

現代日本学演習 V 「実践的統計分析」

## 第12講 ダミー変数

田中重人 (東北大学文学部准教授)

[テーマ] 回帰分析での名義尺度変数の利用

### 1 2値変数の使いかた

たとえば、性別 (男性 =1, 女性 =2 の2値変数) を独立変数とする場合、そのまま使ってもよい。この場合には、「性別」の回帰係数は、値がひとつ増えることの効果をあらわすので、そのまま男性と女性の差を表すことになる。特に、独立変数がひとつだけのばあいには、男女の平均値の差が回帰係数と等しくなる。

ただし、2値変数を独立変数とするときには、一方を0、他方を1にしたほうが結果の解釈が簡単になるので、そういう変数をつくって使うことが多い。

```
recode Q1_1 ( 2=0 ) ( else=copy ) into MALE.
```

このような、特定の条件を満たす場合に1、それ以外の場合にゼロをとるような変数のことを「ダミー変数」(dummy variable) という。

[例題1] 性別による平均値の差の分析と、性別のダミー変数を投入した回帰分析をおこない、結果の対応関係を考える。特に、「定数」の値が何を意味するか考えること。

### 2 3つ以上の値のある変数をダミー変数に変換する

ある変数が  $k$  個の値を持つとする。この変数を回帰分析で使いたい場合は、 $k-1$  個のダミー変数を作成する。たとえば、学歴を3区分した変数 EDU がある場合、つぎのようにして、ふたつのダミー変数を作る。

```
recode EDU ( missing=sysmis ) ( 1 = 1 ) ( else = 0 ) into EDU_1.  
recode EDU ( missing=sysmis ) ( 3 = 1 ) ( else = 0 ) into EDU_3.
```

この例では、EDU=2 の場合にはダミー変数をつくっていないことに注意。

[例題2] これらふたつのダミー変数を投入した回帰分析の結果と、学歴による平均値の差の分析を照合して、どの数値がどれに対応しているかを考える。