appendix(付録)

1.　ロジスティック曲線の論理式（微分方程式の解法）の説明

このモデル式 $\dfrac{dy_t}{dt} = ay_t(K - y_t)$ の微分方程式を解いてみる。
両辺を $\dfrac{1}{(K-y_t)y_t}$ 倍かつ dt 倍すると $\dfrac{dy_t}{(K-y_t)y_t} = adt$ になるが，この式の両辺に
積分記号 $\int$ を導入すると $\int \dfrac{dy_t}{(K-y_t)y_t} = \int adt$ であり $c($ 積分定数 $)$ がつき，次式となる。

$$\int \frac{dy_t}{(K-y_t)y_t} = a\int dt + c = at + c \qquad (1.1)$$

ここで積分をするために左辺の $\int dy_t$ 積分記号部分を外し部分分数の型にすると，次式となる。

$$\frac{1}{(K-y_t)y_t} = \frac{A}{(K-y_t)} + \frac{B}{y_t} = \frac{Ay_t + B(K-y_t)}{(K-y_t)y_t} \qquad (1.2)$$

未知数 A, B を求める。
上式の右辺の分子を展開すると，次式となる。

$$\frac{1}{(K-y_t)y_t} = \frac{A}{(K-y_t)} + \frac{B}{y_t} = \frac{Ay_t + B(K-y_t)}{(K-y_t)y_t} = \frac{Ay_t + BK - By_t}{(K-y_t)y_t} \qquad (1.3)$$

分子を考え変形すると，$Ay_t + BK - By_t = y_t(A-B) + BK = 1$ は，右辺が 1 になるのは
$A - B = 0$ より，$BK = 1$ であり，$B = \dfrac{1}{K}$，$A = \dfrac{1}{K}$ となり，上式の中辺を，この記号で置き換えると，
次式となる。

$$\frac{1}{(K-y_t)y_t} = \frac{A}{(K-y_t)} + \frac{B}{y_t} = \left\{ \frac{\frac{1}{K}\times 1}{(K-y_t)} + \frac{\frac{1}{K}\times 1}{y_t} \right\} = \frac{1}{K}\left\{ \frac{1}{(K-y_t)} + \frac{1}{y_t} \right\} \qquad (1.4)$$

ここで上式に $\int$ 積分記号を導入すると，次式となる。

$$\int \frac{1}{(K-y_t)y_t} = \int \frac{1}{K}\left\{ \frac{1}{(K-y_t)} + \frac{1}{y_t} \right\} \qquad (1.5)$$

上式の右辺の一部 $\dfrac{1}{(K-y_t)}$ の置き換え積分を考えると t で微分すると $\dfrac{dx}{dt} = -1$ となるので，次式
となるが,

$$\int \frac{1}{(K-y_t)y_t} = \int \frac{1}{K}\left\{ \frac{1}{y_t} + \frac{1}{(K-y_t)} \times (-1) \right\} = \frac{1}{K}\left\{ \log y_t - \log(K-y_t) \right\} = at + c$$

この式は $\dfrac{1}{K}\log\left\{ \dfrac{y_t}{K-y_t} \right\} = at + c$ となる。

ここで $Kc = c^{'}$ と置き両辺の e をとると，次式となる。

$$\log\left\{\frac{y_t}{K-y_t}\right\}=Kat+c^{'}$$

両辺の e をとると，

$$=\frac{y_t}{K-y_t}=e^{Kat+c^{'}}=\frac{y_t}{K-y_t}=e^{c^{'}}\cdot e^{Kat}$$

y_t について求めると，

$$y_t=(K-y_t)e^{c^{'}}\cdot e^{Kat}=Ke^{c^{'}}\cdot e^{Kat}-y_te^{c^{'}}\cdot e^{Kat}$$

$$y_t+y_te^{c^{'}}\cdot e^{Kat}=Ke^{c^{'}}\cdot e^{Kat}$$

となる。

$$y_t(1+e^{c^{'}}\cdot e^{Kat})=Ke^{c^{'}}\cdot e^{Kat}$$

$$y_t=\frac{Ke^{c^{'}}\cdot e^{Kat}}{1+e^{c^{'}}\cdot e^{Kat}}$$

となる。

$$y_t=\frac{K\times\dfrac{e^{c^{'}}\cdot e^{Kat}}{e^{c^{'}}\cdot e^{Kat}}}{\dfrac{1}{e^{c^{'}}\cdot e^{Kat}}+\dfrac{e^{c^{'}}\cdot e^{Kat}}{e^{c^{'}}\cdot e^{Kat}}}=\frac{K}{\dfrac{1}{e^{c^{'}}}\times\dfrac{1}{e^{Kat}}+1}=\frac{K}{e^{-c^{'}}\cdot e^{-Kat}+1}$$

