碩士學位 請求論文

層化任意抽出에서의 最適標本設計

指導教授 金 益 贊

濟州大學校 教育大學院

數學教育專攻

金 鎭 求

1990年 8月

層化任意抽出에서의 最適標本設計

이 論文을 教育學 碩士學位 論文으로 提出함.

濟州大學校 教育大學院 數學教育專攻

提 出 者 金 鎭 求

指導教授 金 益 贊

1990年 8月 日

金鎭求의 碩士學位 論文을 認准함.

1990年 8 月 日

主審 ______________________ 印

副審 ______________________ 印

副審 ______________________ 印

濟州大學校敎育大學院

目　　次

<Abstract>

Optimum Design in Stratified Sampling

Kim jin-gu

Mathematics Education Major

Graduate School of Education,

Cheju National University, Cheju Korea

Supervised by Professor ***Kim Ik-Chan***

It is very important to decide the size of a sample,which is connected with the precision of the sample, in Sampling theory

In this thesis, Proportional allocation, Optimum allocation and Neyman allocation were chosen to decide the size of a sample, which is connected with the Cost function in Stratified Random Sampling. They were then summarized, investigated and compared to one another. In a further study, the method that minimizes the variance of mean differences between estimated strata should be studied, in the case of strata L=2, L>2. In addition to this, the Optimum design method, which minimizes the variance in case of stated cost, should be studied. This optimum design method is used for analytical comparison, in a 2x2 contingency table, which decides two factors into each of two classes.

1. 序論

標本論에 있어서 標本의 精度와 관련된 표본의 크기를 決定하는문제는 應用統計 分野에서 있어서 가장 基本的인 主題라 할 것이다.

L개의 層으로 分類된 母集團에서 각 層마다 最適의 標本크기를 決定하는 문제는 오래전부터 研究되어 왔다.

특히 標本을 抽出하기 위한 費用函數가 주어졌을때 그 精度를 극대화하기 위하여 抽出되어야 하는 標本의 크기,또는 정해진 標本의 精度에 대하여 그 費用을 最小化하는 標本의 크기를 決定하는 最適割當의 문제는 Stuart(1954), Rao(1973)등 많은 研究가 있었다.

한편, 각층에 標本을 配分하는 方法으로서 標本크기를 層의 크기에 比例的으로 配分하는 比例配分과 표본당 抽出費用이 모든 層에서 같도록 놓았을때의 最適割當의 특수한 경우에 적용되는 Neyman의 配分法등을 생각할 수 있다.

본 研究에서는 層化標本抽出에 있어서 費用函數와 관련된 표본 크기를 決定하는 方法으로서 上記한 세가지 方法을 要約,檢討,比較하고 나아가서 母集團 분布 또는 母平均이 미리 알려졌다고 假定된 서로 다른 두 層간의 比較分析 方法을 摸索하였다. 즉 2개의 층에서 推定된 층의 平均 사이의 差의 分散을 最小化하기 위한 標本크기를 決定하는 것이다.

이 平均差의 分散을 最小化하는 方法이 $L > 2$ 개의 層으로 一般化된

다.

한편, 두개의 要因이 다시 두개의 部類로 分類되는 경우의 最適配分을 생각하였다. 즉 두 要因의 각 부류들 사이에 동일한 精度가 요청된 경우에, 주어진 費用을 最小化하는 標本크기를 決定하는 方法을 2x2 分割表에 의하여 糾明하여 보았다.

1.1 推定値의 性質

크기 N 인 母集團의 각 要素들을 $y_1, y_2, \ldots, y_N$으로 표기하고 ,이 母集團에서 抽出한 크기 n인 특성치들을 $y_1, y_2, \ldots, y_n$이라고 하면 母集團과 標本의 平均과 分散은 다음과 같이 定義 된다. 여기서 각 標本의 특성치들은 母集團의 그것들과는 서로 다른 것임을 나타낸다.

$$母平均 \quad : \quad \bar{Y} = \frac{y_1+y_2+\cdots+y_N}{N} = \frac{\sum_{i=1}^{N} y_i}{N}$$

$$標本平均 \quad : \quad \bar{y} = \frac{y_1+y_2+\cdots+y_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n}$$

母分散 : $S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$

標本分散 : $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$

이 標記와 定義에 따라서 다음의 定理들이 성립한다

定理1.1-1 標本平均 $\bar{y}$ 는 母平均 $\bar{Y}$ 의 不偏推定値이다.

(證明)

$$E\bar{y} = \frac{1}{NC_n} \sum_{1}^{NC_n} \bar{y} = \frac{1}{\frac{N!}{n!(N-n)!}} \frac{1}{n} \sum_{1}^{NC_n} (y_1 + y_2 + \cdots + y_n)$$

$$= \frac{n!(N-n)!}{N!} \frac{1}{n} \frac{(N-1)!}{(n-1)!(N-n)!} (y_1 + y_2 + \cdots + y_N)$$

$$= \frac{1}{N} (y_1 + y_2 + \cdots + y_N) = \bar{Y}$$

定理1.1-2 單純任意抽出에서의 標本平均 $\bar{y}$의 分散은

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \frac{S^2}{n}\frac{N-n}{N} = \frac{S^2}{n}(1 - f) \quad \cdots\cdots(1.1)$$

$$(\text{단, } f = \frac{n}{N})$$

(證明)

$$n(\bar{y}-\bar{Y})=n(\frac{y_1+y_2+\cdots+y_n}{n} - \frac{n\bar{Y}}{n})$$

$$=(y_1-\bar{Y})+(y_2-\bar{Y})+\cdots+(y_n-\bar{Y})$$

$$E\{n(\bar{y}-\bar{Y})\}^2=n^2E(\bar{y}-\bar{Y})^2 =E\{(y_1-\bar{Y})+(y_2-\bar{Y})+\cdots+(y_n-\bar{Y})\}^2$$

$E(y_1+y_2+\cdots+y_n)=\frac{n}{N}(y_1+y_2+\cdots+y_N)$이므로

$$E\{(y_1-\bar{Y})^2+(y_2-\bar{Y})^2+\cdots+(y_n-\bar{Y})^2\}=\frac{n}{N}\{(y_1-\bar{Y})^2+(y_2-\bar{Y})^2+\cdots+(y_N-\bar{Y})^2\}$$

$$E\{(y_1-\bar{Y})(y_2-\bar{Y})+\cdots+(y_{n-1}-\bar{Y})(y_n-\bar{Y})\}=\frac{n(n-1)}{N(N-1)}\{(y_1-\bar{Y})(y_2-\bar{Y})+\cdots+(y_{N-1}-\bar{Y})(y_N-\bar{Y})\}$$

따라서 $n^2E(y-Y)^2$

$$=\frac{n}{N}[\{(y_1-\bar{Y})^2+(y_2-\bar{Y})^2+\cdots+(y_N-\bar{Y})^2\}$$

$$+\frac{2(n-1)}{N-1}\{(y_1-\bar{Y})(y_2-\bar{Y})+\cdots+(y_{N-1}-\bar{Y})(y_N-\bar{Y})\}]$$

$$=\frac{n}{N}[(1-\frac{n-1}{N-1})\{(y_1-\bar{Y})^2+(y_2-\bar{Y})^2+\cdots+(y_N-\bar{Y})^2\}$$

$$+\frac{n-1}{N-1}\{(y_1-\bar{Y})+(y_2-\bar{Y})+\cdots+(y_N-\bar{Y})\}^2]$$

$$=\frac{n}{N}\frac{N-1-(n-1)}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$$

