

計量分析の表を読む

小林 久高

KOBAYASHI Hisataka

1 はじめに

大学で学部生向けの社会学の講義をする際、言葉や図だけで説明するよりも、数量的なデータ分析の結果の表をもとに解説したほうが良いと思うことがよくある。過去の重要な研究や新発見についての表をもとに説明していけるならば、より説得的に議論が展開できるからだ。しかしそれはなかなか難しいことが多い。というのは、常識的に理解できる平均や比率だけでなく、やや複雑な分析から重要な発見が導かれていることも多々あるからだ。そんな分析結果の表をもとに授業をすると、これまで統計学などを学習してきた学生は理解できるが、そうでない学生にはチンプンカンプンということになりかねない。そういうわけで、言葉や図で結果を要約し「統計学を勉強してからこの論文の分析結果を読むとよくわかるよ」などと言うことになる。

しかし、これでは学生は「なるほど！」という気持ちにもなれず面白くもないはずだ。大学での勉強は何といっても学生の「なぜそんなことが言えるのか？」という疑問や「こうすればもっとうまく分析できるんじゃないか！」といったアイデ

ア創出の喜びに関わっている。こういったことを感じるためには、データ分析の結果を見て、論文の著者がどんな証拠をもとにどう考えたかを追体験できるほうがずっといいこと間違いなしなのである。

そうはいうものの「〇〇社会学」という講義は統計学やデータ解析の講義ではなく、分析法の解説にそんなに時間をさくわけにはいかない。そこで、2~3回程度の講義で、学生が社会学的な計量分析の結果の表などを読めるようにするための基本的な資料を作れないかと考えた。

考えてはみたものの、この作業はなかなかやっかいだ。というのは、簡単すぎると学生には全然分析のイメージがわからないし、誤解することも多くなる。逆に詳しく書きすぎると2~3回の授業ではどうていおさまらないし、数学アレルギーの者は途中で放棄しかねない。このバランスをとるのは実に難しいのである。

そんな思いを持ちながら作成したのが以下の資料である。資料が想定している読者は、これまで大学で統計学を全く学んでこなかった学生である¹。

¹ この原稿を執筆しているとき、1冊の本のことが頭をよぎった。それは安田三郎著『社会調査の計画と解析』のことである。この本は安田先生の社会調査関係の論文を編集したものであり、所収の各論文から筆者はさまざまな示唆を得た。しかしこれまで、どういう真意で先生がこれを編集されたのかを考えることはあまりなかった。

今回、この本の「まえがき」を読み直し、「実は、この本はこれから経験的研究を進めていこうとする者への〈贈り物〉だったのだ」と強く感じた。本稿も精神としてはその流れに属する。今後も、学生や院生に役立つようなものがあれば、少しずつ「セミナー」に掲載していきたいと思っている。

2 基本的なこと

2.1 変数

(1) 変数と値

異なる値をとりうるものを変数という。物の重さや体積は異なる値をとりうるので変数である。身長や体重も異なる値をとりうるので変数である。性別は男女という値をとる変数だ。

(2) 量的変数と質的変数

変数には量的な変数と質的な変数がある。量的な変数とは、年齢や収入といった変数のことであり、質的な変数（カテゴリカルな変数）とは職業や学歴といった変数である。量的な変数と質的な変数は、平均を出して意味があるかどうかということで区別できる。量的変数である年齢や収入の平均には意味がある。しかし、質的変数である職業に、専門職なら1、ホワイトカラーなら2などと値を与えて平均を出しても意味はない。

量的変数を質的変数にしたり質的変数を量的変数にしたりできる場合がある。たとえば年齢を年少、生産年齢、老齢に区分することは量的変数から質的変数への変換である。質的変数としての学歴も、教育年数からとらえれば量的変数になる。人びとの行政への満足度を調べるとき、「満足、やや満足、やや不満、不満」という形で答えを聞くことがある。この行政への満足度という変数は厳密には質的な変数と考えられるが、それぞれに4点から1点までを与えて量的変数として分析することもできる。

(3) 変数と命題

社会学における命題は1つの変数の分布や複数の変数の関係について述べている。「日本の勤労者の年間平均所得は〇〇円だ」という命題は勤労者の所得という1変数の分布の特徴について述べて

いる。「常勤職は非常勤職より所得が多い」という命題は、「常勤と非常勤」という2つの値をとる職業という変数と（さまざまな値をとる）所得という変数の関係について述べている。

(4) 相関関係と因果関係

一方の変数と他方の変数に関連があるとき相関関係があるという。また、一方の変数が他方の変数に影響を与えるとき因果関係があるという。変数間に相関関係があるからといって、必ずしも因果関係があるとは言えない。

(5) 独立変数と従属変数

因果関係において、影響をあたえる変数を独立変数（説明変数）という。影響を受ける変数は従属変数（被説明変数）である。

2.2 平均・分散・標準偏差・相関係数

(1) データの例

量的な変数を分析する際に押さえておかなければならない概念として、平均、分散、標準偏差、相関係数がある。仮想的な表1のデータをもとに説明していこう。

表1 xとyのデータ

ケース番号(i)	得点(x)	得点(y)
1	80 (x ₁)	70 (y ₁)
2	70 (x ₂)	60 (y ₂)
3	40 (x ₃)	30 (y ₃)
4	40 (x ₄)	20 (y ₄)
5	20 (x ₅)	20 (y ₅)
平均	50 \bar{x}	40 \bar{y}
分散	480 s_x^2	440 s_y^2
標準偏差	21.909 s_x	20.976 s_y
相関係数	0.957 r_{xy}	

(2) 平均

n 個の値があるとき、その分布の平均は次式で表される。平均はデータの分布の中心傾向を示し、分布を代表する値だ。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{5} (80 + 70 + 40 + 40 + 20) = 50$$

