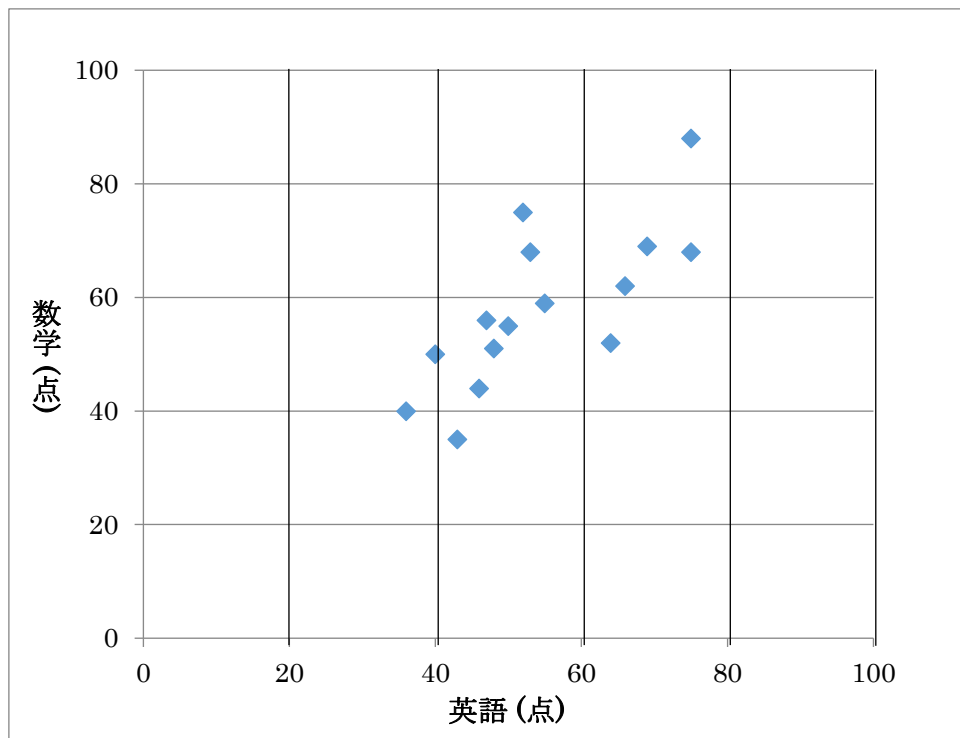
また、この報道にはもうひとつの誤りがあります。それは、「**見せかけの相関**」の可能性を検討していないことです。つまり、読書にも学力にも影響するような「**第三の要因**」があるかもしれないのに、そのことを考慮していないのです。

「**第三の要因**」としては、たとえば、「**子どもに対する親の関心の高さ**」などが考えられます。子どもに対する関心が高い親は、子どもを勉強するように促すでしょうし、同時に子どもに本を買い与えたりもするでしょう。その「**関心の高さ**」こそが両方の変化を同時に引き起こしているにもかかわらず、**あたかも、読書と学力の間に相関関係があるかのように見えてしまう。これが「見せかけの相関」です。**

間違っ**て、見せかけの相関を因果関係と解釈してしまうと誤った判断のもとになります。**

ワーク1

●散布図



相関関係の有無について

比較的強い正の相関がある。

相関係数

0. 7 5

ワーク2 (関係の可能性を下記に2例挙げておきましたが、これだけが正解とは限りません)

可能性6 家庭学習時間が長い → 数学の得点が高い
↙ 英語の得点が高い

英語の得点が高い

英語の得点が高い

可能性5 英語の得点が高い → 勉強に対する自信が → 数学の得点が高い
付いてきていて、ヤル気になった。

付いてきていて、ヤル気になった。

振り返り <リフレクションシートの記入及び回収>

データ分析力 リフレクションシート

年		組		番号		氏名	
---	--	---	--	----	--	----	--

1 日時： _____月_____日（ ） _____校時

2 振り返り

Q1：「統計を使った嘘」にはどのようなパターンがあるか、書いてください。

--

Q2：「因果関係」、「相関関係」とはそれぞれどのような関係のことか、簡潔にわかりやすく説明してください。

因果関係とは
相関関係とは

以下、該当する番号のマークを塗りつぶしてください。

Q3 今回の課題に、あなたはどのように関わりましたか。 ①積極的にできた ②まあまあ取り組めた ③うまく参加できなかった	①	②	③			●
Q4 今回の授業の難易度はどうでしたか。 ①易しかった ②ちょうど良かった ③難しかった	①	②	③			●
Q5 今回の授業の時間設定はどうでしたか。 ①時間が余ってしまった ②ちょうどよかった ③時間が足りなかった	①	②	③			●



