

調査における統計手法

巻頭言

回帰分析はデータ解析にからむ多くの分野で用いられる重要な方法である。なかでも、直線回帰は回帰分析という概念を学習するときにはじめに登場する重要な役割を担っている。物理学科の学生であった頃に、実験の結果に直線を当てはめるといふ実験実習があった。たとえば、 x を温度とし、 y を金属棒の長さとする。一定の温度に保ちながら棒の長さを数回測り、この温度での測定値の平均と標準偏差を求める。これを、別の温度でも行い、測った温度ごとに、グラフ用紙に平均値を表す点とそれを中心として、 y 軸方向に測定値の標準偏差を表す直線を描く。これで、温度を変えたときの、棒の長さの測定値の平均値と標準偏差がグラフ上に描かれる。最後に、各温度 (x) での棒の長さ (y) の平均値と標準偏差を表す直線を概ねカバーするような直線を描いて、温度 (x) による金属棒の長さ (y) を示す直線、 $y = a + bx$ を求める。この求め方が最小 2 乗法とほぼ同じ結果になると教わった。

直線の傾き b は“棒の長さの変化/温度変化”であるから、1 度当たりの金属棒の膨張率 ($\text{mm}/^\circ\text{C}$) を表す。したがって、 b の逆数、 $1/b$ は“棒の変化当たりの温度変化” ($^\circ\text{C}/\text{mm}$) を表す、と多くの学生は思っていたに違いない。これは、正確ではないが、ほぼ正しい。物理学の実験では多くのデータがほとんど直線の上に乗るような法則を確かめる実験を行っているからである。しかし、これが

統計数理研究所統計思考院特命教授 馬場 康維

生物学的なデータになるとどうか。多くの人の身長体重を計測し身長と体重の関係を求めたとする。 x を身長 (cm)、 y を体重 (kg) とし、身長から体重を推測するために、最小 2 乗法で求めた式を $y = a + bx$ とする。逆に体重 (y) から身長 (x) を推測するための式、 $x = g + hy$ を最小 2 乗法で求めたたとすると、 $h = 1/b$ にはなっていない。これは、最小 2 乗法により求めた身長から体重を推測する式を、逆の関係にそのまま使ってはいけないことを示している。

物理実験のケースは相関が 1 になるはずの法則を確かめる実験であり、直線関係そのものが重要である。一方、身長体重のケースは、ある身長の人々の平均体重はこれくらいという目安を与える直線を求めている。この場合は、そこからのずれ = 個体差に注目しており、同じ直線の当てはめでも違った意味をもっている。

調査でも様々な統計的な手法が使われている。つくられた当初は適用の条件を厳密に検討したはずのものでも、ソフトウェアが手軽に使えるようになり、多くの分野で、多くの人が使っているうちに、手法を使うための条件が忘れられ手法だけが一人歩きしているものもあるに違いない。時には、原点に立ち返り、手法を見直すことも必要であろう。