(3) 分散

n 個の値があるとき、その分布の分散は次式で表される。分散はデータの散らばりを表す値だ。

$$\begin{aligned} s_x^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{5} \left\{ (80-50)^2 + (70-50)^2 + (40-50)^2 \right. \\ &\quad \left. + (40-50)^2 + (20-50)^2 \right\} \\ &= 480 \end{aligned}$$

分散は「値と平均の差の 2 乗」の総和（これを偏差平方和という）をケース数 n で割ったものだ。それは、当該データの各ケースの値がどの程度中心傾向である平均値から離れているのかを（ケース数で割っている）「平均的に」示す指標である。ケース数で割っている、ケース数の異なるさまざまなデータにおいても、分散を用いて散らばりの程度が比較できる。

分散を n ではなく n-1 で割って出すこともある。これを特に不偏分散という。

$$s_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

(4) 標準偏差

分散の正の平方根を標準偏差という。

$$s_x = \sqrt{s_x^2} = \sqrt{480} = 21.909$$

変数が cm で測られた身長だとすると、分散は cm² の単位、標準偏差は cm の単位である。標準偏差はもとの変数の単位と同じだから、変数の値と関連づけて考えやすい。

(5) 相関係数

2 つの変数間の関係の強さは相関係数 r で表されることがある。相関係数は次の式で求められる。

$$\begin{aligned} r_{xy} &= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \\ &= \frac{\frac{1}{5} \left\{ (80-50)(70-40) + (70-50)(60-40) \right. \\ &\quad \left. + (40-50)(30-40) + (40-50)(20-40) \right. \\ &\quad \left. + (20-50)(20-40) \right\}}{21.909 \times 20.976} \\ &= 0.957 \end{aligned}$$

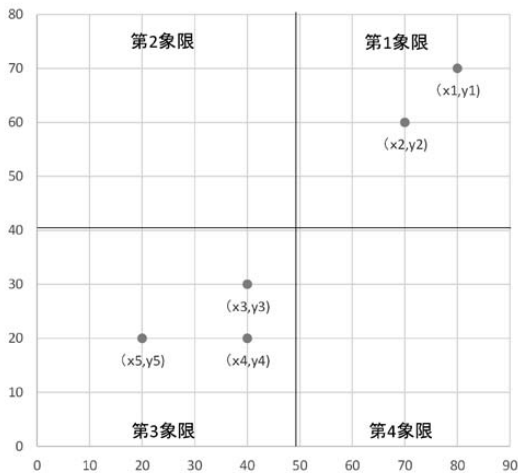
相関係数の値は、一方の変数の値が大きくなれば他方の変数の値も大きくなる時（正の相関）1 に近づき、一方の変数の値が大きくなれば他方の変数の値が小さくなる時（負の相関）-1 に近づく。両者にこういった関係がなければ 0 になる。1 や -1 に近い値のとき強い相関があるといい、0 に近い値のとき相関は弱いという。

なぜ上のような式で関連の強さを表せるかを簡単に述べる。先の表 1 をグラフで表すと図 1 が得られる。ここで中心にある太めの垂直軸と水平軸はそれぞれ x の平均と y の平均を表している。

図を見ると、x が大きな値になるほど y も大きな値になっており、x と y の間には正の相関があることがわかる。この図から予想できるように、正の相関がある場合、データは第 1 象限と第 3 象

限に多く存在し、第2象限や第4象限にはあまりない。負の相関の場合には逆に、データは第2象限と第4象限に多く存在し、第1象限と第3象限にはあまりない。両変数の間に相関がない場合には4つの象限にまんべんなくデータが存在する。

図1 変数xと変数yの得点



ここで相関係数の分子にある次の式の値がどうなるかを考えてみよう。

$$(x_i - \bar{x})(y_i - \bar{y})$$

図1の第1象限にある (x_1, y_1) の場合、この式 $(x_1 - \bar{x})(y_1 - \bar{y})$ の前のカッコ内の値も後のカッコ内の値も正になり、掛け合わせた式全体の値は正となる。 (x_3, y_3) のように第3象限にあるときには前のカッコ内も後のカッコ内も負になるが、掛け合わせた全体はこちらも正となる。 (x_i, y_i) が仮に第2象限や第4象限にあるならばこの式の2つのカッコ内の値は符号が異なり全体で負になる。

次に、この式の総和である次の式について考えよう（これは偏差積和とよばれるものだ）。

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

正の相関がある場合は、相対的に第1、第3象限に多くデータが存在するためこの総和の値は正の大きな値になる。逆に、負の相関がある場合には第2、第4象限に多くのデータが存在するためこの総和の値は負の大きな値になる。相関がほぼない場合には4つの象限にまんべんなくデータがあるため、総和は正負が相殺されて0に近くなる。

相関係数の式はこの総和の式を中心としたものだ。あとnで割ったり、2つの標準偏差で割ったりしており、それで初めて-1~+1の範囲をとる相関係数の式になるのだが（傾きのある直線上にすべてのデータがあるときは+1か-1になる）、中心にあるのはこの総和の式なのである。

2.3 素点と標準得点

データ全体の平均が0、標準偏差が1（すなわち分散が1）になるように値を変換したものを標準得点という（表2）。標準得点は素点から平均を引き標準偏差で割るという作業で得られる。1氏（ケース番号1）のxについての標準得点は、 $(80 - 50) \div 21.909 = 1.369$ となる。

表2 素点と標準得点

ケース番号(i)	x		y	
	素点	標準得点	素点	標準得点
1	80	1.369	70	1.430
2	70	0.913	60	0.953
3	40	-0.456	30	-0.477
4	40	-0.456	20	-0.953
5	20	-1.369	20	-0.953
平均	50	0.000	40	0.000
分散	480	1.000	440	1.000
標準偏差	21.909	1.000	20.976	1.000

