

重回帰分析の応用的手法

— 交互作用項ならびに統制変数を含む分析 —

前 田 和 寛*

1. はじめに

何らかの事象・対象に対して調査・研究する場合には、多かれ少なかれ結果に対する予測や仮説を設定する。そしてその予測や仮説と測定したデータとが一致するかどうかを検査していく。

一般に調査を実施してデータを収集する場合、母集団すべてを測定することは困難であり、母集団より適切なサンプルを抽出して、そのサンプルを元にして母集団のデータを推測する推測統計学の手法が用いられる。そして推測された母集団の特徴と予測・仮説とが一致するかを統計学的検定によって示していく。したがって調査や研究においては、統計学的な検証がかなり重要な位置を占める。

仮説検証に用いられる分析としては、分散分析を代表とする「主に差を吟味する分析」と、重回帰分析を代表とする「主に関連を吟味する分析」とがあり、それぞれにメリット・デメリットがある。分散分析のメリットは計算が比較的容易であり、特に「交互作用効果」について検討が容易な点である。デメリットは、独立変数（原因変数）がカテゴリー変数でなければならず、本来のデータでは連続変数であったものを高群・低群といったように分割して分析しなければならない点である。このように連続変数をカテゴリー変数に変換することは、情報量の減損を招き、データを十分に反映した結果が得られない場合がある。具体的には、100点満点のテストで50点未満を低群、50点以上を高群とした場合、0点でも49点でも同じ低群として扱われることとなる。あるいは49点と51点というたった2点の差でも、0点と100点の差でも、同じ差として扱われるのである。

一方重回帰分析のメリットは説明変数（原因変数）が連続変数のまま検討が可能であり、また変数の取

捨選択が統計的に容易にできる点である。デメリットは、通常の分析モデルでは「交互作用効果」が難しいという点が以前より取り上げられていた。交互作用効果とは、「ある変数の値が変化すると、他の変数の関連が変化する」という効果である。たとえば、英語の勉強時間と英単語の暗記とに関連があるとすると、一般に正の関連が認められるだろうが、ここで他の変数として「年齢」を考慮した場合、年齢が若ければ勉強時間に比例して英単語の暗記も上昇しやすいだろうが、逆に年齢が高いとたとえ勉強時間を多くとったとしても、なかなか英単語を暗記しにくいだろうという可能性が考えられる。このような交互作用効果は現象の理解において非常に重要なポイントとなるのだが、通常重回帰分析では検証が難しい。

では、連続変数で測定したデータを用いて、交互作用効果を検討できないのだろうか。この問題について、近年重回帰分析で交互作用効果を検討する方法が開発され、海外の研究において広まってきている。しかし国内においてはその手法を紹介する文献はなく、未だ十分に浸透していないのが現状である。そこで本稿では、重回帰分析の応用的手法である交互作用効果を検討する方法を紹介し、その拡張として、さらに統制したい変数も含める場合について解説し、その手法をさらに拡張する。

2. 交互作用効果を含んだ重回帰分析のモデル

最もシンプルな重回帰モデルといえば、基準変数（結果変数, Criterion) Y に対して、2つの説明変数（原因変数, Predictor) X と Z を含む重回帰モデルで表される：

* 総合生活デザイン学科

$$\hat{Y} = b_1X + b_2Z + b_0 \dots\dots(1)$$

この重回帰モデルでは、単にXとZそれぞれの主効果しか含まれておらず、交互作用効果による基準変数の変動は含まれていない。XとZの交互作用効果を含む重回帰モデルは、(2)式のようになる：

$$\hat{Y} = b_1X + b_2Z + b_3XZ + b_0 \dots\dots(2)$$

(1)式と比較すると、(2)式では新たにXとZの積で表されたXZという項が追加されている。つまり、新たにこのXZという項を追加することで重回帰分析において交互作用効果を検討することができるのである。ではなぜこれで交互作用効果を検討できるのであるのか。

