現代日本論演習／比較現代日本論研究演習 I「統計分析の基礎」

第7講　クロス表の解釈

田中重人 (東北大学文学部准教授)

[テーマ]　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。SPSS の出力にあてはめて確認してみるとよい。

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 (Σ の中身) の平方根

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

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 φ 係数、% との関連

SPSS では $\sqrt{\chi^2/N}$ の値を Phi (ファイ $= \phi$) として出力する。m=2 のときは、$V = |\ \phi\ |$ である。ϕ 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、ϕ 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　論文等のための表の書きかた

配布資料 参照

比較現代日本論研究演習 I ／ 現代日本論演習

(田中重人)　授業資料

表の書きかた

表 1　性別と性別による不公平感との関連

性別	性別による不公平				
	「大いにある」	「少しはある」	「ない」	合計	（人）
男性	36.0	50.5	13.5	100.0	(111)
女性	27.3	56.8	15.9	100.0	(132)
合計	31.3	53.9	14.8	100.0	(243)

Cramer's V=0.094.　無回答=7.

表 2　県や市町村の部課長以上の役人に知り合いがいる比率の男女差

性別	%	（人）
男性	46.0	(113)
女性	27.6	(134)
合計	36.0	(247)

ϕ=0.191.　無回答=3.

人に見せる表
- カテゴリーの並べ順や行列の組み合わせをわかりやすく
- 変数とカテゴリーの命名
- 表のタイトルとして適切なものをつける

タイトル、表本体、注釈を読めばそれだけでわかるように書くこと

書くべき要素
- 各セルの行（または列）％
- 行（または列）合計の度数と「100.0%」
- 列（または行）合計の％
- 全体の度数
- Cramer の V（または ϕ）
- 欠損数とその原因

行→列の因果を想定するのがふつうだが、列→行でもよい。方向は、合計の「100.0」で区別する。

全度数が 1000 人以下であれば、％は小数第 1 位まで
Vや ϕ などの係数は小数第 3 位まで

2 列表の場合は 1 列の％だけ示してもよい

縦罫線はなるべく引かない
文字列は左揃え、数字は小数点揃えが基本