標準得点はいわゆる偏差値のようなものだ。実は、偏差値は平均 50、標準偏差 10 になるように素点を変換したある種の標準得点なのである。偏差値を使えば 200 点満点の国語テストの 80 点と 150 点満点の数学テスト 70 点とでは、どちらの成績が相対的に（全体の中で）よいのかを比べたりできる。偏差値と同様、標準得点への変換によって、諸変数の得点が（単位が異なっても）比較できるようになる。

2.4 母集団と標本

社会調査で集められたデータの多くは、調べたい集団全体（母集団）のものではなく、そこから得られた標本に関するものだ。われわれは標本の平均や比率などをもとに、母集団の平均や比率などを推定する。たとえば 3000 人の有権者のデータから有権者全体の自民党支持率を推定したりするのである。また、「標本で男女で自民党支持率に差がある」ということから「母集団でも男女で自民党支持率に差がある」と言えるかどうか、といったことを検討したりする（これを検定という）。

2.5 以下でとりあげる文献

以下、小林,2008「共同性の精神的基盤と社会階層」（文献 A）を基本に、小林,2000「政治意識と政治参加の動態」（文献 B）を補足的に利用して計量分析について解説していく。文献 A での利用部分は 3 節までであり、文献 B では 1 つの表のみ利用する。両文献のこれらの箇所では、社会学の計量論文に見られる基本的な分析技法が使用されているので、これらの理解を通して、計量分析を行

っている多くの社会学の文献が読めるようになると思われる。

文献 A は同志社大学学術リポジトリで公開されているので、ダウンロードして適宜参照されたい。URL は次の通りである。

<https://doors.doshisha.ac.jp/duar/repository/ir/14588/031000870001.pdf>

文献 B は、書籍の性格上、多くの大学図書館に所蔵されているはずである²。

3 計量分析の初歩的理解

3.1 単純集計

文献 A の記述は次のような表の紹介からはじまる（表 3）。

表 3 単純集計表（文献 A の表 1）

		そう思う	どちらかといえばそう思う	どちらともいえない	どちらかといえそう思わない	そう思わない	合計	度数
他者信頼	たいていの人は信用できる	5.2	25.0	38.6	15.6	15.5	100.0	2711
	機会があれば、たいていの人は自分のために他人を利用する	7.4	24.3	41.3	14.7	12.4	100.0	2625
	たいていの人は自分のことだけを考えている	16.0	40.4	28.2	9.3	6.1	100.0	2706
弱者配慮	チャンスが平等にあたえられるなら、競争で貧富の差がついてもかまわない	25.0	48.5	—	15.5	11.0	100.0	2617
	今後、日本で格差が広がってもかまわない	2.0	4.7	19.7	28.2	45.4	100.0	2694

² 本稿にかかわることではないが、文献 B の表 1 の「4 次元イデオロギー図式」には重大な誤植がある。文章を読めばどこに間違いがあるのかはわかると思うが、気に

なる読者は小林（2002）や小林（2010）の記述が正しいので参照されたい（誤植は文献 B の図 5 にもある）。

これは5つの質問（すなわち変数）についての回答（すなわち値）の分布を示したものであり、5つの単純集計表をまとめた表ということになる。度数の欄にはケース数（調査対象の数）が示されている。ケース数は「度数」「実数」「N」「n」と表記されることもある。この表の読み方については解説の必要はなからう。

3.2 尺度の構成と主成分分析

(1) 尺度の構成とは

この表をもとに、文献Aでは他者信頼得点と弱者配慮得点という新たな指標が構成される。

複数の変数を合成して新たな指標を構成することを尺度構成という。尺度の構成を行うのは信頼性を高めるためだ。数学のテストをするとき、1問だけからなるテストより、多くの問いのあるテストの方がその人の数学の能力を正確に（＝高い信頼性で）測れる。それと同じで、複数の問いから知りたいものを測るほうが信頼性が高くなるのである³。

もっとも単純な尺度構成の方法は、関連する諸変数の得点を足すという方法だ。すなわち、弱者配慮得点でいうと、「チャンスが平等ににあたえられるなら、競争で貧富の差がついてもしかたがない」「今後、日本で格差が広がってもかまわない」の賛成から反対までに1点から5点を与えてそれを足し、その合計得点を弱者配慮得点にするのだ。

(2) 主成分分析

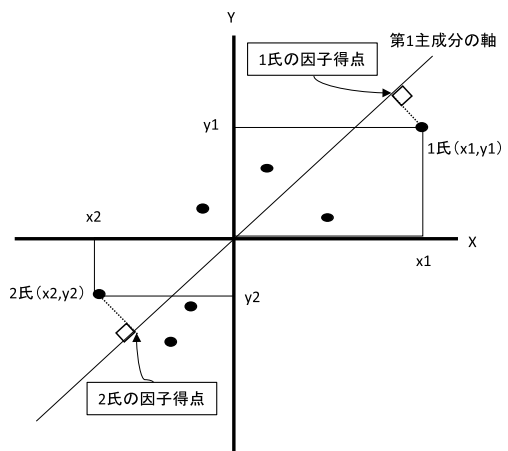
単純に足すというのはわかりやすくいい方法なのだが、文献Aでは「主成分分析」を使った別の方法が使われている。「弱者配慮得点」について

の記述は次の通りだ。

「弱者への配慮に対応する2つの質問「チャンスが平等にあたえられるなら、競争で貧富の差がついてもしかたがない」「今後、日本で格差が広がってもかまわない」を主成分分析にかけ、第1主成分の因子得点を弱者配慮スコアとする（寄与率は63.73%）。

ここで行われていることを図で示したものが図2である。すなわち、X「チャンスが…」とY「今後、日本で…」という2軸（ともに標準得点化）の代わりになる第1主成分の軸を数学的に導き出し、その軸を物差しとしてその軸の中心からの（符号付きの）距離で各ケースの得点（因子得点・主成分得点とも言う）を算出しているのである。この得点は普通、平均0、標準偏差1になるように標準得点化される。

