

# 統計實務基礎理論(I)

(講義教材)

1982

經濟企劃院 調查統計局

本 教材는 UNDP 의 後援하에 '82. 5. 17 ~ '82. 6. 18

간 SIAP 와 調査統計局이 共同으로 實施하는 統計要員訓

練教材로 作成한 것임.

# 目 次

|                                 |    |
|---------------------------------|----|
| ✓ 第 1 章 統計와 그 <u>利用</u> .....   | 11 |
| 1.1 統計의 意義 .....                | 11 |
| 1.2 統計의 利用과 重要性 .....           | 13 |
| 1.2.1 概 要 .....                 | 13 |
| 1.2.2 生産過程으로서의 統計形態 .....       | 14 |
| 1.2.3 統計利用의 一般的인 經驗 .....       | 15 |
| ✓ 1.2.4 政府機關에 對한 統計의 重要性 .....  | 19 |
| 1.2.5 公共統計制度 .....              | 20 |
| 1.2.6 各政府機關의 統計利用 .....         | 23 |
| 1.2.7 政府의 統計需要에 對한 实例 .....     | 26 |
| 1.2.8 特殊한 事例 .....              | 30 |
| 1.2.9 金融 및 企劃部署의 特別한 統計需要 ..... | 31 |
| ✓ 1.2.10 統計의 其他利用者 .....        | 32 |
| 1.2.11 統計作成者의 需要에 對한 評價 .....   | 34 |
| 1.2.12 特殊統計의 諸利用 .....          | 36 |
| 1.2.13 統計作成을 위한 相互協力 .....      | 37 |
| ✓ 1.3 用語의 基本概念 및 定義 .....       | 39 |
| <br>                            |    |
| 第 2 章 統計調査의 企劃 및 設計 .....       | 44 |
| 2.1 概 要 .....                   | 44 |
| 2.2 새로운 統計調査遂行의 諸過程 .....       | 45 |

|          |                     |    |
|----------|---------------------|----|
| 2.3      | 새로운 統計調査의 起源        | 46 |
| 2.4      | 새로운 統計調査의 企劃과 設計    | 47 |
| 2.4.1    | 目的의 樹立              | 48 |
| 2.4.2    | 行政的 準備              | 49 |
| 2.4.2(a) | 法律規定                | 49 |
| 2.4.2(b) | 組織, 管理 및 調整         | 50 |
| 2.4.2(c) | 豫算                  | 51 |
| 2.4.2(d) | 資源                  | 54 |
| 2.4.2(e) | 調査体系                | 57 |
| 2.4.2(f) | 調査時期                | 57 |
| 2.4.2(g) | 日程表                 | 58 |
| 2.4.2(h) | 公表 및 弘報             | 63 |
| 2.4.3    | 技術的 準備              | 66 |
| 2.4.3(a) | 概念 및 定義             | 66 |
| 2.4.3(b) | 調査形態, 統計活動 및 資料蒐集方法 | 69 |
| 2.4.3(c) | 应答者の 範圍, 対象 및 選定    | 70 |
| 2.4.3(d) | 標本の 抽出              | 71 |
| 2.4.3(e) | 指針書 및 調査票類          | 72 |
| 2.4.3(f) | 데이터 쉬트의 設計          | 73 |
| 2.4.3(g) | 에디팅, 資料處理 및 公表      | 75 |
| 2.4.3(h) | 募集 및 訓練             | 76 |
| 2.4.3(i) | 事前檢討                | 77 |

|                         |     |
|-------------------------|-----|
| 第3章 資料蒐集方法              | 79  |
| 3.1 概要                  | 79  |
| 3.2 資料出處                | 80  |
| 3.3 統計的 觀察方法            | 81  |
| 3.3.1 実査：個人面接           | 83  |
| 3.3.2 実査：測定 및 物理的 觀察    | 86  |
| 3.3.3 自計式調査：郵便調査에 依한 蒐集 | 87  |
| 3.3.4 個人面接과 自計式調査方法의 差異 | 89  |
| 3.3.5 行政記錄으로 부터 拔萃      | 92  |
| 3.4 統計利用의 諸形態           | 94  |
| 第4章 資料蒐集活動              | 98  |
| 4.1 概要                  | 98  |
| 4.2 現地에서의 資料蒐集          | 99  |
| 4.2.1 現地調査 節次           | 100 |
| 4.2.2 指導員의 役割           | 110 |
| 4.2.3 現地業務에 있어서의 基本問題点  | 114 |
| 4.2.4 面接技術              | 121 |
| 4.2.5 現地測定 및 觀察의 技術     | 129 |
| 4.2.6 地圖의 使用과 作成        | 130 |
| 4.3 郵便調査                | 133 |
| 4.3.1 郵便調査의 体系          | 133 |
| 4.3.2 資料蒐集의 管理          | 134 |

|       |                                |     |
|-------|--------------------------------|-----|
| 4.3.3 | 目錄의 作成과 備置                     | 137 |
| 4.4   | 既存記錄으로 부터의 統計產出                | 138 |
| 4.4.1 | 記錄의 形態                         | 139 |
| 4.4.2 | 資料의 制約과 質的改善                   | 141 |
| 4.4.3 | 編輯計劃                           | 143 |
| 第5章   | 標本抽出의 原理                       | 147 |
| 5.1   | 概要                             | 147 |
| 5.2   | 概念과 定義                         | 152 |
| 5.3   | 標本抽出方法                         | 159 |
| 5.4   | 任意標本抽出의 種類                     | 163 |
| 5.4.1 | 單純任意抽出                         | 163 |
| 5.4.2 | 系統抽出                           | 164 |
| 5.4.3 | 層化抽出                           | 166 |
| 5.4.4 | 集落抽出                           | 168 |
| 5.4.5 | 2段抽出                           | 168 |
| 5.4.6 | 多段抽出                           | 169 |
| 5.4.7 | 크기에 比例한 確率抽出                   | 170 |
| 5.5   | 標本誤差                           | 171 |
| 5.6   | 標本調査에 있어서 實際적으로 特別히 考慮되어야 할 事項 | 172 |
| 5.6.1 | 亂數表의 使用                        | 172 |
| 5.6.2 | 標本の 交替                         | 179 |

|       |                   |     |
|-------|-------------------|-----|
| 5.6.3 | 標本의 代替            | 181 |
| 5.6.4 | 標本의 維持            | 182 |
|       |                   |     |
| 第 6 章 | 誤差의 源泉            | 183 |
| 6.1   | 概 要               | 183 |
| 6.2   | 非標本誤差와 偏倚         | 185 |
| 6.2.1 | 応答誤差              | 185 |
| 6.2.2 | 偏倚                | 187 |
| 6.3   | 非標本誤差의 源泉         | 190 |
| 6.3.1 | 企劃段階에서의 誤差        | 190 |
| 6.3.2 | 資料蒐集過程의 誤差        | 195 |
| 6.3.3 | 資料處理過程의 誤差        | 202 |
| 6.4   | 農業統計에 있어서의 特別한 問題 | 206 |
|       |                   |     |
| 第 7 章 | 誤差의 管理            | 218 |
| 7.1   | 概 要               | 218 |
| 7.2   | 改善된 設計와 企劃        | 218 |
| 7.2.1 | 定義와 分類            | 218 |
| 7.2.2 | 事前調査와 파일로트 씨베이    | 220 |
| 7.3   | 改善된 資料蒐集          | 223 |
| 7.3.1 | 訓 練               | 223 |
| 7.3.2 | 監 督               | 224 |
| 7.4   | 資料의 確認과 內容檢査      | 225 |

|          |                                      |     |
|----------|--------------------------------------|-----|
| 7.4.1    | 內容檢査의 種類                             | 225 |
| 7.4.2    | 內容檢査의 原則                             | 229 |
| 7.4.3    | 統計報告書의 內容檢査節次                        | 231 |
| 7.4.4    | 合計段階에서의 內容檢査                         | 236 |
| 7.5      | 事後調査                                 | 239 |
| 7.6      | 改善된 資料處理方法                           | 242 |
|          |                                      |     |
| 第8章      | 資料處理                                 | 246 |
| 8.1      | 概要                                   | 246 |
| 8.2      | 資料處理組織의 計劃                           | 247 |
| 8.3      | 資料處理 作業                              | 250 |
| 8.3.1    | 檢討 및 誤謬修正                            | 250 |
| 8.3.2    | 符號記入                                 | 250 |
| 8.3.2(a) | 特性의 코딩                               | 250 |
| 8.3.2(b) | 變數의 코딩                               | 251 |
| 8.3.2(c) | 地域 코드構造(Hierarchical code structure) | 252 |
| 8.3.2(d) | 調査票上에 주어진 코드                         | 254 |
| 8.3.3    | 資料의 轉換                               | 256 |
| 8.3.4    | 資料의 分類集計                             | 259 |
| 8.3.4(a) | 手分類                                  | 259 |
| 8.3.4(b) | 텔리(Tally)表示法                         | 261 |
| 8.3.4(c) | 手分類의 限界                              | 262 |
| 8.3.5    | 演算作業                                 | 263 |



|                                  |     |
|----------------------------------|-----|
| 8.3.5 (a) 四捨五入 .....             | 263 |
| 8.4 機械를 利用한 資料處理 .....           | 265 |
| 8.4.1 計算機 .....                  | 266 |
| 8.4.2 穿孔카드 시스템 .....             | 268 |
| 8.4.2 (a) 機械集計 .....             | 268 |
| 8.4.2 (b) 에즈 (Edge) 穿孔카드 .....   | 269 |
| 8.4.3 컴퓨터에 依한 資料處理 .....         | 271 |
| 8.4.3 (a) 컴퓨터의 크기 및 價格推移 .....   | 271 |
| 8.4.3 (b) 컴퓨터의 原理 .....          | 272 |
| 8.4.3 (c) 컴퓨터 프로그램 .....         | 274 |
| 8.4.3 (d) 資料의 入力과 出力 .....       | 275 |
| 8.4.3 (e) 컴퓨터에 依한 資料處理 .....     | 278 |
| 8.4.3 (f) 結 論 .....              | 282 |
| <br>                             |     |
| 第9章 統計資料의 提供과 公表 .....           | 284 |
| 9.1 概 要 .....                    | 284 |
| 9.2 테블러 (Tabular) 방식의 資料提供 ..... | 284 |
| 9.2.1 統計表의 構成要素 .....            | 285 |
| 9.2.2 統計表의 形態 .....              | 287 |
| 9.2.3 統計表 作成要領 .....             | 293 |
| 9.2.4 表側과 表頭上의 項目配列 .....        | 296 |
| 9.3 函表作成 .....                   | 297 |
| 9.3.1 目的 .....                   | 297 |