$m=e^{-c^{'}},a=Ka$ と置くと，予測値 y_t は, 上限値 k, 傾き（成長の速さ）a, 定数項 m より，次式

$$y_t=\frac{K}{1+me^{-at}} \qquad (1.6)$$

これがロジスティック曲線である。

2.　ロジスティック曲線の解法（ホテリングの最小二乗法）理論式の説明

(1)　まず，ホテリングの最小二乗法を解くのには，ロジスティック曲線 $y_t=\dfrac{K}{1+me^{-at}}$ 式の y_t を t で微分する。

これを解く $(*)$ 商の微分公式は $\left(\dfrac{f(x)}{g(x)}\right)^{'}=\dfrac{f^{'}(x)g(x)-f(x)g^{'}(x)}{\{g(x)\}^2}$ であり，$\dfrac{dy_t}{dt}=\dfrac{K}{1+me^{-at}}=\dfrac{f(t)}{g(t)}$ と置くと，

$f(t)=K$ は，$f^{'}(t)=0$ である。

$g(t) = 1 + me^{-at}$ は，$g^{'}(t) = -ame^{-at}$ である。

以上より，この微分は，次式となる。

$$\frac{dy_t}{dt} = \frac{K}{1+me^{-at}} = \frac{f^{'}(x)g(x) - f(x)g^{'}(x)}{\{g(x)\}^2} = \frac{0 \times (1+me^{-at}) - K \times (-ame^{-at})}{(1+me^{-at})^2}$$
$$= \frac{-K(-ame^{-at})}{(1+me^{-at})^2} = \frac{a \cdot K \cdot me^{-at}}{(1+me^{-at})^2}$$

(2)　次に $y_t = \dfrac{K}{1+me^{-at}}$ 式を 1 階微分した ロジスティック曲線の式の変形を考える。

$$\frac{dy_t}{dt} = \frac{a \cdot K \cdot me^{-at}}{(1+me^{-at})^2} \qquad (2.1)$$

(2.1) 式の分子を考える。

この式の分子の式である $a \cdot K \cdot me^{-at}$ に，$aK - aK = 0$ を導入し aK で括ると，次式となる。

$$a \cdot K \cdot me^{-at} = aK + a \cdot K \cdot me^{-at} - aK = aK\left[(1+me^{-at}) - 1\right]$$

ここで (2.1) 式を考えると，次式となるが,

$$\frac{dy_t}{dt} = a\left\{\frac{K\left[(1+me^{-at}) - 1\right]}{(1+me^{-at})^2}\right\} = a\left\{\frac{\left[K(1+me^{-at}) - K\right]}{(1+me^{-at})^2}\right\}$$
$$= a\left\{\frac{K(1+me^{-at})}{(1+me^{-at})^2} - \frac{K}{(1+me^{-at})^2}\right\} = a\left\{\frac{K}{(1+me^{-at})} - \frac{K}{(1+me^{-at})^2}\right\}$$
$$= a\left\{\frac{K}{(1+me^{-at})} - \frac{K}{(1+me^{-at})} \times \frac{1}{(1+me^{-at})}\right\}，\text{ここで } \frac{K}{(1+me^{-at})} \text{ を括り出す。}$$
$$= a\left\{\left(\frac{K}{1+me^{-at}}\right) \times \left[1 - \left(\frac{1}{1+me^{-at}}\right) \times \frac{K}{K}\right]\right\}$$
$$= a\left\{\left(\frac{K}{1+me^{-at}}\right) \times \left[1 - \frac{\frac{K}{1+me^{-at}}}{K}\right]\right\}$$