図2 主成分軸と因子得点（主成分得点）



っているかということは信頼性の問題である。

³ ある指標で対象を測定するとき、測りたいものを測っているのかということは妥当性の問題であり、正確に測

このような手続きの結果、各ケースについて新たに弱者配慮得点が算出され、データは表 4 のようになる。最右列の弱者配慮得点 (h) は平均 0、標準偏差 1 の新しく構成された変数である。そしてこの新変数について、他の変数の値ごとに (たとえば男女別で) 平均を求めたり、別の変数との相関係数を求めたりすることができる。

表 4 各ケースの弱者配慮得点

ケース番号	チャンスが平等にあたえ…	今後、日本で格差が広がって…	弱者配慮得点 (因子得点)
1	x_1	y_1	h_1
2	x_2	y_2	h_2
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
n	x_n	y_n	h_n
平均	\bar{x}	\bar{y}	0
標準偏差	s_x	s_y	1

上の記述の中には「寄与率」というものがあった。これは、もともと 2 軸で測られていた各データの点をこの 1 本の軸で測っていいかどうかを示す値と考えればいい。x と y の相関が大きければこの値は 1 (すなわち 100%) に近づく。値が小さすぎると、1 本の軸でデータの位置を測るのは、少し無理があるということになる⁴。

「他者信頼得点」については、もとの変数が「たいていの人は信用できる」「機会があれば、たいていの人は自分のために他人を利用する」「たいていの人は自分のことだけを考えている」と 3 つある。この場合は 3 次元の空間に存在するデータ各点の

中を突き通すような軸が第 1 主成分の軸になる⁵。

文献 A ではこれら以外に「経済地位得点」ならびに「社会関係得点」という指標も構成されている。経済地位得点は世帯収入と財産を用いて上と同じ主成分分析を用いた手続きで作られている。

社会関係得点は「過去 1 年間、家にどのような人を招いているか」という質問への回答 (親せきの人、職場や仕事関係の人、近所の人、学校時代の友人、同じサークルや団体に加入している人、その他の友人や知人からの複数選択) の○の数である。

3.3 平均の分析 1 (1 要因分散分析)

(1) 抽出された表

文献 A の表 2 は、性別、年齢、学歴、世帯収入、職業のそれぞれの値 (男女等) ごとに、他者信頼得点と弱者配慮得点の平均をまとめたものだ。したがって、その表は 12 個の表をまとめた 1 つの大きな表である。この中から学歴と弱者配慮の関係、ならびに学歴と他者信頼の関係を示す表を選び、それらをもとに解説しよう (表 5)。

表 5 平均の表 (文献 A 表 2 の抜粋)

	学歴	平均	度数	標準偏差	p
弱者配慮	中学	0.080	(378)	0.996	0.000
	高校	0.083	(1425)	1.000	
	大学	-0.198	(743)	0.974	
他者信頼	中学	-0.092	(411)	1.043	0.086
	高校	0.003	(1436)	0.991	
	大学	0.044	(738)	0.991	

⁴ 第 1 主成分の寄与率のとりうる範囲は「1/変数数」～1 である。

⁵ データが (縦・横・高さ 45 度に) 傾けたラグビーボール (の内部) のように分布しており、ボールの長軸が第

1 主成分軸になるとイメージするといいい。3 変数の相関が小さければボールは球形に近づき第 1 主成分の寄与率は小さくなる (最小値は 1/3)。

(2) 検定

1) 学歴→弱者配慮

表 5 の平均の列より、学歴によって弱者配慮得点の平均値が異なることがわかる。中学卒・高校卒に比べて大学卒では弱者配慮得点の平均が低くなっているのである。

しかし、これは標本の話である。われわれは母集団でも差があると言えるかどうかを知りたいのである（差があると言いたい）。そのためには「1 要因分散分析」を行う必要がある⁶。

まず次のような帰無仮説と対立仮説を立てる。

[帰無仮説] 母集団において中学卒、高校卒、大学卒の弱者配慮得点の平均に差はない。

[対立仮説] 母集団において中学卒、高校卒、大学卒の弱者配慮得点の平均に差がある⁷。

言いたいのは対立仮説の方だ⁸。

表の p (有意確率) の欄には「帰無仮説が正しい場合に、今回の標本のような平均値の差が出る確率は p 以下である」ということが示されている。 p 値が小さければ、帰無仮説を棄却 (否定) して対立仮説を採択する。どのくらい小さければ帰無仮説を棄却して対立仮説を採択するかという基準を有意水準といい、通常 0.05 以下という基準が設定される (これを有意水準 5% と表現する)。

表よりこの値は 0.000 となっている⁹。だから、母集団で弱者配慮得点の平均に差がないとき、標

本でこのような平均の差が出る確率は 0.000 以下だ。したがって次のような結論を導ける。

「有意水準 5% で帰無仮説は棄却され、母集団において中学卒、高校卒、大学卒の弱者配慮得点の平均に差があると言える」。

2) 学歴→他者信頼

次に他者への信頼について。表 5 の平均の列より、学歴によって他者信頼得点の平均値が異なることがわかる。すなわち、中学卒、高校卒、大学卒となっていくにしたがって、他者信頼得点の平均は高くなっていくのである。しかし、これは標本の話である。われわれは母集団でも差があると言えるかどうかを知りたいのである。

ここで帰無仮説と対立仮説を立てる。

[帰無仮説] 母集団において中学卒、高校卒、大学卒の他者信頼得点の平均に差はない。

[対立仮説] 母集団において中学卒、高校卒、大学卒の他者信頼得点の平均に差がある。

ここで p 値を見る。表でのこの値は 0.086 となっている。だから、母集団で他者信頼得点平均に差がないとき、標本でこのような平均の差が出る確率は 0.086 以下だ。この値は有意水準の 5% より大きいので、次のような結論になる

