

現代日本学演習 V 「実践的統計分析」

第11講 固定因子と共変量

田中重人 (東北大学文学部准教授)

[テーマ] 一般線型モデルの出力の読みかた

1 前回宿題について

1.1 疑似相関

X_1 が X_2 と Y の両方に影響をもたらしているが、 X_2 は Y に対しては影響力を持たない場合を考える

X_1	X_2
X_1	Y
X_2	Y

このとき、 X_1 を無視して X_2 と Y だけの関連を分析すると、 X_2 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。

1.2 媒介効果

X_1 によって X_2 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。

X_1	X_2	Y
-------	-------	-----

このとき、 X_1 は Y に直接 影響を与えているのではなく、 X_2 が「媒介効果」をもたらしていることになる。

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。これは相関比 の2乗に等しいので、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。

SPSS の出力では、「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 (有意確率) が表示される。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 パラメータ推定値

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、「パラメータ推定値」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（有意確率の列をみて判断してもよい）。

ただし、SPSSは、固定因子については、その変数の中のいちばん大きい値のカテゴリーを「基準」としてあつかい、その基準と他のカテゴリーとの効果の差を係数として表示している。これらの効果全部（たとえば5カテゴリーの変数であれば、4つの効果があることになる）でどの程度従属変数に影響力を与えているかについては、「被験者間効果」として表示される分散分析表を見たほうがよい。

2.3 平方和の分解

各独立変数の効果の大きさは、「被験者間効果の検定」の表の「平方和」の列に表示される。「修正総和」の行の数値が、従属変数の平方和（分散×（ケース数 - 1））である。各変数の行の平方和をこれで割ると、従属変数に対する相対的な影響力の大きさがわかる。

分散分析表のよみかたについて復習

3 推定周辺平均

「オプション」の「推定周辺平均」「平均値の表示」に、カテゴリー別平均値を表示したい独立変数をえらぶ。

他の固定因子については基準カテゴリー、共変量には平均値を代入して計算した値が表示される

「主効果の比較」をチェックする（「Bonferroni」で信頼区間を調整）と、どのカテゴリー間に有意な違いがあるかを比較できる。（教科書 p. 205–207 「多重比較」の項参照）

4 欠損値処理とケース数

一般線形モデルでは、欠損値は listwise で処理される。このため、多くの変数を投入すると、ケース数が小さくなるので注意。

5 課題

固定因子1つ、共変量1つ以上の一般線型モデル分析をおこない、その結果について、上記の「モデル全体の評価」から「欠損値処理とケース数」までの各項目について解釈する。