

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7講 連関係数 (6/5)

田中重人 (東北大学文学部准教授)

[テーマ] 期待度数、残差、連関係数の導出

### 1 前回課題について

「再」マークのついている人は、再提出 (ISTU で火曜正午まで)

- 「行」と「列」の区別
- 行% と列% の使い分け: 原因→結果に対応
- 論文等に表を載せる場合は、どちらか一方だけを書けばよい
- 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 (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 で火曜日正午まで。

### 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

$\chi^2$  (chi-square): 標準残差の平方和

**クラメールの連関係数  $V$ :**  $\chi^2$  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

6/19 中間試験 (当初の予定より 1 週間おくれ)。出題範囲は、6/12 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。

表1 性別不公平感の男女差

性別	大いにある	少しはある	ない	合計
男性	40 (36.0)	56 (50.5)	15 (13.5)	111 (100.0)
女性	36 (27.3)	75 (56.8)	21 (15.9)	132 (100.0)
合計	76 (31.3)	131 (53.9)	36 (14.8)	243 (100.0)

帯グラフ用データ

性別	大いにある	少しはある	ない
女性	36	75	21
男性	40	56	15

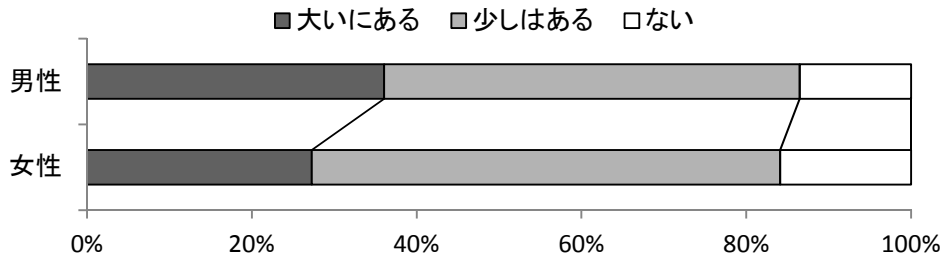


図1 性別不公平感の男女差

度数ポリゴン用データ

性別	0	1	2	3	4
男性	0	36	51	14	0
女性	0	27	57	16	0

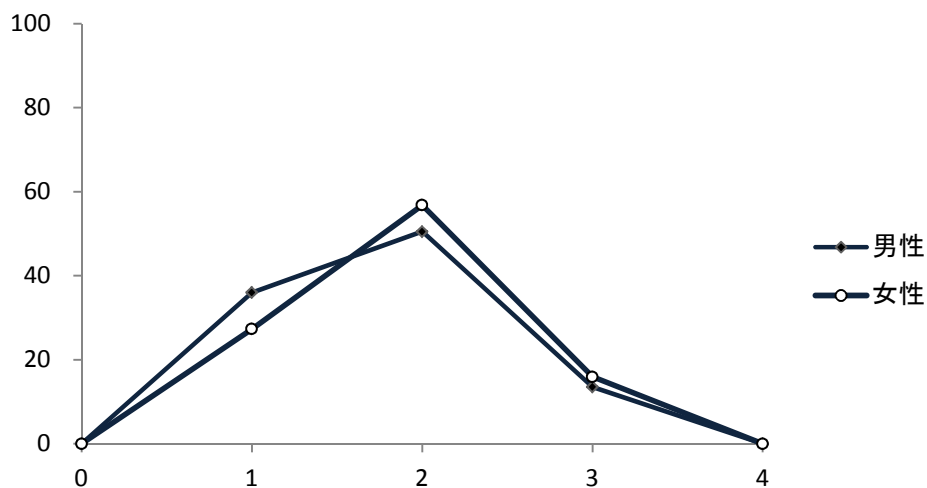


図2 性別不公平感の男女差