⁶ 「一元配置分散分析」ともいう。

⁷ 「差がある」は「全部異なる」という意味ではなく、「3つのうち少なくとも1つが異なる」の意味。

⁸ 検定は通常、帰無仮説を否定して対立仮説を採択するという方法をとる。対立仮説を直接検討せずこのような迂回的方法をとるのは、対立仮説を直接検討することが難しいからである。なぜ難しいかというと、学歴によ

って平均に差がないということは、中学卒平均=高校卒平均=大学卒平均という等式で表されるが、差があるということはこのような単純な等式では表せず、そこから確率を明らかにすることは困難だからだ。簡単に言うと「差がある」といってもいろいろな差があり、これを一挙に検定することは不可能なのだ。

⁹ 0.000 は 0 ではなくこの桁までは 0 であるということ。

「有意水準 5%で帰無仮説は棄却されず、母集団において中学卒、高校卒、大学卒の他者信頼得点の平均に差があるとは言えない」¹⁰。

しかし、次のように有意水準を変えれば対立仮説が採択できる。

「有意水準 10%で帰無仮説は棄却され、母集団において中学卒、高校卒、大学卒の他者信頼得点の平均に差があると言える」。

簡単に言えば、間違いを許容する程度を上げれば、言いたい対立仮説を主張できるようになるということである。有意水準（または有意確率）は必ず論文に記載されている。読者はそれを見て結論がどの程度信用できるかを判断することができるのである。

3.4 重回帰分析

(1) 重回帰分析とは

独立変数 x_1, x_2, \dots, x_m から従属変数 y の値を予測するような 1 次式を立てて分析する方法を重回帰分析といい、独立変数が 1 つの場合は単回帰分析、複数の場合は重回帰分析という。

独立変数が 1 つの場合、予測式は次のように直線を表す式になる。

$$y = a + b_1x_1$$

独立変数が 2 つの場合は、平面を表す式となる。

$$y = a + b_1x_1 + b_2x_2$$

独立変数が m 個ある場合は、次の式になる¹¹。

$$y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$$

(2) 重回帰分析の結果を示す表

弱者への配慮を従属変数とする重回帰分析の結果は次のような表で示される（表 6）。表の「非標準化係数」と「標準化係数」の列にはこの重回帰分析の基本的な情報が示されている。

表 6 弱者への配慮の重回帰分析（文献 A 表 4）

	非標準化係数	標準誤差	標準化係数	t	p
(定数)	0.577	0.202		2.849	0.004
性別（女0）	-0.362	0.059	-0.180	-6.147	0.000
年齢	0.004	0.003	0.052	1.748	0.081
教育年数	-0.027	0.012	-0.077	-2.329	0.020
世帯収入	0.000	0.000	-0.086	-2.915	0.004
専門	-0.169	0.093	-0.063	-1.818	0.069
管理	-0.135	0.113	-0.038	-1.193	0.233
ホワイト	-0.199	0.071	-0.093	-2.788	0.005
農業	-0.057	0.130	-0.013	-0.441	0.659
N	1196				
R2 乗	0.070				
p	0.000				

職業ダミー基礎：ブルー

(3) 非標準化係数をもとにした回帰式

1) 回帰式

非標準化係数（これを（偏）回帰係数という）の列で述べられていることは、次の回帰式でそれぞれのケースの弱者配慮得点が予測できるということである。

¹⁰ ここで「差がない」（帰無仮説の採択）ではなく「差があるとは言えない」（対立仮説の非採択）と述べていることに注意してほしい。一般的な検定において、「帰無仮説が棄却できない」ときに言えるのは対立仮説の非採択であり、帰無仮説の採択ではない。有意確率は「帰無仮説が正しい場合に当該結果が生じる確率」であり、帰無仮

説自体の成立・不成立の確率とは関係がない。それゆえ、有意確率をもとに、帰無仮説の採択についての議論はできないのである。採択・非採択という用語は、通常、対立仮説のみに適用される用語である。

¹¹ 定数項 a や係数 $b_1 \sim b_m$ はデータをもとに「最小二乗法」とよばれる手法で数学的に導き出される。

$$\begin{aligned}
 y(\text{弱者配慮得点}) &= 0.577 \\
 &- 0.362x_1(\text{性別}) + 0.004x_2(\text{年齢}) \\
 &- 0.027x_3(\text{教育年数}) - 0.000x_4(\text{世帯収入}) \\
 &- 0.169x_5(\text{専門}) - 0.135x_6(\text{管理}) \\
 &- 0.199x_7(\text{ホワイト}) - 0.057x_8(\text{農業})
 \end{aligned}$$

たとえば教育年数が1年のびると、(ほかの変数に変化がないとき)弱者配慮の得点は0.027下がることになる¹²。

2) ダミー変数

重回帰分析は、基本的には複数の量的な独立変数から量的な従属変数を予測するものだ。しかし、上の回帰式には性別と職業という質的な変数が含まれている。これは、「ダミー変数」を使って質的な変数を回帰式に組み込んでいるのである。表の中のダミー変数は、「性別」、ならびに「職業」(専門・管理・ホワイト・農業)に対応している。

ダミー変数は「変数の値の数-1個」あればいい。性別という変数は男と女の2個の値をとるので、1個のダミー変数 x_1 (男1、女0)を作る。これで性別ダミーの基礎(女)に比べ、男がどの程度弱者配慮得点を高める(低める)効果を持つかがわかる。表を見ると、男であることは女であることに比べ、弱者配慮得点を引き下げる効果をもつことがわかる(-0.362)。

職業は5カテゴリー(専門・管理・ホワイト・農業・ブルー)あったので、職業のダミー変数は4つになる(x_5, x_6, x_7, x_8)。これらのダミー変数は、それぞれ専門・管理・ホワイト・農業を表し、ブルーを基礎としている。

