

明石工業高等専門学校 正会員 大橋 健一
徳島大学 工学部 正会員 青山 吉隆

1. はじめに

電子計算機の進歩とともに、大量のデータを取り扱う多変量解析は、その使用が容易となり、土木工学においても各分野でよく用いられている。本研究では、多変量解析の一手法である林の数量化理論Ⅱ類の適用と信頼性について考察を行なったものである。

Ⅱ類分析は定性的説明変数(アイテムカテゴリー)から外的基準の分類を行なう手法で、一般にある無限母集団からサンプリングしたデータについて、外的基準が最も説明できるようにカテゴリーに得点を与える。このときの説明の程度が相関比で、得点がカテゴリースコアである。そして、数少ないサンプル数でも、アイテムカテゴリーに反応がありさえすれば解析も可能で、外的基準と説明変数の因果関係を定量化できる点で便利である。その反面、アイテムカテゴリーの取り方と、サンプル数の取り方によって、得られる相関比・スコアが変化することが考えられる。一方、解析に必要なデータ数をサンプリング理論から解析的に求めることも難しいと思われ、本考察では、モンテカルロシミュレーションを用い、簡便化したケースについて相関比・スコアの変動を分析した。

2. モンテカルロ法によるサンプリング

数量化理論の計算は、各カテゴリー間のクロス表 $f_{lm}(jk)$ 、外的基準と各カテゴリーのクロス表 $g_{lm}(jk)$ から行なわれ、アイテム間に相関がないと仮定した場合、 $g_{lm}(jk)$ を決定すればスコアも求まる。すなわち、ある母集団からサンプリングによって $g_{lm}(jk)$ を決定することに他ならない。母集団における $g_{lm}(jk)$ の割合を母確率 P_{ijk} とした場合、 N 個のサンプルを抽出することは、母確率に従って一様乱数から N 個のデータ(1個の外的基準・ m 個のアイテムカテゴリー)を決定することと同一である。ここで、 P_{ijk} は外的基準 i の j アイテム k カテゴリーの母確率で、 $\sum_j P_{ijk} = 1$ とする。

3. 信頼性に影響する要因

数量化分析はカテゴリーに得点を与えて外的基準の

分類を行なうため、分析結果の信頼性に影響する要因として、相関比の高低、サンプル数、アイテム数、総カテゴリー数を取り上げた。

信頼性を示す指標として、母確率における真のスコアと各要因を組み合わせた実験値を比較すればよいのであるが、アイテムカテゴリーを各実験ごとに変化させたため、分析結果としてのスコアも各実験間で比較することは困難となった。このため、数量化分析の精度を示す相関比に着目し、母確率における真の相関比と各実験の相関比との差を信頼性の指標とした。ここで、真の相関比との差が零になっても必ずしも真のスコアに一致するとは限らないが、相関比により全体的な傾向の把握が可能になるものと思われる。また実験の効率を上げるために $L_{16}(2^4)$ の直交表から、

表-1 要因と水準

要因	A サンプル数	B 相関比	C 総カテゴリー数	D アイテム数
水準0	50	0.500	8	2
水準1	200	0.800	12	3

4 要因と6交互作用の影響 表-2 分散分析表

要因	変動	自由度	分散比
A	0.225	1	37.92**
B	0.184	1	30.93**
A×B	0.043	1	7.22**
C	0.078	1	13.07**
A×C	0.032	1	5.43*
B×C	0.002	1	0.39
D	0.005	1	0.86
A×D	0.003	1	0.46
B×D	0.008	1	1.37
C×D	0.000	1	0.04
e	1.834	309	
計	2.414	319	

表-3 要因効果 (全平均 0.0415)

要因	A**	B**	C**	D
水準0	0.0265	0.0240	-0.0156	0.0040
水準1	-0.0265	-0.0240	0.0156	-0.0040

表-4 交互作用効果

要因	A×B**	A×C*	B×C	A×D
(0,0)	0.0116	0.0100	-0.0027	0.0029
(0,1)	-0.0116	0.0100	0.0026	0.0029
(1,0)	-0.0116	-0.0101	0.0028	-0.0029
(1,1)	0.0116	-0.0101	-0.0027	-0.0029

