

# ロジスティック回帰入門

FooQoo

2020 年 1 月 14 日

## 1 概要

ロジスティック回帰は以下の式で予測を行う。

$$p = \frac{1}{1 + \exp(-(a_1x_1 + a_2x_2 + \cdots + a_nx_n + b))} \quad (1)$$

$x_i$  でどんな値をとっても目的変数が 0 から 1 の間に収まる。 $a_i$  に対する  $x_i$  は偏回帰係数である。偏回帰係数は、最小二乗法または最尤法によって求める。式 (4) による式変形によって最小二乗法を適用できる。

$$1 - p = \frac{\exp(-(a_1x_1 + a_2x_2 + \cdots + a_nx_n + b))}{1 + \exp(-(a_1x_1 + a_2x_2 + \cdots + a_nx_n + b))} \quad (2)$$

$$\frac{p}{1 - p} = \exp(a_1x_1 + a_2x_2 + \cdots + a_nx_n + b) \quad (3)$$

$$\ln \frac{p}{1 - p} = a_1x_1 + a_2x_2 + \cdots + a_nx_n + b = l \quad (4)$$

$\ln \frac{p}{1 - p} = y'$  とすれば、重回帰分析と同じように最小二乗法を用いて偏回帰係数を求めることができる。 $l$  のことをロジットと呼ぶ。ロジットに対して指数をとるとオッズが得られる。

$$\exp(l) = \frac{p}{1 - p} \quad (5)$$

オッズは、事象が発生する確率と発生しない確率の比になっている。

## 2 説明変数の解釈

$e^{a_1}, e^{a_2}$  をオッズ比と呼ぶ。オッズ比が大きいほど、目的変数に対する影響度が高いといえる。ただし、オッズ比から倍率の解釈はできない。例えば、喫煙本数が 1 日に 1 本増えると、不健康になるリスクが 1.36 倍になるような解釈である。またオッズ比が 1 を下回ることがあるが、この場合は逆数を計算することで、その反対の解釈を行うこともできる。

説明変数が 1 つの場合は、ロジスティック回帰のオッズ比とクロス集計から算出されるオッズ比は一致する。ロジスティック回帰の説明変数が 2 つ以上のオッズ比を「調整したオッズ比」、1 つを単にオッズ比と呼ぶ。