教師用資料 10章 データ分析力【授業デザイン】

今回の授業の位置づけ

○探究スパイラル(課題設定→情報収集→分析と整理→まとめ・表現)の過程にあつて、課題に対して立てた仮説の検証に向けて、収集したデータの真偽や妥当性を判断する際に駆使するのが「データ分析力」である。

ねらい

○「収集したデータは正しいのか?」「正しいとしても裏付けは確かなのか?」「サンプルの採り方は客観的かつ公平か?」「数値同士の関係の有無は?」「数値が対応しているように見えるが見せかけではないか?」これらの問いをぶつけて検証する、いわゆる「クリティカル・シンキング」(批判的な捉え方)の視点を身につける。

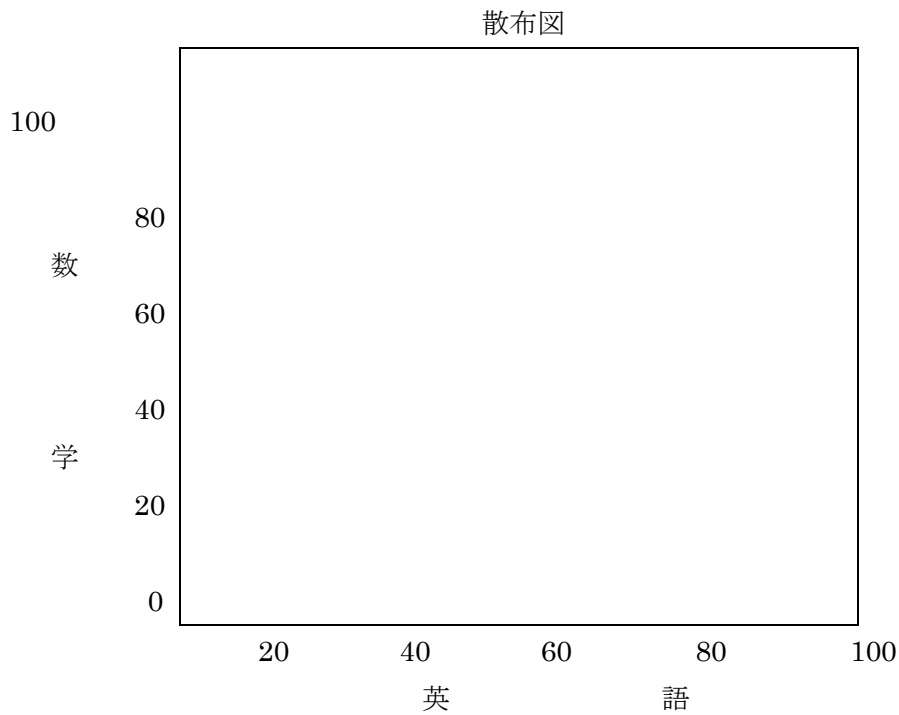
※クラスを6つのグループに分けておいてください。

※『数学I』の教科書を用意させておいてください。

※電卓(もしくはスマホの電卓機能)をあらかじめ用意させてください。

時間	内容・留意点
各自読み 15分	「1『統計を使った嘘』の3つのパターン(1)(2)(3)」、「2相関関係と因果関係」を各自読む。 ※「テキストの1ページから4ページまでを各自15分間で黙読して、内容を理解してください。このあと、見せかけの因果関係を考えたり、事柄の関係を見抜く問題を実際に解いたりするので、そのつもりで内容理解に努めてください。」と指示してください。
【課題1】 20分 ワーク1 15分 ワーク2 7分 ワーク3 8分	「見せかけの因果関係」を各自考える グループ内で発表し合って、グループの最優秀作を決める。 各グループの最優秀作を発表し合って、クラスの最優秀作を決める。 ※ワーク3では先生が司会進行をお願いします。
【課題2】 ワーク1 15分	①各自で報道の間違いの可能性を考える (10分) ②解答例を聴き、要点をメモする (5分) ※上記②については、先生が解答例をかいつまんで読み上げてください。 【前半終了】
【課題3】 40分	散布図をつくり、相関係数を求め、相関関係の有無を答える (40分) ※「グループで協力して頑張ってください。役割分担していないと、時間内にできないかもしれないので、皆で相談してから始めてください。なお、数学Iの教科書は適宜みても構いません。」と指示してください。 ※途中での時間の区切りは不要です。 ※終了時間直前に、解答例の教室掲示をお願いします。既に作業完了したグループから解答例を求められたら、そのつど見せてください。 ※計算の途中で四捨五入するのではなく、途中計算は全て分数で行うと良いです。 最後に四捨五入すると計算にとまどいません。
振り返り 10分	リフレクションシートに記入 ※シートの回収をお願いします

課題3 ワーク1 解答用シート



相関関係の有無について

相関係数

(解答は時間終了時に担当の先生から教室掲示されます)

ワーク2 解答用シート

英語の得点と数学の得点との「関係」の可能性を図式化する。