は、統計

的に意味のある影響は表われていない。表3と表4に各要因の効果を示した。

4. サンプル数 N と数量化分析

相関比からみた数量化分析の信頼性に影響する要因が明らかとなったが、総カテゴリ数と比較してサンプル数 N が少ない場合、 N の影響は非常に大きいものである。そこで、他の要因も固定し N を連続的に増加した場合、相関比やスコアがどのように変化するものか分析した。ただし、アイテム数3、総カテゴリ数10(2・3・5)で、母確率による真の相関比が0.899(ケース1)と0.787(ケース2)の2ケースの各 N について、30回繰り返した平均と標準偏差である。ケース1の相関比を図-1、スコアの1例(2アイテム3カテゴリ)を図-3に、ケース2の相関比を図-2、スコアを図-4に示した。

これらの結果より、相関比は N の増加につれて真の相関比の上側から指数的に真値に近づいている。また標準偏差も N の増加につれて指数的に減少し、零に近づくものと思われる。そして相関比の真値は高いほど早く真値に近づき、相関比の低いほど遅くなる。スコアについては、かなりバラつきがみられるものの下側から指数的に真値に近づき、標準偏差は指数的に減少して零に近づくものと思われる。またカテゴリの所属するアイテムの分類数が少ないほど早く真値に近づいており、分類数が多いほど多くのサンプル数を必要とする。ちなみに、母確率 P_{ijk} の小さなカテゴリは真値に近づくのが遅くなるものと思われる。そして、

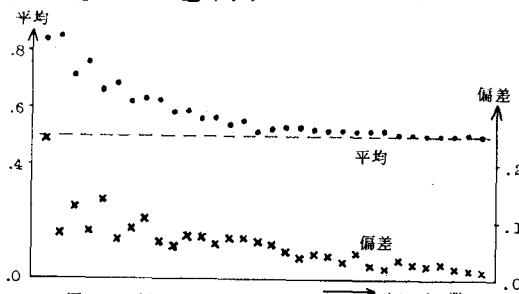


図-1 相関比 ケース 1

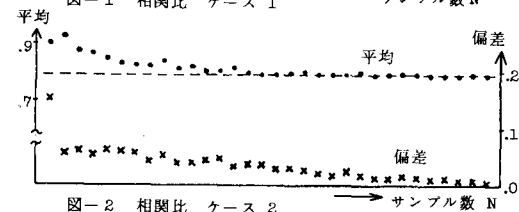


図-2 相関比 ケース 2

スコアに対する相関比の真値の高低の影響は、相関比の場合と同様である。またアイテムの重みを示すレンジについては、カテゴリスコアの結果からわかるように、 N の増加とともに下側から真値に近づく。

一方、任意のサンプル数 N に対する相関比とカテゴリスコアの分布型をみるために、正規分布と対数正規分布の適合度を調べた結果、ほとんどの相関比・スコアとも危険率5%で正規分布に従っている。

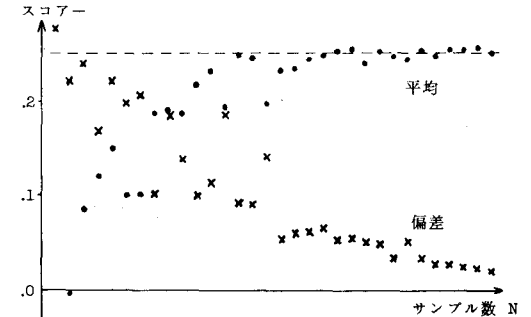


図-3 スコア、ケース 1

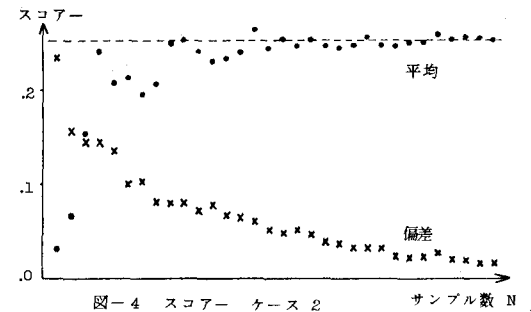


図-4 スコア ケース 2

5. おわりに

以上の結果より、数量化分析は、相関比・サンプル数・カテゴリ数などの条件により得られる結果も真値から一様に変化している。特に、大量データの収集が困難な場合とか、数量化分析が計量分析モデルの一部を構成し政策判定効果なども調べる場合、このような分析が必要になるものと思われる。

なお本考察では母確率を既知としたが、一般には未知であり、母確率を求めるために調査解析するもので収集されたサンプル数内において、解析に用いる N を色々と変化させ相関比・スコアの変動状態から真の相関比・スコアを推定しなけければならないであろう。またアイテム間には相関がないとして考察してきたが一般には何らかの相関関係が存在しており、この場合の分析も今後の課題である。