この式の中を $y_t = \dfrac{K}{1+me^{-at}}$ で置き換えると, 次式となる。

$$\frac{dy_t}{dt} = ay_t\left(1 - \frac{y_t}{K}\right)$$

ここで両辺を $\dfrac{1}{y_t}$ 倍すると，次式となる。

$$\frac{1}{y_t} \cdot \frac{dy_t}{dt} = a - \frac{a}{K}y_t$$

ここで $dy_t = \triangle y_t, dt = \triangle t$ と置くと，次式となるが,

$$\frac{1}{y_t}\cdot\frac{\triangle y_t}{\triangle t}=a-\frac{a}{K}y_t$$

$\triangle t=1$ と置くと $\frac{1}{y_t}\cdot\frac{\triangle y_t}{1}=a-\frac{a}{K}y_t$ は，次式となる。

$$\frac{\triangle y_t}{y_t}=a-\frac{a}{K}y_t \qquad (2.2)$$

さらに，$\frac{\triangle y_t}{\triangle t}=y_t^{'}$，$a=A$(定数)，$-\frac{a}{K}=B$(傾き) と置くと，次式となる。

$$y_t^{'}=A+By_t \qquad (2.3)$$

すなわち，単回帰の 1 次式となり最小二乗法が使えるようになる。なお，A と B の関係から上限値 K を求めると $K=\frac{(-1)\times A}{B}=\frac{-a}{-\frac{a}{K}}=-a\times-\frac{K}{a}$ となる。

(3)　ロジスティック回帰曲線にホテリングの最小二乗法の原理を適用しパラメータ a,b を求めてみる。

(2.2) 式より $\frac{\triangle y_t}{y_t}=a-\frac{a}{K}y_t$(1 次式) であり，残差を求める一般式は,

$\sum e_i^2=\sum\left(y_t-y_t^{'}\right)^2=\sum\left(y_t-a-bt\right)^2$ であるが，今回の方程式は $y_t^{'}=a+bt$ より，次式となる。

$$\sum e_i^2=\sum\left(\frac{\triangle y_t}{y_t}-y_t^{'}\right)^2=\sum\left\{\frac{\triangle y_t}{y_t}-\left(a-\frac{a}{K}y_t\right)\right\}^2$$

最小二乗法を一般の連立方程式で置き換えると，次のようになる。

$$\left\{\begin{matrix}\sum y_i & = & na & + & b\sum x_i\\ \sum x_iy_i & = & a\sum x_i & + & b\sum x_i^2\end{matrix}\right\}\Longrightarrow \text{より，}\left\{\begin{matrix}\sum\frac{\triangle y_t}{y_t} & = & na & + & \frac{a}{K}\sum y_t\\ \sum\triangle y_t & = & a\sum y_t & + & \frac{a}{K}\sum y_t^2\end{matrix}\right\}$$

この連立方程式を解いて a, b のパラメータを求めることができる。

パラメータ b を求めてみる。

$$b=\frac{\sum x_iy_i-\frac{\sum x_i\sum y_i}{n}}{\sum x_i^2-\frac{(\sum x_i)^2}{n}}=\frac{n}{n}\times\left(\frac{\sum x_iy_i-\frac{\sum x_i\sum y_i}{n}}{\sum x_i^2-\frac{(\sum x_i)^2}{n}}\right)=\frac{n\sum x_iy_i-\sum x_i\sum y_i}{n\sum x_i^2-(\sum x_i)^2}\Longrightarrow \text{より，}$$

$$b=\frac{n\sum\triangle y_t-\sum y_t\sum\frac{\triangle y_t}{y_t}}{n\sum y_t^2-(\sum y_t)^2}$$

パラメータ a を求めてみる。

$a = \dfrac{\sum y_i - b\sum x_i}{n} \Longrightarrow$ より，$a = \dfrac{\sum \dfrac{\triangle y_t}{y_t} - b\sum y_t}{n}$ となる。

データ数 n の考え方

取り扱うデータ数 n は，実現値 y_t の全てのデータ系列の項数でロジスティック曲線は，0(ゼロ) からのスタート $t_0, t_1, t_2, t_3, \cdots, t_{n'}$ であり，$n^{'} = n+1$ となる。

(4)　ロジスティック曲線のパラメータ m を求めるのには 2 つの公式がある。

変曲点 $\dfrac{K}{2}$ が存在する場合の m を求めてみる。
$y_t = \dfrac{K}{1+me^{-at}}$ を変形すると，次式
$\dfrac{K}{y_t} = 1 + me^{-at}$ となるが，ここで $y_t = y_r$ と考えて変曲点 $y_r = \dfrac{K}{2}$ を分子の
y_t に代入すると $\dfrac{K}{(\dfrac{K}{2})} = 1 + me^{-at} = 2$ となるが，$1 + me^{-at} = 2$ は，$me^{-at} = 2 - 1 = 1$
であり $me^{-at} = 1$ となり，この両辺に対数をとると，$\log m - at\log_e = \log 1$ は,
$\log m = \log 1 + at\log_e$ となり，(*) 対数の公式 $\log 1 = 0, \log_e = 1$ より，$\log m = 0 + at \times 1$ は,
$\log m = at$ となる。
この式より t を求めてみるため両辺を $\dfrac{1}{a}$ 倍すると，次式となる。