그러므로 $$E(\bar{y}_1-\bar{Y})^2=\frac{1}{n^2}\frac{n}{N}\frac{N-n}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$$

$$=\frac{1}{nN}\frac{N-n}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$$

$$=\frac{S^2}{n}(1-f)$$

定理1.1-3 標本分散 s^2 은 母分散 S^2의 不偏推定量이다

(證明)

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}\{(y_i-\bar{Y})-(\bar{y}-\bar{Y})\}^2$$

$$=\frac{1}{n-1}\{\sum_{i=1}^{n}(y_i-\bar{Y})^2-2(\bar{y}-\bar{Y})\sum_{i=1}^{n}(y_i-\bar{Y})+n(\bar{y}-\bar{Y})^2\}$$

$$=\frac{1}{n-1}\{\sum_{i=1}^{n}(y_i-\bar{Y})^2-2n(\bar{y}-\bar{Y})^2+n(\bar{y}-\bar{Y})^2\}$$

$$=\frac{1}{n-1}\{\sum_{i=1}^{n}(y_i-\bar{Y})^2 - n(\bar{y}-\bar{Y})^2\}$$

따라서

$$E(s^2)=\frac{1}{n-1}\{\frac{n}{N}\sum_{i=1}^{N}(y_i-\bar{Y})^2-nE(\bar{y}-\bar{Y})^2\}$$

$$=\frac{1}{n-1}\{\frac{n}{N}\ N-1\frac{1}{N-1}\ \sum_{i=1}^{N}(y_i-\bar{Y})^2 - n\frac{N-n}{nN}\ S^2\}$$

$$=\frac{1}{n-1}\{\frac{n(N-1)}{N}S^2 - \frac{N-n}{N}\ S^2\} =S^2$$

2. 層化任意抽出法

2.1 定義와 性質

N 개의 單位들로 이루어진 母集團을 먼저 N_1, N_2, N_3,....,N_L 單位들의 部分母集團(Subpopulation)으로 分割한다.

이 部分母集團들은 $N_1+N_2+\cdots+N_L= N$ 이 되도록 겹치거나 중복됨이 없이 구성하며, 이때 部分母集團으로 분류하는 조작을 層化(Stratification)라 하고,分割된 部分母集團을 層(Stratum)이라한다.

層들이 決定되었을때 標本은 서로 다른 層들로 부터 獨立的으로 抽出하고 각층에서 추출한 표본의 크기는 n_1, n_2, n_3,...,n_L로 표시하자.각층마다 單純任意抽出 方法에 의해 標本을 抽出하는 것을 層化任意抽出法이라고 한다.

첨자 h는 層을 표시하고 i는 층내에서의 單位이며 다음의 몇가지 記號를 사용한다.

N_h : 單位들의 總數

n_h : 標本에서 單位들의 수

y_{hi} : i번째 單位의 測定値

$$W_h = \frac{N_h}{N} \quad : \text{層加重値}$$

$$f_h = \frac{n_h}{N_h} \quad : \text{標本抽出率}$$

$$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h} \quad : \text{層平均}$$

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \quad : \text{標本平均}$$

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi}-\bar{Y}_h)^2}{N_h - 1} \quad : \text{層分散}$$

層化抽出方法은 실제로 대단히 많이 適用되는 方法중의 하나이다.특히 조사의 精度를 높이기 위하여 層化할 때는

1) 母集團의 크기에 있어서 대단히 큰 차이가 있는 機構,制度, 혹은 機關으로 되어 있는 경우

2) 측정하고자 하는 主變數가 위의 機構,制度,혹은機關과 밀접 한 관계가 있는 경우

3) 위의 제도,기구,기관에 대한 충분한 情報가 있는 경우

單純任意抽出보다 높은 信賴度를 갖는 자료를 얻는 方法으로서 層化抽出이 이용될 수있다.

定義2.1-1 層化抽出에서의 標本平均은

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} = \sum_{h=1}^{L} \bar{W}_h \bar{y}_h \quad \cdots\cdots\cdots\cdots\cdots\cdots (2.1)$$

標本分散은

$$V(\bar{y}_{st}) = \frac{1}{N} \sum^{N} (\bar{y}_{st} - \bar{Y})^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots (2.2)$$

이다

定理 2.1-1. 모든 層에서 標本推定値 $\bar{y}_h$ 가 不偏推定値이면 $\bar{Y}_{st}$ 는 母平均 $\bar{Y}$의 不偏推定値이다.

(證明)

$$E(\bar{y}_{st})=E\left(\frac{\sum_{h=1}^{L} N_h\bar{y}_h}{N}\right)=E\left(\sum_{h=1}^{L} W_h\bar{y}_h\right)=\sum_{h=1}^{L} W_hE(\bar{y}_h)=\sum_{h=1}^{L} W_h\bar{Y}_h) ,$$

$$\bar{Y}=\frac{\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi}}{N}=\frac{\sum_{h=1}^{L} N_h\frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}}{N}=\frac{\sum_{h=1}^{L} N_h\bar{Y}_h}{N}=\sum_{h=1}^{L} W_h\bar{Y}_h$$

$$\therefore \quad E(\bar{y}_{st})=\bar{Y}$$

定理2.1-2. 서로 다른 層들과 獨立的으로 層化抽出된다면 推定値 $\bar{y}_{st}$의 分散은

$$V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L} N_h(N_h-n_h)\frac{S_h^2}{n_h} = \sum_{h=1}^{L} W_h^2\frac{S_h^2}{n_h}(1-f_h)\cdot\cdot(2.3)$$

이다.

(證明)

$$V(\bar{y}_{st})=V(\sum_{h=1}^{L} W_h\bar{y}_h)$$

$$=V(W_1\bar{y}_1+W_2\bar{y}_2+\cdots+W_L\bar{y}_L)$$

$$=W_1^2\ V(\bar{y}_1)+W_2^2V(\bar{y}_2)+\cdots+W_L^2(\bar{y}_L)+2\sum_{h=1}^{L}\sum_{j>h}^{L} W_hW_jCov(\bar{y}_h,\bar{y}_j)$$

모든 標本들이 다른層과 獨立으로 抽出되므로

$$Cov(\bar{y}_h,\bar{y}_j)=0$$

따라서

$$V(\bar{y}_{st})=\sum_{h=1}^{L} W_h^2V(\bar{y}_h)=\sum_{h=1}^{L}W_h^2\frac{S_h^2}{n_h}\ \frac{N_h-n_h}{N_h}=\sum_{h=1}^{L}(\frac{N_h}{N})^2\frac{S_h^2}{n_h}\ \frac{N_h-n_h}{N_h}$$

$$=\frac{1}{N^2}\sum_{h=1}^{L}N_h(N_h-n_h)\frac{S_h^2}{n_h}$$

$$=\sum_{h=1}^{L} W_h^2\ \frac{S_h^2}{n_h}\ (1-f_h)$$

2.2 標本의 配分

標本의 配分에 있어서 고려할 문제는 첫째 標本의 크기 n을 決定하는 것이고 다음으로 n_h를 決定하기 위해 h층에 標本을 配分하는 문제이다.

