

多次元極小標本に対する回帰分析の手法開発 ：環境保全型農業のコスト規定要因の分析への適用

平泉光一

(平成29年1月5日受付)

要 約

ケース数が少なく説明変数が多数ある多次元極小標本は、自由度不足で通常は重回帰分析ができない。しかしながら、多くの説明変数を集約して一つの合成変量をつくることで、多次元極小標本であっても単回帰分析を行うことはできる。この方法による回帰分析を本稿では擬似重回帰分析と呼ぶ。本稿では環境保全型農業のコスト規定要因の分析に擬似重回帰分析を適用した。用いたデータは、農林水産省が行った稲作の環境保全型農業に対する調査結果である。標本は、ケース数が5個で、説明変数の数が7個である。回帰分析の結果は、決定係数が0.9899で、回帰係数のp値が0.000428であり、説明力も信頼性も高いものであった。得られた回帰方程式の解釈は、常識的理解と一致した。擬似重回帰分析は、多次元極小標本になりやすい事例研究のデータの分析に向いているといえる。

新大農研報, 69:21-24, 2017

キーワード：多次元極小標本、擬似重回帰分析、環境保全型農業

1. 諸言

多次元極小標本という用語は一般には使われていないが、本稿では、説明変数（独立変数）の数がケース数（標本サイズ）よりも多い、非常に小さな標本のことを指すことにする。事例研究等で得られるデータは、ケース数が少ない割に、説明変数となりうる要因の数がケース数よりもずっと多いことが普通である。例えば、事例研究では、ケース数が6個しかないのに説明変数となる要因の数が10個以上あるようなデータが得られることがある。事例研究で得られるデータは典型的な多次元極小標本である例が大部分だといってよいであろう。周知のように多次元極小標本に対しては、通常は重回帰分析等の多変量解析を用いることができない。説明変数の数がケース数よりも多くなると自由度不足になって計算ができないからである。しかしながら、調査した情報を出来る限り有効に用いるには、多くの説明変数を犠牲にせずに使えるようにする定量分析の方法が求められる。これまでの研究では、多変量小標本に対して多変量解析を行うには、特殊で複雑な理論と技法が用いられてきた（青嶋と矢田, 2013）。本稿では、自由度不足で重回帰分析ができない多次元極小標本について、多数の説明変数の合成変量を用いて説明変数を1つに集約して、単回帰分析を行う手法を用いることを提案する。この手法を擬似重回帰分析と呼ぶことにする。そして擬似重回帰分析の具体的な適用例として、環境保全型農業（稲作）のコスト規定要因の分析を行うこととする。

2. 方法

標本のサイズを n 、説明変数の数を m とする。説明変数は、その候補となる l 個の変数から目的変数（従属変数） Y との相関係数が高い変数を m 個選択するものとする。 i 番目の説明変数を X_i とする ($1 \leq i \leq m$)。 Y と X_i は間隔尺度以上の変数とする。個々の説明変数 X_i と目的変数 Y との単回帰分析から

得られる回帰方程式は、

$$Y = a_i + b_i X_i$$

である。ここで、 a_i は定数項、 b_i は説明変数 X_i の回帰係数である。 m 個の説明変数の合成変量として、ここでは単純に、

$$\sum b_i X_i$$

を考える。 $\sum b_i X_i$ は n 個ある。擬似重回帰分析では、この合成変量を使って、次の単回帰分析を行う。

$$Y = \alpha + \beta \sum b_i X_i$$

ここで、 α は定数項、 β は合成変量の回帰係数である。この単回帰分析において、決定係数が十分に大きく、かつ、 β の p 値が有意に小さい場合、 X_1, X_2, \dots, X_m の説明変数のセット（組み合わせ）が説明力をもっていると考えられる（個々の説明変数の影響力は通常は重回帰分析であれば標準偏回帰係数を計算すれば分かるが、擬似重回帰分析では不明である）。説明変数のセットが説明力をもつようにするには、元の単回帰分析あるいは相関分析において X_i と Y との相関係数がそれぞれ十分に大きい必要がある。

なお、相関分析と回帰分析の計算に当たっては、マイクロソフト社の Excel 2007 と IBM 社の SPSS Statistics 21 を用いた。

3. データ

環境保全型農業のコスト規定要因の分析を行うために、農林水産省統計部（2004）による「環境保全型農業（稲作）推進農家の経営分析調査報告」（以下、単に「調査報告」と呼ぶ）のデータを用いることとした。「調査報告」では平成14年産について平成15年に調査した結果を載せている。「調査報告」では有機栽培、無農薬・無化学肥料栽培、無農薬栽培、無化学肥料栽培、減農薬又は減化学肥料栽培の5つの栽培形態の取組区分を調査対象としている。上記の各取組区分の調査対象農家数は、それぞれ71戸、73戸、49戸、66戸、80戸である。調査事例は、道府県別のリストから栽培形態別に無作為に抽出されている。