$$t = \frac{1}{a}\log m$$

変曲点 $\dfrac{K}{2}$ が存在しない場合の m を求めてみる。
$y_t = \dfrac{K}{1+me^{-at}}$ は，$\dfrac{K}{y_t} = 1 + me^{-at}$ より，$me^{-at} = \dfrac{K}{y_t} - 1$ となるが両辺に e^{at} を掛けると，次式となる。

$$m = \left(\frac{K}{y_t} - 1\right)e^{at} \qquad (2.4)$$

なお，いくつかの実現値 y_t の値に対応する t の値より，複数の m 値を計算し，その平均値を求めてみる。これは例えば t_8, t_{12}, t_{16} などさまざまの時点から m の値を計算し平均を求めみると，次のようになる。

$$m_8 = \left(\frac{K}{y_{t_8}} - 1\right)e^{at_8}, \quad m_{12} = \left(\frac{K}{y_{t_{12}}} - 1\right)e^{at_{12}}, \quad m_{16} = \left(\frac{K}{y_{t_{16}}} - 1\right)e^{at_{16}}$$

いま求めた，この 3 つの平均値 m_8, m_{12}, m_{16} の計算は m となる。

$$m = \frac{(m_8 + m_{12} + m_{16})}{3}$$

この m は，$t=\dfrac{1}{a}\log m$ の式を用いる場合の検算として利用する。

3.　マハラノビスの汎の距離の 2 乗の理論式の説明

(1)　2 変量のマハラノビスの汎の距離の 2 乗の理論

まず，サンプル点 $P_i(x_1,x_2)$ の標準化データを考えると，次のようになる。

$$u_1=\frac{x_1-\mu_1}{\sigma_1},\ u_2=\frac{x_2-\mu_2}{\sigma_2}$$

標準化データ $u_1,\ u_2$ を主成分で表わすには P_i 点 $(u_1,\ u_2)$ を $45°(\frac{1}{\sqrt{2}})$ 回転すればよい。
したがって，$Z_1=\dfrac{(u_1+u_2)}{\sqrt{2}},Z_2=\dfrac{(u_1-u_2)}{\sqrt{2}}$ となる。

固有値 λ_1,λ_2 と同じになる主成分の分散は $V[Z_1],V[Z_2]$ であり, 次のようになるが,
標準化データ $u_1,\ u_2$ の分散は，$V[u_1]=1,V[u_2]=1,$ 共分散 $COV[u_1u_2]=\rho$ となる。

また，その期待値は $E[u_1]=0,E[u_2]=0$ である。したがって，第 1 主成分の分散 $V[Z_1]$
を求めてみると，次のようになる。

$$V[Z_1]=E[Z_1-E(Z_1)]^2=E\left[\frac{(u_1+u_2)}{\sqrt{2}}-\frac{(0+0)}{\sqrt{2}}\right]^2=E\left[\frac{(u_1+u_2)}{\sqrt{2}}\right]^2=\frac{1}{2}E[(u_1+u_2)]^2$$

Z_1,Z_2 は標準化データであり，右辺 (カッコ) の中の第 1 項 u_1^2，第 2 項 u_2^2 は，偏差平方和，第 3 項は，偏差積和 $u_1\cdot u_2$ であるが，それぞれを $\dfrac{1}{n}$ で割り平均化をすると，2 つの分散 $+2\times$ 共分散であり，次のようになる。

$$\begin{aligned}&=\frac{1}{2}\times\frac{1}{n}\cdot(u_1^2+u_2^2+2u_1\cdot u_2)\\&=\frac{1}{2}\{V[u_1]+V[u_2]+2COV[u_1u_2]\}\end{aligned}$$

また，$V[Z_2]$ も同じであり，λ_1,λ_2 は，次式となる。

$$\lambda_1=V[Z_1]=\frac{\{V[u_1]+V[u_2]+2COV[u_1u_2]\}}{2}=\frac{(1+1+2\rho)}{2}=1+\rho$$
$$\lambda_2=V[Z_2]=\frac{\{V[u_1]+V[u_2]-2COV[u_1u_2]\}}{2}=\frac{(1+1-2\rho)}{2}=1-\rho$$

