

SPSSによる重回帰分析

村瀬 洋一

1. 重回帰分析とは何か

1.1. 目的と具体例

1) 重回帰分析の目的 — 説明変数Xを複数設定し、被説明変数Yとの関連が強いのが、どの変数なのかを解明すること(村瀬他、2007、pp161-)。

相関係数や2重クロス集計のように、表面的な2変数の関連を見るだけでなく、他の変数の影響を取り除いた後(統制後、コントロール後)の関連を解明するのが目的。

線型の関連(回帰直線)を前提に、関連の強さを測る(p.163散布図を参照)。

説明変数、被説明変数とも連続変数(量的変数)の場合に用いる。説明変数が離散変数(カテゴリー変数)の場合は、分散分析を用いる。なお、両方とも離散変数の場合は、クロス集計やログリニア分析を用いることになる。

なお、Yが2段階回答の場合は、ロジスティック回帰分析(あるいは判別分析、数量化2類など)という別の分析法を用いる。社会調査の4段階回答などは、厳密には連続変数ではないが、3段階以上ならば、連続変数とみなして用いることが多い。データ人数が多い場合(数百人以上)は、このような考え方で、とくに問題はない。

重回帰分析は、あらゆる多変量解析法の基本となるものである。多くの分析法は、Yを一つと、複数のXを設定して、Yへの有意な効果があるかどうかを検討する。その意味で大半の分析法は重回帰分析の発展版である。Yが2値変数の時はロジスティック回帰かロジット分析、Yがある事象が発生する確率の時は生存関数分析(イベントヒストリー分析)ということになるが、基本的な考え方は同じである。重回帰分析は、とても分かりやすい分析法だが、以下の多重共線性に十分に注意して行うことが重要である。

2) 具体例

本人体重と父親体重に関する数百人分のデータがあった場合、それを散布図にして、直線を当てはめ、関連を検討することができる。あるいは、例えば社会調査で「生活全般満足度」に関する問の、4段階回答があった場合に、これをYとして、この原因となっているXを解明することが、重回帰分析の目的となる。変数Yの原因となっている変数は何かを複数考え、複数のXを用いて重回帰を行うことになる。

社会調査の場合、Xとして年齢、教育年数、財産や収入などの属性変数や、自営業かどうかなどの01のダミー変数、その他の心理的変数(態度や意識)を用いることが多い。もし、他の変数の影響を取り除いても(コントロール後でも)、年齢がYと関連していた場合、年齢がYの原因となっている、と考えることができる。実際、高齢の人ほど保守的価値観を持っているため、関連が出ることもある。しかし、高齢の人は低学歴な傾向があるため、表面的にはYと学歴も関連があるように見える。そのような表面的関連でなく、他の変数の影響を取り除いた後の、真の関連を見つけることが分析の目的である。

分析結果として、まず偏回帰係数をいくつか出し、どれが大きいかを確認するとよい。

1.2. 重回帰分析の考え方 (ポーンシュテット・ノーキ, 1990, 第8、第11章などを参照)

説明変数が2つの場合の重回帰分析のモデル

$$\text{標本回帰式 } Y_i = a + b_1 X_{1i} + b_2 X_{2i} + e_i \quad \dots \dots \dots (1)$$

$$\text{標本予測式 } \hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i} \quad \dots \dots \dots (2)$$

Y_i : i番目の個体の、被説明変数Yの値

\hat{Y}_i : i番目の個体の、被説明変数Yの予測値

X_1 、 X_2 : i番目の個体の、説明変数 X_1 、 X_2 の値

a : 切片

b : 偏回帰係数

e_i : 誤差項 (残差項)

$$e_i = Y_i - \hat{Y}_i \quad \leftarrow \text{重要!}$$

bは回帰直線の傾きの大きさを表している (村瀬他, 2007:123の図を参照)。

この数式は、以下の図1のようなモデルを表しているにすぎない。

1.3. 回帰分析のパラメーターの推定と解釈

a 、 b_1 、 b_2 の値の推定 → 最小自乗法 (Ordinary least squares OLS) を用いる。

$\sum e_i^2$ を (誤差の二乗の合計を) 最小にするように、 a 、 b_1 、 b_2 を推定する。

・ 偏回帰係数 (b_1 、 b_2)

他の変数の効果を統制した上で (統計的コントロールの後で)、説明変数が1単位変化した場合、被説明変数がどのくらい変化するかを示す。

・ 標準偏回帰係数 (β_1 、 β_2 ベータ係数、ベータ加重)