コストについては、「調査報告」では集計されていないため、公表されている10a当たり経営費に家族労働の労働費を付加して、副産物価額を引いてから、10a当たり収量で除して60kg当たりコストを5つの取組区分毎に求めた。この労働費については、「調査報告」で公表されている10a当たり労働時間に対して、農林水産省統計部による「平成14年産米及び麦類の生産費」で使われている全国平均の1時間当たりの労賃を掛けて求めた。副産物価額については「平成14年産米及び麦類の生産費」の全国平均と一律に同等とみなした。このコストは大部分が生産費であるが、労働時間に販売・管理労働が含まれているので厳密には生産費以外のコストも一部含まれる。コストの60kg当たりの金額は、有機栽培が24,128円、無農薬・無化学肥料栽培が24,100円、無農薬栽培が23,801円、無化学肥料栽培が19,073円、減農薬又は減化学肥料栽培が13,606円であった。なお、自作地地代と自己資本利子は含めていないので全算入生産費に相当するコスト（フルコスト）ではない。

本研究では、5つの取組区分の全国の集計データのコストおよび各種要因を用いたので、ケース数は5となる。

4. 結果

合成変数に含める説明変数の候補を絞り込むためにコストとの相関分析を行った結果を表1に示した。「調査報告」では様々な項目を調べているが、常識的に考えてコストに影響がありそ

うな変数を10個選択して相関分析を行った（網羅的に相関分析を行ったわけではない）。

相関係数の絶対値の高い順から並べると、1位：「10a当たり労働時間」（生産労働以外に販売・管理労働を含む）、2位：「10a当たり収量」、3位：「堆きゅう肥投入量」、4位：「農協経由率」（主な販売先割合における農協のシェア）、5位：「人力・動力除草機による除草実施農家割合」、6位：「たい肥舎導入農家数割合」（環境保全型農業のために新たに整備した農機具、農用施設等の導入農家数割合におけるたい肥舎の割合）、7位：「作付面積」（調査対象水稻作付面積）、8位：「その他の有機物投入量」、9位：「動力式除草機導入農家数割合」（環境保全型農業のために新たに整備した農機具、農用施設等の導入農家数割合における動力式除草機の割合）、10位：「その他の農機具等導入農家数割合」（環境保全型農業のために新たに整備した農機具、農用施設等の導入農家数割合におけるその他の割合）であった。

このうち、説明変数として上位7個を選んだ。この取捨選択の基準は必ずしも客観的とはいえないが、相関係数が比較的高い0.7以上の変数を残した。相関係数0.7を基準としたのは、説明力を意味する決定係数（相関係数の2乗）がほぼ0.5に達するからである。

7つの変数のそれぞれを説明変数として目的変数であるコストとの単回帰分析を行った結果を表2に示した。表2において標準誤差とt値は回帰係数のものである。X1からX4までの

表1 コストとの相関係数

Xi	説明変数候補	相関係数	p値
X1	10a 当たり労働時間 (hr/10a)	0.963773	0.008232
X2	10a 当たり収量 (kg/10a)	-0.96099	0.009193
X3	堆きゅう肥投入量 (kg/10a)	0.959577	0.009697
X4	農協経由率 (%)	-0.95321	0.012067
X5	人力・動力除草機による除草の実施割合 (%)	0.860664	0.061113
X6	たい肥舎導入農家数割合 (%)	0.759261	0.136557
X7	作付面積 (a)	-0.75781	0.137769
X8	その他の有機物投入量 (kg/10a)	0.691554	0.195844
X9	動力式除草機導入農家数割合 (%)	0.682979	0.203775
X10	その他の農機具等導入農家数割合 (%)	0.628448	0.256177

表2 コストとの単回帰分析の結果

Xi	説明変数	切片	回帰係数	標準誤差	t値
X1	10a 当たり労働時間 (hr/10a)	6392.502	368.5182	58.88234	6.258552
X2	10a 当たり収量 (kg/10a)	125640.1	-233.182	38.74528	-6.01832
X3	堆きゅう肥投入量 (kg/10a)	3759.179	28.5707	4.838109	5.905345
X4	農協経由率 (%)	27736.25	-216.528	39.65115	-5.46082
X5	人力・動力除草機による除草の実施割合 (%)	10138.31	191.7516	65.49522	2.927719
X6	たい肥舎導入農家数割合 (%)	16630.32	637.7627	315.6067	2.020751
X7	作付面積 (a)	28942.32	-58.1448	28.90457	-2.01161

