

Think of an arbitrary line as being given by:

$$(y - \bar{y}) = c(x - \bar{x}) + d$$

Now look at the sum of the squares of the residuals of the n points:

$$\begin{aligned} \Sigma(y_i - c(x_i - \bar{x}) - (d + \bar{y}))^2 &= \\ &= \Sigma((y_i - \bar{y}) - c(x_i - \bar{x}) - d)^2 \\ &= \Sigma(y_i - \bar{y})^2 - 2c(x_i - \bar{x})(y_i - \bar{y}) + c^2(x_i - \bar{x})^2 \\ &\quad + \Sigma d^2 + 2cd(x_i - \bar{x}) - 2d(y_i - \bar{y}) \end{aligned}$$

Use the definitions of s_x^2 , s_y^2 , and r . Also use the fact that $\Sigma(x_i - \bar{x}) = 0$ and $\Sigma(y_i - \bar{y}) = 0$. Then the above sum of squares becomes

$$\begin{aligned} (n-1)(s_y^2 - 2crs_x s_y + c^2 s_x^2) + nd^2 &= \\ \rightarrow (n-1)((cs_x - rs_y)^2 + (1-r^2)s_y^2) + nd^2 \end{aligned}$$

Because squares are always non-negative, this sum is minimized when $d = 0$ and $cs_x - rs_y = 0$. In other words, the line minimizing the sum of the squares of the residuals is

$$(y - \bar{y}) = r \frac{s_y}{s_x} (x - \bar{x})$$

in agreement with our usual formulas.