現代日本論演習／比較現代日本論研究演習 III「実践的統計分析」

第 4 講　相関係数 (1)

田中重人 (東北大学文学部准教授)

[テーマ]　順位相関係数

1　前回課題について

比率の差については、第 2 講資料 の正規分布を利用した信頼区間の幅が x より狭くなる条件を求めればよい。

$$x > 2 \times 1.96\sqrt{\frac{m(1-m)}{n}} \tag{1}$$

$$n > 4 \times 3.84 \times \frac{m(1-m)}{x^2} \tag{2}$$

平均値の差については、前期第 12 講 の平均値の差の信頼区間の公式を当てはめればよい。n 人を半分ずつの人数に分けるとすると、95 %信頼区間の幅は

$$2 \times 1.96 \times \mathrm{SD} \times \sqrt{\frac{1}{n/2} + \frac{1}{n/2}} \tag{3}$$

$$4 \times \frac{1.96\mathrm{SD}}{\sqrt{n}} \tag{4}$$

この幅の半分が x より小さいと有意な差が出ることになるので、

$$x > 4 \times 1.96 \times \frac{\mathrm{SD}}{\sqrt{n}} \tag{5}$$

$$n > 61.45\left(\frac{\mathrm{SD}}{x}\right)^2 = 61.45\frac{1}{\mathrm{ES}^2} \tag{6}$$

第 2 講資料 の「平均値の信頼区間」の幅が x より小さい、と考えて求めても同じ値になる。

※　これは簡略な求め方で、実際には、人数が均等に分かれていなかったり、自由度や併合 SD などの問題があるため、正確に計算するのは相当面倒である。永田 (2003) を参照。

調査規模の目安として、つぎのことをおぼえておくとよい：

- 100 ケースで有意差がでるのは: 比率の 20%以上の差、エフェクト・サイズが 0.4 以上になるような平均値の差
- 400 ケースで有意差がでるのは: 比率の 10%以上の差、エフェクト・サイズが 0.2 以上になるような平均値の差

2　尺度水準と分析法

- 名義×名義 → クロス表
- 名義×間隔 → 分散分析・平均値の比較
- 順序×順序 → 順位相関係数 (rank correlation coefficient)
    - Goodman-Kruskal の　γ
    - Kendall の　τ b
    - Spearman の rs (ρ　と書くこともある)
- 間隔×間隔 → 積率相関係数 (product-moment correlation coefficient)
    - Pearson の r

3　相関係数とは

ふたつの変数どうしが正 (+) の関係にあるか、負 (−) の関係にあるかを、−1 〜 +1 の範囲の値であらわす。

- 無関連のときゼロ
- 完全な関連のとき±1

「相関図」(または「散布図」(scattergram) ともいう) を描いて考えるとよい (教科書 p. 75)。

4　順位相関係数

4.1　Pair

相関図上の任意の2点を直線で結んだとき

- 右上がり → Concordant
- 左上がり → Discordant

それぞれのペアの個数を C, D とする。

4.2　グッドマンとクラスカルの「ガンマ」係数

$$\text{Goodman-Kruskal's } \gamma = \frac{C-D}{C+D} \tag{7}$$

同順位ペアをうまく扱えないので、あまり使われない

4.3　ケンドールの順位相関係数 (タウ b)

- K: x について同順位でないペア数
- L: y について同順位でないペア数

$$\text{Kendall's } \tau_b = \frac{C-D}{\sqrt{KL}} \tag{8}$$

同順位ペアがなければ、Goodman-Kruskal の γ と同じ値になる。

4.4 SPSS コマンド

クロス表の「統計量」オプション →「Kendall のタウ b」を選択

5 課題

(x, y) の値がつぎの組み合わせであるような 6 人の標本があるとする：

(1, 2) (2, 4) (2, 4) (4, 3) (4, 5) (5, 5)

この標本について、Kendall の順位相関係数タウ b を求めよ。

6 次回予習

教科書の第 3 章、第 8 章 7 節を読んでおくこと。

文献

永田 靖 (2003)『サンプルサイズの決め方』朝倉書店.