|        |                 |     |
|--------|-----------------|-----|
| 9.3.2  | 基本的인 圖表         | 297 |
| 9.3.3  | 圖表의 主要한 形態      | 300 |
| 9.3.4  | 가장 適切한 形態의 圖表選定 | 313 |
| 9.4    | 報告書의 解説 및 作成    | 314 |
| 9.4.1  | 概要              | 314 |
| 9.4.2  | 報告書 作成          | 315 |
| 9.5    | 統計刊行物의 作成       | 316 |
| 9.6    | 統計資料의 秘密保護      | 318 |
| 9.7    | 其他의 統計結果 提供方法   | 320 |
|        |                 |     |
| 第 10 章 | 統計分析            | 323 |
| 10.1   | 概要              | 323 |
| 10.2   | 度数分布            | 323 |
| 10.2.1 | 基本概念과 原理        | 323 |
| 10.2.2 | 연속데이터와 이산데이터    | 329 |
| 10.2.3 | 用語의 定義          | 332 |
| 10.2.4 | 累積度数分布          | 333 |
| 10.2.5 | 度数分布圖表          | 335 |
| 10.2.6 | 母集團의 分布         | 342 |
| 10.3   | 平均과 여러位置測定      | 345 |
| 10.3.1 | 平均의 形態          | 345 |
| 10.3.2 | 算術平均            | 348 |
| 10.3.3 | 中位数             | 352 |

|        |                       |     |
|--------|-----------------------|-----|
| 10.3.4 | 最頻數                   | 356 |
| 10.3.5 | 幾何平均                  | 357 |
| 10.3.6 | 平均의 여러形態와 그 使用에 關한 要約 | 359 |
| 10.4   | 散布度                   | 361 |
| 10.4.1 | 基本原理                  | 361 |
| 10.4.2 | 範圍                    | 363 |
| 10.4.3 | 四分位範圍                 | 364 |
| 10.4.4 | 平均偏差                  | 366 |
| 10.4.5 | 標準偏差                  | 368 |
| 10.4.6 | 散布度에 關한 要約            | 373 |
| 10.5   | 標本誤差                  | 374 |
| 10.5.1 | 標準誤差                  | 374 |
| 10.5.2 | 信賴區間                  | 378 |
| 10.5.3 | 結 論                   | 380 |
| 10.6   | 標本데이터로부터 母數推定         | 381 |
| 10.6.1 | 不偏推定量                 | 381 |
| 10.6.2 | 比率推定                  | 385 |
| 10.7   | 推定과 豫測을 위한 統計資料의 利用   | 388 |
| 10.7.1 | 單純線型回歸                | 389 |
| 10.7.2 | 相關係數                  | 396 |
| 第 11 章 | 消費者物價指數 및 其他指數        | 398 |
| 11.1   | 概 要                   | 398 |

|        |             |     |
|--------|-------------|-----|
| 11.2   | 指數의 形態      | 398 |
| 11.3   | 消費者 物價指數    | 399 |
| 11.3.1 | 消費者物價指數의 作成 | 401 |
| 11.3.2 | 消費者物價指數의 管理 | 411 |
| 11.4   | 其他物價指數      | 417 |
| 11.5   | 數量指數        | 418 |
| 11.6   | 金額指數        | 420 |
| 11.7   | 指數의 公式      | 420 |
| 11.8   | 指數公表 및 解析   | 425 |
| 第12章   | 統計制度의 基準    | 430 |
| 12.1   | 概要          | 430 |
| 12.2   | 統計體系        | 430 |
| 12.3   | 統計基準        | 436 |
| 12.3.1 | 概要          | 436 |
| 12.3.2 | 標準定義        | 437 |
| 12.3.3 | 國際分類系       | 439 |
| 12.4   | 統計蒐集의 範圍    | 444 |
| 12.4.1 | 概要          | 444 |
| 12.4.2 | 經濟統計        | 445 |
| 12.4.3 | 非經濟社會統計     | 447 |
| 12.4.4 | 環境統計資料      | 448 |
| 12.4.5 | 統計活動의 構造    | 448 |

# 第1章 統計와 그 利用

## 1.1 統計의 意義

統計라는 말은 거의 모든 사람이 알고 있으나 이의 뚜렷한 意味는 잘 모르고 있다. 가령 10 사람에게 統計의 意義를 質問하면 10個의 다른 對答을 얻을 것이며, 이들中 어느 하나도 完全한 것은 없을 것이다. 統計의 意義를 말로 表現할 수는 없다 하더라도 우리는 日常生活에서 여러 形態로 統計를 利用하고 있다.

우리가 統計에 對한 이야기를 할때 혼돈을 일으키는 理由中의 하나는 통계라는 單語가 여러 意味를 가지고 있기 때문이라 할 수 있다. 이것은 統計에 無關心한 사람에게는 問題가 되지 않을 것이나 統計人으로서의 最小限 그들 意味間의 差異에 對해서는 理解하고 있어야 할 것이다. 大部分의 사람들이 理解하고 있는 統計의 첫번째 意味는 :

' 体系的으로 蒐集된 計數的 事實 '

( 옥스포드 辭典 )

또는 ' 量的資料의 蒐集 ' ( 웹스터의 Third New International 辭典 ) 이라 말할 수 있다.

이러한 意味로 볼때 統計란 恒時 複數的인 意味로 使用되어지며 數로써 이루어지거나 要約된 情報體系로 意味된다. 이와 關聯하여 우리는 統計代身에 흔히 데이터(data)라는 말을 使用

한다.

두번째 意味는 數的資料의 蒐集 및 分析하는 方法에 關係된다고 볼 수 있다.

即 '統計는 大量의 數的資料를 蒐集, 分析, 解説 및 이의 公表를 다루는 科學이다.' (웹스터의 Third New International 辭典)

이러한 意味에서 볼때 統計는 單數的으로 使用되며 이는 研究해야할 主題 또는 分野인 것이다.

세번째의 意味는 標本으로부터 얻은 要約된 資料를 나타내기 위하여 使用되어진다.

統計라는 말이 이러한 세가지 意味에서 使用되는 例는 다음과 같다.

(a) 計數的 意味의 統計

經濟開發을 企劃하는 데에는 農業生産統計를 자주 必要로 하고 있다.

(b) 理論的 意味의 統計

開發途上國家에서는 政府機關에 統計訓練을 받은 사람이 必要하다.

(c) 標本으로부터 얻은 統計

標本으로부터 計算된 特性值를 統計量(Statistic)이라 한다.

이러한 相異한 定義를 모두 記憶할 必要는 없지만 같은 말이 라도 여러 意味가 있다는 것은 알 必要가 있는 것이다. 오늘

날 統計學이란 매우 龐大하고도 複雜한 學問이며 어느 누구도 그것의 모든 意味를 完全히 把握할 수는 없는 것이다.

統計學의 많은 理論은 소위 確率理論이라 稱하는 數學의 特殊 部門에 根柢를 두고 있으나 本課程에서는 이러한 數學的인 面의 理論은 다루지 않고 代身 統計가 政府에 依해 蒐集되는 方法과 統計가 어디로부터 얻어지며 現在 어떻게 利用되고 있는지에 대한 몇가지 方法에 對해 關心을 두고자 한다. 그러나 앞으로 統計學을 보다 깊이 繼續해서 研究하고자 하는 者는 누구나 數學的 理論을 習得하는데 상당한 時間을 消費해야할 것이다.

## 1.2 統計의 利用과 重要性

### 1.2.1 概要

한 나라의 統計體制속에 어떻게든 關여하고 있는 많은 사람들은 때때로 統計의 必要性和 有用性에 대하여 회의를 하지 않으면 안될 경우가 있다. 또한 그들이 하고자 하는일은 때로는 단조롭고 실증나며 無意味하게 보이는 경우가 있다. 그리고 그가 어떠한 일에 從事하는지? 또한 왜 그러한 일을 하고 있는 지?를 自身들에게 說明하는 것조차 어려울 때가 있으며 家族이나 親舊에게 이를 說明하기란 더욱 어려운 것임을 알게 될 것이다. 이러한 사실을 감안하여 볼때 우리는 그들이 業務에 熱意를 가지고 또한, 업무의 價值를 認定하지 않는 限 正直하고도 誠實하게 業務를 遂行하리라고는 期待할 수 없다. 따라서 우리는 統計가 왜 重要한가를 詳細하게 討論하면서 本 教材에 들어

가고자 한다.

### 1.2.2 生産過程으로서의 統計形態

어떤 면에서 우리는 統計制度를 工場이나 其他 生産過程과 어느程度 類似한 것으로 생각할 수 있다. 蒐集된 資料( 調査票를 利用한 実査에 依해 蒐集되었든 다른 手段에 依해 蒐集되었든 간에 )는 生産過程의 原資料로 볼 수 있다. 資料를 審査하고 質을 向上시키기 위하여 遂行해야 할 일은 生産과정에서 原資料를 分析 處理하는 것과 같으며 發刊된 最終報告書나 刊行物 그리고 利用者가 利用할 수 있는 要約된 統計는 生産過程의 最終 産物이다. 그러나 統計制度가 生産過程과 어떤 類似性を 가진다 하더라도 事實上 여러가지 差異점이 있다.

이러한 差異점은 大部分의 生産과정은 모든 從業員들이 하나의 設備 내에서 作業하는데 비하여 이와는 달리 統計機關은 數千名을 分散하여 雇傭하고 있어 그 規模나 分布에 依해서 發生하고 있다. 即 統計從事員들은 本部事務所나 地方事務所 그리고 全國의 小規模 마을까지에도 分布되는 경우가 많은 것이다. 이것은 특히 地方事務所 從事員들은 統計機關이 作成하는 最終産物과의 關聯性에 關心이 적을뿐 아니라 統計利用者와의 關係는 더욱 적다는 것을 意味하고 있다.

釀造場이나 自動車 生産工場의 從事員들은 그들이 生産過程의 어떤 分野에 從事하고 있거나 事務를 遂行하든간에 事實上 그들 會社에서 무엇이 生産되고 그後 生産에 어떤 問題가 發生할 것인가를 알고 있다.



이와 마찬가지로 댐이나 橋梁建設事業에 從事하는 者는 그들 주위에서 進行되고 있는 作業狀況과 무엇이 遂行되고 있는가를 알 수 있으며 그들이 參與하고 있는 事業 全体狀況에 對한 重要性 및 價值를 完全히 評價할 수 있다.

그러나 統計從事者는 이러한 利點을 갖지 못하기 때문에 그들은 自身들이 全体組織의 一部分이고 이에 屬한다고 생각하지 않는 경우가 있다. 따라서 그들은 그들 業務에 對한 一體感을 造成키 위하여 疑問事項에 對하여 많은 質問을 할 수 있는 勇氣를 가지는 것이 바람직하다. 어떤 경우에는 그들이 알아야 할 疑問事項이 많이 있는데 예를 들자면 누가 統計를 利用하고 어떻게 利用하며 왜 利用하는가? 모든 統計는 實際로 다 利用되고 있는지? 우리가 統計蒐集活動을 前부터 해왔기 때문에 그들을 계속 蒐集하는 것인지? 等이다. 우리는 第1章에서 이러한 質問과 其他 類似한 質問에 對해 살펴 보고자 한다.

