

1 手で学ぶ「回帰直線」

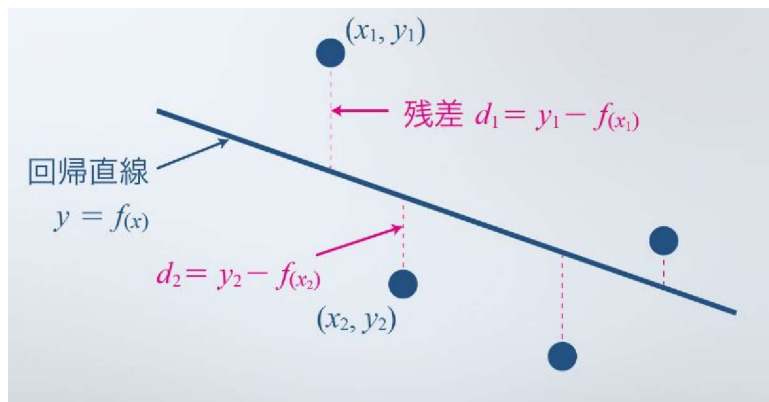
【回帰直線】

説明：海城中高_スプレッドシート実習 より

相関がある散布図において、各プロットに最も近い直線を回帰直線といいます。回帰直線は $y = ax + b$ と表し、 x を説明変数、 y を目的変数、傾き a と切片 b を回帰係数といいます。この方程式を用いることで、未観測のデータの推測や要因の分析をすることができます。

【最小二乗法】

x の値によって求められる直線上の値 y と実際の値との差を残差といいます。各プロットの残差を2乗した値の総和が最小になるような係数を求めることで、回帰式を求めることができます。このような方法を最小二乗法といいます。



このようにして求めた回帰式が、どれくらいの精度でデータに当てはまるのかを表す指標を決定係数 R^2 といい、0 (あまり当てはまらない) ~ 1 (よく当てはまる) の範囲で値をとります。最小二乗法によって得られた回帰直線の場合、相関係数の二乗と決定係数は一致します。

* ----- * ----- * ----- * ----- * ----- *

以下では回帰直線の方程式 $y = ax + b$ の回帰係数 a, b を、2通りの手法で説明してみます。

《1》 微分 (偏微分) を用いて説明するもの

《2》 数学 I の範囲で説明するもの

なお、分散や相関係数などの記号は現在の教科書で使われているものにします。

$$\text{平均 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\text{分散 } s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2, \quad s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - (\bar{y})^2$$

$$\text{共分散 } s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y} \quad \text{標準偏差 } s_x, s_y$$

$$\text{相関係数 } r = \frac{s_{xy}}{s_x s_y}$$

回帰直線 $y = ax + b$ は

$$y - \bar{y} = a(x - \bar{x}),$$

$$a = \frac{s_y}{s_x} \cdot r_{xy} = \frac{s_y}{s_x} \cdot \frac{s_{xy}}{s_x s_y} = \frac{s_{xy}}{s_x^2}, \quad b = \bar{y} - a \bar{x}$$