XとYを標準化した (Z得点にした) 上で求めた回帰係数

説明変数が1標準偏差増えた場合、被説明変数がどのくらい変化するかを示す。

ただし、上記の係数は、その説明変数に固有の値ではない。他の説明変数が変われば、当該の説明変数の係数値も変わる。(久米・飯塚、1987、p.99- を参照)

回帰係数は相関係数とは異なる。他の変数Xの影響を除いた場合の、Yとの関連の強さを表しているのである。

1.4. 決定係数 (R^2)

説明変数Xが、被説明変数Yの分散をどのくらい説明しているかを示す。モデル全体 (回帰式全体) の説明力を表す。レンジは0~1。モデルで分散を完全に説明しているときは1になる。

$$\begin{aligned} R^2 &= \frac{\sum [(Y_i - \bar{Y}_i)^2 - (Y_i - \hat{Y}_i)^2]}{\sum (Y_i - \bar{Y}_i)^2} = \frac{SS_{TOTAL} - SS_{ERROR}}{SS_{TOTAL}} \\ &= \frac{SS_{REGRESSION}}{SS_{TOTAL}} \\ &= \frac{\text{モデルで説明できる分散}}{\text{全分散}} \quad (3) \end{aligned}$$

$$\begin{aligned} & \text{全体平方和 } S S_{\text{TOTAL}} \\ & = \text{回帰平方和 } S S_{\text{REGRESSION}} + \text{残差平方和 } S S_{\text{ERROR}} \dots (4) \end{aligned}$$

(平均値と観測値の距離) = (回帰モデルで説明できる距離) + (観測値と予測値のずれ)

数式の各項が、村瀬他(2007:125)の図ではどの部分になるか、理解すること。図のどの部分が回帰部分か、書き込んでみると良い。

なお、説明変数が2つの場合、標準偏回帰係数と決定係数の間には以下の関係が成り立つ。

$$R^2 = \beta_1^* r_{YX1} + \beta_2^* r_{YX2} \dots (5)$$

r_{YX1} : Yと X_1 の相関

r_{YX2} : Yと X_2 の相関

1.5. 決定係数の有意性検定

R^2 の有意性検定は、F検定によって行う。

$$F = \frac{MS_{\text{REGRESSION}}}{MS_{\text{ERROR}}} = \frac{\text{モデルで説明できる分散}}{\text{モデルで説明できない分散}} \dots (6)$$

$$\text{回帰平均平方 } MS_{\text{REGRESSION}} = \frac{SS_{\text{REGRESSION}}}{\text{自由度}}$$

自由度 : 説明変数の数

$$\text{残差平均平方 } MS_{\text{ERROR}} = \frac{SS_{\text{ERROR}}}{\text{自由度}}$$

自由度 : $N - 1 - (\text{説明変数の数})$

(6)式の意味をよく理解する。F値とはモデルで説明できる部分とできない部分の比を取っている。範囲は0から無限大である。0~1ではない。大きいほど、よく説明している。平均平方MSは、SSを自由度で割ったもの。は、自由度調整済み決定係数については村瀬他(2007:168)を参照。説明変数の数が多く、データ人数に近い場合は、決定係数が無意味に大きくなる。説明変数を増やせば必ず R^2 は大きくなる。しかしデータのケース数が数百人以上など十分にあるならば、通常の R^2 とほぼ同じ値になるので使わなくて良い。

1.6. 多重共線性(マルチコ)

重回帰分析はとても分かりやすく有効な分析法だが、説明変数X同士の相関が高い場合は、重回帰分析を行うことはできない。この点によく気をつけること。説明変数間の相関

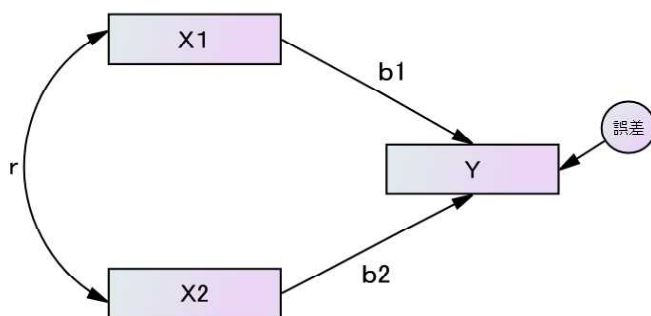


図1.基本的な重回帰分析モデル

(図のr)がとても高い場合、回帰モデルは非常に不安定になる。これは、説明変数の間にすでに別の線形回帰関係が含まれているということであり、その意味でこのような現象を「多重共線性 (multi colinearity)」と言う。経験的に、説明変数間の相関が0.7以上ならば危険であると言われている。