### 1.2.3 統計利用의 一般의인 經驗

먼저 '누가 統計를 利用하는가?'라는 첫번째 質問에 對하여 部分的으로 答辯하고자 한다. 公式統計의 利用者는 많이 있겠으나 政策을 樹立하고 事業을 企劃함에 있어 統計를 利用하는 政府는 하나의 主要利用者이다. 勿論 이것은 우리의 質問에 對한 가장 單純한 對答이며 조금후에 보다 詳細하게 이의 体系的인 檢討를 하기로 하자.

公共統計機關이 作成한 統計를 政府가 어떻게 利用하고 있나를 살펴보기 前에 個人의 日常生活에서 統計를 어떻게 利用하는가를

考慮해 보는것은 도움이 될 것이다. 우리 모두는 우리가 그렇게 하고 있다는 事實을 正確히 알지 못한채 통계를 利用하고 있다. 이를 說明하기 위하여 例를들면 우리가 商店에서 새로운 라디오를 購入할 경우 行하여야 할 마음의 準備를 考慮해 보자. 우선 우리가 蒐集해야할 몇가지 情報가 있는데 即 라디오값과 支払할 수 있는 金額을 알 必要가 있다. 그러나 우리가 신중하다면 이보다 좀더 많은 情報를 蒐集할 것이다. 한 例로, 다른 商店에서의 同種의 라디오 價格이라든가 우리가 購入하고자 하는 다른 物件의 값을 들수 있는데 이것이 오히려 우리에게 重要할지 모르며 라디오를 購入했다 하더라도 支払金額을 더 줄였을 것이다.

이와같이 우리는 物件을 購入할지의 여부에 對한 決定을 내리는데 도움이 될 많은 統計를 가지고 있으며 이들은 우리로 하여금 合理的인 決定을 내릴 수 있도록 해주고 있다.

例를들어 現金으로 라디오를 購入하거나 預金額 및 分割金으로 代金を 支払하는 方法을 들수 있다. 두번째 方法의 長點을 알아보기 위하여 우리는 豫想所得( 다른 物件값을 支払하기 위하여 所得을 얻게 된 後)에 比較하여 알아야할 利子와 定期分割払入金等이 얼마인가 하는 追加情報를 알아야 할 것이다.

우리가 이러한 方法으로 統計의 利用度를 넓혀간다면 現在의 여건( 即 支出할 수 있는 金額과 即時 支出해야 하는 金額)을 考慮하게 되고 우리가 然한 여건을 알아보기 위해 未來를 計劃하게 되는데 이는 매우 重要한 일이다. 우리가 이러한 方法으로

統計를 蒐集한다면 라디오의 購入여부에 대한 敏感한 決定을 할 수 있는 充分한 資料를 가지고 있어야 한다. 그렇게 하는것이 最善의 購入方法일 것이다. 그러나 이러한 去來에 關聯되는 統計를 利用하는 것으로 끝나는 것이 아니고 蒐集할 情報가 더 있는 것이다. 例를들어 分割金 払入期間동안 支出을 위한 豫算을 편성하고 支払狀態를 記錄해야 한다.

記錄方法은 2가지 原則이 있는데 即 첫째는 支払狀態를 点檢하는 分명한 것으로서 最近의 支払狀態를 維持하기 위함이고 둘째는 우리가 겪었던 問題点이나 遲滯行爲를 記錄하는 것인데 이것은 앞으로 있을지 모르는 類似한 去來行爲에 있어 우리에게 도움을 주게 된다. 事實上 以前의 去來行爲를 위한 類似한 記錄을 維持한다면 라디오 購入에 關한 決定을 내림에 있어 過去의 經驗을 基礎로 計劃을 樹立할 수 있다.

아마 우리는 다음과 같은 것에 스스로 自問自答할 수 있을 것으로 본다. 2년전 割賦制로 自轉車를 購入했을 때 8個月에 걸쳐 代金を 支払할 計劃을 세웠으나 다른 經費를 위한 充分한 收入이 없어서 12個月에 걸쳐 支払하게 된 경우이다. 이러한 經驗은 우리가 豫想한 期間以上에 걸친 支払計劃을 마련함으로써 現在의 去來行爲에 대한 우리의 決定을 修正토록 하게 한다.

우리가 討議해 온바와 같이 아주 單純한 狀況下에서도 決定을 내리는 過程에는 利用 可能한 많은 統計들이 있음을 위에서 볼 수 있었다. 우리는 모든 数字를 記錄하고 正確한 計算을 해둘 필요가 반드시 있는 것은 아니지만 최소한 마음의 準備는 해두

어야 할 것이다. 勿論 다음과 같은 決定을 내리는데 있어서 이와같은 統計가 항상 利用될 必要는 없다. 卽 우리가 라디오 값을 支払할 充分한 돈을 가지고 있으면 이러한 問題에 對해 더 以上 考慮해볼 必要없이 商店에 갈수있고 購入署名을 할 수 있다. 이와같은 경우에 우리는 現在를 위한 生活哲學을 採択하여 未來에 이를 活用하거나 앞으로 보상 받을지의 여부를 即刻的으로 判斷할 수 있다는 決定을 내리게 된다. 비록 大部分의 사람들이 어떠한 事實이나 數學에 對한 注意깊은 研究에 根拠한 決定이 直觀에 根拠한 決定보다 좋다고 믿고 있다하더라도. 決定을 함에 있어 어느 方法이 좋은가는 判斷의 問題인 것이다. 그러나 이에 더해야할 하나의 重要한 點이 있는데 이는 事實을 根拠로 한 決定은 그 事實이 正確하고 믿을 수 있어야 한다는 것이다. 例를들어 未來의 所得水準에 關한 資料가 매우 不正確하다면 우리가 내린 決定이 옳지 않다 하더라도 결코 놀라운 일은 아니다. 우리는 사람들이 統計를 利用하는 原則的인 方法을 우리가 알 수 있도록 實例를 充分히 들어왔다. 勿論 統計를 利用할수 있는 다른 方法도 있으나 主要利用方法을 要約하면 다음과 같다.

(a) 決定을 내림에 있어

- 現在狀況을 研究하고
- 過去의 經驗을 指針으로 삼아
- 앞으로 發生할 것에 對한 計劃을 樹立한다.

(b) 事業이나 活動의 進行過程을 把握한다.

#### 1.2.4 政府機關에 對한 統計의 重要性

政府機關( 그리고 統計의 其他 主要 利用者 )이 如何하고 있는 立場은 우리가 들은 간단한 實例와 많은 類似點을 가지고 있으며 이들의 統計利用도 根本적으로 같을 것이다. 政府機關은 願하기만 한다면 짐작, 추측 또는 직관에 따라 政策決定을 할수 있다. 그러나 통계를 이용함으로써 政府는 보다 좋은 政策을 決定할 수 있으며 統計가 正確한限 이를 基礎로 政策決定을 하는 경우 커다란 오류를 犯하지 않을 것이다. 더우기 政策機關은 個人과는 規模로 보아 分명한 差異가 있기 때문에 合理的이며 論理的인 方法으로 政策을 決定하도록 하고 있다. 個人이 決定을 하게 되는 경우에는 考慮해야할 要因들이 거의 없을 것이며 直觀的인 決定이 좋을수도 있다. 그러나 政府機關은 考慮해야할 많은 要因들을 가지고 있는 大規模 事業을 管理하고 있다. 우리는 公務員들이 그들 全部를 한눈에 把握하여 直觀的인 좋은 決定을 내릴수 있는 能力을 期待할 수는 없다. 政府 決定은 數百萬 달러를 消費하는 事業으로서 그 影響力은 매우 크며 그들의 決定은 全國民의 雇傭, 生活方式 및 福祉에 影響을 주게 된다. 이러한 要因들은 매우 重要하므로 事實上 世界各國政府는 구체적이며 効率的인 公共統計制度를 樹立할 必要性을 가지고 있는데 이 制度는 政策決定을 내리기 위해 要求되는 많은 情報를 提供하고 組織的이고 体系的이며 信賴性있는 方式으로 社會 여러 分野의 成長과 開發을 監視할 것이다. 사실상 한나라가 더욱 高度로 發達함에 따라 그 나라의 統計制度는 더 公公

해진다.

### 1.2.5 公共統計制度

우리는 이미 韓國家의 公式統計와 公共統計制度에 關해 많은 言及을 해왔다. 우리는 政府가 統計를 어떻게 利用할 것인가를 詳細히 考慮하기 前에 이러한 用語가 意味하는 바를 明白히 定義해 두어야 할 것이다.

이와 關聯하여 우리는 用語의 3 가지 定義中에서 첫째는 統計的으로 蒐集된 數的 事實의 “통계”에 對해 언급하고자 한다. 公式統計란 政府가 蒐集, 公表하는 數的 事實이며 여기에는 中央政府, 地方政府 및 法政當局이 包含될 수 있다.

公共統計制度에 關해 言及해 보면 統計業務는 여러 統計 및 諸 統計活動間에 存在하는 여러 水準과 諸形態의 關係로 여러 政府 機關에 分布되어 있다. 이것은 政府가 뚜렷한 目的없이 非組織的인 方式으로 統計를 蒐集하지않고 可能한 限 体系的으로 蒐集하고 있음을 시사해주고 있다. 本 教材의 마지막 章에서 우리는 各國政府가 여러 方法으로 公式統計를 蒐集하고 있음을 알게 된다. 一部 國家에는 그 나라의 統計資料 大部分을 蒐集할 責任이 있는 中央統計局이 있으며 여타의 國家에서는 大部分의 資料가 다른 여러 部處에 依해서 蒐集되고 統計局은 이러한 活動을 調整하고 있으며 公式統計의 모든 蒐集者들이 利用하는 定義 分類 및 其他 諸基準을 設定하고 있다. 따라서 統計局의 調整 役割은 매우 重要하며 우리가 이른바 統計制度에 關해 이야기할

수 있도록 해주고 있다. 그러나 統計局이 없다면 같은 統計를 여러 機關에서 作成하려 할 것이며 전혀 다른 定義를 利用할 것이다. 이의 結果로 重複의 努力을 하게 되며 한 機關에서 作成된 統計는 다른 機關에서 作成된 統計와 서로 相衝하게 된다. 이러한 경우는 몇년전만해도 많은 나라에 있었으며 아직도 어느 정도는 있다고 본다.

