

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。

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

$$2 \times \frac{3.92SD}{\sqrt{n}} \quad (1)$$

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、

$$x > 3.92 \frac{SD}{\sqrt{n}} \quad (2)$$

$$n > 15.37 \left(\frac{SD}{x} \right)^2 \quad (3)$$

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。

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 r_s (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、 $-1 \sim +1$ 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

$$\text{Goodman-Kruskal's } \gamma = \frac{C - D}{C + D} \quad (4)$$

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

$$\text{Kendall's } \tau_b = \frac{C - D}{\sqrt{KL}} \quad (5)$$

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。