Ch 9.4

SNEHINH SEN

Given, sample data are
$X = \{x_{ij} : 1 \le i \le k, 1 \le j \le n_i\}$

& $\bar{x} = \text{mean}(X) = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}$

Let $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$.

Hence

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left((x_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{x}) \right)^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + (\bar{x}_i - \bar{x})^2 + 2(x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x})$$

But observe,

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) = SSE \quad \& \quad \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2$$

$$= \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2$$

[As summand is independent of j]

$$= SST$$

Finally, $\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x})$

$$= \sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} x_{ij} (\bar{x}_i - \bar{x}) - \sum_{j=1}^{n_i} \bar{x}_i (\bar{x}_i - \bar{x}) \right)$$

$$= \sum_{i=1}^{k} \left(\left(\sum_{j=1}^{n_i} x_{ij} - n_i \bar{x}_i \right) (\bar{x}_i - \bar{x}) \right)$$ [As $\bar{x}_i - \bar{x}$ is ind. of j]

$$= \sum_{i=1}^{k} (n_i \bar{x}_i - n_i \bar{x}_i)(\bar{x}_i - \bar{x})$$ [By defn]

$$= \sum_{i=1}^{k} 0 (\bar{x}_i - \bar{x}) = 0.$$

Hence, adding everything up, we get

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2 = SSE + SST + 0$$

$$= SSE + SST$$

as desired.

[Note: this is very much similar to law of total variance, also called law of conditional variance, that is

$$Var(X) = E(\underbrace{Var(X|Y)}_{\text{corresponds to SSE}}) + Var(\underbrace{E(X|Y)}_{\text{corresponds to SST}})$$
]