公共統計制度(Official statistical system)의 理念은 이러한 問題를 解決하기 위한 보다 現代的인 接近方法을 反映해 주고 있다. 附錄1에는 公共統計制度가 아시아의 한 國家에서 어떻게 組織되었는가를 보여주는 圖表가 있다. “公共”이란 말은 公表된 資料는 絶對적으로 正確함을 意味하는 것으로 國民이 알고 있지만 반드시 그렇지는 않다. 即 公表된 統計는 統計機關이 統計를 作成함으로써 正確하다고 볼수 있지만 遂行의 모든 過程에서 誤差가 發生할 위험성이 있으며 어떠한 統計도 모든 利用을 위해서는 完全하지 못하다. 그것은 오차를 最小化 하려는 政府統計士이나 職員들의 責任이지만 그들은 誤差를 完全히 除去하지는 못할 것이다. 綜合的인 誤差管理問題는 매우 重要하며 우리는 誤差가 어떻게 發生하며 또한 이를 위해 무엇을 해야 할 것인가를 檢討하기 위하여 本 教材의 2個의 章을 할애하여 說明할 것이다.

公式統計란 보통 사람과 그들의 여러 活動事項을 취급하고 있는데 이러한 活動事項을 直接 觀察한다는 것은 간혹 不可能한 경우가 있어 이때는 다른 手段에 의해 요구된 資料를 얻어야 할 것이다.

우리가 볼 수 있는 바와같이 우리는 一部分의 單位만으로 부터 統計를 蒐集하고 이를 利用하여 全体를 推計해서 얻은 結果를 利用한다. 이러한 것은 公表된 最終結果가 唯一한 接近값이 되는데 그것은 알지 못하는 참값에 대한 推計값이 되는 것임을 意味한다.

大部分의 統計理論은 推計값을 얻는데 關係하고 있으며 不確實한 條件下에서 決定을 내리기 위해 이를 利用하고 있다. 大部分의 國家에서 政府는 特別法에 依하여 統計局을 設置하고 統計를 作成하도록 권한을 부여하고 있다. 이러한 法律은 統計人으로서 하여금 사람이나 組織에 대하여 必要한 資料를 提供할 것을 要求할 수 있도록 해주고 있다. 만일 그들이 質問에 不応하거나 資料提供을 거절할때는 기소를 당하고 벌금을 물게 된다.

그러나 統計法은 同時에 個人이나 事業체로 부터 蒐集한 資料는 秘密로 할 것을 規定하고 있다. 即 統計局의 權限을 받은 사람 以外에는 누구에게도 伝해질 수 없고 다른 政府機關의 職員에게도 伝해질 수 없도록 하고 있다.

예를들어 事業체가 統計局에 提供한 資料는 税金賦課에 利用될 수 없다. 이러한 保護規定이 없는 限 統計人에게 正確한 資料를 提供해 주는것은 期待할 수 없다. 이러한 秘密規定이 널리 알려진다는 것도 重要한 事實이다. 統計資料는 單位의 그룹에 對한 結果가 나타나도록 集計되어야 하는데 이는 統計의 어떤 利用者라도 個人이나 事業체에 對한 資料를 알지 못하도록 注意깊게 行해져야 한다.

조금後 우리는 本章에서 公式統計의 作成者(即 政府統計人과 職



員)의 입장에서 본 統計의 利用에 關해 살펴볼 것이다. 그러나 우선 우리는 政府나 여타 조직들이 어떻게, 왜 統計를 利用하는가에 對하여 考慮할 必要가 있다.

#### 1.2.6 各 政府機關의 統計利用

소위 政府가 統計의 主要利用者라고 말할때 政治家(특히 政策樹立과 責任을 지고있는 長官)나 行政家(政治家가 決定한 政策履行의 責任있는 政府 各級機關의 公務員)의 兩者에 對하여 言及하고자 한다.

우리는 個人이 統計를 利用하는 方法은 이미 要約했으며 政府의 利用도 根本적으로 같음을 明記하였다. 그러나 政府에 對한 統計의 重要性때문에 이미 言及한 바에 對하여 구체적으로 살펴보고 政府의 統計利用을 4個의 項目으로 나누어 說明하고자 한다.

##### (a) 一般的인 行政目的

한 國家를 每日 管理하는것은 매우 老大하고도 複雜한 活動이다. 이때문에 여러 部처가 있기 마련인데 各部처는 各自의 責任分野인 農業, 教育, 保健, 産業 및 其他의 機能을 가지고 있다. 그들이 業務를 遂行하기 위하여 첫째, 생각하고 計劃할 수 있는 知識의 體系로서 둘째 變動指標와 行動에 對한 案内者로서 믿을 수 있고 關聯性있는 統計와 最近의 統計資料를 많이 必要로 하고 있다.

두가지의 簡單한 實例는 統計의 重要性을 잘 나타내주고 있는

데 가령 農水産部가 그나라의 여러 地域에서 生産된 糧穀統計를 가지고 있다면 生産増大를 위하여 特別支援을 必要로 하는 低生産地域을 알아내기 위하여 이러한 統計를 利用할 수 있으며 또한 地域에서 生産量이 갑자기 줄었을 경우 原因을 규명할 行動을 취할 수 있는 것이다. 事例로 그 地域이 병충해에 의해 被害를 입을 경우 이를 撲滅해야할 必要가 있는 것이다. 마찬가지로 保健社会部가 한 地域社会에 関한 正規 疾病報告書를 가지고 있다면 그 地域社会의 여러 地域에 関한 統計는 問題地域을 가려 내는데 利用될 것이다. 이러한 統計는 방역계획, 보건서비스, 保健教育등의 必要性을 나타내 주고 있다.

#### (b) 政策樹立

現在의 問題에 對한 決定과 行動을 취하는 일은 차치하고 政府는 未來를 위한 法律을 制定하고 政策을 決定해야 한다. 政策이 일단 決定되면 상당기간동안 効力を 維持하게 되며 앞으로 어떤 種類의 行政行爲가 취해져야 할 것인가를 決定하게 될 것이다. 여러 政策에 대한 例를들면 6~12歲의 兒童을 위한 義務教育, 都市를 벗어난 農村地域의 産業開發促進, 經濟作物을 재배하는 農夫에 對한 支援, 未來의 政府賃金水準決定등을 들수 있다. 政策樹立에 있어 統計의 役割은 政策樹立家가 現在의 狀態를 把握하고 未來의 目標를 設定하며 나아가 이러한 目標를 達成할 指針을 마련하기 위하여 過去에 무엇이 일어났는가를 알수있게 하는 것이다. 우선 政策問題에 直接 關聯되어 있는 이러한 部門의 統計를 分析함으로써 政治家나 行政家가 몇개의 代案을 選択

하는데 도움을 주게 될 것이다. 進展된 分析을 通하여 어떠한 特別政策의 可能한 惠沢을 얻을 수 있으며 이렇게 함으로써 特殊한 環境에서 最善의 것을 選擇할 수 있게 된다.

(c) 開發計劃

經濟開發을 위해 努力하는데 있어 大部分의 國家들은 通常 5 個年 期間의 定期的인 開發計劃을 準備하고 있다. 이러한 開發計劃은 企劃家들이 그 國家가 期間內에 達成할 수 있다고 생각되는 目標을 나타내 주고 있으며 目標達成에 接近해질 것이라는 意味를 보여 주고 있다. 計劃을 進展시키는 데에는 制限된 資源의 利用에 對한 많은 競爭的인 주장들이 있을 것이다. 그 國家는 보다좋은 保健, 教育, 熟練勞動力, 效率的 生産, 道路, 市場에의 接近 및 여타 많은 것들을 願하게 된다. 그러나 特定 期間內에 모든 것을 다 하는것은 可能하지 않다. 目標은 現實的이어야 하고 우선 순위는 가장 重要한 것부터 行해질 수 있도록 決定되어야 한다. 이러한 것은 모두가 매우 複雜한 過程인데 여기에서 良質의 統計는 重要한 役割을 할 수 있다.

(d) 公報活動

政府는 國民이 選擇했기 때문에 國民에게 그들이 무엇을 하고 있는가에 對해 提報해야할 責任이 있다. 統計資料는 國家가 目標을 達成함에 있어 얼마의 進展이 이루어 졌으며 그 國家의 여러 部門은 다른 目標에 關聯하여 어떻게 發展했나를 보여주고 있다. 그러므로 政府가 統計數字를 公表할 때는 그 나

라의 現況에 對하여 國民에게 알려주고 있다.

### 1.2.7 政府의 統計需要에 對한 實例

이제 2 個의 政府機關 即 教育과 高速道路 關係機關에 對해 살펴보자. 上記 機關들은 直面해야할 많은 政策的 考慮事項中의 하나를 實例로 取하고 있고 一般的이고도 올바른 政策을 決定하기 위하여 어떠한 統計가 必要한가를 검토하게 된다. 政府機關이 그들의 日常的 機能을 遂行함에 있어 必要한 各 統計의 多樣性을 모두 알아본다는 것은 쉬운 일이 아니다.

(a) 文敎部는 한나라의 教育制度의 發展과 管理에 責任을 지고 있다. 政策을 決定할 하나의 例를 든다면 國民學校와 中學校 敎師로서 教育받아야 할 人員數가 얼마며 새로운 敎師訓練院( 輩出 되어야 할 자질있는 많은 敎師를 收容할 )이 施設을 갖추어야할 必要性의 有無를 들 수 있다. 이러한 質問에 近接하게 答辯할 수 있는 統計에는 다음과 같은 事項들이 있다.

- 現在 利用可能한 敎師數, 每年 輩出되는 새로 資格을 얻은 敎師數 및 은퇴나 모종의 理由로 敎師職을 떠난 敎師數
- 國民學校, 中學校의 敎師統計把握, 地域別( 어느 地域에서 어떤 敎師를 必要로 하는가를 보여줌 ) 敎師統計의 把握
- 學校級別, 地域別 學生數, 最近의 學生動向( 이 資料는 學生數와 敎師數間에 直接的인 關係가 있기 때문에 變化한다. 即 學生數가 많을수록 많은 敎師가 要求된다 )
- 地域別 就學年齡 以前の 兒童數( 앞으로 몇년간에 就學할 새

로운 學生數를 豫想케 해줌 )

- 出生統計 : 이는 앞으로 繼續 作成되어야 하며 教師의 모집과 訓練은 長期間이 걸리기 때문에 訓練計劃에 關係되는 決定은 미리 해두어야 함.
- 姓別 移民統計 : 教育받아야 할 사람數가 農村地域으로 부터 都市로의 移住者를 包含한 한나라 또는 特定地域의 流出人口에 依해 영향받을 것인지의 여부를 나타내줌
- 學生對 教師比率 : 適切한 教育水準에 이르게 되어 요구되고 있는 比率과 比較됨 ( 現在의 學生數가 適切한가를 나타내줌 )
- 職業別 失業統計 ( 教師로서 資格을 얻은 사람이 있는 가를 보여줌 但, 現在로서 教師職을 얻지 못한 者 )
- 最近 教師訓練院에서 訓練받은 教師의 合格 및 不合格統計 ( 現在 教師訓練을 받은 者中에서 누가 資質있는 者인가를 보여주며 擴大된 計劃이 被教育者로서 能力있는 教師의 關心을 끌수 있는가의 여부를 나타내 줌 )
- 새로운 研修院을 設立하고 새로운 建築事業을 위한 豫想費用을 나타내 주는 統計. 이는 利用할 수 있는 總建築資金統計와 比較되고 있음 ( 建築 우선 순위를 定할 수 있게 하고 研修院에 充分한 資金이 할당된 것인지의 여부를 決定해줌 )
- 建築資材의 價格統計 ( 完成되기 까지 建築事業費가 얼마나 引上될 것인가를 나타내 줌 )

이러한 것은 한 政府機關이 많은 政策中の 하나를 이룩하는데 있어 도움이 되는 광범위한 統計目錄이다. 우리가 作成하는 모든 統計가 實際로 다 利用되지는 않을 것이다. 即 一部資料는 要求된 形態로 利用되지 않는 경우가 있다. 그러나 적어도 이러한 例와 아래 사항은 政府機關이 利用할 수 있는 統計의 範圍를 말해주고 있다. 即 하나의 決定을 내림에 있어서 도움이 되는 여러가지 統計에 對하여 考慮해야 할 것이다.