その人の職業が専門ならば、その人のダミー変

数(x_5, x_6, x_7, x_8)の値は(1,0,0,0)だ。その人の職業が管理ならば、その人のダミー変数の値は(0,1,0,0)だ。その人の職業がホワイトならば、その人のダミー変数の値は(0,0,1,0)だ。その人の職業が農業ならば、その人のダミー変数の値は(0,0,0,1)だ。そして最後に、その人の職業がブルーならば、その人のダミー変数の値は(0,0,0,0)である。

これで、職業ダミーの基礎(ブルー)に比べ、それぞれの職業がどの程度弱者配慮得点を高める(低める)効果を持つかがわかる。表を見ると、ここにある4つの職業はすべてブルーに比べ、弱者配慮得点を引き下げる効果をもつことがわかる。特に、ホワイトであることはブルーに比べ、弱者配慮得点を引き下げる効果が大きい(-0.199)。

(4) 標準化係数をもとにした回帰式

各変数のデータを標準得点にし、それで回帰分析を行った結果が「標準化係数」の列に示されている(これを標準(偏)回帰係数という)。

次の式は、標準化係数による回帰式である。標準化係数の回帰式では定数項はなくなる。

$$\begin{aligned}
 y(\text{弱者配慮得点}) & \\
 &= -0.180x_1(\text{性別}) + 0.052x_2(\text{年齢}) \\
 &- 0.077x_3(\text{教育年数}) - 0.086x_4(\text{世帯収入}) \\
 &- 0.063x_5(\text{専門}) - 0.038x_6(\text{管理}) \\
 &- 0.093x_7(\text{ホワイト}) - 0.013x_8(\text{農業})
 \end{aligned}$$

単位が異なる変数をもとにした非標準化係数では、従属変数への影響力がどの独立変数で大きいかはわからない。たとえば、万円で測られている

¹² 偏回帰係数は他の独立変数の影響を除いた当該変数に

よる従属変数への効果を示すものである。

世帯収入は数百という大きな値をとるが、教育年数は大学卒でも16年であり、値の桁数が違う。この違いは非標準化係数の値に反映して、世帯年収の係数(0.000)は教育年数の係数(-0.027)よりずっと(絶対値で)小さくなってしまっている。

影響力の大きさを判断するためには標準得点化された変数をもとにした標準化係数を見る必要がある。標準化係数を見ると、世帯収入(-0.086)の方が教育年数(-0.077)よりも弱者への配慮を引き下げる効果が大きいことがわかる。

(5) 回帰係数の検定

表6の「標準誤差」「t」「p」の列の情報は母集団と標本の関係に関わり、検定や推定に利用される。ここでは、pについてのみ解説する。

表での多くの回帰係数は0ではないが、これは標本での話である。われわれは、母集団での回帰係数(これを母回帰係数という)が0でないということが知りたいのである¹³。

ここでまず、それぞれの回帰係数について、帰無仮説と対立仮説を立てる。

[帰無仮説] 母回帰係数は0。

[対立仮説] 母回帰係数は0でない。

帰無仮説は母集団で当該変数は従属変数に何の影響も与えていないということを意味している。われわれが言いたいのは対立仮説の方だ¹⁴。

そこで表の一番右の列のpを見る。この列には、「帰無仮説が正しい場合に、今回の標本のような

回帰係数の値が出る確率がp以下である」ということが示されている。p値が小さければ、帰無仮説を棄却して対立仮説を採択する(通常0.05以下でそうする)。帰無仮説の棄却(=対立仮説の採択)によって、当該変数が母集団で弱者配慮得点に影響していると言える。

表6をもとに有意水準5%で判定すると、性別、教育年数、世帯収入、職業が母集団でも弱者への配慮に効果をもつということになる。

職業については、基礎のブルーに比べホワイトが弱者への配慮を低める効果があると言えることから(p=0.005)、職業という変数が弱者への配慮に影響すると言えることになる。年齢は有意水準5%では効果をもつとは言えないが、有意水準10%では効果をもつと言える。

(6) 回帰モデルの検討

とりあげられた諸変数を独立変数とする回帰式が何らかの説明力を持つのかどうかの検討のことを回帰モデルの検討という。表にあるR2乗は「決定係数」と呼ばれるもので、これらの独立変数からなるモデルで弱者配慮得点がどの程度説明できるかを示すものだ。まったく説明できなければ0、完全に説明できれば1になる。表からはこの回帰モデルで弱者への配慮が説明できる割合は7%であることがわかる¹⁵。

しかし、このことは標本での話だ。母集団でのこのモデルに説明力はないかもしれない。母集団でこのモデルが全く説明力をもっていないこととすべての母回帰係数が0であることは同等なので

¹³ 回帰係数が0であるということは、その変数が従属変数に何の効果ももっていないということだ。

¹⁴ この検定は非標準化係数の検定でありかつ標準化係数の検定でもある。

¹⁵ この値は決して大きいと言えないが、全く説明できていないわけではない(0%ではない)。しかし、文献Aの著者としては少々なさけないというのが本当のところだ。

次のような帰無仮説と対立仮説を立てる。

[帰無仮説] すべての母回帰係数=0。

[対立仮説] 「すべての母回帰係数=0」というわけではない。

言いたいのは対立仮説の方だ。

結果は表の一番下の p のところに示されており、この値が小さければ、帰無仮説を棄却して対立仮説を採択することになる（通常 0.05 以下でそうする）。表 6 の p は 0.000 なので、帰無仮説は棄却され、この回帰モデルが母集団でも説明力をもつという結論になる。

3.5 相関係数と偏相関係数

(1) 相関係数とその検定

文献 A には経済的地位ならびに社会関係が他者信頼や弱者配慮とどう関係するのかを相関係数と偏相関係数で示した表がある（表 7）。

相関係数を示す左の 2 列を見てみると、経済的地位が高ければ他者信頼は強まり、弱者配慮は弱まることがわかり、社会関係が広げれば他者信頼は強まり、弱者配慮も強まることがわかる。