また，主成分 $Z_{1,}Z_2$ の期待値は $E[Z_{1,}]=0,E[Z_2]=0$ で，その共分散は $COV[Z_{1,}Z_2]=0$ であり, 次式である。

$$COV[Z_{1,}Z_2]=E[Z_1-E(Z_1)]\cdot E[Z_1-E(Z_1)]=E\left[\frac{(u_1+u_2)}{\sqrt{2}}\right]\times E\left[\frac{(u_1-u_2)}{\sqrt{2}}\right]$$
$$=\frac{u_1^2+u_1u_2-u_1u_2-u_2^2}{2}=0$$

なお，主成分の平均は $\bar{Z}_1=0, \bar{Z}_2=0$ であり，Z_1, Z_2 の標準化は $\left[\dfrac{(Z-\bar{Z})}{\sigma[Z]}\right]^2$ であり，

$\left[\dfrac{(Z_1-\bar{Z}_1)}{\sigma[Z_1]}\right]^2=\dfrac{Z_1^2}{V[Z_1]}$,　$\left[\dfrac{(Z_2-\bar{Z}_2)}{\sigma[Z_2]}\right]^2=\dfrac{Z_2^2}{V[Z_2]}$ より，

主成分の共分散 $COV[Z_1,Z_2]=0$（無相関）であるので, 重心 $(0,0)$ と $Z_1=\dfrac{(u_1+u_2)}{\sqrt{2}}$ の 2 乗 Z_1^2 と $Z_2=\dfrac{(u_1-u_2)}{\sqrt{2}}$ の 2 乗 Z_2^2 を，その分散 $V[Z_1], V[Z_2]$ で割り標準化したものの和がマハラノビスの汎の距離の 2 乗 D^2 となる。なお，マハラノビスの汎の距離の 2 乗 D^2 は, 標準化データ u_1，u_2, 共分散 ρ より, 次式となる。

$$
\begin{aligned}
D^2 &= \frac{Z_1^2}{V[Z_1]}+\frac{Z_2^2}{V[Z_2]}=\frac{(u_1+u_2)^2}{(\sqrt{2})^2}\cdot\frac{1}{1+\rho}+\frac{(u_1-u_2)^2}{(\sqrt{2})^2}\cdot\frac{1}{1-\rho}\\
&=\frac{(u_1+u_2)^2}{2(1+\rho)}+\frac{(u_1-u_2)^2}{2(1-\rho)}\\
&=\frac{(1-\rho)(u_1+u_2)^2+(1+\rho)(u_1-u_2)^2}{2(1+\rho)(1-\rho)}\\
&=\frac{u_1^2+u_2^2-2\rho u_1u_2}{1-\rho^2} \qquad (3.1)
\end{aligned}
$$

4.　2 変量についてのウォード法の理論式の説明

統合クラスター $t(i,j)$ と他のクラスター g の距離を統合前のクラスターにより計算するものである。すなわち，クラスター $i(n_i)$ と $j\ (n_j)$ が統合されクラスター $t(n_t)$ が作られるものと定義すると，

表 4.1　データ

サンプルNo.	x_1	x_2
1	x_{11}	x_{21}
2	x_{12}	x_{22}
:	:	:
i	x_{1i}	x_{2i}
:	:	:
j	x_{1j}	x_{2j}
:	:	:
n	x_{1n}	x_{2n}

(4.1) 式より，クラスター t と他のクラスター g の距離 $\triangle d_{tg}$ は，偏差平方和の増分 S_{tg} であり（表 4.1）のデータより，次式により定義される。なお,$\bar{x}_{1i}$ と $\bar{x}_{2i}$ は l クラスの平均であり，$\bar{x}_{1j}$ と $\bar{x}_{2j}$ は m クラスターの平均である。なお，平均 $\bar{x}_{1i}$ と $\bar{x}_{2i}$ は l のクラスターであり，平均 $\bar{x}_{1j}$ と $\bar{x}_{2j}$ は m のクラスターであり，次のようになる。

$$(\bar{x}_{1i}, \bar{x}_{2i}) = \frac{\sum_{l=1}^{n_i} x_{1li}}{n_i}, \frac{\sum_{l=1}^{n_i} x_{2li}}{n_i}, \quad (\bar{x}_{1j}, \bar{x}_{2j}) = \frac{\sum_{m=1}^{n_j} x_{1mj}}{n_j}, \frac{\sum_{m=1}^{n_j} x_{2mj}}{n_j}$$