(b) 高速道路(主要道路) 管理機關은 한나라의 主要道路를 建設하고 管理할 責任이 있다. 이 機關이 내려야 할 代表的인 決定은 市를 通過하는 主要迂廻路를 建設할 것인가 또는 都心地를 경유하는 貫通路를 建設할 것인가의 여부인 것이다. 그러나 考慮해야 할 많은 要素들이 있으며 이들의 一部를 적어보면 다음과 같다.

- 高速道路의 交通事故統計

現在의 交通問題의 심각성을 評價하고 새로운 도로가 生命과 財産을 保護할 것인지를 豫想할 수 있도록 함.

- 交通量 資料

새로운 道路에 의해 節約될 時間에 對한 評價를 할 수 있음.

- 市地域의 小売業統計

이 統計는 많은 서비스業体나 가게 등이 交通路의 變更에 따른 販賣損失로 閉業하게 될지 모르기 때문에 細密한 研究가 必要함. 따라서 一部業体の 移動을 위한 政府의 補助

金에 對하여 考慮해야할 必要가 있음.

- 都市勞動力에 關한 細部資料

이 資料는 事業을 위한 必要勞働者가 地域別로 모집될 수 있는가를 把握 可能토록 해주고 있는데 이러한 資料가 없으면 他地域에서 轉入된 勞働者를 收容할 宿泊施設費를 考慮해야 함.

- 道路施設物의 利用, 除去 및 輸送費에 關한 資料

- 人口 및 農業等에 關한 統計

通行路를 따라 土地가 道路建設을 위해 收容될 경우 얼마의 生産量이 減少되고 얼마의 人口가 影響받을 것인가를 決定하기 위한 統計임.

이러한 資料는 高速道路(主要 道路) 管理機關이 考慮해야할 完전한 것은 아니다. 한例로 이 機關은 한地域의 기후조건, 新設로의 洪水被害事實 및 그 都市의 公해도(새로운 高速道路가 建設되면 減少될지 모르지만)等에 關한 많은 統計資料를 必要로 할 것이다.

우리가 들은 2가지의 例는 敏感하고도 타당성있는 政府決定을 위해 얼마나 많은 統計資料가 必要한가를 보여주기에 充分해야 한다. 지금까지 우리는 各政府機關을 위한 政策決定만을 取扱했지만 이러한 機關들이 每年 해야할 必要있는 類似한 研究事業은 數百, 數千個가 있을 것이며 그들 모두가 政策決定의 土台가 되는 統計를 要求할 것이다. 단지 2個의 機關에 對해서만 言及해 왔는데 農林, 保健, 貿易 및 産業, 鉦業, 港灣, 民間航空, 經濟開發

移民(流入) 및 關稅部門等의 政府機關과 또는 地方政府와 같은 다른 機關에 對해 살펴보다도 類似한 狀況임을 알 수 있을 것이다. 이러한 狀況에서 正確한 統計資料의 廣範圍한 利用은 政府機關이 遂行할 必須的인 것이다.

#### 1.2.8 特殊한 事例

지금까지 우리가 討議한 것은 政府機關 作成統計가 그 의 經常的 機能의 一部에 對한 決定을 내리는데 도움이 되는 必要性에 關해 重點을 두어왔다. 우리가 言及해야할 또다른 重要한 考慮事項은 非常 또는 災難사태에 統計가 종종 必要하다는 것이다. 이때 決定은 매우 迅速히 취해져야 하는데 그렇다하여 요구되는 모든 統計가 迅速히 利用될 수 있는 것은 아니다. 따라서 關聯資料는 可能한 限 迅速히 蒐集할 必要性이 있는 것이다. 이러한 경우에 統計는 標本調査를 통해 얻어져야함은 거의 確實하다. 即 全体에 對하여 把握할 時間的 餘裕가 許容되지 않기 때문에 必要한 資料를 얻기 위하여 母集團의 一部에 對해서만 把握하고 있다. 標本調査는 統計의 가장 重要한 源泉이며 이의 利用은 增加되고 있다.

우리는 第5章과 本 教材의 다음 部門에서 標本에 關하여 詳細히 討議할데지만 잠시 緊急事態의 한例를 들어보고 많은 問題들이 어떻게 제기되며 어느 對策이 만족스러운 것인가를 알아보자. 한나라의 한地域이 農耕地에 相當한 被害를 주는 심한 地진으로부터 고통받고 있다고 생각해 보자. 이 지역은 食糧이



나 其他 生必品 不足으로 고통을 받기 쉽다. 即時 提起될 問題의 一部는 다음과 같다.

- 얼마나 많은 사람들이 지진으로 被害를 받았는가?
- 얼마나 많은 農耕地가 유실되었는가?
- 豫想되는 食糧 不足量은 얼마인가?
- 食糧이 다시 地域的으로 生産되기 까지 얼마의 時日이 걸릴 것인가?
- 얼마나 많은 사람이 他地域으로 移住되어야 할 것인가?
- 얼마의 비축식량이 그 나라의 다른 地域으로부터 供給될 것인가?
- 얼마의 식량이 輸入되어야 할 것인가?
- 어떠한 交通手段이 利用될 것인가?
- 救護活動費用은 얼마인가? 等이다.

이러한 問題와 其他 이와 類似한 問題에 對해 만족스럽게 答辯하거나 救護活動을 잘 計劃하려면 方대한 統計資料의 配列이 要求될 것임은 分明하다.

#### 1.2.9 金融 및 企劃部署의 特別한 統計需要

우리는 政府의 統計需要가 얼마나 많은 가를 알아보기 위하여 充分한 實例를 들었다. 그러나 이러한 問題를 끝마치기에 앞서 政府의 다른 機關보다 統計를 보다 많이 必要로 할 2個의 特殊機關에 對해 特別히 言及하고자 한다.

이들 機關은 普通 金融部處와 企劃部處로 알려져 있는데 그나

라의 金融企劃과 行政의 責任을 맡고 있다. 以前의 實例에서 言及한 바와같이 이들 部處는 할당받은 豫算의 範圍內에서 各其의 活動分野에 對한 政策을 決定한다. 그러나 金融 및 企劃部處의 長官과 高位 經濟諮問官들은 이를 全國的인 次元에서 淸급해야만 한다. 그들은 政府가 늘리고자 하는 總通貨量(이렇게 하기 위해서는 그 나라의 諸分野에 對한 税金種類와 其他料金, 課稅水準 및 課稅效果等을 檢討해야 한다)과 이 通貨가 어떻게 쓰여질 것인가(各 部處와 法政當局이 提示한 金融的 提案을 評價하고 우선 順位를 결정하며 이에 따라 資金을 할당하는 것을 意味함)를 決定해야만 한다. 이것은 巨大한 課題인데 事實上 社會의 諸分野를 망라하는 廣範圍한 統計資料를 要求하고 있다.

일찌기 檢討한 바와같이 個人과 政府의 主要利用者 特히 企劃部處間의 統計需要上의 差異는 分明 規模上의 差異인 것이다. 個人은 決定을 내림에 있어 考慮할 事項이 거의 없으므로 어떠한 事實이나 統計數字없이도 直觀的으로 合理的인 決定을 내릴수 있다. 그러나 감당할 수 없을 程度의 많은 考慮를 해야하는 企劃部處는 그렇게 할 수는 없는 것이다.

#### 1.2.10 統計의 其他利用者

政府以外的 統計利用者에는 여러 사람과 機關들이 包含된다. 企業家들은 그들의 事業問題와 展望에 關聯되는 資料를 要求하고 原資材의 價格과 在庫量, 賃金率, 輸出入, 其他等에 關한 具體的인 資料는 그들에게 特히 重要하다. 統計는 그들이 새로

은 商品을 生産하고 販賣함으로써 또한 다른 地域에서 새로이 事業을 開始함으로써 그들의 事業을 擴張할 것인가의 与否, 그들이 政府로 부터의 支援 例를들어 對外競争에 對하여 그들을 保護하기 위한 關稅惠沢을 必要로 하고 있는지의 与否, 海外市場에 對하여 그들의 事業을 開始 또는 擴張할 것인가의 与否(即 커다란 效果를 거두기 위해서는 廣告費를 써야 함), 그들이 事業을 有利하게 運營하기 위해서 價格을 引上해야 하는지의 与否, 教育制度가 그들이 앞으로 모집할 必要가 있을 것으로 보는 熟練工을 배출하고 있는지의 与否等を 決定하고 그들이 事業을 效果的으로 運營함에 있어 그들에게 매우 重要視되는 類似性있는 考慮事項을 決定하는데 있어 도움을 줄 것이다.

國際機構는 또한 重要한 統計利用者 이다, UN 統計局 (United Nations Statistical Office) 과 亞細亞太平洋 經濟社会 委員會 (ESCAP)와 같은 機構는 많은 國內外的 關心事에 關한 統計를 얻기 위하여 일하고 있으며 많은 統計 刊行物을 發刊한다. 이러한 資料는 그들이 여러 分野에서 이룬 進척사항과 이웃나라 또는 해결해야할 類似한 問題를 가지고 있는 나라가 達成한바를 比較하는데 있어 政府에 크게 도움이 되고 있다. 이러한 機構들은 援助資金이 어떻게 分配되고 어느 나라가 가장 큰 도움을 要하고 있는지를 決定함에 있어 그 機關에 協調하는 것은 價值있는 일이다.

亞細亞開發銀行(Asian Development Bank)과 같은 國際金融機構 역시 重要한 統計利用者이다. 그들은 各國이 스스로 資金을

充當할 수 없는 大規模開發事業借款을 提供하고 있다. 그러나 그들은 事業을 위한 金融支援을 同意하기 前에 그 事業이 實踐力이 있고 支援價值가 있는가를 確認하기 위하여 도움이 될 많은 統計를 必要로 할 것이다. 그들은 또한 그들이 資金支援한 事業의 進척사항을 알아볼수 있는 統計를 要求하게 될 것이다.

