

参考表 2007年「賃金構造基本統計調査」による賃金関数

| | | | | | |
|---------------|------------------------|------------------------------|---------------------------|--------------------|------------------------|
| [所定内賃金(千円)] = | 134.3 | - 1.334 × [年齢] | + 10.706 × [勤続年数] | + 47.136 × [男性ダミー] | + 148.640 × [大卒・院卒ダミー] |
| | (13.072) | (-4.436) | (15.621) | (8.650) | (22.911) |
| | + 59.721 × [高専・短大卒ダミー] | + 11.392 × [1,000人以上規模企業ダミー] | + 6.753 × [100~999人規模ダミー] | | |
| | (9.204) | (1.700) | (1.057) | | |
| | | | | 自由度調整済み決定係数=0.836 | () 内は t 値 |

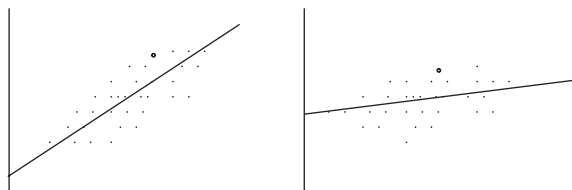
夏になりましたので、今回は怪奇（回帰）のお話をしたいと思います。年齢や勤続年数、学歴といった働く人々の属性別の賃金を調査し、その構造を明らかにするものに厚生労働省「賃金構造基本統計調査」があります。筆者などは、その発表があるとすぐ、その基本的な結果表を使って参考表のような回帰式を求めたくになります。

回帰式は、式の左辺にあるデータの実際の値と右辺の式で算出される値との差（誤差）が最小になるように右辺の各係数を求めたものです。参考表の右下にある決定係数が回帰式のあてはまりの良さを示すもので、今回の回帰式はいつもよりあてはまりがよいとはいえませんが、それでも80%強のあてはまりがあります。右辺の各項に適当な数値（たとえば、42歳で勤続18年、男性で……）を入れれば、そうした属性を持つ人（ここではパートタイム労働者などではない一般の雇用者）の所定内賃金が、当たらずといえども遠からずの値で知ることができるといことになります。ダミーには、それに該当すれば「1」が入り、該当しなければゼロ、すなわちその項はないものとして計算します。統計表も便利ですが、この式も使いこなせればかなり使えます。また、男性ダミーの係数をみれば「47.136」となっていますが、これは、年齢や勤続、学歴などが同じであっても、女性の所定内賃金は、月で5万円ほど男性よりも低い傾向が平均的にはみられると解釈することもできます。

さて、今回とくに解説しておきたいのが、回帰式右辺の各項の下に書き添えてある（ ）内の数値で、「t 値」と呼ばれるものです。数学的・統計的に厳密な解説は専門書に譲るとして、ここでは、統計ユーザーとしての筆者がイメージ的に理解しているところを紹介して、読者の参考にしていただきたいと思います。

イメージ図をご覧ください（図）。左右2つのグラフがあり、それぞれの点が得られたデータを示してしましましょう。縦軸が賃金で横軸が何らかの属性（年齢など）であるとイメージしていただいても結構です。ただし、このグラフはあくまで仮のものです。このデータを使って回帰式を推定して、それぞれグラフの直線が求められたとします。

図 t 値理解のイメージ図



さて、いつぞやお話ししましたように、統計とは一部のデータから全体を推し量るものといえます。得られたデータはあくまで一部のデータでしかありません。ある意味で「たまたま」得られたデータなのです。したがって、今回は得られてはいるけれども、たとえばグラフの白マルで示したデータが得られなかった可能性もあるわけです。そうであったら、回帰式の直線はどうなるでしょうか。左側のグラフの場合、傾きは若干小さくなるでしょうが右上がりの直線のままでしょう。しかし、右側のグラフの場合は微妙で、もしかしたら水平に、ひょっとしたら右下がりになってしまうかもしれません。そうすると、この回帰式で2つのデータ（変数）の関係を示すのはなんとも心許ないというほかありません。この「心許ない」ということを、統計用語では「統計的に有意でない」といいます。

「t 値」は、この「心許ない」かどうかを示すものです。普段の実務では、大体「2」以上あれば「大丈夫」と考えてよいでしょう。ただし、「2」未満だからといって必ずしも駄目というわけでもありません。「心許なさ」とうまく付き合っていくのも人生の醍醐味です。

もう少し先に進んで、p 値というものがあります。t 値から回帰式の係数が「ゼロ」になってしまう確率を求めたものです。上述の「2」以上というのは、「ゼロ」になるのは20回のうち1回程度の低い確率しかないというものです（5%基準）。ちなみに、冒頭の回帰式で、「1,000人以上規模ダミー」の t 値は1.700ですが、この場合「ゼロ」になる確率（p 値）は9%です。10回に1回もありません。これを「使える」と考えるかどうかは、分析者の判断にかかっています。

（労働政策研究・研修機構 主席統括研究員 浅尾 裕）