ここで変量 $\bar{x}_1(\bar{x}_{1i}, \bar{x}_{1j})$ および $\bar{x}_2(\bar{x}_{2i}, \bar{x}_{2j})$ のクラスター i と j を統合したときのクラスター t の距離の加重平均 ($weighted\ average$) は,

$$\bar{x}_{1t} = \frac{n_i\bar{x}_{1i} + n_j\bar{x}_{1j}}{n_i + n_j}, \qquad \bar{x}_{2t} = \frac{n_i\bar{x}_{2i} + n_j\bar{x}_{2j}}{n_i + n_j}$$

となる。

$$S_{tg}(\triangle d_{tg}) = \frac{n_t n_g}{n_t + n_g}\left[(\bar{x}_{1t} - \bar{x}_{1g})^2 + (\bar{x}_{2t} - \bar{x}_{2g})^2\right] \tag{4.1}$$

1)　(4.1) 式の右辺の [括弧] の中の第 1 項は，$\bar{x}_{1t} = \dfrac{n_i\bar{x}_{1i} + n_j\bar{x}_{1j}}{n_i + n_j}$ より偏差平方和は，次式

$$(\bar{x}_{1t} - \bar{x}_{1g})^2 = \left[\frac{n_i\bar{x}_{1i} + n_j\bar{x}_{1j}}{n_i + n_j} - \bar{x}_{1g}\right]^2 = \left[\frac{n_i\bar{x}_{1i} + n_j\bar{x}_{1j}}{n_i + n_j} - \frac{\bar{x}_{1g}}{1}\right]^2 \tag{4.2}$$

となる。

(4.2) 式の右式の第 2 項を $\dfrac{n_i + n_j}{n_i + n_j} = 1$ 倍すると,

$$= \left[\frac{n_i\bar{x}_{1i} + n_j\bar{x}_{1j}}{n_i + n_j} - \frac{(n_i + n_j)\bar{x}_{1g}}{(n_i + n_j)}\right]^2 = \left[\frac{n_i\bar{x}_{1i} + n_j\bar{x}_{1j} - n_i\bar{x}_{1g} - n_j\bar{x}_{1g}}{n_i + n_j}\right]^2$$

$$= \left[\frac{n_i\left(\bar{x}_{1i} - \bar{x}_{1g}\right) + n_j\left(\bar{x}_{1j} - \bar{x}_{1g}\right)}{n_i + n_j}\right]^2 = \left[\frac{\left\{n_i\left(\bar{x}_{1i} - \bar{x}_{1g}\right)\right\}^2 + \left\{n_j\left(\bar{x}_{1j} - \bar{x}_{1g}\right)\right\}^2}{\left(n_i + n_j\right)^2}\right]$$

となるが $\dfrac{1}{\left(n_i + n_j\right)^2}$ を括りだすと，次式

$$= \frac{1}{\left(n_i + n_j\right)^2}\left[\left\{n_i\left(\bar{x}_{1i} - \bar{x}_{1g}\right)\right\}^2 + \left\{n_j\left(\bar{x}_{1j} - \bar{x}_{1g}\right)\right\}^2 + 2n_i n_j\left(\bar{x}_{1i} - \bar{x}_{1g}\right)\left(\bar{x}_{1j} - \bar{x}_{1g}\right)\right]$$

となり，右辺の第 1 項と第 2 項の n_i^2 を前にだすと，次式

$$= \frac{1}{\left(n_i + n_j\right)^2}\left[n_i^2\left(\bar{x}_{1i} - \bar{x}_{1g}\right)^2 + n_j^2\left(\bar{x}_{1j} - \bar{x}_{1g}\right)^2 + 2n_i n_j\left(\bar{x}_{1i} - \bar{x}_{1g}\right)\left(\bar{x}_{1j} - \bar{x}_{1g}\right)\right] \tag{4.3}$$

となる。

公式より,

$A^2(A-C)^2+B^2(B-C)^2+2AB(A-C)(B-C) = (A^2+AB)(A-C)^2+(B^2+AB)(B-C)^2-AB(A-B)^2$
$A = n_i, A^2 = n_i^2, (A - C) = (\bar{x}_{1i} - \bar{x}_{1g}), B = n_j, B^2 = n_j^2, (B - C) = (\bar{x}_{1j} - \bar{x}_{1g})$

