

1. はじめに

林の数量化理論Ⅱ類の分析結果は、カテゴリースコアと相関比で一般に示される。ところが、数量化理論の適用に当たっては、サンプル数の違いなどによって、分析結果が変化しており、得られる情報に差異が生ずることが考えられる。一方、Ⅱ類分析の結果は、アイテムカテゴリー間のクロス表と、外的基準とアイテムカテゴリーのクロス表から得られるものであり、相関比・スコアの決定に必要なサンプル数をサンプリング理論等から直接決定することも困難と思われる。

このため本研究では、簡単化したケースについて、モンテカルロシミュレーションによりサンプル数を色々と変化させ、相関比・スコアがどのように変動するか、またある許容誤差を与えた場合の必要なサンプル数について考察を行った。

2. モンテカルロ法によるサンプルの抽出

数量化分析は、母集団から無作為に抽出したサンプルに対して分析するものであり、このような操作は、アイテムカテゴリーと外的基準の母集団確率(真値)に従って一様乱数から電算機内でサンプルを発生させることと同一と考えられる。そこで図-1に示すように、アイテム間には相関がないものとして、 N 個のサンプルの外的基準とアイテムカテゴリーもモンテカルロ法で決定する。

PN_t : 外的基準 t に属する母集団の確率
 N_t : 外的基準 t に属するサンプル数 ($N_t = N \cdot PN_t$)
 N : サンプル数 ($N = \sum_{t=1}^T N_t$)
 M : アイテム数 T : 外的基準の数

3. 相関比・スコアの変動

母集団から任意のサンプル N 個を無作為抽出し、それぞれの分析から得られる相関比・スコアは、 N を一定としても変化するのである。またサンプル数 N を変化させることによって相関比・スコアは変化する。そこで、サンプル数 N を一定として、繰り返し400回分析して相関比・スコア・レンジの分布状態を調べ、次いで、サンプル数 N を増していった場合の相関比・スコアの変動を解析した。具体例としては、表-1

示すように外的基準 t 、相関比 r の2つのケースで、アイテム数、総カテゴリー数は同一とした。

表-2は、相関比・10個のスコア・3個のレンジ(合計14個)に正規分布を当てはめた χ^2 検定の結果である。

$N=20$ 付近では、総カテゴリー数に比してサンプル数が非常に少なく非現実的なため、非正規分布と仮定して正規分布とはみなせない。一方、現実的なサンプル数と思われる

200、500付近では正規分布と仮定できる。

次に、サンプル数 N を増加していき、それぞれのサンプル数で30回繰り返し分析した相関比の平均と標準偏差、スコアの平均と標準偏差を求めた。ケース1の10個のスコアの平均が変化していく様子も示したのが図-2である。相関比・スコアとも、サンプル数の増大とともに真値に近づいている。

4. 相関比とサンプル数

数量化分析の結果から情報として用いるのは主カテゴリースコアであるが、サンプル数とスコアの変動を個々に解析することは困難であり、分析全体の適合度を示す相関比とサンプル数の関係を調べる。

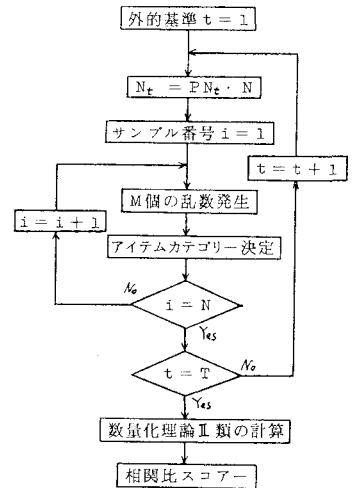


図-1 モンテカルロ法によるサンプリング

表-1 分析例

ケース	母相関比	総カテゴリー数
1	0.499	10 (2,3,5)
2	0.787	10 (2,3,5)

表-2 正規分布の適合度

	N	正規分布でない (5%)	正規分布でない (1%)
ケース1	20	8 (6)	8 (6)
	50	5 (9)	5 (9)
	200	2 (12)	1 (13)
	500	3 (11)	2 (12)
ケース2	20	10 (4)	7 (7)
	50	6 (8)	6 (8)
	200	4 (10)	3 (11)
	500	1 (13)	1 (13)

このため、相関比の平均・偏差とサンプル数 N との関係は次のように仮定して、回帰分析を行った。

$$\eta_1 = (1 - 0.499) \exp(-0.258N^{0.444}) + 0.499 \quad (0.973)$$

$$\sigma_1 = 0.602 \exp(-1.02N^{0.168}) \quad (0.964)$$

$$\eta_2 = (1 - 0.787) \exp(-0.224N^{0.478}) + 0.787 \quad (0.943)$$

$$\sigma_2 = 1.21 \times 10^{-5} \exp(-11.0N^{0.0325}) \quad (0.946)$$

η_1 はケース1の相関比、 σ_1 はケース1の標準偏差、 η_2 、 σ_2 はケース2である。相関比は $N=0$ のとき1となり、 N の増加とともに指数的に真の相関比に近づき、また相関比の標準偏差は零に近づいている。回帰の相関係数を右端に示しているが、全て0.9以上となった。

