1) Fact: $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$.

Proof.

$$
\begin{aligned}
&\sum_{i=1}^{n}(x_i - \overline{x}) \\
=&\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} \\
=&\sum_{i=1}^{n} x_i - n\overline{x} \\
=&n\overline{x} - n\overline{x} \\
=&0
\end{aligned}
\tag{0.1}
$$

□

2) The computational formular for sample variance:

$$
\begin{aligned}
&\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\
=&\frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2) \\
=&\frac{1}{n-1}\{\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2x_i\overline{x} + \sum_{i=1}^{n} \overline{x}^2\} \\
=&\frac{1}{n-1}\{\sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + n\overline{x}^2\} \\
=&\frac{1}{n-1}\{\sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + n\overline{x}^2\} \\
=&\frac{1}{n-1}\{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\} \\
=&\frac{1}{n-1}\{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2\}
\end{aligned}
\tag{0.2}
$$

3) Chebyshev's rule: let k¿1. For any data set, the proportion of observations within k standard deviation of the mean lying in the interval $(\overline{x} - ks, \overline{x} + ks)$ is at least $1 - \frac{1}{k^2}$.

Proof. It is equivalent to prove the proportion of the tail observations outside the interval $(\overline{x} - ks, \overline{x} + ks)$ is less than or equal to $\frac{1}{k^2}$. Suppose the proportion of the tail is greater than $\frac{1}{k^2}$. Then the number of points is greater than $n\frac{1}{k^2}$. Let A denote the set of observations outside the interval $(\overline{x} - ks, \overline{x} + ks)$.

We have

$$\begin{aligned} s^2 =& \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\ \geq& \frac{1}{n-1}\sum_{x\in A}(x_i - \overline{x})^2 \\ \geq& \frac{1}{n-1}\sum_{x\in A}(ks)^2 \\ \geq& \frac{1}{n-1}\frac{n}{k^2}(ks)^2 \\ \geq& \frac{n}{n-1}s^2. \end{aligned} \tag{0.3}$$

□

This is a contradiction. Thus the proportion of tail set A has to be less than or equal to $1/k^2$.