以外에도 統計利用者가 많이 있는데 이들은 經濟企劃으로 부터 歷史研究分野에 이르기 까지 目的에 必要한 資料를 要求하게 되며 政治家, 言論人, 歷史家, 經濟學者, 教師, 市場研究會社, 保險會社, 醫師, 圖書館員等이 包含된다.

#### 1.2.11 統計作成者의 需要에 對한 評價

위의 說明에서 우리는 利用者의 觀點에서 統計의 需要를 살펴왔다. 卽 特別한 決定을 내릴수 있도록 하기 위하여 어떠한 統計資料가 要求되는 가를 檢討하였다. 이제 우리는 統計作成者의 立場에서 같은 問題를 檢討해 볼 수 있다. 換言하면 우리(統計人)는 어떠한 統計가 우리 業務를 遂行하기 위하여 要求된다고 생각할 것인가 라고 質問할 수 있다. 그後 우리는 作成되는(作成할 수 있는) 統計의 範圍를 생각하게 되고 누가 이 統計를 利用하며 이 統計는 어떻게 利用될 것인가를 決定하게 될 것이다. 이것이 바로 그들이 利用可能한 資金範圍內에서 政府와 其他利用者를 위하여 가장 값진 統計를 提供하는 경우 統計人과 그의 高位官吏가 해야만할 일인 것이다. 이것은 그들이 먼저 우선순위를 定하고 이에 따라 作成된 統計의 利用範圍

를 定해야만 한다는 것을 뜻하고 있다. 分明 統計人은 利用者 와 자주 接觸해야 하고 그들이 統計需要가 어떠한 것인지 알고 자 하는바를 完全히 把握하여야 한다. 統計人은 서비스를 提供 하고 利用者가 要求하는 바에 대해 對答해 주고 있다. 統計資料의 實際需要는 이러한 需要에 對處할 公式統計 서비스의 限界를 훨씬 超過하고 있는것은 各國의 共通된 事實이다. 그러므로 統計當局의 高位官吏들이 各種 統計需要에 對한 注意깊은 評價를 하고 最高의 우선 順位를 가질 것으로 判斷되는 統計를 提供하는데에 그들의 努力을 기울여야 함은 매우 重要한 일이다.

本章에서 일찌기 우리가 統計에 關해 質問받게 될 것으로 불 質問目錄을 作成했을 때 우리는 一部統計가 要求됐기 때문이 아니라 過去부터 作成되어왔고 그들을 蒐集하는데 익숙해졌기 때문에 作成되고 있지않나 하고 質問했었다. 이에 對한 答辯으로서 統計資料가 아직도 必要한가의 與否에 對한 생각을 많이 하지않 고도 活動體系가 잘 樹立될 수 있기만 하다면 資料蒐集을 繼續 해야할 일이 있다는 것은 分명한 事實이라고 우리는 말할 수 있다. 어떠한 效率의이고도 잘 組織化된 서비스體系가 繼續的인 檢 討下에서 그 組織의 活動을 維持할 수 있도록 하는 것은 追加 資料의 提供을 위한 統計利用者들의 切실한 要求인 것이다. 사 실상 蒐集된 어떠한 統計도 널리 利用되고 있지않다고는 불 수 없다. 現在 作成되고 있는 資料中에서 一部를 拔萃하여 利用한 다거나 이러한 資料를 必要로 하는 政府機關, 企業 및 個人의 수 효가 얼마인가를 把握하는 것이 可能해야 할 것이다.

事實上 統計의 尙大한 需要에 比하여 統計資料가 不足한 것을 감안할 때 現在 作成되고 있는 어떠한 統計도 그나름대로 유용하게 利用되고 있는 것으로 볼 수 있다.

#### 1.2.12 特殊統計의 諸利用

우리는 이미 統計利用者가 特殊目的을 위하여 必要한 여러 統計를 보았다. 이제 統計의 一面을 살펴보고 이러한 統計의 利用은 무엇인가 하고 質問해 보자. 이에 對해 出生統計를 가지고 說明하고자 한다.

첫째 出生統計(또는 死亡統計)는 總人口(普通은 한나라 또는 한地方)의 變化狀態와 人口成長率을 把握하기 위하여 利用된다. 이러한 資料는 여러사람 特히 特殊專門家에게 아주 必要하다. 即 住宅當局은 사람들을 入住시킬 住宅建設計劃을 樹立하기 위한 資料를 要求하게 되고 한나라의 食糧供給計劃에 責任을 맡고 있는 公務員은 食糧生産이 적당한지, 이에따라 農業生産增大를 장려할 計劃을 樹立해야 하는지 또는 보다 많은 食糧輸入이 必要한지의 與否를 評價할 수 있도록 하기위해 現在와 未來의 人口를 알고자 할 것이다. 出生統計는 다른 統計와 關聯하여 널리 利用되고 있는데 一例로 保健 및 福祉機關은 1,000 名당 어린이의 死亡數인 유아사망율의 趨勢를 把握하기를 願하고 있다. 일찌기 指摘했지만 教育當局에서는 5~6歲의 豫想되는 新學校 登錄者에 對한 計劃을 樹立하는데 必要한 統計를 알고자 할 것이다. 以外の 統計利用者로서 兒童 保護 및 病院施設을 計劃함에는 保健當局이, 家族計劃運動의 實施結果를 評價함에

있어서는 家族計劃 機關이, 그리고 製品을 販賣할 市場을 評價하는데는 兒童服 및 其他品目を 生産하는 製造会社 등이 包含되고 있다.

米穀生産, 家畜數字 및 小売商店에서의 商品價格과 같이 現在 作成되는 統計資料의 다른 側面을 採択하고, 그러한 資料의 利用者 名簿作成을 試圖해 보는것은 하나의 흥미로운 일이다. 統計는 그들 自身에 바로 有用한 것은 아니나 重要한 措置를 취하기 위해서는 다른 統計와 關聯하여 흔히 利用되고 있음을 상기해 보라. 예를들어 여러 機關들이 自身の 權利로서 유용한 統計인 米穀生産 統計를 알고자 願할지 모르나 尙 平均生産量을 把握하기 위해서는 地域資料와 關聯하여 그것을 알고자 할 것이다.

即 그들은 總飲食物 生産을 把握하기 위해서는 다른 農作物의 生産資料와 關聯해서 또는 必要한 輸入量과 輸出을 위한 加工剩餘農産物을 評價하기 위해서는 人口 및 消費率과 關聯해서 이를 利用하고자 할 것이다.

#### 1.2.13 統計作成을 위한 相互協力

우리는 統計가 國家의 經濟社會開發을 성공적으로 管理하는데 있어서, 또한 私企業이나 研究의 여러 目的을 위해서 그리고 一般적인 情報를 얻기 위해서 매우 重要한 것임을 보아왔다. 統計를 利用하여 취해진 決定의 重要性은 統計가 時期適切하고 信憑性이 있어야 함을 強調하고 있다. 信憑性이 없는 統計는 오류를 야기시킬 수 있는데 한例로 米穀生産의 不正確한 測

定은 正確한 豫想 收穫量을 把握할 수 없게 한다. 農作物이 國民需要에 充足함을 나타내주는 統計가 있다면 (以後 事實이 아닌 것으로 判明됨) 莫大한 費用을 들여 米穀을 導入함으로써 部門的으로는 解決될지 모르나 심각한 欠收현상을 招來하게 될 것이다. 이와같은 統計의 커다란 오류는 모든 사람의 生計水準에 影響을 주게 된다.

분명히 統計人들은 그들의 業務를 注意깊게 誠實히 遂行할 커다란 責任을 지고 있다. 適切한 教育 및 訓練을 받는 것은 차치하고라도 모든 統計從事員은 그들이 할 수 있는 限 誠實하게 統計作成組織에 있어 그들의 任務를 遂行해야만 할 것이다. 이에는 2 가지의 協力해야할 事項이 있는데

첫째 全体の 統計作成機關은 統計利用者, 統計人 및 調査票를 作成하거나 調査에 參與함으로써 基礎資料를 提供하는 公務員들간의 相互協力を 要求하고 있다. 統計利用者들은 作成된 統計가 그들의 需要와 聯關이 있도록 하기 위하여 그들이 무엇을 必要로 하는지를 알아야만 하고 統計人들과의 緊密한 接觸을 해야 한다. 統計人들은 統計적으로 무엇이 可能한가 (또는 不可能한가) 를 알아보기 위하여 統計利用者를 도와야 하며 適切한 時間內에 合理的인 費用으로 良質의 統計를 얻기위해 그들의 모든 技術을 利用해야만 하고 國民은 調査上の 質問에 對한 応答이 國家的인 重要한 事業過程에 깊이 聯關되고 있음을 알아야만 한다.

둘째, 相互 緊密한 關係를 가지고 일해야 하는 各界各層의 統計從事者間에는 強力한 協同精神이 있어야 하며 그들 各自는 그



들의 任務를 誠實히 遂行하여야 한다. 어떤 主要한 統計事業을 推進하기 위해서는 調查設計者, 行政家, 監督者, 教官, 코오더 (Corders), 에디터 (Editors), 테블레터 (Tabulators), 調查員 및 其他 많은 要員들을 必要로 하고 있다. 組織의 各部門은 全体努力活動의 一部門으로서의 各其 自己分野의 役割을 알지 못하는 限 많은 組織構成員들의 誠實하고 주의깊은 모든 作業活動도 不拘하고 다른 階層의 不注意와 無誠意한 作業으로 因하여 파멸을 招來하게 될 지 모른다. 그리고 組織의 어느 部分도 스스로는 성공할 수 없게 된다. 即 統計組織의 各 部門은 其他의 다른 部門에 全的으로 依存해야만 한다. 모든 部門이 共同으로 作業하여 성공을 거두는 경우에만 最終 統計結果가 主要目的을 위한 價值있는 것이 되며 그 目的을 위하여 統計는 利用될 것이다.

### 1.3 用語의 基本概念 및 定義

우리가 統計科目을 研究하려면 特殊한 用語나 概念의 定義를 選定할 必要가 있는 것이다. 이러한 用語들은 每日 使用함에 있어 상당히 広範圍한 意味를 지니게 되고 여러 여건하에서 널리 利用되고 있을 것이다. 統計에 있어서 우리는 우리가 이야기 하고 있는 바에 대해 알고 있는 用語를 항상 利用할 수 있도록 그 意味를 아주 명백하게 定義해 두어야만 한다. 우리는 本 教材에서 이에 對해 좀더 알아야 함으로 우리가 直面하게 되는 새로운 定義를 해두어야 할 것이다. 그러나 새로운 主題를 討議하기 前에 一部用語의 基本概念 및 定義를 設定해야 할

必要가 있다. 統計資料는 量的인 것일수도 있고 質的인 것일수도 있는데 卽 다시 말하면 數字로 나타낼 수도 있고 特殊한 個體 또는 單位의 特質 또는 屬性으로 나타낼 수 있다. 예를들면 年齡 및 性에 關한 資料를 蒐集하는 境遇 年齡은 量的 資料가 될 것이다. 卽 그것은 年, 月에 따라 달라지는 個人의 年齡을 나타내주는 一連의 數字인 것이다. 두번째의 것 卽 性은 質的인 資料가 되는데 우리는 各 사람에 對해 男性 또는 女性이라고 기록하게 된다. 統計資料는 統計單位의 特性을 觀察한 結果이다. 3 가지의 用語 卽 觀察(Observation), 特性(Characteristic), 및 單位(Unit), 正確하게 말하면 統計單位(Statistical unit)는 各 各 統計에 있어 特別한 意味를 가지고 있는데 이들을 차례로 하나씩 定義하고 例를들어 살펴 보고자 한다.