5. 相関比の信頼度

相関比はサンプル数の増加により真値に近づいており、真値からの許容誤差 δ を与えて、誤差内に入る相関比の割合を示したのが表-3である。Ⅱ類分析も適用する外部条件によって許容誤差は決まるものと思われるが、ここでは、絶対誤差 δ_1 と相対誤差 δ_2 を次のようにした。ケース1の絶対誤差 $\delta_1 = 0.05$ 、相対誤差 $\delta_2 = 0.05 \times 0.499 = 0.025$ 、ケース2の $\delta_1 = 0.05$ 、 $\delta_2 = 0.05 \times 0.787 = 0.039$ である。真の相関比が高ければケース2は、同じサンプル数でも信頼度が高い。また絶対誤差 $\delta_1 = 0.05$ とした場合、95%が許容範囲に入るサンプル数は、ケース1が約900、ケース2が約250であり、相関比の真値の大小によって数量化分析に要するサンプル数も大きく影響

表-3 相関比の信頼度(%)

ケース	1 (0.499)		2 (0.787)	
許容誤差	0.05	0.025	0.05	0.039
$N=25$	10.88	5.21	33.19	25.85
50	19.54	9.46	50.48	40.06
100	35.60	17.82	74.03	61.92
200	57.98	31.02	92.22	83.36
500	84.99	52.73	99.52	97.35
1000	96.52	70.80	99.99	99.76
2000	99.84	88.52	100.00	99.98
3000	100.00	95.85	100.00	100.00

響されている。

6. おわりに

本研究より、数量化理論Ⅱ類も適用する場合の一般的な結論はまだ得られていないが、相関比など母集団の特性によって必要とされるサンプル数に大きな違いがあることが明らかとなった。

なお、モンテカルロ法を用いることは膨大な計算時間を要するために大きな欠点もある。また相関比で信頼度も示したが、スコアに着目して信頼度を求めたり、アイテム間の相関を考慮することも必要である。このような方法で現実問題の信頼性も分析する場合、真の相関比・スコア・母集団確率は当然未知であり、収集したサンプル N_0 を母集団とみなして、同様な分析をし、 N_0 付近での相関比・スコア・安定度から、信頼性も判断しなければならないであろう。

【参考文献】

大橋・青山「数量化理論Ⅱ類の適用に関する一考察」
土木学会第35回学術講演要集Ⅳ、1980

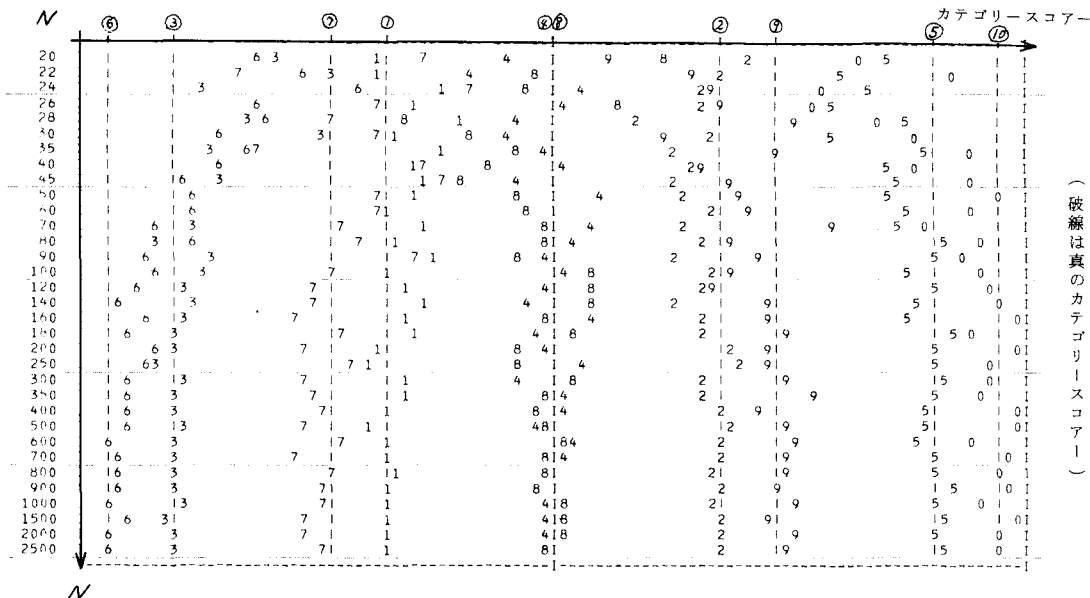


図-2 サンプル数 N とカテゴリースコア (10個) (ケース1 母相関比 0.499)