Sum of Corrected Squares of x_i $\qquad S_x^2 = \frac{1}{(n-1)} \Sigma (x_i - \bar{x})^2$ = Sample Variance of x

The other forms of S_{XX} and S_{XY}:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} = \sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i\,(x_i - \bar{x}) = (n-1)S_x^2$$

Corrected Sum of Cross Products of x_i and y_i

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y} = \sum_{i=1}^{n} (x_i - \bar{x})\,(y_i - \bar{y})$$

$$= \sum_{i=1}^{n} y_i\,(x_i - \bar{x}) = \sum_{i=1}^{n} x_i\,(y_i - \bar{y})$$

$$\boxed{\bar{x} = \frac{\sum^n x_i}{n}}$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^{n} x_i (x_i - \bar{x}) - \bar{x} \sum_{i=1}^{n} (x_i - \bar{x})$$

$$= \sum_1^n x_i (x_i - \bar{x}) - \bar{x} \cancel{\left[\sum_1^n x_i - n\bar{x} \right]}^{\,0} = \sum_{i=1}^{n} x_i (x_i - \bar{x})$$

$$= \sum_1^n x_i^2 - \bar{x} \underbrace{\sum_1^n x_i}_{n\bar{x}} = \sum_1^n x_i^2 - \bar{x}(n\bar{x})$$

$$\sum_1^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i (y_i - \bar{y}) - \bar{x} \boxed{\cancel{\sum_1^n (y_i - \bar{y})}^{\,0}} = \sum_1^n x_i (y_i - \bar{y})$$

$$= \sum_1^n x_i y_i - \bar{y} \sum_{i=1}^{n} x_i = \sum_1^n x_i y_i - \bar{y}\, n\bar{x}$$

Ex 2: Consider the following data set, in which the variables of interest are x = commuting distance (miles) and y = commuting time (min).

$\Sigma x_i y_i = 1765$

x	5	10	15	20	25
y	8	16	22	23	31

$\Sigma x_i = 75$, $\Sigma x_i^2 = 1375$

$\Sigma y_i = 100$, $\Sigma y_i^2 = 2294$

$$\Rightarrow \bar{x} = \frac{\sum_1^n x_i}{n} = \frac{75}{5} = 15$$

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{100}{5} = 20$$

$$S_{xx} = \Sigma x_i^2 - n\bar{x}^2 = 1375 - (5 \times 15^2) = 250$$

$$S_{xy} = \Sigma x_i y_i - n\bar{x}\bar{y} = 1765 - (5 \times 15 \times 20) = 265$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{265}{250} = 1.06$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 20 - (1.06 \times 15) = 4.1$$

$$\hat{y} = 4.1 + 1.06x$$ ← Fitted model.

$\Rightarrow$ As commuting distance increases by 1 mile, the average commuting time increases by 1.06 min.

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ $\bar{y} = \frac{\sum_1^n y_i}{n}$

Section 2.2.2: The properties of the least squares estimators:

1. They are linear functions of y_i. That is, they can be expressed as linear combinations of y_i.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_1^n (x_i - \bar{x}) y_i}{S_{xx}} = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_{xx}} \right) y_i = \sum_{i=1}^{n} c_i y_i \quad \text{where } c_i = \frac{x_i - \bar{x}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{\sum_1^n y_i}{n} - \left(\sum_{i=1}^{n} c_i y_i \right) \bar{x} = \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{x} c_i \right) y_i \quad ✓$$

2. They are unbiased estimators of regression coefficients.

   A statistic T is an unbiased estimator of parameter θ if $E(T) = \theta$

$$E(\widehat{\beta_0}) = \beta_0$$

$$E(\widehat{\beta_1}) = \beta_1$$

Proof:

$$E(\hat{\beta}_1) = E\left[\sum_1^n c_i y_i \right] = E\left[\sum_1^n \left(\frac{x_i - \bar{x}}{S_{xx}} \right) (\beta_0 + \beta_1 x_i + \varepsilon_i) \right]$$

by assumption

$$= \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_{xx}} \right) E\left[\beta_0 + \beta_1 x_i + \varepsilon_i \right] = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_{xx}} \right) \left[\beta_0 + \beta_1 x_i + \cancelto{0}{E(\varepsilon_i)} \right]$$

$$= \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_{xx}} \right) \beta_0 + \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_{xx}} \right) \beta_1 x_i = \frac{\beta_0}{S_{xx}} \cancelto{0}{\sum_1^n (x_i - \bar{x})} + \frac{\beta_1}{S_{xx}} \underbrace{\sum_1^n x_i (x_i - \bar{x})}_{S_{xx}}$$

$$E(\hat{\beta}_1) = \beta_1 \quad \blacksquare$$

At home: Show $E(\hat{\beta}_0) = E(\beta_0)$

and read section 2.2 (chapter) of textbook.