이상적인 配分으로서의 基準은 便宜性과 正確性이고 이외에 最小의 費用으로 最大의 精度를 얻을수 있는 基準을 생각한다

즉, 推定量의 精度는 그것이 分散에 의해 推定되므로 주어진 分散에 대하여 費用을 最小化 하든지 혹은 주어진 費用에 대하여 分散을 最小化하는 基準을 생각한다.

여기서 標本抽出費用은 標本設計費用,프레임作成費用, 面接者의 訓練費用 ,資料處理費用과 公的인 支出費등 標本抽出시 드는 전반적인 費用을 포함한다.

分析的인 目的을 위해 이를 標本의 크기에 대한 函數와 固定費用으로 나누어 다음과 같이 費用函數를 定義하자.

즉, $$C = C_o + \sum_{h=1}^{L} C_h n_h \quad \cdots\cdots\cdots\cdots(2.4)$$

단, C_o : 固定費用

C_h : h層으로 부터 抽出單位當 費用.

2.2-1 比例配分

각 層으로 부터 抽出率이 同一하게 되도록하는 配分方法이 比例配分이며 抽出率

$$f = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_L}{N_L} = \frac{n}{N}$$ 로 둔다.

즉, h 層의 標本의 크기 n_h를 母集團의 加重値 $W_h = \frac{N_h}{N}$ 에 比例的으로 配分하는 方法이다

따라서 $$n_h = n\frac{N_h}{N} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (2.5)$$

이때의 母平均 $\bar{Y}$ 의 不偏推定量 $\bar{y}_{st}$ 는

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} \frac{n_h}{f} \frac{y_h}{n_h}}{\sum_{h=1}^{L} \frac{n_h}{f}} = \frac{1}{n}\sum_{h=1}^{L} n_h \bar{y}_h$$

$$= \frac{1}{n}\sum_{h=1}^{L} n_h \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} = \frac{1}{n}\sum_{h=1}^{L}\sum_{i=1}^{n_h} y_{hi} \cdots\cdots\cdots (2.6)$$

이고, 分散은 $V(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}$ 에 式(2.5)를 代入하여

$$V(\bar{y}_{prop}) = \frac{N-n}{N}\sum_{h=1}^{L} \frac{N_h}{N} \frac{S_h^2}{n} \cdots\cdots\cdots\cdots\cdots (2.7)$$

이 된다.

2.2.2 最適配分

層化抽出法에서 각 層에서의 標本의 크기 n_h 의 값들을 決定함에 있어서 費用에 대하여 $V(\bar{y}_{st})$ 이 最小가 되도록 하든지, 정해진 $V(\bar{y}_{st})$ 에 대하여 그 費用이 最小가 되도록 n_h를 決定하는 方法이 最適配分法이다.

定理 2.2-1 $C = C_o + \Sigma C_h n_h$ 와 같이 線型 費用函數를 가지는 層化抽出法에서는 n_h가 $W_h S_h/\sqrt{C_h}$ 에 比例할 때 정해진 費用 C 에 대하여 $\bar{y}_{st}$ 의 分散은 最小가 되고, 分散 $V(\bar{y}_{st})$가 정해질 때에는 費用이 最小가 된다.

(證明)

$$C=C_o+\sum_{h=1}^{L} C_h n_h$$

$$V=V(\bar{y}_{st})=\sum_{h=1}^{L}\frac{W_h^2 S_h^2}{n_h}(1-f_h)$$

$$=\sum_{h=1}^{L}\frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^{L}\frac{W_h^2 S_h^2}{n_h}\frac{n_h}{N_h}$$

$$=\sum_{h=1}^{L}\frac{W_h^2S_h^2}{n_h}-\sum_{h=1}^{L}\frac{W_h^2S_h^2}{N_h}$$

$$C'=\sum_{h=1}^{L}C_hn_h=C-C_o,$$

$$V'=\sum_{h=1}^{L}\frac{W_h^2S_h^2}{n_h}=V+\sum_{h=1}^{L}\frac{W_h^2S_h^2}{N_h}$$

라 놓으면,

$$V'C'=(V+\sum_{h=1}^{L}\frac{W_h^2S_h^2}{N_h})(C-C_o)=(\sum_{h=1}^{L}\frac{W_h^2S_h^2}{n_h})(\sum_{h=1}^{L}C_hn_h)$$

이다.

Cauchy-Schwarz 不等式에 의하여

$$V'C'=(\sum_{h=1}^{L}\frac{W_h^2S_h^2}{n_h})(\sum_{h=1}^{L}C_hn_h)\geq(\sum_{h=1}^{L}W_hS_h\sqrt{C_h})^2$$

이것은 n_h을 어떻게 택하더라도 $V'C'$를 $(\sum W_hS_h\sqrt{C_h})^2$보다 더 작게 만들수 없으며

$$\frac{\sqrt{C_hn_h}}{\frac{W_hS_h}{\sqrt{n_h}}}=\frac{n_h\sqrt{C_h}}{W_hS_h}$$:一定할 때 等號가 성립하여

最小値가 된다는 것을 보여준다.

$$\frac{n_h \sqrt{C_h}}{W_h \ S_h} = k \text{ 라 하면 }, \quad n_h = \frac{W_h \ S_h \ k}{\sqrt{C_h}} \text{ 이 되어}$$

$$n = \sum_{h=1}^{L} n_h = \sum_{h=1}^{L} \frac{W_h S_h k}{\sqrt{C_h}} \text{ 이다}$$

따라서

$$\frac{n_h}{n} = \frac{\dfrac{W_h S_h k}{\sqrt{C_h}}}{\Sigma \dfrac{W_h S_h k}{\sqrt{C_h}}} = \frac{\dfrac{N_h S_h}{\sqrt{C_h}}}{\Sigma \dfrac{N_h S_h}{\sqrt{C_h}}} \quad \cdot\cdot(2.8)$$

만약 費用 C 가 정해진다면.

$C = C_o + \Sigma C_h n_h$에

$$n_h = \frac{N_h S_h / \sqrt{C_h}}{\Sigma (N_h S_h / \sqrt{C_h})}$$ n를 代入하여 n에 대하여 정리한다.

즉, $$C - C_o = \Sigma C_h n_h = \Sigma C_h \frac{N_h S_h / \sqrt{C_h}}{\Sigma (N_h S_h / \sqrt{C_h})} n = \frac{\Sigma (N_h S_h \sqrt{C_h})}{\Sigma (N_h S_h / \sqrt{C_h})} n$$

$$n = \frac{(C-C_o)\,\Sigma\,(N_hS_h/\sqrt{C_h})}{\Sigma\,(N_hS_h\sqrt{C_h})} \cdots\cdots\cdots\cdots\cdots\cdots (2.9)$$

만약 分散 V가 정해진다면.

$$V = \sum_{h=1}^{L}\frac{W_h{}^2S_h{}^2}{n_h} - \sum_{h=1}^{L}\frac{W_h{}^2S_h{}^2}{N_h}$$ 에

$$n_h = \frac{N_hS_h/\sqrt{C_h}}{\Sigma(N_hS_h/\sqrt{C_h})}$$ n를 代入한 다음 n에 대하여 정리한다.