何らかの重回帰式を書き、そこに $X_1 = 2X_2$ という値を代入して計算してみるとよい。例えば、

$$Y_i = 2X_{i1} + 3X_{i2} \quad \text{に代入するとどうなるだろうか。}$$

多重共線性に注意するために、回帰分析を行う際には、まず分析前に、説明変数 X 間の相関行列を見て、相関がととても強いものがあれば片方は説明変数から除くことが必要。どのような説明変数の組み合わせがもっとも適しているかを明らかにするために、変数選択の方法が数種類ある。

1.7. 回帰分析を行う上での注意 (久米・飯塚, 1987:193-)などを参照)

説明変数が多ければ多いほど、決定係数 R^2 は必ず大きくなる。しかし、決定係数が大きい回帰式が良いモデルというわけでは、まったくない。

上記のように、説明変数の間に相関が強いと、回帰分析はできない。極端なことをいえば、説明変数が2つで、その間の相関が1ならば、2つの説明変数は同じものなのだから、どちらか1つを回帰モデルに入れれば良いのである。モデルはシンプルほど良い。

決定係数 R^2 がどれくらい大きければ良いと見なすのは難しい問題だが、社会調査データでは、測定誤差などのノイズが多いので、0.20を越えればとても大きいと言うことが多く、値が0.10以下でも、モデル全体のF値が有意ならば(統計的に意味があるならば)良い。心理学や経済学で用いるデータでは0.50以上になることが多いが、社会調査データでは、そのような例は少ない。

なお、重回帰分析を行うためには、ある程度のデータ数が必要である。通常、数百人以下のデータで重回帰分析を行うことはない。係数が有意かだけでなく、関連の大きさを知りたい場合は、少なくとも200人くらいは必要である。数百人以下の時は、自由度調整済み決定係数を用いた方がよい。

◆問 重回帰分析の説明変数 X は連続変数だが、離散変数も説明変数として用いることができる。どのようにすれば良いか。

答 if文を使って、0、1型の変数(ダミー変数)を作ればよい。
2カテゴリーの時 DMY (0,1) 例 男なら1、女なら0

3カテゴリーの時 QW (0,1) 例 ホワイトカラー、ブルーカラー、農業
QB (0,1) 3カテゴリーの場合、ダミー変数を2つ作ればよい。

3カテゴリーの時に、ダミー変数を3つ作らないように注意。
基準カテゴリーを一つ決めて、それ以外のものについてダミー変数を作る。基準カテゴリーはできる限り均質で分かりやすいものにすることが分析上のコツ。

1.8. ダミー変数を作るSPSSシンタックスの例

COMPUTE文やIF文を使って、ダミー変数を作ることができる。

書式 IF (条件式) 条件があった場合の命令文 ピリオド

例 職業の変数から、自営かどうかの0、1型変数を作成
COMPUTE SELFEMPL=0. ←新変数SELFEMPLを作成(変数名は好きな名前でもよい)
IF (Q33=6 OR Q33=7) SELFEMPL=1.
IF (Q33=1) SELFEMPL=1.

2. SPSSによる分析

2.1. SPSSの操作 (村瀬他, 2007: 174-179参照)

シンタックスを書くか、もしくはSPSSで、画面上の「分析」をクリックし、「回帰」→「線形」を選ぶ。

シンタックス例

```
REGRESSION                                ← (REGだけでも動く)
  /DEPENDENT Q6B
  /METHOD=ENTER   AGE Q2 Q3 Q4A .
```

解説

```
/DEPENDENT      この行にYになる変数名を書く
/METHOD=ENTER この行にXとなる変数名を複数書く
/MISSING LISTWISE 欠損値がある人のデータは取り除く (この行は省略可能)
```

分析実行前に必ず欠損値処理をすること。年齢や教育年数、自営ダミーなど、基本属性もXとして入れるとよい。変数は逆転するなど方向を事前にそろえる。

2.2. 作業手順

まず、被説明変数Yを1つ決め、さまざまなXを入れて自分の好きなモデルを考える。

はじめは数個の説明変数Xを入れ、少しずつ増やしてみると良い。ただし、最終的には、すべてのXをいれたモデルを検討すること。一部のXだけを入れた分析結果を、いろいろ出して表にしても、とくに意味はない。