まず(1)式では、XにおけるYへの効果 (b_1) は、Zがどのような値をとろうが変化しない。また、ZにおけるYへの効果 (b_2) も同様に変化しない。なぜなら、 b_1 はZの効果を経験的に統制したXの効果であり、 b_2 もまたXの効果を経験的に統制したZの効果であるからである。

一方(2)式は、XにおけるYへの効果 (b_1) は、Zの値によって変化する。それは、(2)式を次の(3)式のように変形することで見えてくる：

$$\hat{Y} = (b_1 + b_3Z)X + (b_2Z + b_0) \dots\dots(3)$$

この(3)式は(2)式の右辺をXについてまとめたものである。これはXにおけるYへの一次モデルだとみなすことができ、 $(b_1 + b_3Z)$ はXにおけるYへの効果、 $(b_2Z + b_0)$ は定数を表す。このXの効果には b_3Z が含まれており、これは「Xの効果はZによって変化する」ことを表す。つまり、 b_3 が交互作用効果を示すこととなる。もし b_3 が効果なし、つまり0となる場合には、Xの効果は b_1 のみとなり、主効果によってのみXがYに対して効果を持つこととなる。この b_3 は、まさにXとZを掛け合わせた項であるXZの偏回帰係数である。

3. 具体的な分析方法

ここでは、実際に交互作用効果を含む重回帰分析のテクニックを説明する。

3-1. 変数の作成

交互作用効果を含む重回帰のモデルを分析するには、XZという項、すなわちXZという変数を作る必要がある。この変数は、変数Xと変数Zを単純に掛け合わせることで作成できる。そして、説明変数として、X、Z、そしてXとZを掛け合わせたXZという3つの

変数を投入することとなる。

しかし、ここで重大な問題が発生する。変数XZはXとZで作られた変数であるため、XとXZ、あるいはZとXZの相関が非常に高くなってしまおうという問題である。相関が高い変数を説明変数として含めて重回帰分析を実施すると、多重共線性という問題が発生する。この多重共線性という問題が発生すると、パラメータの誤差が非常に大きくなって結果が不安定となり、結果が妥当でないものとなる。

そのため、このように交互作用効果を含む重回帰分析を実施するためには、あらかじめ変数XとZに「中心化」という処理を行う必要がある。中心化とは、その変数の平均値を0となるように変換するものである。XとZを中心化しておき、その中心化した変数を使って交互作用項XZを作成すると、中心化していない場合と比較して、主効果の項と交互作用の項との相関がかなり低下する。このテクニックにより、主効果の項と交互作用の項との相関に起因する多重共線性を回避することができる。数学的解説は、Aiken & West(1991)を参照されたい。

なお、この中心化ではなく標準化(平均値を0、標準偏差を1)を行っても、今後の分析の有意性は全く変化しない。下位検定などを実施する際には標準化の方が簡単になるが、グラフ化する際に元のデータ範囲に復元したい場合には逆に手間がかかることとなるため、検討したい目的によって選択するとよいだろう。また、この中心化という操作は、あくまで「主効果の変数と交互作用の変数の相関を抑えるためのテクニック」であり、「多重共線性を回避するテクニック」ではない。つまり、XとZに高い相関があり、それぞれを中心化してもその相関は一切変化しないことに留意していただきたい。

3-2. モデルの検討

変数の作成が終了したので、次の段階である全体モデルを分析する。重回帰分析にはいくつか投入方が存在するが、交互作用効果を検討する場合には「強制投入法」が適当だろう。ただし、1次の項(主効果)をStep 1に、交互作用項をStep 2にて投入するような、次数によって階層を分けた階層的重回帰分析を実施するのが望ましい。これは、どの次数まで投入したモデルが適切かを判断するためであり、いたずらに次数を増加させることを回避する上で有効である。本稿ではこの階層的重回帰分析については触れず、また変数の数が少ないため3つの変数全てを同時投入したモデルにて解説する。