즉 $$V + \sum_{h=1}^{L}\frac{W_h{}^2S_h{}^2}{N_h} = \sum_{h=1}^{L}\frac{W_h{}^2S_h{}^2}{n_h} = \sum_{h=1}^{L} W_h{}^2S_h{}^2 \frac{\Sigma(N_hS_h/\sqrt{C_h})}{(N_hS_h/\sqrt{C_h})}\frac{1}{n}$$

$$= \sum_{h=1}^{L} W_hS_h\frac{N_h}{N}\, S_h\sqrt{C_h}\;\frac{\Sigma(N_hS_h/\sqrt{C_h})}{N_hS_h}\frac{1}{n} = \sum_{h=1}^{L} W_hS_h\sqrt{C_h}\;\sum_{h=1}^{L}\left(\frac{N_h}{N}\,\frac{S_h}{\sqrt{C_h}}\right)\frac{1}{n}$$

$$= \sum_{h=1}^{L} W_hS_h\sqrt{C_h}\;\Sigma(W_hS_h/\sqrt{C_h})\frac{1}{n}$$

$$(V + \frac{1}{N}\sum_{h=1}^{L} W_hS_h{}^2)\; n = \sum_{h=1}^{L} W_hS_h\sqrt{C_h}\,\Sigma(W_hS_h/\sqrt{C_h})$$

$$\therefore \quad n = \frac{\sum_{h=1}^{L} W_h S_h \sqrt{C_h} \; \Sigma (W_h S_h / \sqrt{C_h})}{V + \frac{1}{N} \sum_{h=1}^{L} W_h S_h^2} \quad \cdots\cdots\cdots\cdots (2.10)$$

$$(\text{단}, \; W_h = \frac{N_h}{N} \;)$$

한편, $V(\bar{y}_{st})$를 最小化하는 n_h는 Lagrange의 乘數法에 의해서도 해결될 수 있다. 즉,

$$\phi = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h} + \lambda (\sum_{h=1}^{L} C_h n_h - C) \text{ 으로 두면,}$$

$$\frac{\partial \phi}{\partial n_h} = 0$$

따라서

$$-\frac{N_h^2}{N^2} \frac{S_h^2}{n_h^2} + \lambda C_h = 0$$

$$\therefore \quad n_h = \frac{1}{\sqrt{\lambda}} \frac{N_h S_h}{N \sqrt{C_h}}$$

이들 L 개의 層을 모두 합하여

$$\Sigma n_h = n = \frac{1}{\sqrt{\lambda}} \Sigma \frac{N_h S_h}{N \sqrt{C_h}}$$

또 $$\sqrt{\lambda} = \frac{1}{n} \quad \frac{1}{N} \Sigma \frac{N_h S_h}{\sqrt{C_h}}$$

따라서 $$n_h = \frac{N_h S_h / \sqrt{C_h}}{\Sigma (N_h S_h / \sqrt{C_h})} n \quad \cdots\cdots\cdots\cdots\cdots\cdots (2.11)$$

2.2.3 Neyman 配分

각 층에 대한 費用 C_h가 같은 경우, 즉, $C = C_o + C_h \sum_{h=1}^{L} n_h = C_o + C_h n$

따라서 $n = \dfrac{C - C_o}{C_h}$ 로서 標本의 크기 n이 固定된다.

즉, $$\begin{cases} V(\bar{y}_{st}) = \sum^{L} \left(\dfrac{N_h}{N}\right)^2 \dfrac{N_h - n_h}{N_h} \quad \dfrac{S_h^2}{n_h}, \\ n_1 + n_2 + \cdots + n_L = n \quad : \text{固定} \end{cases}$$

그러면,

$$n_h = \frac{N_h S_h}{\Sigma N_h S_h} n \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (2.12)$$

이 되고

標本의 크기 n은 $N_h S_h$에 比例的으로 配分된다.이 方法은 層內의 變動이 커서 그자료들이 서로 異質的일때는 더욱 큰 n_h의 값을 갖게 되는 特徵이 있다.

2.2.4 標本크기의 決定과 效率性比較

標本의 크기 n을 決定하는데 推定値의 標本分散에 의한다. 즉,

$$V(\bar{y}_{st}) = \Sigma\left(\frac{N_h}{N}\right)^2 \frac{N_h-n_h}{N_h} \frac{S_h^2}{n_h} \cdots\cdots\cdots\cdots (2.13)$$

한편,

$$\left.\begin{aligned} &\text{比例配分} : n_h=\frac{N_h}{N} n \\ &\text{最適配分} : n_h=\frac{N_hS_h/\sqrt{C_h}}{\Sigma(N_hS_h/\sqrt{C_h})} n \\ &\text{Neyman配分} : n_h=\frac{N_hS_h}{\Sigma N_hS_h} n \end{aligned}\right\} \cdots\cdots (2.14)$$

이고 이들 각 式에 대한 分散은 式(2.14)를 式(2.13)에 代入함으로서 얻을 수 있다.

$$V(\bar{y}_{prop})= \frac{1}{N} \Sigma\frac{N_hS_h^2}{n} - \frac{1}{N^2} \Sigma N_hS_h^2 \cdots\cdots\cdots\cdots (2.15)$$

$$V(\bar{y}_{opt})= \frac{1}{N^2} \frac{1}{n}(\Sigma N_hS_h\sqrt{C_h})(\Sigma\frac{N_hS_h}{\sqrt{C_h}}) - \frac{1}{N^2} \Sigma N_hS_h^2 \cdot (2.16)$$

$$V(\bar{y}_{Ney}) = \frac{1}{N^2} \frac{(\Sigma N_h S_h)^2}{n} - \frac{1}{N^2} \Sigma N_h S_h^2 \cdots\cdots\cdots (2.17)$$

이제 標本의 크기 n을 決定하기 위하여

$d^2 = z^2 V(\bar{y}_{st})$ (단 d: 要求된 精度 z:信賴水準)$\cdots\cdots$(2.18)

로 두고,精度 d_o와 信賴水準 z_o가 주어진다고 假定하면,

$$V(\bar{y}_{st}) = \frac{d_o^2}{z_o^2} = D^2(\text{要求分散}) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (2.19)$$

이 된다.

따라서 각 경우의 標本의 크기 n은 다음과 같이 구할 수 있다.

$$n = \frac{N \Sigma N_h S_h^2}{N^2 D^2 + \Sigma N_h S_h^2} \qquad (\text{比例配分}) \cdots\cdots (2.20)$$

$$n = \frac{(\Sigma N_h S_h \sqrt{C_h})(\Sigma N_h S_h / \sqrt{C_h})}{N^2 D^2 + \Sigma N_h S_h^2} \qquad (\text{最適配分}) \cdots\cdots (2.21)$$

$$n = \frac{(\Sigma N_h S_h)^2}{N^2 D^2 + \Sigma N_h S_h^2} \qquad (\text{Neyman配分}) \cdots\cdots (2.22)$$

한편, 抽出率 $\frac{n_h}{N_h}$가 아주 작거나 fpc=1일때 각분모의 두번째항은 생략할 수 있다.