YもXも量的変数しか使えないことに注意。まず、YとXすべてについて度数分布を見る。極端に偏っていたり、欠損値が多いものなどを確認する。例えば、5段階回答で、ほぼ全員が1と答えている場合は、分析はできない。統計の基本は分布を確認することである。

以下の注意点にもあるが、分析の前に必ず欠損値処理を行う。また、変数の方向をそろえる (回答方向を逆転した新変数を作るなどする)。

良いモデルを得るために多重共線性に注意せよ。まず、事前に説明変数間の相関行列を見てみること。

2.3. 結果のまとめと解釈

分析結果は、学術論文では以下のような形式の表にまとめる。図の方が一般向けには分かりやすい。各説明変数の偏回帰係数は有意か、モデル全体の説明力はどうか、なぜそのような結果が出たのかなどについて検討し、結果の解釈や考察を行うこと。

この例では3つのモデルについて表している。Yとの相関係数rは別途分析すること。

重回帰分析の結果 表のまとめ方の見本
(数字は架空例)

表2.1. 関係の資源保有の規定因に関する重回帰分析結果 1995年××調査男性

説明変数 () 内は変数のレンジ	地方議会議員		町内会役員		企業の経営者	
	β	r	β	r	β	r
年齢(20-69)	0.03	0.14**	0.03	0.11**	-0.01	-0.02
学歴(教育年数 6-17)	-0.01	0.00	-0.10	-0.07**	0.05	0.03
世帯資産(保有財産数 0-20)	0.15	0.14**	0.12	0.12**	0.14	0.14**
居住地域都市度(1-8)	-0.18	-0.12**	-0.08	-0.06**	-0.01	0.00
地域移動経験の有無(1,0)	-0.42**	-0.06**	-0.35**	-0.05**	0.16	0.00
組織内の役職(1-6)	0.13	0.05**	0.07	0.01	0.34**	0.12**
従業先企業規模(1-7)	-0.01	0.00	0.07	0.03*	-0.24**	-0.13**
本人職業威信スコア(26.7-83.5)	0.01	0.02	0.00	0.00	0.01	0.00
父職業威信スコア(23.4-87.3)	-0.02	0.00	-0.04	-0.02	-0.03	0.00
父学歴(教育年数 6-17)	0.00	0.00	0.00	0.00	0.00	0.00
本人職 自営ノンマニュアル(1,0)	0.68**	0.05**	0.72**	0.05**	0.15*	0.00
本人職 自営マニュアル(1,0)	0.29	0.00	0.36*	0.02	-0.08	0.00
本人職 農業(1,0)	0.42*	0.03*	0.85**	0.07**	-0.79**	-0.13**
R-square	0.27**		0.21*		0.16*	
Adjusted R-square	0.21		0.18		0.12	
N	381		324		356	

注 被説明変数は、××の場合4、××の場合0
縦1列が1つの回帰式を表し、点線は標準偏回帰係数と相関係数
説明変数のうち、レンジが(1,0)のものはダミー変数。職業ダミー変数の基準はその他の職業
** 1%水準で有意 * 5%水準で有意

注意点

表だけを見て、第3者が分かるのが大原則である。
表タイトルも的確に分かりやすく。表タイトルは表の上に書き、表番号をつける。
説明変数についての説明を、表の下に注で書く。
通常、縦1列が1本の重回帰式になる。この例では3本の重回帰分析の結果を1つの表にまとめている。
縦1列での小数点の位置をそろえる。
説明変数間の相関行列も、別途表にすると良い。
有効桁は2桁でよい。あまり細かい数字を書いても誤差を考えると意味がない。

SPSS出力をそのまま使ってはいけない。適切な形式の表に直すこと。出力をエクセルポートしてエクセル形式などで保存した後に、エクセルで読み込んで有効桁などを合わせ表にする。四捨五入したい範囲のセルをマウスで囲んで「書式」をクリックし、「セルの書式設定」→「ユーザー定義」を選択すると、数字を小数点以下2桁等に揃えることができる。エクセルで、そのように作った表を、ワード等にて、オブジェクトとして貼り付ける。ワードの画面上で「挿入」をクリックして、オブジェクトを選ぶ。