中心化した X , Z , XZ を説明変数, Y を基準変数とする重回帰分析を実施した例として, 以下に結果の例を示す:

$$\bar{X}=0(s_x=0.50), \bar{Z}=0(s_z=1.50)$$

$$\hat{Y}=0.3X+4.0Z+3.5XZ+2.4 \dots\dots(4)$$

この例で, s は各変数の標準偏差を表し, またこの結果では, Z の主効果と XZ という交互作用項が有意であったとする。つまり, 交互作用効果が認められたこととなり, その効果を詳細に検討する下位検定を実施する必要がある。

3-3. 下位検定

(4)式の結果を元に, 交互作用効果の下位検定を実施することとする。下位検定とは, 「ある変数がどのような値の時に, もう一方の効果がどのようなようになるのか」を分析することである。ここでは例として, 「 Z がどのような値のときに, X の効果がどのようなようになるのか」について下位検定を実施していく。

まず(4)式を(5)式のように X についてまとめると, (5)式ようになる:

$$\hat{Y}=(0.3+3.5Z)X+(4.0Z+2.4) \dots\dots(5)$$

先述したように, この(5)式は X の Y への回帰とみならずことができ, また X の効果は $(0.3+3.5Z)$ で表される。したがって, この X の係数にあたる $(0.3+3.5Z)$ を X の単純傾斜(simple slope)といい, 特定の Z の値をこの(5)式に代入することで, それぞれの Z の値での X の Y の回帰直線を求めることができる。

では, Z の値としてはどのようなものが妥当なのだろうか。原則としては自由であるが, 数学的なバイアスを回避するために, 「 Z の平均」, 「 Z の平均から1SD上」, 「 Z の平均から1SD下」という $Z_M \cdot Z_H \cdot Z_L$ の値がCohen & Cohen(1983)によって提案されている。特定の値を使用したいという理由がない限り, このガイドラインに従うべきであろう。

今回の例では, Z の標準偏差は1.50であるため, この値を用いてそれぞれ $Z_M \cdot Z_H \cdot Z_L$ に基づく X の Y への回帰直線を算出すると以下ようになる:

$$\begin{aligned} Z_H: \hat{Y} &= \{0.3+3.5 \times (1.50)\} X + \{4.0 \times (1.50) + 2.4\} \\ &= 5.55X + 8.4 \end{aligned}$$

$$\begin{aligned} Z_M: \hat{Y} &= \{0.3+3.5 \times (0)\} X + \{4.0 \times (0) + 2.4\} \\ &= 0.3X + 2.4 \end{aligned}$$

$$\begin{aligned} Z_L: \hat{Y} &= \{0.3+3.5 \times (-1.50)\} X + \{4.0 \times (-1.50) + 2.4\} \\ &= -4.95X - 3.6 \end{aligned}$$

この3つの回帰式をグラフにしたものが図である。このように, Z の値が減少していけば X と Y の関連は負から0, そして正へと変化していくという交互作用効果が見てわかるだろう。

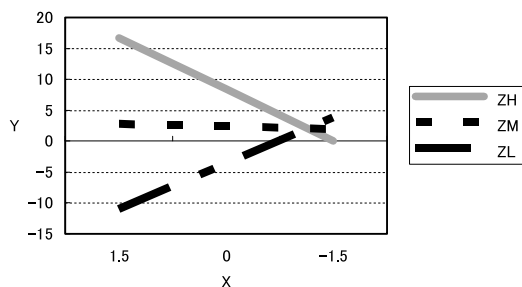


図 交互作用効果

3-4. 単純傾斜の有意性の検定

(4)式の例を使って, 5.50, 0.3, -4.95という三つの単純傾斜が算出されたが, これら単純傾斜は統計的に有意な効果を持つのだろうか。この3つのうち Z_M である0.3はまさに X の主効果であるため, 全体モデルの分析における有意性検定がそのまま該当する。それ以外の単純傾斜である5.50と-4.95という値については, 別途検討する必要がある。