(4.3) 式は，公式より展開すると，

$$(\bar{x}_{1t}-\bar{x}_{1g})^2=\frac{1}{(n_i+n_j)^2}\left[\left(n_i^2+n_in_j\right)(\bar{x}_{1i}-\bar{x}_{1g})^2+(n_j^2+n_in_j)(\bar{x}_{1j}-\bar{x}_{1g})^2-n_in_j(\bar{x}_{1i}-\bar{x}_{1j})^2\right]$$

となる。

さらに，$\dfrac{1}{(n_i+n_j)}$ を [括弧] の中に入れると，次式

$$=\frac{1}{(n_i+n_j)}\left[\frac{\left(n_i^2+n_in_j\right)}{(n_i+n_j)}(\bar{x}_{1i}-\bar{x}_{1g})^2+\frac{(n_j^2+n_in_j)}{(n_i+n_j)}(\bar{x}_{1j}-\bar{x}_{1g})^2-\frac{n_in_j}{(n_i+n_j)}(\bar{x}_{1i}-x_{1j})^2\right]$$

となるが，この式の右辺の [括弧] の中の第 2 項，第 3 項の分子を n_i で括りだし整理すると，

$$=\frac{1}{(n_i+n_j)}\left[\frac{n_i(n_i+n_j)}{(n_i+n_j)}(\bar{x}_{1i}-\bar{x}_{1g})^2+\frac{n_j(n_j+n_i)}{(n_i+n_j)}(\bar{x}_{1j}-\bar{x}_{1g})^2-\frac{n_in_j}{(n_i+n_j)}(\bar{x}_{1i}-\bar{x}_{1j})^2\right]$$

は，次式

$$=\frac{1}{(n_i+n_j)}\left[n_i(\bar{x}_{1i}-\bar{x}_{1g})^2+n_j(\bar{x}_{1j}-\bar{x}_{1g})^2-\frac{n_in_j}{(n_i+n_j)}(\bar{x}_{1i}-\bar{x}_{1j})^2\right] \qquad (4.4)$$

となる。

(4.4) 式の右辺の [括弧] の中は，$u=\left[n_i(\bar{x}_{1i}-\bar{x}_{1g})^2+n_j(\bar{x}_{1j}-\bar{x}_{1g})^2-\dfrac{n_in_j}{(n_i+n_j)}(\bar{x}_{1i}-x_{1j})^2\right]$

より，

$$(\bar{x}_{1t}-\bar{x}_{1g})^2=\frac{1}{(n_i+n_j)}\cdot u \qquad (4.5)$$

である。

(4.1) 式の右辺の第 2 項を展開すると偏差平方和は，次式

$$(\bar{x}_{2t}-\bar{x}_{2g})^2=\left[\frac{n_i\bar{x}_{2i}+n_j\bar{x}_{2j}}{n_i+n_j}-\bar{x}_{2g}\right]^2 \qquad (4.6)$$

となる。

2)　1) と同様に，

$$=\frac{1}{(n_i+n_j)^2}\left[\left(n_i^2+n_in_j\right)(\bar{x}_{2i}-\bar{x}_{2g})^2+(n_j^2+n_in_j)(\bar{x}_{2i}-\bar{x}_{2g})^2-n_in_j(\bar{x}_{2i}-\bar{x}_{2j})^2\right]$$

となるが $\dfrac{1}{(n_i+n_j)}$ を [括弧] の中に入れ整理すると，次式

$$= \frac{1}{(n_i+n_j)}\left[n_i\left(\ \bar{x}_{2i}-\bar{x}_{2g}\right)^2+n_j\left(\bar{x}_{2j}-\bar{x}_{2g}\right)^2-\frac{n_in_j}{(n_i+n_j)}\left(\bar{x}_{2i}-\bar{x}_{2j}\right)^2\right] \qquad (4.7)$$

となる。

(4.7) 式の右辺の [括弧] の中は，$v=\left[n_i\left(\ \bar{x}_{2i}-\bar{x}_{2g}\right)^2+n_j\left(\bar{x}_{2j}-\bar{x}_{2g}\right)^2-\dfrac{n_in_j}{(n_i+n_j)}\left(\bar{x}_{2i}-\bar{x}_{2j}\right)^2\right]$

より，

$$(\bar{x}_{2t}-\bar{x}_{2g})^2=\frac{1}{(n_i+n_j)}\cdot v \qquad (4.8)$$

である。

(4.1) 式は偏差平方和の増分 $\triangle S_{tg}$ であり，次式

$$\triangle S_{tg}=\frac{n_tn_g}{n_t+n_g}\left[(\bar{x}_{1t}-\bar{x}_{1g})^2+(\bar{x}_{2t}-\ \bar{x}_{2g})^2\right]$$