重回帰分析の結果 図のまとめ方の見本
(数字は架空例 Xが4つある場合の例)

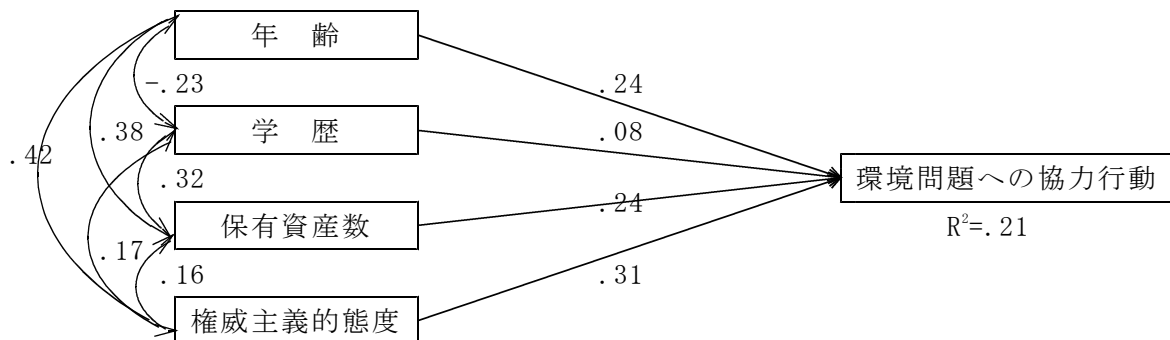


図2. 向環境行動の規定因に関する重回帰分析結果 1998年××調査
数字は標準化係数

図を書く際の注意点

因果関係の流れは、左から右へ、原因から結果となるようにする。

実在する変数（実在変数、観測変数）は四角で表現（因子は楕円で表現）。

決定係数 R^2 も必ず書く。

図のタイトルは図下に書く。数字の説明も忘れずに書く。

説明変数間の相関も書くこと（相関係数の分析で出力）。

なおMSワードでは、画面上の「挿入」をクリック、図形ボタンが表示、四角ボタンなどをおして図を書くが良い。塗りつぶし無しの四角などを使う。

2.4. SPSS出力の見方について

モデル全体の決定係数は R 二乗の値を、各変数の標準偏回帰係数（standardized estimate）はベータを見れば良い。各値の有意水準（有意確率、危険率）も見ること。

R 二乗の有意水準（モデル全体の有意水準）は、 F 値の有意水準を見ればよい。これが0.05未満ならば、 R 二乗が誤差である確率は5%未満なので、このモデルを採用して良い。

結果をまとめる時は、以下の3. 注意点に気をつける。

2.5. 変数選択

重回帰分析に説明変数を複数入れ、その後、どの変数を採用するのが適切かを検討することができる。このことを変数選択という。初めは、10個の説明変数でモデルを作り、その後、説明変数を5個くらいに絞るなどするとよい。

調査データの場合、とくに変数選択をせず、強制投入法としてすべての X を用いることも多い。変数選択法に確実なものはない。使うとしても参考程度であろう

・STEPWISE

既存のモデルをもとに、次に新しい変数を入れるか、あるいはモデルに既に入っている変数を落とすかを逐次的に行う。

・RSQUARE

候補となる説明変数のすべての組み合わせについて、回帰式と変数選択のために提案されている各種統計量を計算する。

3. 分析時の注意点

3.1. 分析の前に必ず欠損値処理をすること

多くの場合、欠損値は9か99。SPSSの場合、missing valuesコマンドを用いる。

回答が2桁の場合、欠損値99である。全変数の度数分布を見て確認するとよい。

3.2. 分析の前に変数の向きを必要に応じて逆転し、わかりやすく設定する

分析を行う前に、原則として、すべての変数を、数字が大きいほど肯定になるように直すこと。数字が小さいほど肯定となる変数が混ざっていると、とても分かりにくい。

シンタックスのデータ定義文の後で、以下のようなCOMPUTE文を書けばよい。

例 そう思う1 --- そう思わない4 → そう思う4 --- そう思わない1

Q4a を逆転し、新変数NEW4A (好きな名前をつける) を作る

```
COMPUTE NEW4A = 5-Q4a .
```

3.3. 用いる変数について

重回帰分析で用いる変数は、XもYもすべて連続変数(量的変数)であることに注意。名義尺度の変数はダミー変数以外は使えない。また、変数内でカテゴリー合併などをする必要はない。むしろ、なるべく回答の段階は細かい方が、連続変数に近くなるのでよい。