表 7 相関と偏相関の表（文献 A 表 7）

		相関		偏相関	
		経済地位	社会関係	社会関係 制御	経済地位 制御
他者 信頼	p	0.085	0.071	0.070	0.077
	N	(1332)	(2586)	(1270)	(1270)
弱者 配慮	p	-0.142	0.027	-0.149	0.062
	N	(1308)	(2547)	(1270)	(1270)

しかし、これは標本での話である。それぞれの相関係数について母集団でも相関があると言える

のかどうかを見るためには、次の帰無仮説と対立仮説をもとにした相関係数の検定が必要である。

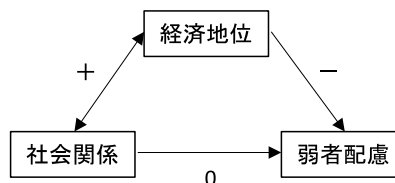
[帰無仮説] 母相関係数は 0。

[対立仮説] 母相関係数は 0 でない。

ここで p 値を見てみると他の相関は母集団でもあると言えるが、社会関係と弱者配慮の相関については $p=0.174$ になっており、「母集団で相関があるとは言えない」ということになる（表 7 の左の網掛け部分に注目）。

経済地位と社会関係の間に相関があることを考慮すると（文献 A には $r=0.155$, $p=0.000$ との記載がある）、母集団で見られる関係は図 3 のようなものだと考えられるのである。

図 3 社会関係・経済地位・弱者配慮の関係



(2) 偏相関係数とその検定

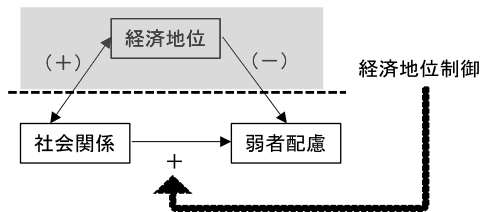
しかし、ここで次のような疑問がわく。すなわち、社会関係の広い人は経済地位も高い。それゆえ社会関係と弱者配慮の相関係数を出すだけでは社会関係が本来もつ弱者配慮への独自の効果はわからないのではないか、という疑問である。

表 7 の右にある偏相関の欄は、こういった疑問を解決するために利用できる。偏相関係数は、ある変数の影響を除いた 2 変数間の相関を表すものだからである。

表 7 の右の網掛け部分を見ると偏相関は 0.062

であり、通常の相関 (0.027) より大きくなっていることがわかる。また、p 値も 0.027 と 5% 以下である。したがって、経済地位の影響を取り除くと、社会関係が広いほど弱者への配慮は強まるということが母集団でも言えるという結論になる (図 4)。

図 4 社会関係と弱者配慮の関係 (経済地位制御)



3.6 平均の分析 2 (2 要因分散分析)

(1) 2 変数から見た平均

文献 A には経済階層と社会関係を組み合わせて他者信頼得点と弱者配慮得点の平均がどうなるかを示した表がある。表 8 はそこから弱者配慮得点の部分だけを抜き出したものである。

この表より、経済階層が高いほど弱者への配慮は低くなることわかる (「 -0.232 、 -0.023 、 0.094 」の比較)。また、社会関係が広いほど弱者への配慮は高まることもわかる (「 -0.016 、 -0.100 」の比較)。そして、どの経済階層においても社会関係の広さは弱者への配慮を高める効果をもつこともわかる (「 -0.192 と -0.283 」、「 0.052 と -0.119 」、「 0.146 と 0.058 」それぞれの比較)。

表 8 平均の表 (文献 A 表 8 抜粋)

経済階層	社会関係	弱者配慮		
		平均	N	標準偏差
上	広	-0.192	(255)	0.944
	狭	-0.283	(195)	1.006
	合計	-0.232	(450)	0.971
中	広	0.052	(246)	0.938
	狭	-0.119	(191)	1.029
	合計	-0.023	(437)	0.982
下	広	0.146	(173)	0.942
	狭	0.058	(248)	1.003
	合計	0.094	(421)	0.978
合計	広	-0.016	(674)	0.951
	狭	-0.100	(634)	1.020
	合計	-0.057	(1308)	0.986

以上のことは標本で言えることだ。母集団でも言えるかどうかを検討するためには「2 要因分散分析」を行う必要がある¹⁶。

(2) 分散分析表

その「2 要因分散分析」の結果を示したものが表 9 である¹⁷。

表 9 分散分析表 (文献 A 表 9 抜粋)

		平方和	自由度	平均平方	F 値	有意確率
弱者 配慮	主効果 (結合)	27.256	3	9.085	9.533	0.000
	経済階層	24.975	2	12.487	13.103	0.000
	社会関係	4.328	1	4.328	4.541	0.033
	交互作用	0.475	2	0.238	0.249	0.779
	モデル	28.710	5	5.742	6.025	0.000
	残差	1240.793	1302	0.953		
合計		1269.503	1307	0.971		

この表の経済階層の行を見てみよう。そこには「経済階層上の母平均 = 経済階層中の母平均 = 経

¹⁶ 「二元配置分散分析」ともいう。

¹⁷ 分散分析を少し勉強したことのある読者の中にはこの表の平方和の計算が合わないと思う者がいるかもしれない

い。この 2 要因分散分析はアンバランスデザインによるもので、平方和は Type III で算出されている。それでこのような結果になるのである。

済階層下の母平均」という帰無仮説の下での検定結果が示されている。有意確率は 0.000 となっているので有意水準 5% で (0.1% でも) 帰無仮説は棄却される。したがって、母集団においても経済階層が弱者への配慮に影響するということになる。

