# 【数量化１類】

説明変数が「定性データ」の場合の重回帰分析。

「定性データ（量的データ）」とは、アンケートの「不満・やや不満・普通・やや満足・満足」といった、数値ではないデータのこと。

|  |  |
| --- | --- |
| 数値化1類のデータ | |
| 目的変数 | 定量データ |
| 説明変数 | 定性データ（量的データ、カテゴリデータ） |

## 【例】定性データをカテゴリデータに変換

|  |  |  |
| --- | --- | --- |
| 満足度 | 性別 | 文理 |
| 5 | 男 | 理系 |
| 3 | 女 | 理系 |
| 4 | 男 | 文系 |
| 1 | 女 | 文系 |
| 2 | 男 | 文系 |

※目的変数＝満足度、説明変数＝性別・文理

次のように説明編巣を0以上の整数で数値化する。

|  |  |
| --- | --- |
| 説明変数（定性データ） | 説明変数（カレゴリデータ） |
| 性別 | 男：0, 女：1 |
| 文理 | 文系：０,　理系：１ |

これで定性データを定量データにする変換が完了

|  |  |  |
| --- | --- | --- |
| 満足度 | 性別 | 文理 |
| 5 | 0 | 1 |
| 3 | 1 | 1 |
| 4 | 0 | 0 |
| 1 | 1 | 0 |
| 2 | 0 | 0 |

## 【数量化1類　＝　重回帰分析　＋　ダミー変数】

数量化1類はダミー変数を導入した重回帰分析と等価になる。

ダミー変数を用いて先ほどの例を表にすると以下になる。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 満足度 | 男 | 女 | 文系 | 理系 |
| 5 | 1 | 0 | 0 | 1 |
| 3 | 0 | 1 | 0 | 1 |
| 4 | 1 | 0 | 1 | 0 |
| 1 | 0 | 1 | 1 | 0 |
| 2 | 1 | 0 | 1 | 0 |

※１なら真、0なら偽になる。

上記のように、カテゴリデータが各カラムになることをダミー変数という。

**カテゴリデータを用いて、機械学習を行うと定量データからカレゴリーデータに人為的に変換しているため、機械学習をする際に意図しない意味をデータに与えてしまうことがあり上手くいかないことがあるため、ダミー変数に変換して機械学習を行うのが好ましい。**