6.7 Prove the relationship in (6.45), which relates the pairwise difference between examples in a dataset with the raw-score expression for the variance.

Solution. To prove:

$$\frac{1}{N^2}\sum_{i,j=1}^{N}(x_i - x_j)^2 = 2\left[\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N}x_i\right)^2\right]$$

$$L.H.S = \frac{1}{N^2}\sum_{i,j=1}^{N}(x_i - x_j)^2$$

$$= \frac{1}{N^2}\sum_{i,j=1}^{N} x_i^2 + x_j^2 - 2x_ix_j$$

$$= \frac{1}{N^2}\left(\sum_{i,j=1}^{N} x_i^2\right) + \left(\sum_{i,j=1}^{N} x_j^2\right) - \left(\sum_{i,j=1}^{N} 2x_ix_j\right)$$

$$= \frac{1}{N^2}\left(N\sum_{i=1}^{N} x_i^2\right) + \frac{1}{N^2}\left(N\sum_{j=1}^{N} x_j^2\right) - \frac{2}{N^2}\left(\sum_{i,j=1}^{N} x_ix_j\right)$$

$$= \frac{2}{N}\left(\sum_{i=1}^{N} x_i^2\right) - \frac{2}{N^2}\left(\sum_{i,j=1}^{N} x_ix_j\right)$$

$$= \frac{2}{N}\left(\sum_{i=1}^{N} x_i^2\right) - \frac{2}{N^2}\left(\sum_{i=1}^{N} x_i\sum_{j=1}^{N} x_j\right)$$

$$= \frac{2}{N}\left(\sum_{i=1}^{N} x_i^2\right) - \frac{2}{N^2}\left(\sum_{i=1}^{N} x_iN\mu\right)$$

$$= \frac{2}{N}\left(\sum_{i=1}^{N} x_i^2\right) - \frac{2\mu}{N}\left(\sum_{i=1}^{N} x_i\right)$$

$$= \frac{2}{N}\left(\sum_{i=1}^{N} x_i^2\right) - 2\mu\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)$$

$$= \frac{2}{N}\left(\sum_{i=1}^{N} x_i^2\right) - 2\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)$$

$$= \frac{2}{N}\left(\sum_{i=1}^{N} x_i^2\right) - 2\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2$$

$$= 2\left[\frac{1}{N}\left(\sum_{i=1}^{N} x_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2\right]$$

$$= R.H.S$$

Thus proved.