その下の社会関係の行には「社会関係広の母平均＝社会関係狭の母平均」という帰無仮説の下での検定結果が示されている。有意確率は 0.033 なので、有意水準 5% で帰無仮説は棄却される。したがって、母集団でも社会関係が弱者への配慮に影響することがわかる¹⁸。

3.7 クロス表の分析 (カイ二乗検定)

これまで文献 A を用いて計量分析について初歩的な解説をしてきた。しかし、社会的な計量分析において最もよく利用されるクロス表の分析はここにはなかった。そこで、文献 B の中から 1 つ表を選んで解説することにする。

文献 B には投票行動と人々の社会的属性の関係について記載したかなり大きな表が掲載されている (文献 B の表 4)。ここから、人々の現住地への居住年数と前の衆議院議員選挙での投票・棄権を示したものが表 10 である。

表を見ると居住年数と投票行動の間にはかなりきちんとした関係があることがわかる。すなわち、居住年数が長くなるほど投票に行くようになるのである。

表 10 居住年数と投票 (文献 B 表 4 抜粋)

居住年数	投票	棄権	計	実数	有意確率
3年未満	72.1	27.9	100.0	(104)	0.000
3年以上	76.7	23.3	100.0	(279)	
10年以上	81.1	18.9	100.0	(514)	
20年以上	88.1	11.9	100.0	(1400)	

ただしこれは標本での話だ。母集団でもそんなことが言えるのかどうかを検討するためにはいわゆる「カイ二乗検定」(χ^2 検定) をする必要がある。この検定における帰無仮説と対立仮説は次のものだ。

[帰無仮説] 母集団において居住年数のカテゴリーと投票・棄権のカテゴリーは独立。

[対立仮説] 母集団において居住年数のカテゴリーと投票・棄権というカテゴリーは独立ではない。

検定の結果は「有意確率」のところに示されており、p 値は 0.000 となっている。このことは母集団で居住年数と投票・棄権に関係がない場合、このような関係のある標本が生じる確率は 0.000 以下であるということを示している。

したがって、有意水準 5% で帰無仮説は棄却され、母集団でも居住年数と投票するか棄権するかということには関係があると言える (有意水準 1% でも 0.1% でも言える)。

¹⁸ 経済階層と社会関係の行の下には「交互作用」と書かれた行がある。これは経済階層と社会関係が相伴って弱者配慮に独自の効果をもつかどうかを調べるためのものだ。いわば「食い合わせ」の効果である。ここでの帰無仮説は「母集団における社会関係の広狭による弱者配慮得点の平均の差は経済階層の上中下で異なる」とい

うものだ。検定によると有意確率は 0.779 となっているので、母集団で経済階層と社会関係の交互作用はあるとは言えないということになる。これが棄却されなかったことは「どの経済階層においても社会関係の広さは同じように弱者への配慮を強める可能性が高い」ということを暗示している。

3.8 まとめ

以上、いくつかの基本的な計量分析の手法について述べてきた。ここで紹介した平均についての分析（分散分析）、クロス表の分析、相関や偏相関の分析、重回帰分析、主成分分析などは計量分析を伴う社会学の多くの論文で見られる分析法である。また、母集団と標本の関係をもとにした検定も普通になされている。本稿でこれらについての初歩的知識を得た後、各自それぞれの分析法についてさらに理解を深め、最終的には自身で計量分析を行えるようになってもらいたい。

ただ同時に、こういった分析法は技法にすぎず、それを超えるものではないということも記憶しておいてほしい。より重要なのは、分析のテーマであり、なぜそれを分析するのかという自身の問題関心である。本稿で使った文献 A や文献 B が技法を学ぶためのものとしてだけでなく、そういったことを考える際にも役立つならば、筆者としてはとてもうれしい。

4 おわりに

ここまで、社会学の計量分析論文における記述の意味や表の読み方を学生に解説する資料を紹介してきた。読者の中にはこれでは簡単すぎるといった印象を持つ方もいると思うし、逆に難しすぎて使えないと思う方もいると思う。

筆者も最適な答えには到達していないが、現在のところ次のように思っている。(1) 学生が社会学の基礎的な計量論文を読めるようになるにはこの程度の範囲の基礎知識は必要である。そして、この範囲の基礎知識があれば、学生は自分で知識を深めたりさらに範囲を広げたりしていくことができる。(2) 「理解できるように」という思いで内容をやさしくしすぎると、学生は全体についてぼや

っとした印象しか持たず、「そうなるのはなぜだろう」という明確な疑問を持たない。以後の勉強につながるのは「どこがわからないのかわからない」という不明確な疑問ではなく「ここがわからない」という明確な疑問だ。それゆえ解説にはある程度の詳しさが必要である。(3) 2～3 回の授業で完了という時間設定のもとでは、この程度の範囲と詳しきでの解説が限界である。「あれもこれも」という思いや、「ここはもう少し詳しく」という思いは出てくるが、そこはぐっとこらえる必要がある。そうしないと「〇〇社会学」の講義は「データ解析法」の講義になってしまう。

本稿で取り上げたような問題に関して、読者の中にはよりよいアイデアをお持ちの方も少なくないと思う。参考にしてバージョン・アップしたいので、学生・院生・教員を問わずそんなアイデアをお教えいただければありがたい。どうぞよろしくお願いします。

文献

- 小林久高, 2000 「政治意識と政治参加の動態」 間場寿一編『講座社会学 第9巻 政治』東京大学出版会, 43-88.
- , 2002 「漂流する政治意識」 原純輔編『講座社会学 変動 第5巻 流動化と社会格差』ミネルヴァ書房, 233-265.
- , 2008 「共同性の精神的基盤と社会階層」『評論・社会科学』同志社大学社会学会, 87:1-28.
- , 2010 「投票行動と政治意識」日本社会学会社会学事典刊行委員会編『社会学事典』丸善, 852-853.
- 安田三郎, 1970 『社会調査の計画と解析』東京大学出版会.