例題)

1,000개의 書店이 備置된 책수의 規模별로 3개의 集團으로 層化되어 있는데, 이 書店이 하루 平均 顧客수를 推定하고자 한다 각각의 層에 대한 基本的인 資料는 다음과 같다고 할 때, 이 자료를 이용하여 比例配分,最適配分,Neyman 配分에 의하여 각각의 標本의 크기와 推定値의 分散을 구해보면 다음과 같다.

(단,最適配分에서의 각층에 대한 費用 C_1=10,000원,C_2=20,000원,C_3=30,000원이고 要求分散 D=1이다.)

층	N_h	S_h	S_h^2	N_hS_h	$N_hS_h^2$	$N_h^2S_h^2$
小	300	30	900	9,000	270,000	81,000,000
中	500	20	400	10,000	200,000	100,000,000
大	200	10	100	2,000	20,000	4,000,000
	1,000			21,000	490,000	185,000,000

(1) 比例配分인 경우

$$n=\frac{N\Sigma N_h S_h^2}{N^2D^2+\Sigma N_h S_h^2}=\frac{1,000\times 490,000}{1,000^2\times 1^2+490,000}=329$$

각 層의 配分은 $n_h=\frac{N_h}{N}\times n$ 이므로

$n_1=(300/1,000)\times 329=\ 99$

$n_2=(500/1,000)\times 329=165$

$n_3=(200/1,000)\times 329=\ 66$

分散은 $V(\hat{\bar{y}}_{prop})=\frac{1}{N}\Sigma\frac{N_h S_h^2}{n}-\frac{1}{N^2}\Sigma N_h S_h^2$

$$=\frac{1}{1,000}\left(\frac{490,000}{329}-\frac{490,000}{1,000}\right)=1.49-0.49=1$$

(2) 最適配分인 경우

最適配分인 경우에는 각층에 대한 費用이 필요하므로 주어진 資料로 부터 다음과 같이 정리된다.

N_hS_h	$\sqrt{C_h}$	$N_hS_h\sqrt{C_h}$	$N_hS_h/\sqrt{C_h}$
9,000	100	900,000	90
10,000	141.4	1,414,000	71
2,000	173.2	346,400	12
21,000		2,660,400	173

따라서 標本의 크기는 $n=\dfrac{(\Sigma N_h S_h \sqrt{C_h})(\Sigma N_h S_h/\sqrt{C_h})}{N^2 D^2+\Sigma N_h S_h^2}$

$$=\frac{2,660,400\times 173}{1,000^2\times 1^2+490,000}=309$$

이며 각층의 配分은 $n_h=\dfrac{N_h S_h/\sqrt{C_h}}{\Sigma N_h S_h/\sqrt{C_h}}\,n$

$n_1=(90/173)\times 309=161$

$n_2=(71/173)\times 309=127$

$n_3=(12/173)\times 309=\ 21$

그리고 分散은

$$\hat{V}(\bar{y}_{opt})=\frac{1}{N^2}\,\frac{1}{n}(\Sigma N_h S_h \sqrt{C_h})(\Sigma N_h S_h/\sqrt{C_h})-\frac{1}{N^2}\Sigma N_h S_h^2$$

$$=\frac{1}{1,000^2}\times\frac{1}{309}\times 2,660,400\times 173-\frac{1}{1,000^2}\times 490,000$$

$$=1.49-0.49=1$$

(3) Neyman 配分인 경우

Neyman 配分인 경우에는 각층 대한 費用은 같다고 假定하므로 標

本의 크기는 $n=\dfrac{(\Sigma N_h S_h)^2}{N^2D^2+\Sigma N_h S_h^2}=\dfrac{21,000^2}{1,000^2 \times 1^2+490,000}=296$이며

층에 대한 配分은 $n=\dfrac{N_h S_h}{\Sigma N_h S_h}\times n$ 이므로

$n_1=9,000/21,000\times 296=127$

$n_2=10,000/21,000\times 296=141$

$n_3=2,000/21,000\times 296=28$

또한 分散은

$$\hat{V}(\bar{y}_{Ney})=\frac{1}{N^2}\,\frac{(\Sigma N_h S_h)^2}{n}-\frac{1}{N^2}\Sigma N_h S_h^2$$

$$=\frac{1}{1,000^2}\times\frac{21,000^2}{296}-\frac{1}{1,000^2}\times 490,000$$

$$=1.49-0.49=1$$

2.3 單純任意抽出法과 層化任意抽出法과의 比較

層化抽出法과 單純任意抽出法중, 어느 抽出方法이 精度를 더 높게 해 줄것인가를 알아 보고자 한다.

2.3.1 單純任意抽出法과 比例的 層化抽出法과의 比較

이 두 과정의 效率性을 比較하기 위하여 두개의 다른 方法에 의해 얻어진 推定量의 分散을 比較할 필요가 있다. 먼저 單純任意抽出方法에 대한 分散은

$$V(\bar{y}_{ran})=\frac{N-n}{N}\frac{S^2}{n} \cdots\cdots\cdots\cdots(2.23)$$

이다. 만약, n이 N에 비하여 매우 작으면,fpc=1이 되어 式 (2.23)은

$$V(\bar{y}_{ran})=\frac{S^2}{n} \cdots\cdots\cdots\cdots(2.24)$$

이 된다.그리고 N이 1에 비하여 매우 크다면,N-1≒N으로 둘수있어

σ^2≒ S^2이되므로. 母分散과 層內,層間分散 사이의 관계식,즉,

$$\sigma^2=\sigma_w^2+\sigma_b^2 \cdots\cdots\cdots\cdots(2.25)$$

(단. $\sigma_w^2=\frac{1}{N}\sum_{h=1}^{L}N_h\sigma_h^2$, $\sigma_b^2=\frac{1}{N}\sum_{h=1}^{L}N_h(\bar{Y}_h-\bar{Y})^2$)

을 이용하면, 式(2.24)은

$$V(\bar{y}_{ran}) = \frac{1}{n}\,(\sigma_w^2 + \sigma_b^2) \cdots\cdots\cdots\cdots (2.26)$$

와 같이 된다.

한편, 比例的層化抽出法의 分散은

$$V(\bar{y}_{prop}) = \frac{N-n}{N}\,\Sigma\,\frac{N_h S_h^2}{N\ n} \cdots\cdots\cdots\cdots (2.27)$$

그리고 層內分散 $\sigma_w^2 = \frac{1}{N}\sum_{h=1}^{L} N_h \sigma_h^2$은 N_h이 1에 비하여 매우 크다면,

$N_h - 1 \fallingdotseq N_h$ 이되어 $\sigma_w^2 = \frac{1}{N}\sum_{h=1}^{L} N_h S_h^2$ 과 같이 되므로 式(2.27)은

$$V(\bar{y}_{prop}) = \frac{N-n}{N}\,\frac{1}{n}\,\sigma_w^2 \quad (fpc \neq 1) \cdots\cdots\cdots\cdots (2,28)$$

$$= \frac{1}{n}\,\sigma_w^2 \qquad (fpc=1) \cdots\cdots\cdots\cdots (2.29)$$

이다.

따라서 式(2.29)을 式(2.26)에 代入하면,

$$V(\bar{y}_{ran}) = V(\bar{y}_{prop}) + \frac{1}{n}\,\sigma_b^2 \cdots\cdots\cdots\cdots (2.30)$$

이것은 比例的層化抽出法의 分散이 單純任意抽出法의 分散보다 $\frac{1}{n}\sigma_b^2$ 만큼 작게 되고 있어서 層間分散 $\sigma_b^2=\frac{1}{N}\Sigma N_h(\bar{Y}_h-\bar{Y})^2$이 커질때 比例的 層化任意抽出法의 精度가 單純任意抽出法보다 높다는것을 의미 해준다.

