

現代日本学演習 V 「実践的統計分析」

第11講 ダミー変数

田中重人 (東北大学文学部准教授)

[テーマ] ダミー変数を含む回帰分析

1 ダミー変数の課題について

前回の課題の回帰分析の推定結果では、 $Q39g$ の値が次の式で近似されていることになる：

$$Q39g = \text{定数} + B_1 X_1 + B_2 X_2 \quad (1)$$

ただし、

- X_1 は初等教育のものについて1、それ以外は0とする
- X_2 は高等教育のものについて1、それ以外は0とする

推定された係数 (B) それぞれについて、区間推定と統計的検定がおこなわれる
定数と係数は、カテゴリ別の平均とつぎのような関係にある：

初等: $3.60 - 0.71 = 2.89$
中等: 3.60 (= 定数)
高等: $3.60 - 0.01 = 3.59$

2 ダミー変数の使いかた：補足

このように、 k 個の値を持つ変数を回帰分析に投入するときは、 $k-1$ 個のダミー変数に変換して使う。このとき、すべてのダミー変数がゼロになるカテゴリがひとつ出てくることになるが、このカテゴリを「基準」と呼ぶことがある。「基準」のカテゴリの平均値が回帰分析結果の「定数」となり、各係数 (B) は、基準カテゴリとの平均値の差をあらわす。

どのカテゴリを基準にしてもよいのだが、通常は、つぎのどちらかにすることが多い：

- いちばんケース数の多いカテゴリ
- 平均値が最大 (または最小) のカテゴリ

このようなダミー変数は、本来はひとつの変数だったものなので、まとめてどの程度の影響をあたえているかを判断したいことがある。その場合、ダミー変数を投入した結果としなかった結果との間で決定係数 (R^2 乗) を比較して、どれくらい増えたかを見ることがある。

3 ダミー変数をふくむ重回帰分析

ダミー変数も、普通の変数と同様に使ってよいので、複数の独立変数を投入して回帰分析をすることができる。

4 平均値と最小 2 乗解の関係

平均値は最小 2 乗解である (= 偏差平方和が最小になる)

例題: 数直線上の 1, 2 に点がある。任意の点をいろいろ設定して、これらの点までの距離の 2 乗和を求める。

この「距離の 2 乗和」が最小になる点はどこか?