

多変数順位相関係数の構築

九州大学 鈴木讓

1 目的

この報告の目的は、離散変数間の関連性を分析する多変量解析の手法を構築することである。意識調査などで得られるデータは本来離散変数であるが、分析では連続変数を前提とした多変量解析の手法が通常用いられる。これは、2変数の離散変数ではクロス表分析における順位相関係数を用いることができるが、3変数以上の多変数の場合には適切な手法が存在しないことに起因している。そこで本報告では、順位相関係数を拡張し多変数順位相関係数を構築する。

2 方法

ここではクロス表は用いず、変数の個数を k 、標本数を n とし、 n 行 k 列の行列として調査データを表現する。つまり、各標本のデータは k 個の値から構成されているから、これを 1 行として扱えば、標本全体としては n 行 k 列の行列として扱うことができる。

n 個の標本から 2 個の標本を抽出しデータ対を作成する。少なくとも 1 つは座標の値が異なるような対の集合を R とし、 R を次の部分集合 R_1 と R_2 に排他的に分類する。

R_1 : 異なる値を持つ座標が 1 つだけしかない対の集合

R_2 : 異なる値を持つ座標が 2 つ以上ある対の集合

集合 R_2 の各要素（サンプル対）において、値の異なる座標の個数を v とし、集合 R_2 をさらに次の 2 つの部分集合 R_{2P} と R_{2Q} に排他的に分類する。

R_{2P} : v 個の座標の大小関係がすべて揃っている対の集合

R_{2Q} : v 個の座標の大小関係が混在している対の集合

R_1 の要素数を r_1 、 R_{2P} の要素数を p_1 、 R_{2Q} の要素数を q_1 とし、 R_{2P} の各要素に関して v を用いて調整した正相関の得点数を p_2 、 R_{2Q} の各要素に関して v を用いて調整した負相関の得点数を q_2 とし、

多変数順位相関係数 $\omega = \frac{p_2 - q_2}{p_1 + q_1 + r_1}$ と定義する。

3 結果

2変数の場合の ω と、クロス表分析の順位相関係数である Goodman-Kruskal の γ 、および Kendall の τ_b との比較を行う。上述の r_1 は、異なる値を持つ座標が 1 つだけしかないような対の個数であるから、無相関に対応すると考えられる。従って順位相関係数の計算においては、分母には含めるが分子には含めないような計算方法が適切と判断できる。しかしながら、Goodman-Kruskal の γ では同順位を全く考慮していないので分母、分子のいずれにも含まれない。また Kendall の τ_b では分母に含まれてはいるが、 x 座標だけが異なる対の個数、 y 座標だけが異なる対の個数を、正相関の得点数、負相関の得点数に別々に加算し平方根をとり、それらの値を乗じて分母とする、という計算方針がとられており、論理的意味合いが明確ではない。

4 結論

多変数の離散変数間の関連性を分析する手法として、多変数順位相関係数を構築した。2変数の場合、既存の順位相関係数とは計算結果は必ずしも一致しないが、今回用いた計算方法の方が適切だと考えられる。また今回構築した手法では、多重回帰分析とは異なり従属変数と独立変数を区別する必要がなく、その意味でも 2変数の順位相関係数からの自然な拡張と言える。