表3 合成変数 $\sum biXi$ とその内訳

	b1X1	b2X2	b3X3	b4X4	b5X5	b6X6	b7X7	$\sum biXi$
有機栽培	16299.56	-103299	19170.94	-1732.22	14860.75	9757.769	-9768.32	-54711
無農薬・無化学肥料栽培	16992.37	-100734	19942.35	-4655.35	11984.47	3699.023	-5523.75	-58295.3
無農薬栽培	19155.58	-100968	19828.07	-5521.46	13978.69	4591.891	-3953.84	-52888.7
無化学肥料栽培	12444.86	-107497	17542.41	-8791.03	6404.503	3507.695	-7907.69	-84296
減農薬又は減化学肥料栽培	7853.122	-110994	9428.332	-13273.2	6788.007	0	-12850	-113048

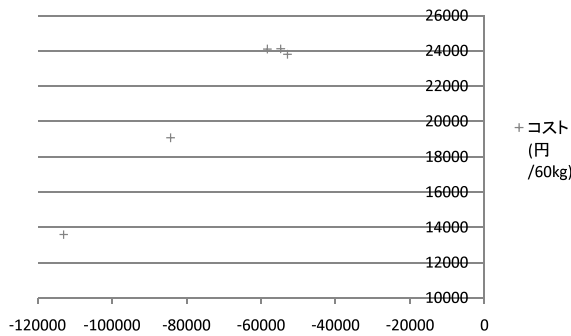


図 合成変量 $\Sigma biXi$ とコストとの関係

説明変数は t 値の絶対値が 5 以上あった。

合成変量 $\Sigma biXi$ の計算の過程を表 3 に示した。

コスト Y を目的変数として合成変量 $\Sigma biXi$ を説明変数とする単回帰分析（擬似重回帰分析）の結果は、

$$Y = 33845.87 + 0.177628 \Sigma biXi \\ (787.52) \quad (0.010327)$$

であった（括弧内は標準誤差である）。回帰係数の p 値は 0.000428（0.1%水準で有意）であった。決定係数 R^2 は 0.989961（自由度調整なし）であった。

図に合成変量とコストの関係を示した。概ね直線上にあることがわかる。

5. 考察

合成変量 $\Sigma biXi$ とコスト Y との決定係数が 1 に近くて非常に高かったばかりでなく、標本サイズが非常に小さいにもかかわらず p 値も低く、0.1%水準で有意であった。このことから、X1 から X7 までの説明変数のセット（組み合わせ）による説明力は十分に高く、また、推定の信頼性も高いことを示している。ただし、今回の回帰分析の回帰係数 β の p 値である 0.000428 は、個々の変数との相関係数が最も高かった「10a 当たり労働時間」の p 値である 0.008232 よりも小さくなっていたが、常に β の p 値が、最も相関係数の高い変数の p 値より低くなる保証はないと思われる。説明変数の組み合わせによっては、 β の p 値が最も相関係数の高い変数の p 値より低くなることもあろう。ともあれ、今回のように、単純な方法でも説明力と信頼性の高い説明変数のセットを見いだせたことは、多次元極小標本の擬似重回帰分析の有用性を例示できたと考えられる。

上述した 7 個の説明変数による回帰分析の結果は常識的な理解とも合致する。1) 単位面積当たり労働時間が大きくなってコストが高くなることは、環境保全型農業では従前より指摘されている問題点である。2) 有機農業に典型的に見られるように環境保全型農業では、単収が下がりやすく、環境保全型農業に本格的に取り組むほど単収は低下し、コストの上昇をもたらすことになる。3) 堆きゅう肥投入量が説明変数となっているのは、堆きゅう肥をつくるのに手間がかかるか、または、その購入費用が増すかであって、いずれにせよコスト上昇要因であるからである。4) 農協経由率の低さは環境保全型農業による生産物の販路の開拓と販売の手間の煩雑さを意味していて、そのことがコストに反映していると考えられる。5) より本格的な環境保全型農業では、除草剤が使えなくなるので、人力や動力除草機による除草作業を実施することになるが、その実施割