2.3.2 比例的 層化抽出法과 Neyman層化抽出法과의 比較

Neyman 配分에 의한 層化抽出法에 대한 分散은

$$V(\bar{y}_{Ney})=\frac{1}{N^2}\frac{(\Sigma N_hS_h)^2}{n}-\frac{1}{N^2}\Sigma N_hS_h^2 \cdots\cdots\cdots\cdots(2.31)$$

앞의 定義에 의하면 $V(\bar{y}_{prop}) > V(\bar{y}_{Ney})$임은 분명하다.

이제 $V(\bar{y}_{prop})$ 와 $V(\bar{y}_{Ney})$ 사이의 差이를 調査 檢討한다.

우선 比例的 配分에 의한 分散은

$$V(\bar{y}_{prop})=\frac{N-n}{N}\Sigma\frac{N_hS_h^2}{Nn} \quad (fpc\neq 1) \cdots\cdots\cdots\cdots(2.32)$$

$$=\Sigma\frac{N_hS_h^2}{Nn} \quad (fpc=1)\cdots\cdots\cdots\cdots(2.33)$$

fpc=1일때 $V(\bar{y}_{prop})$는 다음과 같이 變形하여 쓸 수 있다.

$$V(\bar{y}_{prop})=\Sigma\frac{N_hS_h^2}{Nn}+\frac{(\Sigma N_hS_h)^2}{nN^2}-\frac{(\Sigma N_hS_h)^2}{nN^2}$$

$$=V(\bar{y}_{Ney})+\{\Sigma\frac{N_hS_h^2}{Nn}-\frac{(\Sigma N_hS_h)^2}{nN^2}\}$$

$$=V(\bar{y}_{Ney})+\frac{1}{nN}\{\Sigma N_hS_h^2-\frac{1}{N}(\Sigma N_hS_h)^2\}$$

$$=V(\bar{y}_{Ney})+\frac{1}{nN}\{\Sigma N_hS_h^2-\frac{2}{N}(\Sigma N_hS_h)^2+\frac{\Sigma N_h}{N^2}(\Sigma N_hS_h)^2\}$$

$$=V(\bar{y}_{Ney})+\frac{1}{nN}\Sigma N_h\{S_h^2-\frac{2}{N}S_h(\Sigma N_hS_h)+\frac{1}{N^2}(\Sigma N_hS_h)^2\}$$

$$=V(\bar{y}_{Ney})+\frac{1}{nN}\Sigma N_h(S_h-\frac{1}{N}\Sigma N_hS_h)^2$$

$$=V(\bar{y}_{Ney})+\frac{1}{nN}\Sigma N_h(S_h-\bar{S}_h)^2 \quad (\text{단 } \bar{S}_h=\frac{1}{N}\Sigma N_hS_h)$$

즉,

$$V(\bar{y}_{prop})=V(\bar{y}_{Ney})+\frac{1}{nN}\Sigma\ N_h(S_h-\bar{S}_h)^2 \cdots\cdots\cdots\cdots \quad (2.34)$$

S_h과 $\bar{S}_h$의 差가 크면 클수록 $(S_h-\bar{S}_h)^2$의 값이 커져서 $V(\bar{y}_{prop})$와$V(\bar{y}_{Ney})$

의 差가 더욱 커지게 된다. 이것은 層의 크기와 層間變動이 크게 다를때 Neyman 配分의 효과가 높아진다는 것을 말해준다.

이제 有限母集團修正係數(finite population correction factor:fpc) 를 무시하고 층별 抽出率 $\frac{n_h}{N_h}$ 를 무시할 수 있다면, 위에서 分析한 結果는 結論的으로

$$V(\bar{y}_{Ney}) \leq V(\bar{y}_{prop}) \leq V(\bar{y}_{ran}) \quad \cdots\cdots\cdots\cdots \quad (2.35)$$

와 같이 되어 Neyman 配分에 의한 層化抽出法을 이용하면 分散을 감소 시킬 수 있다는 것을 보여 준다.

3. 層化任意抽出에서의 最適標本設計

3.1 一元分類에서의 最適標本設計

서로 다른 層들이 一元分類 형태를 취하고 있을때 層간의 平均比較를 위한 最適設計方案에 대하여 생각하여 보자.

단, 母集團 각층의 크기는 미리 알려져 있다고 假定한다.

만일 2개의 層만 있을 때는 推定된 層平均들 간의 平均差 $(\bar{y}_1-\bar{y}_2)$ 의 分散을 最小化하는 標本크기 n_1, n_2를 구할 수 있다.

$$V(\bar{y}_1-\bar{y}_2)=V(\bar{y}_1)+V(\bar{y}_2)=\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2} \quad \cdots\cdots (3.1)$$

이고

費用函數를 $C=C_0+C_1n_1+C_2n_2$ $\cdots\cdots$ (3.2)

로 두면,

$$VC'=\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)(C_1n_1+C_2n_2) \quad (단, C'=C-C_0) \cdots (3.3)$$

한편, Cauchy-Schwarz不等式

$(\Sigma a_h^2)(\Sigma b_h^2) \geq (\Sigma a_h b_h)^2$에 의하여

$a_h = \dfrac{S_h}{\sqrt{n_h}}$, $b_h = \sqrt{C_h n_h}$ 로 두면, $a_h b_h = \dfrac{S_h}{\sqrt{n_h}} \sqrt{C_h n_h} = S_h \sqrt{C_h}$

$\therefore \quad VC' = (\Sigma a_h^2)(\Sigma b_h^2) \geq (\Sigma S_h \sqrt{C_h})^2$

$$\frac{b_h}{a_h} = \frac{\sqrt{C_h n_h}}{S_h / \sqrt{n_h}} = \frac{n_h \sqrt{C_h}}{S_h} = k$$

$$n_h = \frac{kS_h}{\sqrt{C_h}} \cdots\cdots\cdots\cdots\cdots\cdots (3.4)$$

그리고 $n = \Sigma n_h = \dfrac{kS_1}{\sqrt{C_1}} + \dfrac{kS_2}{\sqrt{C_2}}$ 이다.

$$\frac{n_h}{n} = \frac{kS_h/\sqrt{C_h}}{kS_1/\sqrt{C_1} + kS_2/\sqrt{C_2}} = \frac{S_h/\sqrt{C_h}}{\Sigma(S_h/\sqrt{C_h})}$$

$$n_h = \frac{S_h/\sqrt{C_h}}{\Sigma(S_h/\sqrt{C_h})} n \cdots\cdots\cdots\cdots\cdots\cdots (3.5)$$

$$n_1 = \frac{nS_1/\sqrt{C_1}}{S_1/\sqrt{C_1} + S_2/\sqrt{C_2}}, \qquad n_2 = \frac{nS_2/\sqrt{C_2}}{S_1/\sqrt{C_1} + S_2/\sqrt{C_2}} \cdots (3.6)$$

일때 平均差의 分散은 最小가 된다.

