

現代日本論演習 / 比較現代日本論研究演習 I 「統計分析入門」(2013)

第8講 クロス表の解釈 (6/12)

田中重人 (東北大学文学部准教授)

[テーマ] 連関係数と%の関係、クロス表の整形と解釈の方法

1 前回課題について

2×3 のクロス表なので、 $m=2$ である。このため、 $m-1=1$ となるので、結局 $V = \sqrt{\chi^2/N(m-1)} = \sqrt{\chi^2/N}$ である。

式 [4-17] のなかでは、

- 「Pearson のカイ 2 乗」 = 式全体の値
- 「観測度数」 = 「セルの度数の実際の値」
- 「期待度数」 = 「セルの度数の予測値」
- 「残差」 = 分子の括弧の中
- 「標準残差」 = 分数全体 () の中身の平方根

添字を使った表記 (たとえば $n_{11}, n_{1.}, n_{.1}$ など) も覚えておくとよい。

連関係数 V は

- 独立 (無関連) のとき最小値 0
- 関連が強くなるほど大きくなる
- 完全関連のとき最大値 1 をとる

「完全な」関連とは? 教科書 p. 115

2 連関係数の解釈

連関係数は、「モデル」と「データ」の乖離を表した値と解釈できる

- 特定のモデル (この場合、独立の状態) の下で予測される値 (この場合、期待度数) を求める
- 実際のデータの値 (この場合、観測度数) と比較する
- モデルから予測される値と実際のデータの値の違いを集計する

- 0～1の範囲の係数になるように調整する（ちがうデータ間で比較しやすくなる）

多くの統計手法で、このタイプの「係数」が使われる。

連関係数 V の大きさの評価は主観的な問題であり、対象とする変数の性質によって基準が変わる。非常に大雑把には、つぎのような値を目安にするとよい：

- 0.1 未満 関連はない（無視してよい程度）
- 0.2 程度 弱い関連
- 0.3 程度 そこそこの関連
- それ以上 強い関連

3 連関係数と 係数、%との関連

SPSS では $\sqrt{\chi^2/N}$ の値を Phi (ファイ =) として出力する。 $m=2$ のときは、 $V = | \quad |$ である。係数については別の計算方法がある（教科書 p. 110）。

2×2 クロス表においては、係数と%の差との間には、一定の関係がある。

- 教科書 pp. 112–114 の記述を読んで、どのような関係があるか理解する
- 実際のデータで 2×2 クロス表を出力して、確認してみよう

4 今後の予定

6/19 中間試験。出題範囲は、今日の授業内容まで。持ち込み可（ただし通信・相談禁止）。コンピュータで解答を作成して、ISTU で提出。試験後は、通常通り授業。

6/26 休講（集中講義のため）