合が高くなると、労働費や農機具費を押し上げることになる。6) たい肥舎導入農家数割合が高いほどコスト増大につながっているのも環境保全型農業ならではのといえよう。7) 一般に作物栽培において作付面積が大きいほど生産コストは減少する。環境保全型農業の場合はこの減少の程度は顕著ではないが、作付面積拡大による生産コストの低減が観察された。今回の相関係数分析と擬似重回帰分析の結果は、環境保全型農業の主要なコスト規定要因を反映しているとみなせよう。

注目すべきは、環境保全型農業の 5 つの取り組み区分（有機栽培、無農薬・無化学肥料栽培、無農薬栽培、無化学肥料栽培、減農薬又は減化学肥料栽培）には、コストの較差の序列があって、その序列が合成変量 $\Sigma biXi$ によって、ある程度連続的に説明できることがわかったことである。環境保全型農業の成立条件等を論じた胡（2007）に対する書評で富岡（2008）は、「有機農業と減農薬減化学肥料栽培を『同じ範疇の中での程度の差』と見てよいのかという点」を論点として提起している。「著者（＝胡氏）は減農薬減化学肥料栽培から有機農業までを『環境保全型農業』にとらえ、前者から後者へと取り組みが『高度化していく』としている。評者（＝富岡氏）の感覚では、有機農業と減農薬減化学肥料栽培の間は、連続よりも断絶の方が大きいように思われる。」と批判的に評している。本稿の結果は、胡（2007）のように有機農業をその他の環境保全型農業と連続的に捉えることを否定していないと考えられる。

多次元極小標本の擬似重回帰分析は、統計学的にあまり価値がないとされてきた例数が少数で説明変数が多数である、いわば「横長のデータ」に対するアプローチである。例数が多数で説明変数が少数である、いわば「縦長のデータ」に対してのみ統計学的アプローチが有用であるのではなく、「横長のデータ」に対してもデータを活用した統計学的アプローチが可能であることを本稿では提起した。今回の分析データは比較的大きな規模の全国調査のデータを用いたが、事例研究で得られる多次元極小標本に対しても、擬似重回帰分析のアプローチが有効であるケースが十分にありうると思われる。

本稿では、擬似重回帰分析の一つの方法を示したが、多数の説明変数を集約する合成変量の構成法は本稿で提示した方法に限らないはずである。例えば、尺度の違う変数を共通に扱えるように平均が 0、分散（＝標準偏差）が 1 になるように変数変換したうえで、相関係数に基づいたウェイトで合成変量を構成する方法等も考えられる。本稿で提示した合成変量の構成法は各説明変数のウェイトのかかり方を統制できない方法であるので、改善の余地があるといえる。別の方法で合成変量を構成した場合、決定係数や p 値がどのくらい変わるかは今後の課題としたい。

6. 引用文献

- 青嶋誠・矢田和善. 2013. 高次元小標本における統計的推測. *数学*, 65(3) : 225-247
- 胡柏. 2007. *環境保全型農業の成立条件*. 農林統計協会, 東京.
- 農林水産省統計部. 2004. *環境保全型農業（稲作）推進農家の経営分析調査報告*. 農林統計協会, 東京.
- 富岡昌雄. 2008. 胡柏著『環境保全型農業の成立条件』, *農林業問題研究*, 44(2) : 396-397.

Development of Regression Analysis Method for Multidimensional Very Small Sample : An Application for Analysis of Cost Determinants in Sustainable Agriculture

Koichi HIRAIZUMI

(Received January 5, 2017)

Summary

In general, multidimensional very small sample, which has few cases and many explanatory variables, usually can not perform multiple regression analysis due to lack of degrees of freedom. However, it is possible to perform single regression analysis even for multidimensional very small sample by summarizing many explanatory variables to create one synthetic variable. Regression analysis by this method is called pseudo multiple regression analysis in this paper. In this paper, pseudo multiple regression analysis was applied to analysis of cost determinants of sustainable agriculture. The data are the results of the survey on rice cultivation in sustainable agriculture conducted by the Ministry of Agriculture, Forestry and Fisheries. In the sample, the number of cases is five and the number of explanatory variables is seven. Regression analysis results showed that the coefficient of determination was 0.9899, the P-value of the regression coefficient was 0.000428, and the interpretability and reliability were high. Interpretation of the obtained regression equation was consistent with commonsense understanding. It can be said that pseudo multiple regression analysis is suitable for analysis of data of case studies which tend to be multidimensional very small sample.

Bull.Facul.Agric.Niigata Univ., 69:21-24, 2017

Key words : Multidimensional very small Sample, Pseudo multiple regression analysis, Sustainable agriculture