한편, L개층이 L>2인 경우의 最適設計는 서로 다른 층간의 平均 比較를 위하여 要求되는 精度의 量에 따른다.

예를들어 $V(\bar{y}_h-\bar{y}_i) \leq V_{hi}$가 되는 條件을 갖는 $\frac{L(L-1)}{2}$ 개의 集合에 의거해서 費用을 最小化하도록 할 수 있다. 여기서 V_{hi}의 값들은 층 h와 i 들간의 충분한 比較를 위하여 要求되는 精度에 의해 選擇 된다.

S_h 와 C_h 가 크게 다르지 않을때 $\frac{L(L-1)}{2}$ 개 쌍들의 層간의 差의 平均 分散을 最小化하기 위하여 다음과 같이 생각하였다.

$$\bar{V} = \frac{1}{{}_LC_2}\{V(\bar{y}_1-\bar{y}_2)+V(\bar{y}_1-\bar{y}_3)+\cdots+V(\bar{y}_1-\bar{y}_L)$$
$$+V(\bar{y}_2-\bar{y}_3)+V(\bar{y}_2-\bar{y}_4)+\cdots+V(\bar{y}_2-\bar{y}_L)$$
$$+\cdots+V(\bar{y}_{L-1}-\bar{y}_L)\}$$

$$= \frac{1}{L(L-1)/2}[(L-1)\{V(\bar{y}_1)+V(\bar{y}_2)+\cdots+V(\bar{y}_L)\}]$$

$$= \frac{2}{L}\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \cdots + \frac{S_L^2}{n_L}\right)$$

$$= \frac{2}{L} \sum_{h=1}^{L} \frac{S_h^2}{n_h} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3.7)$$

이를 最小화 하기 위하여 費用

$$C = C_0 + \Sigma C_h n_h \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3.8)$$

가 정했을때 $L=2$개의 경우와 마찬가지로 $n_h \propto S_h/\sqrt{C_h}$ 한다는 것을 適用하면,

$$n_h = \frac{S_h/\sqrt{C_h}}{\sum_{h=1}^{L} (S_h/\sqrt{C_h})} n \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3.9)$$

일때 $\bar{V}$ 는 最小가 된다. 또 式(3.9)를 式(3.8)에 代入함으로서

$$n = \frac{(C - C_o)\Sigma(S_h/\sqrt{C_h})}{\Sigma(S_h\sqrt{C_h})} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3.10)$$

임을 알 수 있고, 式(3.10)을 式(3.9)에 代入함으로서

$$n_h = \frac{S_h}{\sqrt{C_h}} \frac{C - C_o}{\Sigma S_h\sqrt{C_h}} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3.11)$$

이되어 式(3.11)를 式(3.7)에 代入함으로서 $\bar{V}_{(min)}$ 를 구할 수 있다 즉,

$$\bar{V}_{(min)} = \frac{2}{L} \sum_{h=1}^{L} \frac{S_h^2}{(S_h/\sqrt{C_h})\{(C - C_o)/\Sigma S_h\sqrt{C_h}\}}$$

$$= \frac{2}{L} \frac{(\Sigma S_h\sqrt{C_h})^2}{C-C_o} \cdots\cdots\cdots\cdots (3.12)$$

이다.

3.2 두가지 要因의 分析調査를 위한 最適設計

대부분의 標本調査에 있어서 根本目的은 有限母集團내의 몇가지 분야들간에의 比較이다.

특히 α 와 τ 두가지 要因이 각각 두가지 部類로 나누어져 2x2分割表로 표시되는 경우를 생각한다.이러한 分割表에서의 두要因 比較 分析調査를 위하여 주어진 費用에 대해서 分散을 最小化하는 最適設計方案을 본절에서 해결하려고 한다.

이제 α의 i번째 部類의 요소와 τ의 j번째 部類의요소를 (i,j)로 표시하면 각 要因에 대한 두가지 部類는

$$\left.\begin{aligned} D_\alpha &= W_{.1}(\bar{Y}_{11}-\bar{Y}_{21})+W_{.2}(\bar{Y}_{12}-\bar{Y}_{22}) \\ D_\tau &= W_{1.}(\bar{Y}_{11}-\bar{Y}_{12})+W_{2.}(\bar{Y}_{21}-\bar{Y}_{22}) \end{aligned}\right\} \cdots\cdots\cdots\cdots (3.13)$$

에 의해서 比較될 수 있다.

여기서 N_{ij} : (i,j) 요소의 母集團의 크기

$$W_{ij} = \frac{N_{ij}}{N}$$

$$W_{.j} = \Sigma_i W_{ij}$$

$$W_{i.} = \Sigma_j W_{ij}$$

$\bar{Y}_{ij}$: (i,j) 요소내의 母平均

$N = \Sigma_i \Sigma_j N_{ij}$: 母集團의 크기

式(3.13)의 不偏推定量이

$$\left.\begin{aligned} \hat{D}_\alpha &= W_{.1}(\bar{y}_{11}-\bar{y}_{21})+W_{.2}(\bar{y}_{12}-\bar{y}_{22}) \\ \hat{D}_\tau &= W_{1.}(\bar{y}_{11}-\bar{y}_{12})+W_{2.}(\bar{y}_{21}-\bar{y}_{22}) \end{aligned}\right\} \cdots\cdots\cdots\cdots (3.14)$$

($W_{.j}=\Sigma_i W_{ij}$, $W_{i.}=\Sigma_j W_{ij}$, $W_{ij}=n_{ij}/n$: (i,j)요소의 標本比率

$\bar{y}_{ij}$: (i,j)요소의 標本平均)

임을 쉽게 알 수있다

이제 $\hat{D}_\alpha$와 $\hat{D}_\tau$에 대하여 同一한 精度가 要求되는 것으로 한다면 그 目的函數를 두 推定量의 分散의 합의 平均으로서

$$\bar{V} = \frac{1}{2}\{V(\hat{D}_\alpha) + V(\hat{D}_\tau)\} \cdots\cdots\cdots\cdots (3.15)$$

으로 定義할 수있다.

여기서 fpc 를 생략함으로서

$$\hat{V}(D_{\alpha})+\hat{V}(D_{\tau})=w_{.1}^{2}\left(\frac{s_{11}^{2}}{n_{11}}+\frac{s_{21}^{2}}{n_{21}}\right)+w_{.2}^{2}\left(\frac{s_{12}^{2}}{n_{12}}+\frac{s_{22}^{2}}{n_{22}}\right)$$

$$+w_{1.}^{2}\left(\frac{s_{11}^{2}}{n_{11}}+\frac{s_{12}^{2}}{n_{12}}\right)+w_{2.}^{2}\left(\frac{s_{21}^{2}}{n_{21}}+\frac{s_{22}^{2}}{n_{22}}\right)$$

$$=(w_{.1}^{2}+w_{1.}^{2})\frac{s_{11}^{2}}{n_{11}}+(w_{.2}^{2}+w_{1.}^{2})\frac{s_{12}^{2}}{n_{12}}$$

$$+(w_{.1}^{2}+w_{2.}^{2})\frac{s_{21}^{2}}{n_{21}}+(w_{.2}^{2}+w_{2.}^{2})\frac{s_{22}^{2}}{n_{22}}$$

$$=\sum_{i}\sum_{j}(w_{.j}^{2}+w_{i.}^{2})\frac{s_{ij}^{2}}{n_{ij}}$$

따라서.

$$\bar{V}=\sum_{i}\sum_{j}\frac{g_{ij}^{2}}{n_{ij}} \quad \cdots\cdots\cdots\cdots\cdots \quad (3.16)$$

(단, $2g_{ij}=(w_{.j}^{2}+w_{i.}^{2})s_{ij}^{2}$)

한편, 費用函數를

$$C=C_0+\sum_{i}\sum_{j}C_{ij}\,n_{ij} \quad \cdots\cdots\cdots\cdots\cdots \quad (3.17)$$

로 定義하면 $C'=C-C_0=\Sigma\Sigma C_{ij}n_{ij}$로 두었을때 最適設計配分 方式이 適用될 수있다.