であるが,

(4.1) 式は，(4.5) 式と (4.8) 式により構成されているが，その 2 つの式の係数は,
$n_t=n_i+n_j$ であり $\dfrac{1}{n_i+n_j}=\dfrac{1}{n_t}$ より，次のようになる。

$$=\frac{n_tn_g}{n_t+n_g}\cdot\frac{1}{n_t}\left[u+v\right]=\frac{n_g}{n_t+n_g}\left[u+v\right]$$

そこで u,v を要約した u,v を具体的に表すと，次式

$$=\frac{n_gn_i}{n_t+n_g}\left[(\bar{x}_{1i}-\bar{x}_{1g})^2+(\bar{x}_{2i}-\bar{x}_{2g})^2\right]+\frac{n_gn_j}{n_t+n_g}\left[(\bar{x}_{1j}-\bar{x}_{1g})^2+(\bar{x}_{2j}-\bar{x}_{2g})^2\right]$$

$$-\frac{n_g}{n_t+n_g}\cdot\frac{n_in_j}{n_i+n_j}\left[(\bar{x}_{1i}-\bar{x}_{1j})^2+(\bar{x}_{2i}-\bar{x}_{2j})^2\right] \qquad (4.9)$$

$$=\frac{1}{n_t+n_g}\{(n_gn_i\left[(\bar{x}_{1i}-\bar{x}_{1g})^2+(\bar{x}_{2i}-\bar{x}_{2g})^2\right]+n_gn_j\left[(\bar{x}_{1j}-\bar{x}_{1g})^2+(\bar{x}_{2j}-\bar{x}_{2g})^2\right]$$

$$-n_g\cdot\frac{n_in_j}{n_i+n_j}\left[(\bar{x}_{1i}-\bar{x}_{1j})^2+(\bar{x}_{2i}-\bar{x}_{2j})^2\right]\}$$

ここで右辺の { 括弧 } の中第 3 項の n_gn_i に $1=\dfrac{n_i+n_g}{n_i+n_g}$ と第 5 項の n_gn_j に $1=\dfrac{n_j+n_g}{n_j+n_g}$ を
掛けると,

$$=\frac{1}{n_t+n_g}\{(n_i+n_g)\,\frac{n_gn_i}{n_i+n_g}\left[(\bar{x}_{1i}-\bar{x}_{1g})^2+(\bar{x}_{2i}-\bar{x}_{2g})^2\right]+(n_j+n_g)\,\frac{n_gn_j}{n_j+n_g}\left[(\bar{x}_{1j}-\bar{x}_{1g})^2+(\bar{x}_{2j}-\bar{x}_{2g})^2\right]$$

$$-n_g \cdot \frac{n_i n_j}{n_i + n_j}\left[(\bar{x}_{1i} - \bar{x}_{1j})^2 + (\bar{x}_{2i} - \bar{x}_{2j})^2\right]\}$$

となる。

$\triangle S_{tg}$ は, クラスターの増分距離より $\triangle S_{ig}, \triangle S_{jg}, \triangle S_{ij}$ をもちいて表すと，次式

$$\triangle S_{ig} = \frac{n_g n_i}{n_i + n_g}\left[(\bar{x}_{1i} - \bar{x}_{1g})^2 + (\bar{x}_{2i} - \bar{x}_{2g})^2\right], \triangle S_{jg} = \frac{n_g n_j}{n_j + n_g}\left[(\bar{x}_{1j} - \bar{x}_{1g})^2 + (\bar{x}_{2j} - \bar{x}_{2g})^2\right]$$

$$\triangle S_{ij} = \frac{n_i n_j}{n_i + n_j}\left[(\bar{x}_{1i} - \bar{x}_{1j})^2 + (\bar{x}_{2i} - \bar{x}_{2j})^2\right]$$

$$\triangle S_{tg} = \frac{1}{n_t + n_g}\left[(n_i + n_g) \triangle S_{ig} + (n_j + n_g) \triangle S_{jg} - n_g \triangle S_{ij}\right]$$

となる。

なお，次式により統合するたびに分類距離行列を更新させてゆく。

$$S_{tg} = \frac{n_i + n_g}{n_t + n_g} S_{ig} + \frac{n_j + n_g}{n_t + n_g} S_{jg} - \frac{n_g}{n_t + n_g} S_{ij}$$