

第9講 多変量解析入門

田中重人 (東北大学文学部准教授)

[テーマ] 多変量解析の種類と、重回帰分析の基本的な考えかた

1 前回課題について

対応のあるデータの場合、平均値の差の信頼区間を求める際の数式の標準誤差 (standard error) を、相関係数を用いて調整する。この点が、通常の (対応のない) 平均値の場合と異なる。

→ 相関図を描いて考えてみよう

2 対応のある分析について: 結果の書きかた

2.1 個々の結果を表示する十分なスペースがある場合

クロス表 (または相関図) をいちいち示すのが基本 (別紙参照)。各セルには、度数と 全体での%を書く。統計量などは表の下に書く。必要な統計量は分析法によって違うので注意。

- 対応のある t 検定 → 相関係数、平均値の差、有意水準 (対応のある検定であることを明記)
- 符号検定 → $x > y$ ケースと $x < y$ ケースの比率、有意水準

2.2 スペースがあまりない場合

対応のある t 検定であれば、各変数の平均と SD の表をのせる。表の下に、人数、相関係数、平均値の差、有意水準 (対応のある検定であることを明記) を書く。

符号検定であれば、 $x > y$, $x = y$, $x < y$ 各ケースの比率の表をのせる。表の下に、有意水準 (符号検定であることを明記) を書く。

2.3 多数の変数間の関連を示す場合

ハッセ図 (Hasse diagram) が使える。平均値などの高い順に変数を並べ、有意な差がある変数どうしを線でむすぶ。具体例は別紙参照。

3 多変量解析とは

3つ以上の変数をつかう分析法を「多変量解析」(multivariate analysis) という。次の2種類に分けられる (大野, 1998, p.48-56)。

- 類似関係型: 似た変数同士をまとめたり、潜在因子を取り出したりするもの。因子分析, クラスタ分析など
- 因果関係型: 原因と結果の関係を追究するもの。回帰分析, 分散分析, 一般線型モデルなど

この授業では後者だけをあつかう。

4 用語

従属変数 (dependent variable): 結果になる変数のこと。通常、ひとつの分析についてひとつだけ。「目的変数」「被説明変数」ということもある。

独立変数 (independent variables): 原因になる変数のこと。ひとつの分析に複数あってよい。「説明変数」ということもある。

従属変数と独立変数は、しばしば Y と X であらわされる

5 課題 1

Q39g (……指導者や専門家……) と Q1.1a (満年齢), Q6.1(学歴) の関連を確認し、これらの間の因果関係についてどのようなことがいえそうかを考える。

ただし、学歴の変数は次のように 3 分割すること：初等 (1,2,12); 中等 (3-5,13); 高等 (その他)

6 第 3 変数の統制 (control)

複数の要因が影響を与えていると想像される場合、因果関係を確定するには、ある変数の効果を「一定に保った」状態をつくったうえで、別の変数の効果を推定する必要がある。

実験の場合: 被験者の割り当ての時点で統制する (無作為割り当てなど)

観察の場合: 分析の段階で、多変量解析をおこなう

たとえば、データセットを学歴で 3 分割して、年齢と Q39g との相関分析を行ってみるだけでも、かなりのことがわかる。このような発想を洗練させたものが多変量解析である。

7 一般線型モデル

SPSS で、「分析」→「一般線型モデル」→「1 変量」を選択。

変数は次のように指定する

- 従属変数 = Q39g
- 固定因子 = 学歴 (3 分割)
- 共変量 = 満年齢

「オプション」で「記述統計」「パラメータ推定値」を指定

8 結果の読みかた

Q39g の値は、つぎの式で近似できることになる

初等教育: $Q39g = \text{切片} + B1X1 + B2$

中等教育: $Q39g = \text{切片} + B1X1 + B3$

高等教育: $Q39g = \text{切片} + B1X1$

9 課題 2

上記の一般線型モデルを変形して、年齢だけ、学歴だけを独立変数とする分析をそれぞれおこなう。結果出力のパラメータ推定値を比較して、なにがわかるかを考察。水曜正午までに ISTU に提出。

文献

大野高裕 (1998) 『多変量解析入門』同友館。

三土修平 (1997) 『初歩からの多変量統計』日本評論社。