

ダミー変数と回帰係数

九州大学 鈴木讓

1 目的

この報告では、線型回帰分析におけるダミー変数とその回帰係数について考察する。一般に線型回帰分析においては、二値変数であるダミー変数に関しても、連続変数と同様に標準偏差と平均値を用いて標準化回帰係数ベータを計算する。そしてベータの絶対値を比較することによって独立変数が従属変数に与える相対的な影響度を分析している。また、回帰係数の有意性に関しては、ダミー変数の場合も連続変数と同様に t 検定によって各回帰係数ごとにその有意性を評価している。しかしながら、このようなダミー変数の取り扱いには問題がある。この報告の目的は、これらの問題点を指摘し、より適切な分析手法を具体的に提案することである。

2 方法

まず、ダミー変数の回帰係数に関する問題点の整理を行う。問題点は、記述統計におけるものと、推測統計におけるものに二分される。記述統計における問題点は、回帰係数の標準化に関するものであり、これは更に概念上の問題と計算上の問題に分けられる。推測統計における問題点は、有意性の扱いである。

次に、それぞれの問題点について論理的、実証的観点からの考察を行う。回帰係数の標準化に関しては、そもそも標準化とは何かを問い直し、標準偏差を用いる意味を吟味する。そして、実際にダミー変数の標準化回帰係数を計算する際に生じる一種のパラドックスについて考察する。回帰係数の有意性に関しては、レファレンス・カテゴリーの設定との関係について考察する。

3 結果

回帰係数の標準化とは、本質的には変数の変化の基準値をどのように設定するかである。連続変数の場合には、どのような値の変化も論理的に可能であるが、ダミー変数の場合には、値の変化は 0 から 1 しかあり得ない。従ってこの場合には、標準偏差を用いた演算は意味がないことが分かる。

回帰係数の有意性に関しては、2つの問題点が指摘できる。1つは、レファレンス・カテゴリーの設定によって個々の回帰係数の p レベルが変化することであり、今ひとつは、ダミー変数のグループをまたがって回帰係数の有意性を比較することの意味が不明確なことである。従って、レファレンス・カテゴリーの設定に依存せず、しかも、ダミー変数のグループを1つの単位とする指標が必要である。

4 結論

- ・ダミー変数の回帰係数を標準偏差を用いて標準化することは、概念上意味がない。
- ・ダミー変数が 0, 1 の二値で設定されていれば、その回帰係数はすでに標準化された回帰係数と見なすことができる。
- ・ダミー変数の標準化回帰係数を、連続変数の標準化回帰係数と比較することは、標準化の基準値が異なるのでそもそも意味がない。
- ・異なるダミー変数のグループ間での比較は、回帰係数ではなくグループ自体を1つの単位としてとらえ、その回帰係数の値の範囲を用いて従属変数への影響の程度を比較すべきである。
- ・ダミー変数の有意性は、個々の回帰係数を t 検定により評価するのではなく、ダミー変数のグループに対する F 検定により行われるべきである。