残差平方和

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

単回帰の最小二乗和推定量の計算は単純な式変形で導くことができる。ポイントは，未知数についての 2 次式であると見て式変形を行うことである。まずは式の中でより単純な$\hat{\beta}_0$の方について整理して平方完成を行い，残った部分を$\hat{\beta}_1$について平方完成する。式変形の途中で，x や y の算術平均，標本分散，標本共分散の定義式を利用する。

$$\begin{aligned}
\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left(-\hat{\beta}_0 + y_i - \hat{\beta}_1 x_i\right)^2 &= \sum_{i=1}^{n} \left\{\hat{\beta}_0^2 - 2\left(y_i - \hat{\beta}_1 x_i\right)\hat{\beta}_0 + \left(y_i - \hat{\beta}_1 x_i\right)^2\right\} \\
&= \sum_{i=1}^{n} \hat{\beta}_0^2 - 2\sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 x_i\right)\hat{\beta}_0 + \sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 x_i\right)^2 \\
&= n\hat{\beta}_0^2 - 2\left\{\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i\right\}\hat{\beta}_0 + \sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 x_i\right)^2 \\
&= n\hat{\beta}_0^2 - 2n\left(\bar{y} - \hat{\beta}_1 \bar{x}\right)\hat{\beta}_0 + \sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 x_i\right)^2 \\
&= n\left\{\hat{\beta}_0 - \left(\bar{y} - \hat{\beta}_1 \bar{x}\right)\right\}^2 + \sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 x_i\right)^2 - n\left(\bar{y} - \hat{\beta}_1 \bar{x}\right)^2
\end{aligned}$$

$\hat{\beta}_0$の平方完成ができたので，後半の部分を$\hat{\beta}_1$について平方完成する。

$$\begin{aligned}
\sum_{i=1}^{n} \left(y_i - \hat{\beta}_1 x_i\right)^2 - n\left(\bar{y} - \hat{\beta}_1 \bar{x}\right)^2 &= \sum_{i=1}^{n} \left(y_i^2 - 2\hat{\beta}_1 x_i y_i + \hat{\beta}_1^2 x_i^2\right) - n\bar{y}^2 + 2n\hat{\beta}_1 \bar{x}\bar{y} - n\hat{\beta}_1^2 \bar{x}^2 \\
&= \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - 2\hat{\beta}_1 \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right) + \hat{\beta}_1^2 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \\
&= n\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2\right) - 2n\hat{\beta}_1 \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}\right) + n\hat{\beta}_1^2 \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2\right) \\
&= ns_y^2 - 2ns_{xy}\hat{\beta}_1 + ns_x^2\hat{\beta}_1^2 = n\left(s_y^2 - 2s_{xy}\hat{\beta}_1 + s_x^2\hat{\beta}_1^2\right) = n\left\{s_x^2\left(\hat{\beta}_1^2 - 2\frac{s_{xy}}{s_x^2}\hat{\beta}_1\right) + s_y^2\right\} \\
&= n\left\{s_x^2\left(\hat{\beta}_1 - \frac{s_{xy}}{s_x^2}\right)^2 + s_y^2 - \frac{s_{xy}^2}{s_x^2}\right\}
\end{aligned}$$

$\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n} {y_i}^2 - \bar{y}^2$などを使うのがコツである。
以上より，$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$，$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2} = r_{xy} \cdot \frac{s_y}{s_x}$ のときに残差平方和は最小値をとる。これで，標本データから予測式を求めることができる。

$$\hat{y}_i = \bar{y} - r_{xy} \cdot \frac{s_y}{s_x} \cdot \bar{x} + r_{xy} \cdot \frac{s_y}{s_x} \cdot x_i = r_{xy} \cdot \frac{s_y}{s_x}(x_i - \bar{x}) + \bar{y}$$

これは平面における直線の式なので**回帰直線**ともよぶ。式から分かるように，この直線は点$(\bar{x}, \bar{y})$を通る。また，x と y の標準偏差が等しいとき，回帰係数は積率相関係数に一致する。