觀察(Observation)이란 特別한 源泉으로 부터 資料를 얻고 그것을 記錄하는 過程이다. 普通 觀察은 器具(測定手段)를 利用하거나 또는 利用하지 않고서 或者에 依해 遂行된다. 가장 보편적인 觀察方法의 몇가지는 다음과 같은데 卽 測定, 計算, 個人的 判斷, 인터뷰에 基礎한 応答, 既存資料의 複写, 自計記錄器具의 利用등을 들 수 있다. 觀察은 統計單位의 特性으로 이루어진다.

特性(Characteristic)이란 말은 統計單位의 特色과 性質을 뜻하며 勿論 單位는 우리가 觀察하게 되는 全部 또는 一部에 對한 여러가지의 特性을 가지고 있다. 한例로 警察當局이 警察을 蒐集하는 境遇 適格條件을 가지는 사람을 選定하기 위하여 지원

자의 여러 特性을 觀察해야만 할 것이다. 卽 다음과 같다.

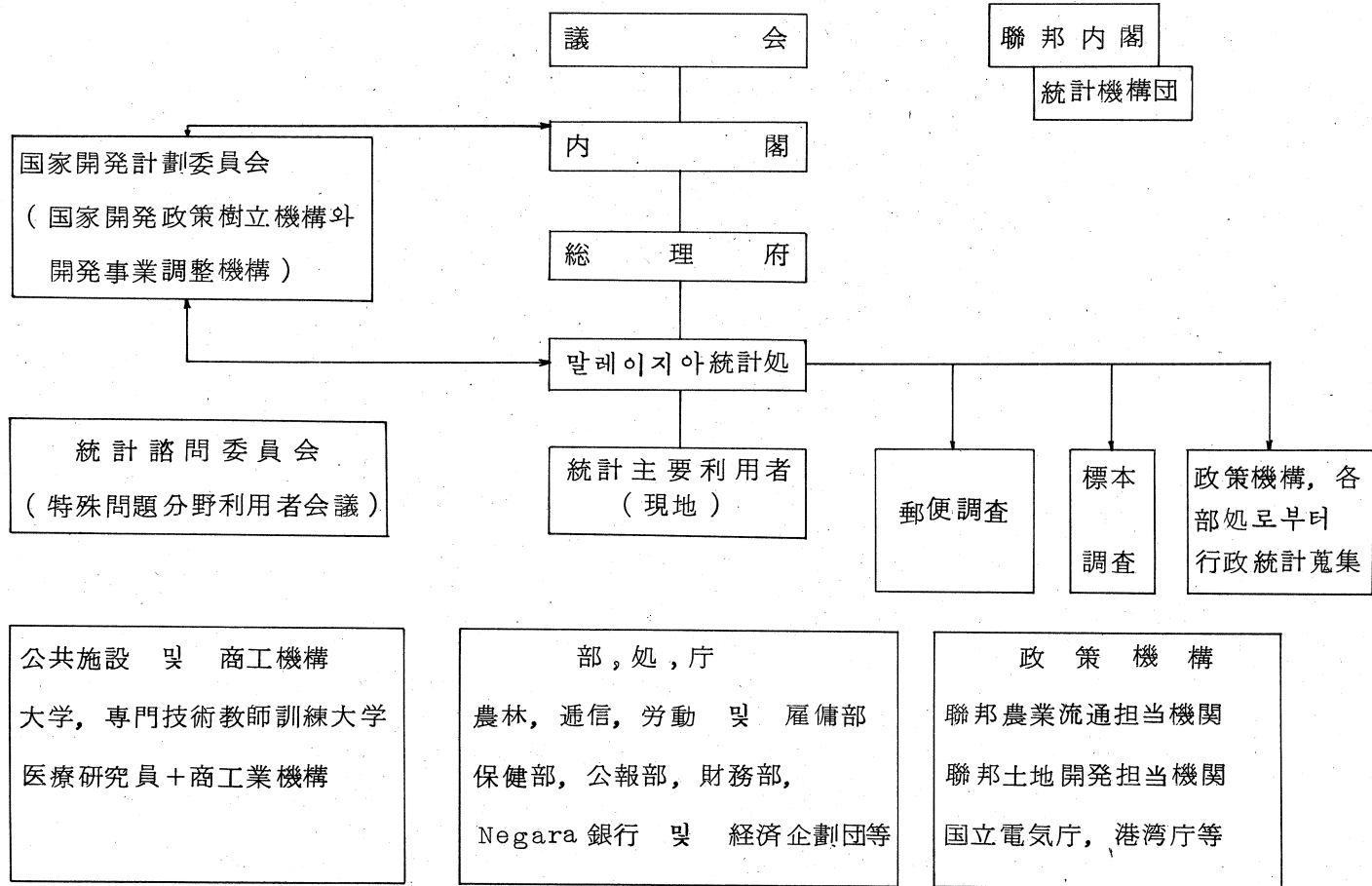
| <u>目 的</u> | <u>單位 ( 또는 個人 )</u> | <u>特 性</u>   |
|------------|---------------------|--|
| 警察募集       | 男子, 18歲以上~<br>30歲未滿 | 性, 年齡, 身長, 體重,<br>視力, 無能力事項, 結婚<br>狀態, 扶養家族數, 學歷<br>刑罰事項 |

特性에는 2 가지 種類가 있는데 첫번째는 變數(Variable)라 불리는 數的 價值인데 例를 들면 所得, 身長, 扶養者數, 耕地面積, 收入 價值等이다. 두번째 境遇의 特性은 質的 形態를 취할 수 있는데 이 경우 그것은 屬性(Attribute)이라 불리운다. 屬性의 例로서 經濟狀態, 性, 職業, 눈의 색깔等이다.

統計單位(Statistical unit)는 觀察되어지는 어떠한 要素 또는 要素의 集團을 말하는데 이는 (이러한 狀態에서) 더욱 조그만 單位로 細分되지 않는다. 統計單位의 例로서 사람, 家口, 家族, 事物, 事物의 集團, 農地, 마을, 나라, 期間等이다. 過去에는 個體(Individual)라는 말이 統計單位에 우선하여 많이 利用되었으나 때로는 滿足할만한 用語가 아닐 때가 있다. 이는 다른 狀況에서 우리가 家口의 各 構成員에 對하여 말하는 경우 期間을 個體로 보기 困難하며 家口(또는 家族)를 個體로서 說明하는 것이 妥當치 않기 때문이다. 統計單位는 보다 좋은 말로서 廣範圍하게 理解되고 있으며 本 教材에서 우리가 利用할 用語인 것이다.

우리는 同一한 單位에서 하나以上的 特性을 觀察할 수 있고 이러한 特性中 一部는 變數(Variable)이고 다른것은 屬性(attribute)일 것이다. 特殊産業의 被雇用者 調査에서 被雇用者年齡, 性, 職業, 教育水準, 賃金 및 俸給 그리고 勤務期間等を 調査할수 있다. 이러한 6個의 特性中에서 첫번째, 5번째, 6번째의 것은 變數인 反面 2번째, 3번째, 4번째의 것은 屬性이다. 同種의 많은 統計에 對한 하나 以上の 特性을 觀察하는것은 一連의 統計 資料를 作成하는 것이다. 어떤 統計調査에서 어떠한 資料가 蒐集되고 어떠한 特性이 觀察되며 어떠한 觀察方法이 利用될 것인가에 對하여 統計單位를 確認해 보는 것은 매우 重要한 일이다.

말레이시아의 統計서비스 組織



## 第2章 統計調査의 設計 및 企劃

### 2.1 概要

센서스나 서베이 같은 統計調査가 本部事務所나 規模가 큰 地方事務所의 高位職級者에 依해 普通 設計되고 있다 하더라도 實際的인 履行은 統計組織의 各階層과 여러 地域에 있는, 또한 여러 가지 방대한 技術을 가지고 있는 一團의 統計人들의 協同을 要求하고 있다. 그런데 이러한 龐大한 技術의 範圍는 統計調査 및 方法에 있어서의 高級訓練과 經驗, 廣範圍한 主題에 대한 知識, 調査票設計, 組織設計, 訓練, 要函作成, 資料處理, 指導監督 및 企劃에 있어서의 技術, 그리고 實査에 까지 이르고 있다. 따라서 統計調査의 企劃과 設計는 모든 統計人들에게 關心있는 分野이다. 全體的인 活動을 評價함으로써 各階層의 從事者들은 그들 自身業務를 把握할 수 있고 그들 從事業務의 複雜性和 相互關聯性을 評價할 수 있으며 그들이 해야 할 重要한 役割에서의 功지를 발휘할 수 있다.

더우거나 地方事務所 職員들은 本部事務所의 高位職級者들이 알지 못할 것으로 생각되는 現地事情에 대해 잘 알고 있으며 특히 보다 멀리 떨어진 地域에 있는 応答者들과의 緊密한 接觸關係를 가지고 있다. 그러므로 그들은 当面하는 있을법한 問題들을 종종 把握할 수 있게 되고 調査의 最初過程에서 応答者에 依해 惹起되는 어려움을 깨닫게 될 것이다. 이와같이 이들은 組織의 設計가 修正되고 改善될 수 있도록 하기 위한 助言과 意見

을 提示해 주는 者로서 設計過程에 있어 重要な 役割을 할 수 있게 된다. 이것은 우리가 1章에서 言及한 統計制度에서의 協力關係를 나타내주는 하나의 例인 것이다.

## 2.2 새로운 統計調査遂行의 諸過程

새로운 統計調査는 다음과 같은 段階의 過程을 必要로 하고 있다.

- (a) 權限獲得 및 財源確保
- (b) 調査의 設計 및 企劃
- (c) 調査의 遂行

- 資料蒐集
- 資料의 質的管理
- 結果處理 및 製表
- 結果分析
- 結果公表

本章에서는 (a)와 (b)項을 다루게 되고 調査遂行의 諸狀況은 뒷章에서 詳細히 다루어 질 것이다. 우리는 本章을 다루어 나가면서 標本家口를 對象으로 遂行되는 社会經濟調査에 입각한 例를 들어 調査의 設計와 企劃의 諸般事項을 說明하고자 한다. 大規模 調査의 設計와 企劃에 있어서의 諸問題와 措置들은 小規模調査와 반드시 같은 것임은 주목되어야 할 것이다. 即 大部分의 問題들은 既存統計調査의 定期的인 檢討와 聯關되어 있는 것이다.

