Technical Note 20.1

To Show that b_1, b_0, and $\hat{Y}$ are Linear Combinations of the Y Scores:

We start by showing that the expression for b_1 can be broken into two components, one of which can be shown to be 0, that is,

$$b_1 = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{SS_X} = \frac{\sum_i (X_i - \overline{X})Y_i}{SS_X} \text{ because } \overline{Y}\sum_i (X_i - \overline{X}) = 0 .$$

Therefore, b_1 can be expressed as a linear combination of the Y's; that is, $b_1 = \sum_i f_i Y_i$ where $f_i = \dfrac{X_i - \overline{X}}{SS_X}$.

Also,

$$b_0 = \overline{Y} - b_1\overline{X} = \frac{1}{N}\sum_i Y_i - \overline{X}b_1$$

$$= \frac{1}{N}\sum_i Y_i - \overline{X}\sum_i f_i Y_i = \sum_i \left(\frac{1}{N} - \overline{X}f\right) Y_i$$

so,

$$b_0 = \sum g_i Y_i \quad \text{where} \quad g_i = \frac{1}{N} - \frac{\overline{X}(X_i - \overline{X})}{SS_X}$$

and

$$\hat{Y}_j = b_0 + b_1 X_j = \sum_i g_i Y_i + X_j \sum_i f_i Y_i$$

$$= \sum_i \left(g_i + X_j f_i\right) Y_i = \sum_i \left(\frac{1}{N} - \frac{\overline{X}(X_i - \overline{X})}{SS_X} + \frac{X_j(X_i - \overline{X})}{SS_X}\right)$$

$$= \sum_i \left(\frac{1}{N} + \frac{(X_j - \overline{X})(X_i - \overline{X})}{SS_X}\right) Y_i$$

so that

$$\hat{Y}_j = \sum_i h_{ij} Y_i \qquad \text{where } h_{ij} = \frac{1}{N} + \frac{(X_j - \overline{X})(X_i - \overline{X})}{SS_X}$$

$$= h_{jj} Y_j + \sum_{i \neq j} h_{ij} Y_i \quad \text{where } h_{jj} = \frac{1}{N} + \frac{(X_j - \overline{X})^2}{SS_X}$$ is the leverage of case j.