

中心を通ると思われる直線とデータの差を何とというか。

【解説】

□解説ビデオクリップ



データの中心を通ると思われる直線と値との差を**残差** (residual)あるいは、**誤差** (error、略語は e)と呼ぶ。データを現実の値とすれば、直線上にある点は推定された値となる。ここでは右図のように縦方向(データ Y)のズレに注目する。その関係を式で表現すれば、以下のようになる。

$$\text{現実値} - \text{推定値} = \text{推定誤差(残差)} e$$

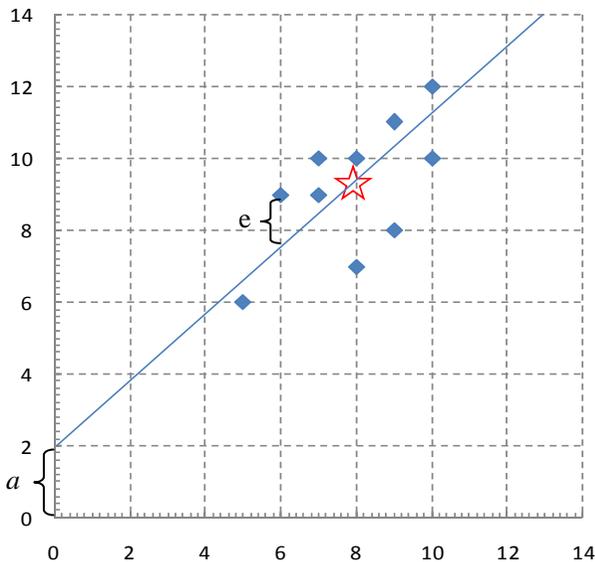
推定誤差 e を 0 に近づけることは、推定誤差が小さくなることなので、中心を通る直線の基準として望ましい。そこで、各推定誤差の和 ($\sum e$) は、以下のような S で表すことができる。

$$S = e_1 + e_2 + e_3 + e_4 + e_5 + e_6 + e_7 + e_8 + e_9 + e_{10}$$

ただし、この式の右辺にある各 e の符号は正と負まちまちである。符号を統一するために各 e を2乗し、符号をプラスに揃える。推定誤差を平方した値 (e^2) の合計 ($\sum e^2$) は次のようになる。

$$S^2 = e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2 + e_6^2 + e_7^2 + e_8^2 + e_9^2 + e_{10}^2$$

この S^2 を**残差平方和** (residual sum of squares) という。



【関連問題】

年 月 日

1. 先の設定で $a=2, b=1$ のとき、10 個のデータの残差を求めなさい。
2. そのときの残差の合計 ($\sum e$) を求めなさい。
3. そのときの残差平方和 ($\sum e^2$) を求めなさい。