

第13講 講義まとめ

田中重人 (東北大学文学部准教授)

[テーマ] 回帰分析および講義の全体について復習とまとめ

1 前回課題について

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：

$$Q39g = \text{定数} + B_1X_1 + B_2X_2 \quad (1)$$

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 標準化係数 (ベータ)

独立変数・従属変数の両方を標準化した場合の係数を「標準化係数」(standardized coefficient) と呼び、 β で表す (SPSS では「標準化係数」と書いている)。独立変数の効果の相対的な大きさをみたい場合に使える (積率相関係数 r とおなじ感覚で評価できる)。

5 回帰分析結果の書きかた

係数の推定値の表だけを書く

- 推定値と標準誤差を書くのがふつう
- ダミー変数の場合、ひとまとまりであることと基準のカテゴリーがわかるように工夫する
- 検定結果をアスタリスク (*) で示す
- 表の下に決定係数 R^2 とその検定結果、人数を示す (人数は、「分散分析」表の「合計」の自由度に 1 を足すとわかる)

6 一般線型モデルのまとめ

- 因果関係とは
- 疑似相関と媒介効果
- 剰余変数の制御 (control)
- 最小二乗法
- ダミー変数

7 発展のための新しいトピック

- 被験者内効果を含むモデル (変量効果 → random effect)
- マルチレベルデータや時系列データ
- 因果関係の分析をめぐる近年の展開 (厳密な無作為化実験への接近)
- 複雑なモデルを推定する方法 (ベイズ統計学とシミュレーション)

表 1 政治効力感の一般線型モデル

独立変数	係数	標準誤差
定数	2.508*	0.336
年齢	0.017*	0.007
男性	0.609	0.322
学歴 (基準：高等教育)		
初等教育	-1.728*	0.351
中等教育	-1.497	0.297

$R^2 = 0.225^*$. $N=239$. * : $p < 0.05$