4段階尺度や順序など、厳密には連続変数ではないが、量的変数と見なして重回帰で使って問題はない。ただし被説明変数Yは、3段階以上が望ましい。

カテゴリー変数(質的変数、離散変数)をXとしたい時は、if文やrecode文を用いて、ダミー変数や連続変数に直すとよい。

例 問44の学歴を、教育年数という連続変数に直す。その他7、無回答9は欠損値

```
COMPUTE          EDU=Q44.
RECODE           EDU(1=6)(2=9)(3=12)(4=13)(5=14)(6=16)(7,9=99).
MISSING VALUES EDU(99).
```

4. 発展版

4.1. 男女別等の分析 – ファイルの分割について

調査データの場合、男女別に分析して結果を出すことが多い。重回帰分析も、多くの場合、男女別にデータを分割した後で分析すると、より明確な結果が出る。男女すべて合わせたデータだと、うまく関連が出ないこともある。

SPSSにはデータ分割機能があるので便利である。以下のシンタックスで分割する。あるいは、SPSSのデータウィンドウで、画面上の「データ」をクリックし、ファイルの分割を選ぶ。データを男女別に2つに分割後に分析すると、2つの分析結果が出る。

```
SORT CASES BY Q47SEX.
SPLIT FILE LAYERED BY Q47SEX.
```

4.2. モデルの作り方 (p. 184)

自分の目的を明確に決めてYとなる変数を1つ設定する。Xとして、心理的変数(意識に関する質問項目など)と社会的変数(年齢、学歴、財産数など基本属性や社会的地位に関連するもの)を入れ、さまざまなモデルを作ってみると良いだろう。

初めは、年齢や学歴など基礎的なXだけを入れたモデルを作り、少しずつXを増やしていくとよい。因果関係をよく考えてXを入れる。

最終的なモデルは、Xとして心理的変数と社会的変数の両方を含むと良い。重回帰分析は、多くのXを同時に投入することに意味がある。別々に入れたモデルはあまり意味がない。ただし多重共線性には注意する。

心理学的なデータは、全員が大学生であることが多く、年齢や学歴や職業ダミー変数をXとして使うことはできず、心理変数間の関連を見ることになる。そのような分析も意味があるが、現実の社会には、様々な社会階層の人がいるのだから、教育年数や年収など、階層に関する何らかの変数もXとして使った方がよい。また、社会学的データにおいて、

心理的変数をXとして使うことも重要である。とくに、態度や価値観（権威主義的態度や伝統主義的態度や性別役割意識）など、無意識や深層心理に近いものと、表面的な意識と、両方を使うことができると、より良いだろう。

4.3. エクセルやワードでの図表の書き方

- ・ワードやエクセル画面の上「挿入」をクリックし、図形描画ボタンを選ぶ
- ・ボタンを押して四角や矢印などをかく。

1) 図を微調整したい時

かいた図を右クリックして書式設定を選ぶ。線の太さや矢印種類などを変更できる。両方向矢印などにすることができる。

2) 図の中に文字を書くには

- ・図形を右クリック → テキストの追加
- ・図形を右クリック → 図の書式設定
 - 「色」ボックスをクリック → 塗りつぶしなし（白でなく透明になる）
- ・あるいは画面上「挿入」をクリックすると「テキストボックス」が出る。設定して文字を書く。その後、微調整は、テキストボックスを右クリックして書式設定を選ぶ

3) エクセルでの罫線の引き方

- ・線を引きたいセルをマウス（またはシフトキー＋矢印）で囲む
- ・囲んだ部分を右クリックして「書式設定」→「罫線」タブを選ぶ
- ・下線ボタンなどを押しOKボタンを押す

4) 曲線矢印の引き方

- ・画面上「挿入」を押し、図形ボタン押し曲線を選ぶ
- ・曲線を引く。真ん中で一度クリックしさらに引く。書き終わるときはダブルクリック
- ・引いた線を右クリックし「書式設定」を選ぶ
- ・矢印「始点や終点のスタイル」を選ぶ

4.4. エクセルで作った表をワードの中に貼るには

- ・まずエクセルで表を作る
- ・ワード画面にて、画面上の「挿入」をクリック
- ・「オブジェクト」を選択し「エクセルワークシート」新規を選ぶ
- ・ワークシートが出てくるので、自分で作った表をはりつける

