ISYE 8803 HDDA

Homework 2

June 10th, 2021

1.

$We\ know\ based\ on\ statistics\ that\ if\ we\ have\ total\ variance\ and\ the\ within\ group\ variance\ than\ the\ between\ group\ variance\ is\ calculated\ as\ follows:$

$$SSB\ =\ SST\ -\ SSW$$

$$= \sum_{i=1}^{t}(x_i - \mu_\tau)^2 - \sum_{i=1}^{t}\sum_{j=1}^{k} w_{ij}(x_i - \mu_j)^2$$

$$= \sum_{i=1}^{t}\left(x_i^2 - 2x_i\mu_\tau - \mu_\tau^2\right) - \sum_{i=1}^{t}\sum_{j=1}^{k}\left(w_{ij}x_i^2 - 2w_{ij}x_i\mu_j - \mu_j^2\right)$$

$We\ can\ then\ factor\ out\ both\ summations\ in\ the\ equation:$

$$= \sum_{i=1}^{t}\sum_{j=1}^{k}\left(\frac{x_i^2 - 2x_i\mu_\tau - \mu_\tau^2}{k} - w_{ij}x_i^2 - 2w_{ij}x_i\mu_j - w_{ij}\mu_j^2\right)$$

$We\ can\ then\ use\ two\ very\ useful\ properties,\ which\ is:$

$$\sum_{i=1}^{t} x_i = t\mu_\tau$$

$Take\ two\ groups:$

$Group\ 1:\ \{1, 2, 3\}$

$Group\ 2:\ \{4, 5, 6\}$

$\mu_\tau = 3.5$

$t\ =\ 6$

$$\sum_{i=1}^{t} x_i = 21$$

$t\mu_\tau =\ 3.5 * 6$

$t\mu_\tau =\ 21$

$Similar\ property\ holds\ for:$

$$\sum_{i=1}^{t} w_{ij} = t_j$$

Now we can replace these terms in the original equation :

$$= \sum_{j=1}^{k}\left(\frac{1}{k}\sum_{i=1}^{t}x_i^2 - \frac{2\mu_\tau}{k}\sum_{i=1}^{t}x_i - \frac{t\mu_\tau^2}{k} - \sum_{i=1}^{t}w_{ij}x_i^2 - 2\mu_j\sum_{i=1}^{t}w_{ij}x_i - \mu_j^2\sum_{i=1}^{t}w_{ij}\right)$$

$$= \sum_{j=1}^{k}\left(\frac{2\mu_\tau(t\mu)}{k} + \frac{t\mu_\tau^2}{k} + 2\mu_\tau(t_j\mu_\tau) - \mu_j{}^2 t_j\right) + \sum_{j=1}^{k}\sum_{i=1}^{t}\left(\frac{x_i^2}{k} - w_{ij}x_i^2\right)$$

Which then simplifies to :

$$= \sum_{j=1}^{k}\left(t_j\left[\mu_j{}^2 - 2\mu_\tau\mu_\tau + \mu_j{}^2 + 2\mu_\tau\mu_\tau - \mu_j{}^2 - \mu_\tau^2\right]\right)$$

Then cancel the remaining terms and factor out the squared result and you have :

$$= \sum_{j=1}^{k}\left(t_j\left[\mu_j - \mu_\tau\right]^2\right)$$

Therefore, it can be shown that the result is equal to the between – cluster variation or SSB