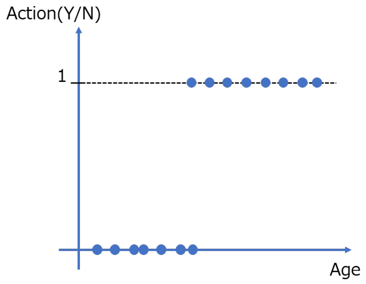
■ロジスティック回帰

確率の答えを2つに分ける回帰分析の手法。
難しく表現すれば、**ロジスティック解析**のモデルは、複数の要因（説明変数）から2値の結果（目的変数）の発生確率を予測する手法、といえます。

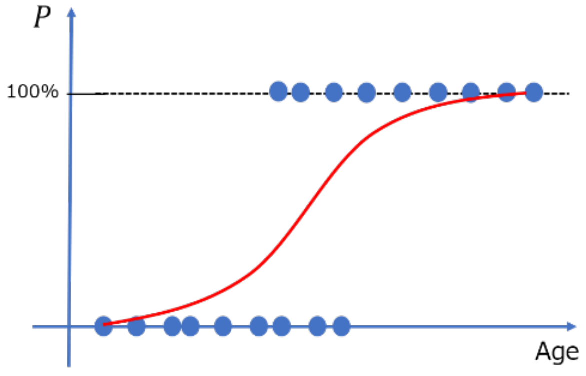
「2値」とは、「賛成・反対」、「合格・不合格」など、答えが2つのどちらかに当てはまるものです。ロジスティック解析では他の回帰分析と違って「-0.2」や「1.2」などの値はでず、常に「0から1」の範囲で判断ができます。

ここで使われるのが「**ロジット変換**」。ロジット変換とは、成功の確率を失敗の確率で割るオッズの自然対数です。このオッズを関数で表したものがロジット関数で、ロジット関数[1]から「ロジスティック回帰モデル」が作成できます。

例：ある商品についての意見でYes/Noで年齢別にアンケートを集計した。



このグラフから、なんとなく年齢と賛否に関係性が見えてくる。
それを「シグモイド関数」を使って、なだらかな線を求める。



[1] ロジット関数

logit（p）=log（p/1-p）

【ロジスティック回帰モデルの計算式】（ロジスティック関数）

p=1/1+exp（-（b0+b1×x1+b2×x2+b3×x3））

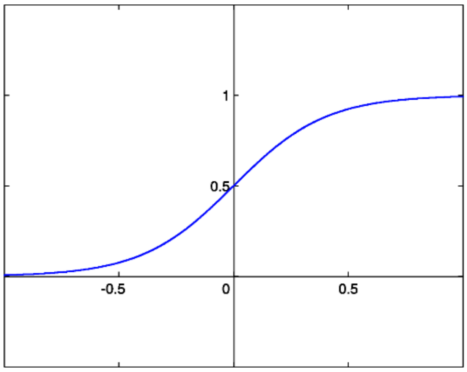
※pは目的変数、x1～x3は説明変数、b0は切片、b1～b3は回帰係数、expは自然対数の底を表します。

＜補足＞シグモイド関数

sigmoid function

$$\varsigma_a(x) = \frac{1}{1+e^{-ax}} = \frac{\tanh(ax/2)+1}{2}$$

で表される実関数である。



グラフは下のように 0＜ y ＜ 1 の範囲で滑らかな曲線となる。

シグモイド関数は x の値に何を入れても出力値が必ず 0 より大きく 1 より小さい値に変換されるため、ちょうど確率のように 0 から 1 の値を取るため二値分類を行うのに当てはまりが良い予測モデルを作成できます。

■回帰分析

変数 x と変数 y の平均的な値の間に関数関係を想定した分析のことを**回帰分析**と言います。
（1 次関数の場合、統計検定 2 級レベル）

x によって y の平均的な値を説明するという関係があるので，x を「**説明変数**」，y を「**被説明変数**」、または、「**目的変数**」と呼ぶこともある。

たとえば、ある会社の広告費を x 億円，売上を y 億円として，x によって y がどのように説明できるかを調べることが考えられます。過去のデータから，おおむね y＝2x＋30 という関係があることがわかれば，広告費を 1 億円増やせば，売上が平均して 2 億円増えると予測できる。

◆単回帰分析と重回帰分析

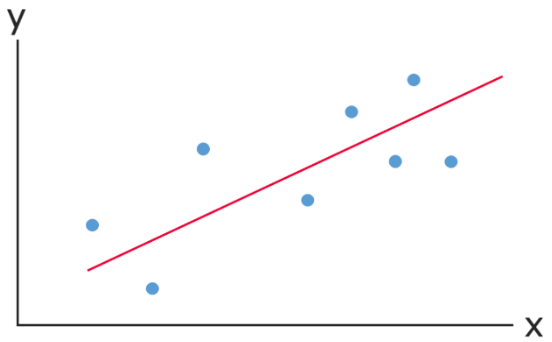
「説明変数」が 1 つのときを**単回帰分析**と言う。

$$y = a + bx$$

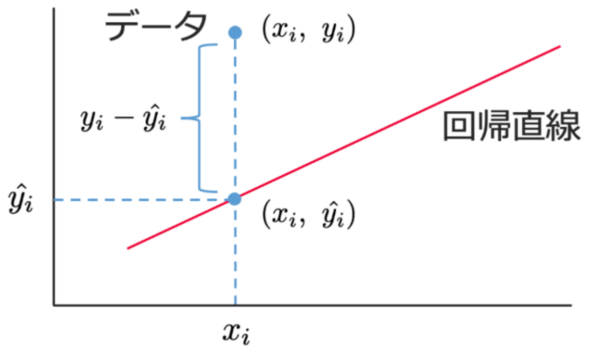
「説明変数」が 2 つ以上のときを重回帰分析と言う。

$$y = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

単回帰分析では、与えられたデータからもっとも適合する直線「**回帰直線**」を見つけ出す。



$$y_i - \hat{y}_i = y_i - (a + bx_i)\ (i = 1, 2, \cdots, n)$$



※詳しくは「統計」の時間で説明します