5. モデル構築の考え方

1) Xは何個くらいがいいのか

とくに基準はないが、社会学的データの場合、普通は年齢、学歴、収入と基礎的な社会意識項目数個の他に、関連する項目を数個、合計10前後のXを入れて分析することが多い。

社会調査データでは、「伝統的価値観」に関する項目など、何らかの基礎的態度と関連する項目を入れた方がよい。あるいは入れたものかないものなど、複数のモデルを作る。

2) 有意でないXを除いた方がいいのか

除く必要はない。関連がないということも、重要な発見。

3) Yと有意な関連があるXはいくつくらいあった方がよいのか

1つでも構わない。すべてがYと無関連だと、他にXを探すべきということになるが、基本的に、無関連でも構わない。

4) 変数選択や、ステップワイズの使い方は

社会学的データの場合、とくに使う必要はない。ただ、X同士の相関が強い場合は、どちらを投入するべきか判断するために、部分的に使うことがある。

5) 4段階回答などは、どの程度、量的と考えるとよいのか

完全な名義尺度を量的変数として使ってはいけない。しかし、ある程度の方向性や順序としての性質があれば、量的変数として考えて問題はない。尺度の種類は理論的には数種類あるが、実際に分析する場合、尺度についてあまり細かく考えても意味がない。データ人数が数百人以上あり、3段階以上の回答ならば、量的変数と考えると現実には問題はない。

6. 論文の構成 ー全般的な分析の流れについて

レポートや論文を作る際には、冒頭で目的（何を明らかにしたいか）と仮説（因果関係を含む文）を明確に書く。そしてYとなる変数を1つ決める。その後、まず因子分析結果や相関行列を出し、全体的な変数間の関連を確認すると良い。可能ならば、いくつかの変数について散布図を作っても良いだろう。その後、因果関係を自分の頭で考えて、何をXにするかを決めてモデルをいくつか作り重回帰分析で因果関係を確認する。その後、さらに用いる変数を絞って、クロス集計やエラボレーションを行うと良い。

論文には、分析結果として、基本的な男女別集計の横棒グラフ等をまず載せ（分布の偏りを確認し、どのような質問項目か読者に分かってもらう）、主な変数間の相関行列（または因子分析）を表にしてから、重回帰分析、主要な変数に関するクロス集計やエラボレーション（3重クロス集計を用いた因果関係の確認）の順で結果を並べることが多い。

7. 課題

自分で自由にテーマを決め、何らかの調査データを用いて、男女別に重回帰分析を行う。結果を、男女別の2つの図にまとめ、自分の解釈を書く。被説明変数Yは、自分が興味ある質問項目を1つ決めればよい。説明変数Xを5個以上入れること。

結果を見て、自分の意見として、結果の解釈を豊富に書くことが重要である。上記「分析時の注意点」に、十分に気をつけること。

参考文献

- ボーンシュテット・ノーキ. 1990. 『社会統計学 ー社会調査のためのデータ分析入門』. ハーベスト社.
- 早川毅. 1990. 『回帰分析の基礎』朝倉書店.
- 市川伸一・大橋靖雄. 1987. 『SASによるデータ解析入門』. 東京大学出版会.
- 石村貞夫. 2001. 『SPSSによる多変量データ解析の手順』東京図書.
- 石村貞夫. 2001. 『SPSSによる統計処理の手順』東京図書.
- 岩井紀子・保田時男. 2007. 『調査データ分析の基礎 ーJGSSデータとオンライン集計の活用』有斐閣.
- 久米均・飯塚悦功. 1987. 『回帰分析』. 岩波書店.
- 栗原伸一・丸山敦史他. 2017. 『統計学図鑑』オーム社.
- 蓑谷千風彦. 1990. 『回帰分析のはなし』東京図書.
- 縄田和満. 1998. 『Excelによる回帰分析入門』朝倉書店.
- 三輪哲・林雄亮. 2014. 『SPSSによる応用多変量解析』オーム社.
- 三宅一郎・山本嘉一郎他. 1986. 『新版SPSS X 基礎編』東洋経済新報社.
- 村瀬洋一他編. 2007. 『SPSSによる多変量解析』オーム社.
- 室淳子・石村貞夫. 2002. 『SPSSでやさしく学ぶ多変量解析』東京図書.
- 岡太彬訓・古谷野亘. 1993. 「多変量解析法の不適切な利用」. 数理社会学会『理論と方法』Vol. 8 No. 2.
- 小塩真司. 2018. 『SPSSとAmosによる心理・調査データ解析 第3版 ー因子分析・

共分散構造分析まで』東京図書.
小塩真司. 2020. 『共分散構造分析はじめの一步 一図の意味から学ぶパス解析入門』ア
ルテ.
佐和隆光. 1990. 『回帰分析』朝倉書店.
田部井明美. 2001. 『SPSS完全活用法 一共分散構造分析(Amos)によるアンケート処理』
東京図書.

