

第10講 回帰分析 (つづき)

田中重人 (東北大学文学部准教授)

[テーマ] 回帰分析の結果を解釈する一般的な手順

1 独立変数の「効果」をどう読むか

1.1 疑似相関

X1 が X2 と Y の両方に影響をもたらしているが、X2 は Y に対しては影響力を持たない場合を考える

X1	X2
X1	Y
X2	Y

このとき、X1 を無視して X2 と Y だけの関連を分析すると、X2 が Y に対して影響を与えているように見えるので、因果関係を見誤ることがある。このような現象を「疑似相関」という。(教科書 168-169 ページも参照。)

1.2 媒介効果

X1 によって X2 が決まり、それによって Y が影響を受けるという因果関係を考える。

X1	X2	Y
----	----	---

このとき、X1 は Y に直接 影響を与えているのではなく、X2 が「媒介効果」をもたらしていることになる。

1.3 直接効果

以上のようなことを考えながら、独立変数間の関連と、従属変数に対する直接的な効果を同時に把握する。

例題: 前回宿題の、3つの回帰分析の結果から、3変数間のどこに「直接効果」「媒介効果」「疑似相関」がみられるかを考える。

課題 1: 自分の興味のある分野で、「媒介効果」と「疑似相関」の例をひとつずつ考えよ。

2 モデルの評価と係数の検定

2.1 モデル全体の評価

一般線形モデルのデータに対するあてはまりのよさをあらわすのが、「決定係数」 R^2 である。決定係数は、分散分析でいう「相関比」の2乗に相当するので、平方根をとれば、と同様の感覚で、「そのモデルによって、従属変数がどの程度説明できているか」を評価できる。回帰分析ではこれを「重相関係数」と呼び、 R であらわす。

PSPP が出力する「分散分析」表では、「母集団においては決定係数がゼロである」(= どの独立変数も、従属変数に対して効果を持たない) という帰無仮説について検定を行った結果が表示される(「有意水準」の欄)。この結果が有意でなければ、モデル全体について、説明力があるとはいえないことになる。

2.2 係数の推定値

決定係数が有意であれば、モデル内の各独立変数の効果について解釈していく。

各変数にかかる係数については、「係数」の表に、95%信頼区間が表示される。このなかにゼロが含まれているかどうかで、5%水準で有意な効果があるかどうかを判断できる（「有意水準」の列をみて判断してもよい）。

2.3 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)。

3 欠損値処理とケース数

回帰分析をふくめ、多変量解析では、投入したすべての変数についてひとつも欠損値のないケースだけを分析に使う。このような欠損値処理方法を、「表単位」(listwise) の欠損値除去という。

この処理の結果として、多くの変数を投入すると、それだけケース数が小さくなるので注意。

4 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示すので、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場合は、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2: 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。

5 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。

課題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。