単純傾斜の有意性を検討するには, それぞれの回帰係数の標準誤差を算出して, 統計量 t 値を算出しなければならない。しかしこれはかなり手間のかかるものであり, 統計学の知識が相当分もとめられる。しかし, 一般の統計パッケージを使用すれば比較的容易に分析することが可能である。

まずは, (a)新しい変数 Z_{above} , Z_{below} を作成する。この前者は Z_H に, 後者は Z_L に対応する変数であり, 以下のように算出する:

$$Z_{above} = Z - (s_z)$$

$$Z_{below} = Z - (-s_z)$$

Z_{above} は高い方だが, Z の標準偏差を引くことに注意していただきたい。標準偏差を用いているのは, 先述の Z の値として $\pm 1SD$ というガイドラインに従ったためである。新たな変数を作成した後に, (b)説明変数 X と, 新たに作った Z_{above} , Z_{below} をそれぞれ掛け

合わせた変数 XZ_{above} , XZ_{below} を作成する。これは、交互作用項として、 Z の代わりに Z_{above} , Z_{below} の変数を用いることで、対応する Z の値での交互作用効果を検討できるようにするためである。

そして最後に、(c) 説明変数として X , Z_{above} , XZ_{above} を投入する重回帰分析と、説明変数として X , Z_{below} , XZ_{below} を投入する重回帰分析を、それぞれ実施する。まず Z_{it} に対応する X , Z_{above} , XZ_{above} の結果について説明すると、全体モデルとほとんど同一の結果が認められるはずである。異なるのは、定数項の値と、 X の係数の 2 点のみである。しかもこの 2 つの値が、先ほど Z_{it} として算出した X の Y への回帰直線の係数と一致するのである。したがって、この分析結果では、 Z_{it} として算出した X の Y への単純傾斜の有意性検定も実施されているので、単純傾斜を検討することができるのである。また Z_{it} に対応する説明変数として X , Z_{below} , XZ_{below} を投入する重回帰分析についても同様である。

以上が、交互作用効果を含む重回帰分析を実施する方法である。これらの方法は比較的簡便でありながら、以前から問題となっていた点をカバーすることができる非常に有効な手段である。

4. 統制変数を含める場合

これまで説明してきたモデルは、説明変数として X , Z , XZ という 2 つの変数の主効果とその交互作用効果を含むものであった。しかしながら基準変数 Y について、 X と Z 以外にも影響を及ぼすであろうと予想される変数がある。そのような時はランダムサンプリングなどデータ測定時における統制処理を行うのが適切であるが、どうしても対処できない場合には統計的に統制することとなる。このように、説明変数として X , Z , XZ を投入する重回帰分析に、新たに統計的に統制する変数である統制変数 C を考慮したモデルは非常に有効的な分析である。本稿ではこの統制変数 C を考慮したモデルの検討方法について紹介する。

4-1. 統制変数、主効果、交互作用効果を含む重回帰モデル

先ほどまでの交互作用効果を含む重回帰モデルに、新たに統制変数を含めると以下ようになる：

$$\hat{Y} = b_1 X + b_2 Z + b_3 XZ + b_4 C + b_0 \dots (6)$$

つまり、新たに $b_4 C$ という項が追加されることとなる。重回帰分析では、投入される変数同士の影響は統

制されるため、このようにモデルに同時投入することで統計的にコントロールされることとなる。しかしながら注意すべきこととして、統制変数 C と説明変数 X や Z との相関がある。重回帰分析では投入される変数同士の相関が高いと多重共線性という問題が発生する。この統制変数を含めた場合も同様であり、 C と X や Z との相関が高いと多重共線性が引き起こされるため、その場合はこのモデルを検討することができなくなる。