즉, $\bar{V}C'=(\Sigma\Sigma\frac{g_{ij}^2}{n_{ij}})(\Sigma\Sigma C_{ij}\ n_{ij})$

$$=\{\Sigma\Sigma(\ g_{ij}\ /\sqrt{n_{ij}})^2\}\{\Sigma\Sigma(\sqrt{C_{ij}\ n_{ij}})^2\}$$

$$\geq(\Sigma\Sigma g_{ij}\sqrt{C_{ij}})^2$$

$$\frac{\sqrt{C_{ij}\ n_{ij}}}{g_{ij}/\sqrt{n_{ij}}}=k \Rightarrow \frac{n_{ij}\ \sqrt{C_{ij}}}{g_{ij}}=k \Rightarrow n_{ij}=\frac{g_{ij}}{\sqrt{C_{ij}}}k$$

그리고

$$n=\Sigma\Sigma n_{ij}=\Sigma\Sigma(\ k\ g_{ij}/\sqrt{C_{ij}}\ )$$

따라서

$$\frac{n_{ij}}{n}=\frac{k\ g_{ij}/\sqrt{C_{ij}}}{\Sigma\Sigma(k\ g_{ij}/\sqrt{C_{ij}})}$$

$$\Rightarrow\ n_{ij}=\frac{g_{ij}/\sqrt{C_{ij}}}{\Sigma\Sigma\ (g_{ij}/\sqrt{C_{ij}})}\ n \cdots\cdots\cdots\cdots\cdots\cdots(3.18)$$

式(3.18)를 式(3.17)에 代入하면,

$$C-C_o=\Sigma\Sigma C_{ij}\frac{g_{ij}/\sqrt{C_{ij}}}{\Sigma\Sigma(g_{ij}/\sqrt{C_{ij}})}n$$ 이 되어

$$n=\frac{(C-C_o)\Sigma\Sigma(g_{ij}/\sqrt{C_{ij}})}{\Sigma\Sigma\ g_{ij}\sqrt{C_{ij}}} \cdots\cdots\cdots\cdots\cdots\cdots (3.19).$$

이 된다.

또 式(3.19)를 式(3.18)에 代入하면.

$$n_{ij}=\frac{g_{ij}}{\sqrt{C_{ij}}}\ \frac{(C-C_o)}{\Sigma\Sigma\ g_{ij}\sqrt{C_{ij}}} \cdots\cdots\cdots\cdots\cdots\cdots (3.20)$$

이 되고. 이때의 最小分散 $\bar{V}_{min}$는

$$\bar{V}_{min}=\frac{(\Sigma\Sigma g_{ij}\sqrt{C_{ij}})^2}{C-C_o} \cdots\cdots\cdots\cdots\cdots\cdots (3.21)$$

이다.

한편, $2g_{ij}^2=(w_{.j}^2+w_{i.}^2)s_{ij}^2=\frac{n_{.j}^2+n_{i.}^2}{n^2}\ s_{ij}^2$에서 $\frac{n_{.j}^2+n_{i.}^2}{n^2}=1$

로 두면. $g_{ij}^2=\frac{s_{ij}^2}{2}$ 이고 式(3.19)와 式(3.20)은 각각 다음과 같이

같이 유도된다.

$$n = \frac{(C-C_o)\sum\sum s_{ij}/\sqrt{C_{ij}}}{\sum\sum s_{ij}\sqrt{C_{ij}}} \cdots\cdots\cdots\cdots\cdots\cdots (3.22)$$

$$n_{ij} = \frac{s_{ij}}{\sqrt{C_{ij}}} \; \frac{(C-C_o)}{\sum\sum s_{ij}\sqrt{C_{ij}}} \cdots\cdots\cdots\cdots\cdots\cdots (3.23)$$

또 이 경우의 $\bar{V}_{min}$ 은

$$\bar{V}_{min} = \sum_i \sum_j \frac{\dfrac{s_{ij}^2}{2}}{\dfrac{s_{ij}}{\sqrt{C_{ij}}} \; \dfrac{(C-C_o)}{\sum\sum s_{ij}\sqrt{C_{ij}}}}$$

$$= \frac{1}{2} \; \frac{\left(\sum_i \sum_j s_{ij}\sqrt{C_{ij}}\right)^2}{C-C_o} \cdots\cdots\cdots\cdots\cdots (3.24)$$

이다.

参 考 文 獻

[1] 金鍾浩. 「標本調査法」. 서울: 自由아카데미 , 1987.

[2] 南宮 坪. 「標本理論」. 서울: 博英社 , 1986.

[3] 劉東善 外.「理工系統計學」. 서울: 集賢社 , 1985.

[4] 李載昌 外.「統計的方法」. 서울: 法經出版社.,1986.

[5] Cochran,W.G. Sampling Techniques. New York: John Wiley & Sons, Inc., 1977.

[6] Kish,L. Survey Sampling. New York: John Wiley & Sons,Inc., 1965.

[7] Rao,J.N.K. On Double Sampling for Stratification and Analytical Surveys.Biometrika., **60** , 1973.

[8] Stuart,A. A Simple Presentation of Optimun Sampling results.Jour. Stat.Soc., **B** , 1954.

[9] Williams,B. A Sampler On Sampling. New York: John Wiley & Sons, Inc., 1978.

[10] Yamane,T. Elementary Sampling Theory. Englewood Cliffs: Prentice Hall,Inc., 1967.

感謝의 글

많은 曖昧模糊함과 獨特한 아름다움과 슬픔과 기쁨을 지닌 삶의 廣大한 空間을 조금이나마 이해할 수 있도록 이끌어주신 數學教育科,數學科 모든 教授님께 感謝를 드립니다.

더욱이 本 論文을 完成하기까지 公休日에도 時間을 내면서 많은 助言과 指導를 해주신 金益贊 教授님께 깊은 感謝를 드립니다.

또한 初.中.高等學校에서 大學院에 이르기까지 저를 가르켜 주신 선생님들, 講義를 같이 받았던 先輩.同僚.後輩 院生들,늘 관심속에 激勵를 아끼지 않으신 黃寅炯 奬學官님, 서울에서 한점 흐트러짐 없이 공부에 專念히 주는 영숙씨등 저에게 直間接으로 도움을 준 모든분들께 紙面을 통해서나마 고마운 마음을 전하고 싶습니다.

끝으로 오늘의 제가 되기까지 오직 犧牲으로 點綴하여 오신 父母님께 이 작은 結實을 올리고자 합니다.

1990년 초여름

김 진 구