抵抗直線

回帰直線は従属変数と独立変数の分散・共分散に基づいて求められるので，外れ値が存在した場合にはそれに引きずられて直感的に考える傾向線が得られないことがある。言い方を変えれば，外れ値に傾向線が近づくため外れ値であることが認識しにくくなる。これに対して，抵抗直線は独立変数の大きさの順に 3 分割し，各区分中の従属変数の中央値の近辺を通る傾向線を求めるので，外れ値の影響を受けにくい（外れ値に対して抵抗性があることから抵抗直線と名付けられた）。また，このような性格を持つので，傾向線から離れた位置にある外れ値を認識しやすくなる。

n 組のデータ（X_i，Y_i）（$i = 1, 2, \ldots, n$）に対して，$Y = a_r X + b_r$ というモデル（抵抗直線）を考える。

抵抗直線の係数 a_r，b_r は以下のような繰り返し計算で求める。

1. n 組のデータを X の小さい順に並べ替え，$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ とする $(X_1 \leq X_2 \leq \ldots \leq X_n)$。
2. n 組のデータを 3 つのグループに分ける。例えば，$n = 3k$ の場合には，以下のように区分する。

$$\begin{cases} (X_1, Y_1), \ldots, (X_k, Y_k) & \text{グループ } I \\ (X_{k+1}, Y_{k+1}), \ldots, (X_{2k}, Y_{2k}) & \text{グループ } II \\ (X_{2k+1}, Y_{2k+1}), \ldots, (X_{3k}, Y_{3k}) & \text{グループ } III \end{cases} \tag{1}$$

3. $a_r = b_r = 0$ とおく。
4. I，II，III グループでの X の中央値を求め，それぞれ M_{x1}，M_{x2}，M_{x3} とする。
5. I，II，III グループでの Y の中央値を求め，それぞれ M_{y1}，M_{y2}，M_{y3} とする。
6. a_r，b_r の修正量 Δ_a，Δ_b を，（2）式により求める。

$$\begin{cases} \Delta_a = \dfrac{M_{y3} - M_{y1}}{M_{x3} - M_{x1}} \\[2ex] \Delta_b = \dfrac{(M_{y1} - \Delta_a M_{x1}) + (M_{y2} - \Delta_a M_{x2}) + (M_{y3} - \Delta_a M_{x3})}{3} \end{cases} \tag{2}$$

7. $\Delta_a \fallingdotseq 0$ ならば，求められた a_r，b_r が抵抗直線の傾きと切片である。
8. 修正量が十分に小さくないときは，（3）式により a_r，b_r を修正する。

$$\begin{cases} a_r \leftarrow a_r + \Delta_a \\ b_r \leftarrow b_r + \Delta_b \end{cases} \tag{3}$$

9. Y_i を残差 $Y_i - (\Delta_a X_i + \Delta_b)$ で置き換え，5. から繰り返す。

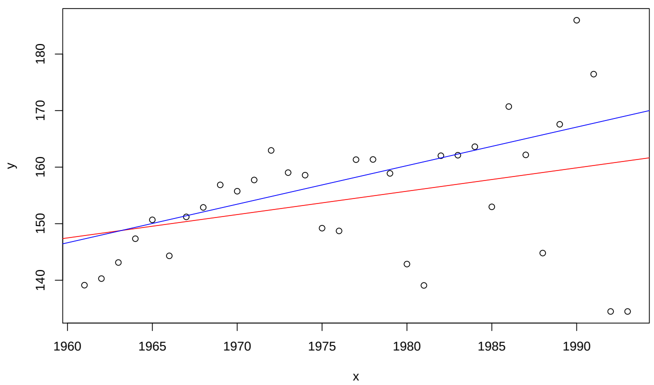

図 1　抵抗直線の例（青が抵抗直線，赤は回帰直線）