4-2. 統制変数を含む重回帰モデルの分析方法

具体的に分析を実施するには、その準備として統制変数 C を中心化しておく必要がある。これは分析結果を得たときに、 X , Z , XZ という本来検討したい変数について図示する際や下位検定を実施する際に、統制変数 C を中心化していなければそれだけ算出が複雑になるためである。実際には、統制変数は中心化しておけば、あとの分析をする際には先述した下位検定の方法でそのまま実施することとなる。これは非常に簡単となるため、必ず実施すべきだろう。

統制変数を含む全体モデルを分析するには、Step 1 として統制変数 C を、Step 2 として主効果の変数 X と Z を、そして Step 3 として交互作用項を投入する階層的重回帰分析を実施する。これにより、各段階で次の次元を投入することに意味があるかどうかを評価することができ、さらに統制変数を投入することに意味があるかを評定することができる。

また、下位検定を実施する際には、中心化した C をそのまま下位検定を検討するモデルに、同じように投入すればよい。結果の読み方などは統制変数を含まない場合と同一である。

4-3. 統制変数が複数ある場合

もし統制変数が C のみではなく、他に複数ある場合にはどうしたらよいのだろうか。たとえば C 以外にも D , E もあわせてコントロールしたい場合には、3 つの統制変数 C , D , E を全て中心化して、Step 1 にて 3 つの変数を投入するようにすればよい。これにより、3 つの変数を全て統計的に統制することが可能である。しかし、後述するがいたずらに変数を増やすことはそれ以外に様々な問題を引き起こすこととなるため、注意が必要であろう。

5. おわりに

本稿では、従来の重回帰分析での問題点を改善する

方法として、「重回帰分析において交互作用効果を含む方法」を紹介し、さらにその応用手法として「統制変数、交互作用効果を含んだ重回帰分析」について解説した。これらの手法は、連続変数にて測定されたデータを解析する上で非常に有用な手段であり、今後より普及していくだろう。本稿が、その一助となれば幸いである。

しかしながら、最後に紹介した「統制変数が複数ある場合」について、いくつか注釈をしておきたい。統制変数を含める手続きは非常に簡便であり、ある現象に対して影響を及ぼす要因は現実には複数あるため、つい多くの変数を投入しがちになってしまう。しかしあまり多くの変数を投入することはモデルとして不適切となることが多い。たとえば、重回帰分析におけるモデル全体の説明率を表す重決定係数は、たとえランダムな変数を追加しても増加する。したがって通常は自由度調整済み重決定係数を採用するが、これは変数の投入数が多ければ多いほど、重決定係数が低下するように算出されるため、本来のモデルの適切性を減じてしまいかねない。また、変数が増加することにより、多重共線性の問題も発生し

やすくなるため、「最低限統制する必要がある変数」のみをコントロールするように心がけるべきである。その基準としては、概念的・統計的両方の観点から研究者や調査者が判断していくべきだろう。

また、今回紹介した応用方法以外にも、3要因以上の交互作用効果を含むモデルや曲線関係、あるいは説明変数にカテゴリ変数を含む手法も開発されている。今後、従来から用いられている重回帰分析に対する見直しを含め、より適切なデータ解析手法を習得し、使用していくことが必要である。

引用文献

- Aiken, L. S., and West, S. G., (1991) *Multiple regression: Testing and interpreting interaction*. SAGE publication.
- Cohen, J., and Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum.
- (受理 平成19年10月9日)

Abstract

Applications of Multiple Regression Analysis: Analysis Including Interaction Term and Control Variables.

Kazuhiro MAEDA*

Recently, multiple regression analysis has been expanded and various methods have been developed in research studies. Specially, it has been come up the technique including interaction term. This article introduces the multiple regression analysis technique including interaction term, and explains the procedure of this post hoc analysis. Additionally, the method of multiple regression analysis including not only interaction term but also control variable is explained, because variable controlling is one of the most important thing on data analysis.

(Received October 9, 2007)

* Department of Comprehensive Human Life Studies