分析実習資料　　2019/06

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、関連が出ることがある。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（ボーンシュテット・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

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

標本予測式　$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i}$　・・・・・・(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。

Σe_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。回帰平均平方

$$R^2 = \frac{\Sigma\left[(Y_i-\overline{Y}_i)^2-(Y_i-\hat{Y}_i)^2\right]}{\Sigma(Y_i-\overline{Y}_i)^2} = \frac{SS_{TOTAL}-SS_{ERROR}}{SS_{TOTAL}}$$

$$= \frac{SS_{REGRESSION}}{SS_{TOTAL}}$$

$$= \frac{\text{モデルで説明できる分散}}{\text{全分散}} \qquad (3)$$

全体平方和 SS_{TOTAL}
＝ 回帰平方和 $SS_{REGRESSION}$ ＋ 残差平方和 SS_{ERROR} ・・・・(4)

（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。

$R^2 = \beta_1^* \ r_{YX1} + \beta_2^* \ r_{YX2}$ ・・・・・・(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。

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

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

自由度：説明変数の数

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

自由度：Ｎ－１－（説明変数の数）

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。

多重共線性に注意するために、回帰分析を行う際には、まず説明変数間の相関行列を見て、相関がとても強いものがあれば、片方は説明変数から除く、といったことが必要である。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、変数選択の方法がいくつか考えられている。詳しくは、村瀬他(2007)などを参照。

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。モデルはシンプルなほど良い。

決定係数R^2がどれくらい大きければ良いと見なすのかは難しい問題だが、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作り、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こつ。

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。

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）。**YもXも量的変数**しか使えないことに注意。

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。

性別ダミー変数を使うか、あるいは男女別にデータを分割してから分析し、２つの結果を比べるなどするとよい。調査データの場合、男女で関連の具合が異なることが多いため、分割した後に分析した方がうまくいくこともある。データ分割後に、分析を実行する。

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.関係的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、点線上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。出力をエクス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）

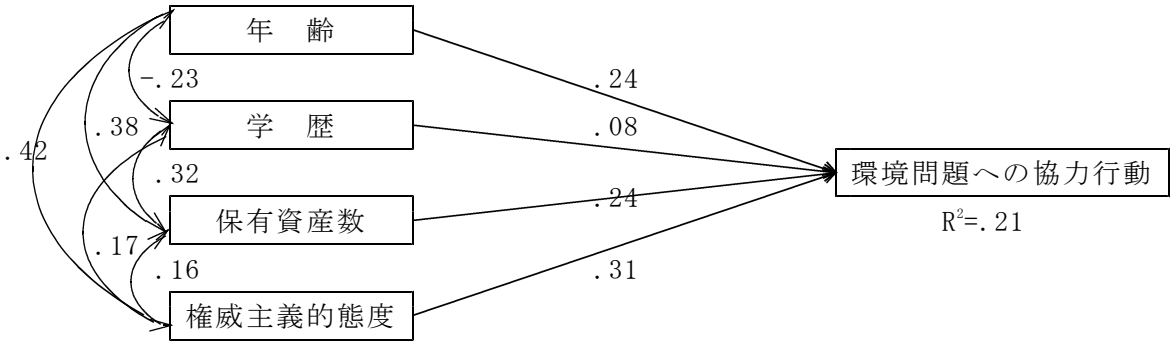

図1.向環境行動の規定因に関する重回帰分析結果　1998年××調査
数字は標準化係数

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と良い。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（standaradized 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、以下の注意点に気をつける。とくに、全変数について欠損値処理をしているかどうか、よく確認すること。

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。

・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。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と便利。データを男女別に２つに分割してから、分析を行うと、男女について２つの分析結果が出る。

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を入れると良い。

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。

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。書き終え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

4.5. 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良い。あるいは入れたものとないものなど、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、量的と考えてよいのか**

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。変数の性質についてあまり細かく考えても意味がない。

５．論文の構成　－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。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イ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）、相関行列（または因子分析）、重回帰分析、主要な変数に関するクロス集計の順で結果を並べることが多い。

６．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。

７．参考文献

ボーンシュテット・ノーキ．1990．『社会統計学　－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 ―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．
蓑谷千凰彦．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．2004．『SPSSとAmosによる心理・調査データ解析 ―因子分析・共分散構造分析まで』東京図書．
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 ―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.

/*****  データ全体を男女別に分割     *****/
SORT CASES  BY Q47SEX.
SPLIT FILE LAYERED BY Q47SEX.
```