

第6講 連関係数とクロス表の解釈

田中重人 (東北大学文学部准教授)

[テーマ] 連関係数と%の関係を理解する

1 前回課題について

- 「行」と「列」の区別
- 行% と列% の使い分け: 原因 結果に対応
- SPSS では「 の %」と表示される (は変数ラベル)
- 論文等に表を載せる場合は、行%か列%どちらか一方、適切なほうだけを書く
- グラフにする場合は、帯グラフ (積み上げ棒グラフ) で合計 100%になるようにするのが標準 (折れ線グラフまたは度数ポリゴンでもよい)
- Excel の「積み上げ棒グラフ」ではカテゴリ順序が逆転するので注意 (もとおりにしたいときは、シート上の順序をいれかえる)
- 列%によるグラフになってしまう場合は、右クリック 「データの選択」で行 / 列を入れ替える
- 「レイアウト」 「線」で「区分線」を指定するとよい。
- 「全体」のグラフは不要

2 今回の課題

「性別」と「性別による不公平」のクロス表を作成する。ただし、「セル」「統計量」オプションで「観測度数」「期待度数」「残差」「標準残差」「カイ 2 乗」「Phi」「Cramer V」の数値を指定すること。

出力と教科書 (pp. 108, 116–117) をもとに、つぎのことを考える：

- 連関係数「Cramer の V」と「Pearson のカイ 2 乗」の間の数学的な関係 [式 4-19]
- 式 [4-17] のなかに、「Pearson のカイ 2 乗」「観測度数」「期待度数」「残差」「標準残差」はどのように表れているか
- 連関係数 V の最小値・最大値はそれぞれいくつか。またどのような場合に最小値・最大値をとるか。

提出は、ISTU で月曜日正午まで。

なお、余力があれば、次のことも考えてみる：

- 2×2 クロス表におけるファイ係数 (: 教科書 p.110 [式 4-10]) は Cramer の V とどのような関係にあるか

3 キーワード

独立 (無関連 = independent): すべての列について行%が等しい (またはすべての列について行%が等しい) 状態

周辺度数 (marginal frequency): クロス表の右端・下端に書く「合計」の度数

期待度数 (期待値 = expected frequency): 周辺度数を固定しておいて、独立な (架空の) クロス表をつくった場合、各セルに入る (と期待される) 度数

観測度数 (frequency): 各セルに入っている実際の度数

残差 (residual): 観測度数 - 期待度数

標準残差 (standard residual): 残差を期待度数の平方根で割ったもの

χ^2 (chi-square): 標準残差の平方和

クラメールの連関係数 V : χ^2 を全度数で割り、セル数を調整したものの平方根

行・列の数が多いクロス表では、各セルの%を比較するのが大変である。また、%の差が大きいように見えても、度数が少ない場合には、実質的には大差ないと考えるべきであるが、そのようなことを判断するのもむずかしい。そこで、まずクロス表全体について「連関係数」を見ることで、行変数と列変数の「連関の強さ」を判断し、そのうえで細かく%を比較していくのが定石になっている。

4 今後の予定

5/28 進度確認。出題範囲は、今週の授業内容まで。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。