付録 シンタックス例

```
***** ダミー変数作成 *****/  
COMPUTE MOTIIE =0.  
IF (Q25 =1) MOTIIE =1.  
  
***** 変数の方向を逆転 *****/  
MISSING VALUES Q7A (9).  
COMPUTE N7A =5-Q7A.  
  
***** 重回帰分析 *****/  
REG  
/DEP Q30A  
/MET=ENTER AGE EDU Q3 Q4 Q5.  
  
***** 分散分析 *****/  
UNIANOVA Q16 BY Q27 NENDAI  
/METHOD=SSTYPE(3)  
/PLOT=PROFILE(NENDAI*Q27)  
/PRINT=DESCRIPTIVE  
/DESIGN=Q27 NENDAI Q27*NENDAI.  
  
***** 欠損値を除き人数を減らす処理 *****/  
SELECT IF age < 99.
```

分析実習時の指導メモ

福島2015調査 シンタックス例

```
/***** GAKUREKI WO KYOUIKU NENSUU HE HENKAN *****/
```

```
COMPUTE EDU=Q43.
```

```
RECODE EDU(1=6) (2=9) (3=12) (4=13) (5=14) (6=16) (7, 9=99).
```

```
MISSING VALUES edu (99).
```

```
COMPUTE SEIMAN =5-Q2A.
```

```
COMPUTE ninman =5-Q2b.
```

```
COMPUTE kuniman =5-Q2c.
```

```
COMPUTE kenman =5-Q2d.
```

```
COMPUTE sigotman =5-Q2e.
```

```
COMPUTE TYONAI =5-Q6A.
```

```
COMPUTE YAKUNIN=5-Q6B.
```

```
COMPUTE GIIN =5-Q6C.
```

```
COMPUTE KUMIAI =5-Q6D.
```

```
compute network = yakunin + giin.
```

```
COMPUTE FUKOHEI =Q13.
```

```
COMPUTE YARITAI =5-Q15A.
```

```
COMPUTE HYOBANO =5-Q15B.
```

```
COMPUTE DANZYO =5-Q15C.
```

```
COMPUTE KENI =5-Q15D.
```

```
COMPUTE axlife =5-Q29A.
```

```
COMPUTE axincom=5-Q29B.
```

```
COMPUTE axjobls=5-Q29C.
```

```
REG
```

```
/DEP axlife
```

```
/METHOD=ENTER sexdmy AGE EDU danzyo yaritai seiman network.
```

重回帰分析の手順について 2020指導メモ
まずデータ読み込み命令文をすべて実行。
あるいはデータファイルを開く。

1 何をYとして使うか決める

2 まず変数の分布をみる

度数分布表を全て出す

分布を確認することは分析の基本！

3 欠損値処理をする。

4 変数を逆転する、方向をそろえる。

5 シンタックスを書いて分析

書く順番に注意

データ読み込み、欠損値処理、compute文などデータ加工、分析命令文。

6 モデルの作り方

Xとしては、年齢や教育年数や自営業ダミー変数など基本属性は入れてみることに。

XとYはすべて連続変数。できるだけ細かい変数を使う。
重回帰の場合は**カテゴリー合併しない。**

```
/***** 逆転した新変数作成          *****/  
compute RSOPINIO = 5-Q22b.
```

```
/***** IF文による自営ダミー変数作成 *****/  
COMPUTE SELFEMPL=0.  
IF (Q33=6 OR Q33=7) SELFEMPL=1.  
IF (Q33=1          ) SELFEMPL=1.
```

```
CRO  
  /TAB= Q22b by RSOPINIO .
```

```
CRO  
  /TAB= Q33 by SELFEMPL .
```

p. 189

```
REG  
  /DEP RSOPINIO  
  /MET=ENTER sexdmy AGE EDU livrate FAMILYNO  
  dmgm slstrat Honer hitennyu tatennyu.
```