Appendix A

Proof that the statistic, $S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$, is the unbiased estimator of the population parameter, σ^2.

 Let $X_1, X_2, \ldots X_n$ be independent and identically distributed random variables, each with expected value, μ, and variance, σ^2. For the entire population, $\sigma = \sqrt{\sigma^2}$ and $\sigma^2 = E\left(X_i - \mu\right)^2$. The variance is the theoretical expected value for the square of the distance from the mean to an individual data point (e.g., deviation). Thus, we want a statistic such that $E\left(S^2\right) = \sigma^2$. Intuitively, we would guess that S^2 is $\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}$ (the average of our sample deviations). Drawing on the algebra of expected values, however, this statistic results in the following expected value:

$$E\left(S^2\right) = E\left(\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\left(\left(X_i - \mu\right) - \left(\overline{X} - \mu\right)\right)^2\right)$$

$$E\left(S^2\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\left((X_i - \mu)^2\right) - 2\sum_{i=1}^{n}(X_i - \mu)(\overline{X} - \mu) + \sum_{i=1}^{n}\left((\overline{X} - \mu)^2\right)\right) = \frac{1}{n}\left(\sum_{i=1}^{n}E\left((X_i - \mu)^2\right) - n \cdot E\left((\overline{X} - \mu)^2\right)\right)$$

Substituting $\sigma^2 = E\left(X_i - \mu\right)^2$ (initial assumption) and $E\left(\overline{X} - \mu\right)^2 = \frac{\sigma^2}{n}$ (from central limit theorem) results in the following:

$$E\left(S^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\sigma^2 - n \cdot \frac{\sigma^2}{n}\right) = \frac{1}{n}\left(n\sigma^2 - \sigma^2\right) = \frac{n-1}{n}\sigma^2 \cdot$$

Thus, the expected value of our statistic, S^2, is a *biased* estimate (a fraction) of σ^2. Because the *n* in the denominator originated from our initial guess, replacing *n* with *n-1* in the original formula would result in an unbiased estimate of our population parameter. So, $S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$ is the statistic that is an unbiased estimator of the desired population parameter, σ^2. ■