

一般用マイクロデータ就業構造基本調査版（仮） ～擬似標本データによる統計演習例～

伊原 一、北林三就（統計センター）、滝澤有美（総務省統計局）

統計センターでは、総務省統計局との共同研究として一般の人が誰でも利用できる一般用マイクロデータの作成と提供を行っている。本稿では、就業構造基本調査に基づく一般用マイクロデータを用いて統計演習を行う場合の例として、相関係数の計算を紹介する。

1. カテゴリー変数の相関係数

就業状態や性別などのカテゴリー変数の相関係数を求める場合は、変数に0と1の値を付与し、二項分布の計算により相関係数を求めることができる。二項分布に従う変数間の相関係数はφ（ファイ）係数とよばれており、下式により求めることができる。

$$\text{相関係数 } \phi = \frac{a d - b c}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

度数表

		被説明変数		
		合計	Y0	Y1
説明変数	合計	a+b+c+d	a+c	b+d
	X0	a+b	a	b
	X1	c+d	c	d

2. 相関係数φの計算

15歳以上人口の性別と就業状態について、一般用マイクロデータ就業構造基本調査版を用いて度数及び相関係数（φ係数）を計算してみると、女性は男性に比べて有業者率が低くなっており、相関係数の検定結果から女性と有業者の相関係数は有意に負値であることがわかる。

度数表（2012年）

	標本度数Σ		加重度数Σw		公表値ベース	
	無業者(0)	有業者(1)	無業者(0)	有業者(1)	無業者(0)	有業者(1)
男(0)	33,507	73,257	16,669,600	36,743,300	16,668,700	36,744,500
女(1)	59,282	55,584	29,725,100	27,675,400	29,725,700	27,676,200
相関係数φ	-0.204848		-0.208407		-0.208424	

一般用マイクロデータ就業構造基本調査版は、比推定補正によりベンチマーク人口（基本項目：都道府県、男女、年齢、就業状態）の基本項目を全てクロスして集計すると公表結果表と一致するように作成されている。ただし、これらの基本項目を部分的にクロスする場合については合計の丸め誤差による影響と、さらに基本項目以外の項目をクロスする場合は標本誤差による影響が加わるため、公表値ベースの相関係数の計算と完全には一致しないことに留意する必要がある。

参考：一般用マイクロデータの利用<<http://www.nstac.go.jp/services/ippan-microdata.html>>