最小二乗法による回帰直線と相関係数の求め方—回帰分析と相関分析の基礎—

① 検量線(calibration curve)

未知試料中のある成分(物質，元素)の濃度を求める（定量分析と言う)には，その成分の存在量または濃度と，その成分の性質に基づいた測定値(吸光度，発光強度，比導電率など)との関係をあらかじめ求めておくことが必要になる．この関係を示す線を検量線と言う．通常，検量線による定量は直線性の得られる範囲で行なう．

② (積率)相関係数(correlation coefficient)

測定点(例えば，濃度と吸光度の関係)が直線にどれくらい一致しているかは，積率相関係数(単に相関係数と呼ぶことが多い) r を計算することにより検討することができる．

例えば，濃度(x)に対する吸光度(y)とすると，相関係数は下式により求めることができる．

$$r = \frac{S(xy)}{\sqrt{S(xx) \cdot S(yy)}}$$

ここで、
$$\begin{cases} S(xy) = \sum (x - \bar{x})(y - \bar{y}) \\ S(xx) = \sum (x - \bar{x})^2 \\ S(yy) = \sum (y - \bar{y})^2 \end{cases}$$

$\bar{x}$: x の平均値　　　$\bar{y}$: y の平均値

r は-1≦r≦1 の範囲の値をとる．r=1 は完全な正の相関を表わし，すべての測定点は正の傾きを持つ直線(回帰直線)上にあることを意味する．同様に r=-1 は完全な負の相関を表わす．0.90<r<0.95 の検量線は中程度，0.95<r<0.99 は良い直線性，0.99<r は優れた直線性を示している．精度を確保するためには， r 値が 0.99 以上の検量線を使用する必要があり，0.90 以下の r 値を示す検量線(回帰直線)を定量に用いてはならない．

③ 回帰直線(regression line)

相関係数の計算により直線性が得られることを確認した上で，次式の最小二乗法により回帰直線を求めることができる．

$$y = bx + a$$
$$b = \frac{S(xy)}{S(xx)}$$
$$a = \bar{y} - b\bar{x}$$

<付録>

最小二乗法の原理

最小二乗法とは，“偏差の平方の和が最小となる”ように，直線の傾き b と切片 a を決定する方法である．ここでは，未知数 a，b は以下のように求める．

測定点を(x_i, y_i)とすると，x_iから y_iを推定したときの偏差ε_i は，次式によって表される．

$$\varepsilon_i = y_i-(a+bx_i) \qquad ①$$

ここで，残差の二乗和を

$$Q=\Sigma\varepsilon_i^2=\Sigma\{y_i-(a+bx_i)\}^2 \qquad ②$$

と置き，これを最小にするように a, b を定める．これを求めるためには，$\partial Q/\partial a$，$\partial Q/\partial b$ を求め，得られた式を 0 とおいて，a と b について解けばよい．

$$(\partial Q/\partial a)_b=-2\Sigma\{y_i-(a+bx_i)\}=0 \qquad ③$$

$$(\partial Q/\partial b)_a=-2\Sigma x_i\{y_i-(a+bx_i)\}=0 \qquad ④$$

この 2 式は次のように変形される．

$$na+b\Sigma x_i=\Sigma y_i \qquad ③'$$

$$a\Sigma x_i+b\Sigma x_i^2=\Sigma x_iy_i \qquad ④'$$

この 2 式から，b と a は次のように求まる．

$$a=\bar{y}-b\bar{x} \qquad ⑤$$

$$b=\Sigma\{(x_i-\bar{x})(y_i-\bar{y})\}/\Sigma(x_i-\bar{x})^2 \qquad ⑥$$

ここで，$\bar{x}$と$\bar{y}$はそれぞれ x，y の平均値である．

<⑥式の導出>

⑤式を④’式に代入

$$(\Sigma y_i/n-b\Sigma x_i/n)\Sigma x_i+b\Sigma x_i^2=\Sigma x_iy_i$$

$$b=\{\Sigma x_iy_i-\Sigma x_i\Sigma y_i/n\}/\{\Sigma x_i^2-(\Sigma x_i)^2/n\}$$

$$=\{\Sigma x_iy_i-\Sigma x_i\Sigma y_i/n-\Sigma x_i\Sigma y_i/n+\Sigma x_i\Sigma y_i/n\}/\{\Sigma x_i^2-(\Sigma x_i)^2/n-(\Sigma x_i)^2/n+(\Sigma x_i)^2/n\}$$

ここで，分子$=\{\Sigma x_iy_i-(\Sigma x_i/n)\Sigma y_i-\Sigma x_i(\Sigma y_i/n)+\Sigma x_i\Sigma y_i/n\}$

$$=\{\Sigma x_iy_i-\bar{x}\Sigma y_i-\bar{y}\Sigma x_i+(n\bar{x})(n\bar{y})/n\}$$

$$=\{\Sigma x_iy_i-\bar{x}\Sigma y_i-\bar{y}\Sigma x_i+n\bar{x}\bar{y}\}$$

$$=\{\Sigma x_iy_i-\Sigma\bar{x}y_i-\Sigma\bar{y}x_i+\Sigma\bar{x}\bar{y}\}$$

$$=\Sigma(x_iy_i-\bar{x}y_i-\bar{y}x_i+\bar{x}\bar{y})$$

$$=\Sigma\{(x_i-\bar{x})y_i-(x_i-\bar{x})\bar{y}\}=\Sigma(x_i-\bar{x})(y_i-\bar{y})$$

分母$=\{\Sigma x_i^2-2(\Sigma x_i)^2/n+(\Sigma x_i)^2/n\}$

$$=\{\Sigma x_i^2-2(\Sigma x_i)(\Sigma x_i/n)+n(\Sigma x_i/n)^2\}$$

$$=\{\Sigma x_i^2-2(\Sigma x_i)\bar{x}+n\bar{x}^2\}$$

$$=\{\Sigma x_i^2-\Sigma(2\bar{x}x_i)+\Sigma\bar{x}^2\}$$

$$=\Sigma(x_i^2-2\bar{x}x_i+\bar{x}^2)=\Sigma(x_i-\bar{x})^2$$

と変形できるので，$b=\Sigma\{(x_i-\bar{x})(y_i-\bar{y})\}/\Sigma(x_i-\bar{x})^2$