1 Giving back to Regression

Now that we have correlation formula, this can be used to simplify our regression formula as well. Recalling the simple linear regression model formula (for both possible lines),

$$\hat{Y}|x = \hat{\beta}_0 + \frac{\mathrm{cov}(X,Y)}{s_X^2}x$$
$$\hat{X}|y = \hat{\beta}_2 + \frac{\mathrm{cov}(X,Y)}{s_Y^2}y$$

where

$$\hat{\beta}_0 = \overline{y} - \frac{\mathrm{cov}(X,Y)}{s_X^2}\overline{x}$$
$$\hat{\beta}_2 = \overline{x} - \frac{\mathrm{cov}(X,Y)}{s_Y^2}\overline{y}$$

Substituting, $r = \dfrac{\mathrm{cov}(X,Y)}{s_X s_Y}$,

$$\begin{aligned}\hat{Y}|x &= \overline{y} - \frac{\mathrm{cov}(X,Y)}{s_X^2}\overline{x} + \frac{\mathrm{cov}(X,Y)}{s_X^2}x \\ &= \overline{y} + \frac{\mathrm{cov}(X,Y)}{s_X^2}(x - \overline{x}) \\ &= \overline{y} + \frac{r s_X s_Y}{s_X^2}(x - \overline{x}) \\ &= \overline{y} + r\frac{s_Y}{s_X}(x - \overline{x})\end{aligned}$$

Similarly,

$$\begin{aligned}\hat{X}|y &= \overline{x} - \frac{\mathrm{cov}(X,Y)}{s_Y^2}\overline{y} + \frac{\mathrm{cov}(X,Y)}{s_Y^2}y \\ &= \overline{x} + \frac{\mathrm{cov}(X,Y)}{s_Y^2}(y - \overline{y}) \\ &= \overline{x} + \frac{r s_X s_Y}{s_Y^2}(y - \overline{y}) \\ &= \overline{x} + r\frac{s_X}{s_Y}(y - \overline{y})\end{aligned}$$

Regression Alternate Form Via Correlation

The true regression lines could also be expressed in terms of correlation as,

$$\begin{aligned}\hat{Y}|x &= \overline{y} + r\frac{s_Y}{s_X}(x - \overline{x}) \\ \hat{X}|y &= \overline{x} + r\frac{s_X}{s_Y}(y - \overline{y}) \qquad (1)\end{aligned}$$

Note above lines are not symmetric, their intercepts and slope differ. Symmetry happens after standardizing the dataset.