

研究員 の眼

当てはまりと滑らかさのバランス 相反するものの中で、どのように折り合いをつけるか？

保険研究部 主任研究員 篠原 拓也
(03)3512-1823 tshino@nli-research.co.jp

自然科学、社会学、経済学などの研究を行う際に、実験や調査を行って、データを収集する場合があります。これをもとに、将来の予測や、仮説の検証などを行うのだが、ここで1つ問題がある。通常、データにはノイズと呼ばれる不要な情報が含まれている。データをそのまま用いると、ノイズが悪影響を及ぼして、予測や検証の精度を下げてしまう場合がある。そこで、ノイズを除去するために、「モデル化」が行われる。これにより、データの背後にあるパターンを捉えることができる。

生命保険では、保険料の計算にあたり、死亡率が用いられる。死亡率は、ある人が今後1年間に死亡する確率を表すもので、通常、性別や年齢ごとに設定される。設定の際は、一定の規模の人々からなる群団の、1年後の死亡者数をもとに、死亡率のデータを収集する。年齢によって、収集できるデータの量が違うため、データの信頼度も異なる。例えば、100歳の高齢者は、人数が限られるため、データの信頼度が低く、ノイズが入り込みやすい。そこで、死亡率のモデル化が行われる。

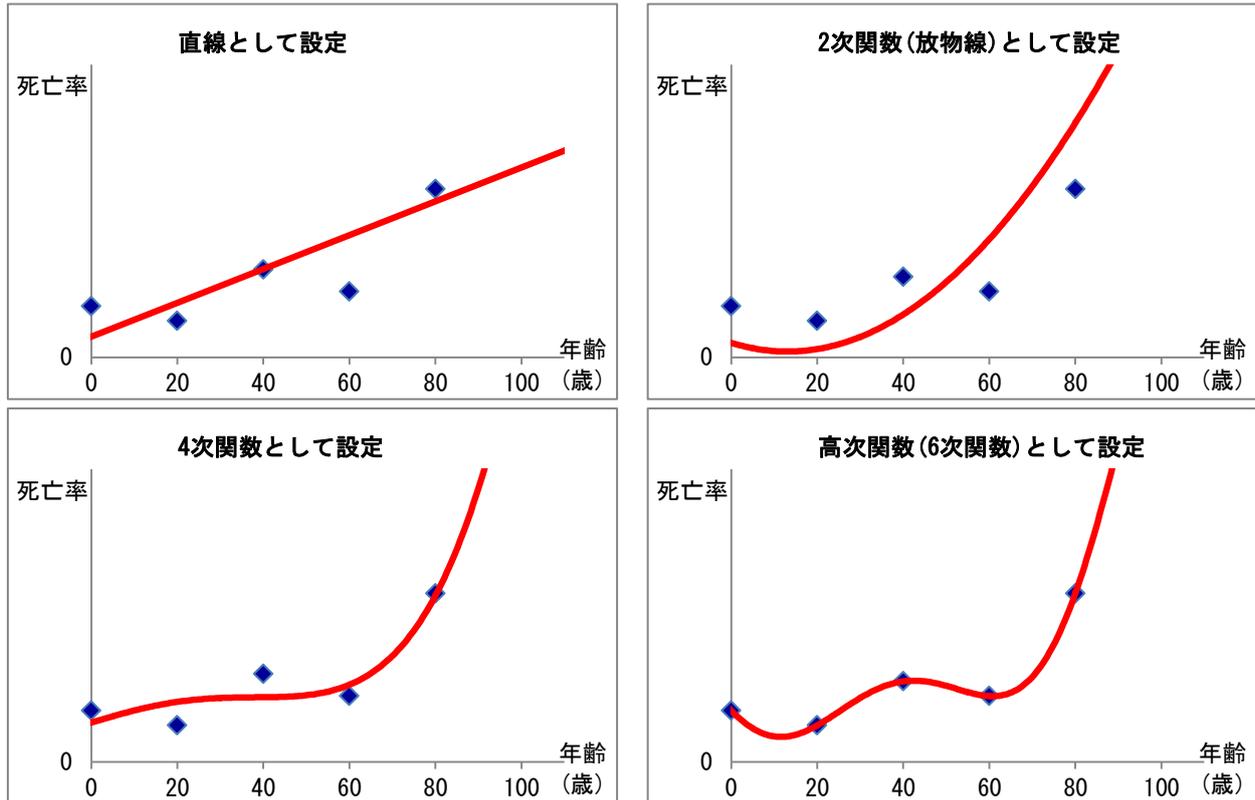
具体的には、年齢を指定すると、その年齢の死亡率を返すような算式を設定する。年齢を横軸、死亡率を縦軸にとったグラフを描くと、これは一つの曲線を表す。そこで、この算式は、死亡率曲線と呼ばれる。どのように死亡率曲線を引くべきかは、生命保険に限らず、将来の人口推計等を扱う数理人口学の分野で重要なテーマであり、多くの人口学者やアクチュアリーが研究を進めている。

死亡率曲線を検討する際には、当てはまり(英語では、fitness)と、滑らかさ(同、smoothness)のバランスを、どのようにとるかが大きな問題となる。

例えば、死亡率曲線を直線として設定しようとする、高齢において死亡率が年齢とともに急激に上昇することを、うまく表現できない。そこで、2次関数、つまり放物線として設定すると、直線よりはましになるが、ずれが残る。それでは3次曲線で、それでもだめなら4次曲線で、…というように、より高次の曲線を考えていくことになる。

死亡率曲線の設定イメージ

(◆ … 得られたデータ、— … 死亡率曲線)



通常、高次関数を使うとデータへの当てはまりは高まる。しかし、単にデータの間をつないだ曲線となり、ノイズが凹凸として残ることもある。これでは、将来の予測や、仮説の検証には使いにくい。

そこで、滑らかさを追求する視点が大切になる。多くのデータを、いかに次数の低い単純な曲線として設定するかという視点である。つまり、どこまで細かさを反映して、どこから細かさを省くべきかという、バランスの問題が重要となってくる。

このバランスをとる上で、ペナルティー要素を用いた評価という考え方がある。通常、曲線の設定は、データへの当てはまり度合いを計量して、評価をしながら行う。この評価の中に、ペナルティー要素を付け加える。ペナルティー要素は、曲線の凹凸が大きいと、罰として大きなマイナスの値をとる。このマイナスを加味した上で、評価が高くなるように、曲線を設定していく。

このようなことは、単に、モデル化の話にとどまらない。例えば、企業経営戦略における「攻め」と「守り」や、ビジネスの交渉での「強硬」と「柔軟」など、2つの相反するもの間で、どのように折り合いをつけるか、ということは、日常でよく問題となる。

その際、ペナルティー要素を用いた評価のような考え方は、有効と考えられるが、いかがだろうか。