### 2.3 새로운 統計調査의 起原

公共統計制度下에서 새로운 統計調査(서베이, 센서스等)의 必要性은 여러가지 方法으로 提起되고 있다. 어떤 경우에는 統計人이 政府나 部處長官으로 부터 直接 要求받게 된다. 그 要求는 特殊한 統計나 統計表를 요구하는 特異한 것일수도 있으며, 오히려 소위 「貧困에 關한 資料」와 같이 모호하게 要求될 수도 있다. 다른 경우에 있어 새로운 統計의 必要性은 統計人 自身에 依해 認識되고 있는데 이미 存在하거나 앞으로 있을 것으로 豫想되는 特別한 需要가 얼마인가를 알고 既存統計制度의 심각한 결함을 把握함으로써 비롯된다.

強調되어야 할 事項은 발단과는 關係없이 新規統計가 主要目的을 위한 資料의 需要에 따라 作成되어야 하는 것이다. 政府와 各部處長官이 必要로 하는 統計需要를 把握하고 그들의 需要에 充足될 統計를 얻기 위하여 그들을 돕는것은 統計人의 主要 責任인 것이다.

일단 需要가 判斷되고 統計가 作成되어야 한다고 政府内에서 合意가 이루어지면 統計人은 遂行해야 할 매우 重要한 몇가지 機能을 갖게 된다. 統計人은 政府 또는 各료급 利用者들이 達成하고자 하는것이 무엇인지, 새로운 資料를 왜 必要로 하는지, 또한 이를 利用하기 위하여 어떻게 提議해야 하는지를 正確하게 判斷하기 위하여 그들과 具體的인 討議를 해야만 한다. 이러한 討議를 함으로써 統計人은 그 資料가 行政記錄 또는 公式統計(例



를 들어 센서스 結果의 보다 具體的인 分析을 通하여 )로서 이  
미 存在하는가의 여부를 알아볼 수가 있다. 萬一 討議도 하지 않  
은 가운데 統計調査가 必要한 경우에는 統計人은 調査項目이 많  
이 要求되는지 단지 몇 가지만 要求되는지의 与否, 센서스로 할지  
標本調査로 할지의 与否, 全国单位로 할지 制限된 地域单位로 할  
지의 与否, 郵便調査方法과 面接調査方法中 어느것이 適切한지 等  
의 与否를 判断할수 있어야 하며, 이러한 狀態에서 統計人은 利  
用者の 需要를 滿足시키기 위해 遂行될 必要性이 있는 統計調査  
에 對한 어느 程度의 "判断力"을 가지고 있어야 한다.

이 경우 統計人은 政府, 財務部 또는 承認當局을 위한 事例를  
準備해야 하고 새로운 統計調査의 目的과 그 性格(센서스 또는  
標本調査, 標本規模, 作成週期等)을 把握해야 하며 最初調査와 繼  
統調査 費用을 推計해야만 한다. 調査가 承認되면(적어도 原則  
의으로) 統計人은 새로운 統計調査의 細部的인 計劃과 設計를 推  
進할 수 있게 된다.

#### 2.4 새로운 統計調査의 企劃과 設計

새로운 統計調査의 企劃과 設計에 對한 여러가지 事項들이  
있는데 2章의 나머지 部門에서 詳細히 說明된다. 그러나 여기  
에서 繼續 討議되겠지만 實際에 있어서 그러한 事項間에는 상당  
한 重複과 相互關聯이 있으며 이러한 事項의 상당수는 다른 事  
項에 依存하고 있으나 現在로서 반드시 設計되고 企劃되어야 하  
는 것이다. 따라서 이러한 事項들은 여러 段階에서 자주 檢討

되고 調整될 것이다.

設計 ( Design )란 무엇이 行해지고 , 가장 經濟的인 方法으로 그 目的을 達成하기 위해 어떻게 行해질 것인가를 決定하는 것과 關係되고 있으며

企劃 ( Planning )이란 設計된것을 履行하기 위한 모든 準備를 갖추는 것과 關係되고 , 各分野의 作業을 누가 , 언제 , 어디에서 遂行하고 어떠한 資料와 準備가 要求되는가에 關聯되고 있다.

우리는 統計人이 調査를 遂行함에 있어 考慮해야 할 모든 要因을 여기에서 全部 다룰수는 없으나 要求되는 細部的인 企劃의 性格에 關한 아이디어를 提供하기 위하여 重要한 몇가지 事項에 對해서만 提供하고자 한다.

#### 2.4.1 目的의 樹立

調査가 最初 原則으로 承認될때 그 目標와 目的에 對한 몇가지의 아이디어를 알수 있다. 그러나 設計의 始作段階에서 目的을 分明히 해두는 것은 바람직한 일이며 이러한 目的은 潛在的인 統計利用者들에 依해서 確認된다. 이것은 그 目的이 企劃 및 履行段階를 통해 維持되는 가를 確認하기 위해 統計人들에게 必要한 것으로 간주될 수 있다.

必要한 것은 다음과 같은 것에 對한 明白한 說明이다.

- (a) 調査의 目的
- (b) 調査의 範圍
- (c) 目的에 合당한 統計表

이러한 段階에서 생긴 오류는 必要없는 統計를 作成케 하기 때문에 目的을 分明히 해두는것이 必要하다. 例를들어 住宅不足을 把握하기 위한 統計가 要求된다면 政府가 無住宅者에게 住宅을 供給하고, 不良住宅地域을 整備하며, 새로운 住宅需要에 對한 供給을 위해 住宅建設産業을 支援할 意圖가 있는지를 명백히 해 두어야만 한다. 이들 各其의 目的을 위해서 서로 다른 統計가 要求되고 있다.

調査의 範圍는 目的에 따라 主로 決定될 것이다. 우리의 關心은 全國的이나 또는 特別히 不利한 集團이나 地域에 局限할 것인가 이다. 要求되는 統計表는 目的과 範圍가 明白할 때만이 決定될 수 있다. 이러할때, 統計의 諸作業이 始作되며 統計人은 무엇을 할 것인가를 알게 되고 要求되는 統計表를 作成하게 될 統計調査의 企劃과 設計를 始作할 수 있다.

#### 2.4.2 行政的 準備

##### 2.4.2(a) 法律規定

調査를 위한 適切한 法律規定이 있는지를 알아봐야 하는데 萬一 없는 경우 法令이나 行政命令을 制定하고 統計調査를 할수 있도록 統計人에 權限을 부여하는 것과 같은 形式으로 法律規定이 마련 되어야만 한다. 그렇지 않으면 자발적으로 遂行할 수 밖에 없는 것이다. 法律規定이 마련되어 있다면 統計從事者는 이를 명백하게 해 두어야만 할 것이다.

#### 2.4.2(b) 組織, 管理 및 調整

統計調査가 成功的으로 遂行되려면 그 事業에 必要한 人員을 配置하고 그들의 各가지 責任을 明白하게 定해둘 必要가 있다. 事業의 設計와 最初 企劃을 遂行할 指導者와 몇몇 補助要員을 任命하는 것은 大規模 事業에 있어 종종 必要한 것이다. 勿論 그들은 調査遂行에 必要한 職員構成 및 責任의 形態(即 組織)를 決定할 수 있게 될 것이다.

管理(Control)란 事業指導者가 計劃된 活動이 計劃에 따라 滿足스럽게 履行되는지의 與否를 항시 알수 있게 하는 記錄體系를 發展시키는 것을 意味한다. 類似한 管理體系를 準備해 두면 모든 必要한 活動이 計劃된대로(例로서 適正의 調査員이 募集되고, 訓練받고, 資料를 調達하고, 業務를 報告하고, 日日業務를 遂行하고, 要約報告를 하고, 完成調査票를 回送했는지 等) 遂行되고 있는가를 把握할 수 있게되며 以後의 履行段階에서 必要로 하게 된다. 大規模 統計調査를 위해서 이러한 管理體系는 잘 整備될 것이나 調査의 規模와 關係 없이 이러한 措置는 事業의 成功的 遂行을 위해 必要한 것이다. 이러한 管理體系가 亦是 設計되어야 하는것은 統計調査의 設計에 있어 必須的이다.

設計, 企劃 및 履行의 調整(Coordination)은 統計調査의 各分野가 各기 다르기 때문에 매우 重要하다. 事業規模가 클 境遇 調整을 위하여 한명의 高位職 官吏를 任命하는 것이 必要할 것이다. 統計調査에 參與하고 있는 大部分의 專門家들은 調査設計, 分類, 訓練, 要圖作成, 電算處理等과 같이 좁은 分野에만 責任을 지고

있다. 그러나 아마도 그들중 어느 한 사람의 作業은 다른 사람의 作業과 相互緊密하게 聯關되어 있을 것이다. 이러한 理由때문에 調整은 매우 重要的 것이다. 調整을 担当하는 者는 全體의 計劃, 時間計劃의 諸狀況, 다른 集團의 責任等を 完全히 알게 될 것이다. 모든 集團이 그들 作業과 關係되는 即 作業進度에 뒤지지 않게 하고 障礙要因을 確認, 除去하며 일어나는 심각한 事態를 事業指導者에게 알려주기 위해서는 다른 集團業務의 諸般狀況을 把握하고 있는가를 確認해야 하는데 이는 調整하는 者의 責任인 것이다.

#### 2.4.2(c) 豫算

豫算目的上 統計調査의 諸經費는 政府計定을 위해 利用되는 會計年度에 따라 쓰여지고 配分되어야 한다. 總經費에는 賃金 및 俸給, 旅費, 裝備 및 資料, 賃貸料等の 經費가 包含되어야 하는데 一般的으로 經費는 豫想보다 큰 것이다. 그러므로 豫算編成은 責任있는 그러나 融通性 있는 現實的인 方法으로 遂行되어야 한다.

資金이 일단 그 事業에 쓰여지면 豫算項目別 經費의 使用記錄이 維持되어야 하며 承認된 經費를 超過하지 않아야 한다. 必要的인 경우에는 새로운 資金割當이 公式적으로 모색되어야 하는데 豫算項目内에서 割當된 資金을 다른 目的으로 使用하는 것은 一般的으로 許容되지 않음을 알아야 한다. 따라서 豫算項目이 넓을수록 資金使用의 融通性도 커지며 豫算規定은 반드시 制限手段이 아니라 權能을 賦与하는 手段이 되어야 한다. 이렇게 豫算해 보면 틀림

이 없을 것이다.

社会經濟調査 豫算計劃 (× 地域)

A. 家 口 数 : 5,572 家口

B. 実査要員数 : 575 名

a. 面接調査員 : 428

b. 指 導 員 : 147

C. 豫 算

1. 準備事項

1.1 調査票, 実査要員構成 및 調査準備 ( 타이핑 등 包含 )

50 mm × \$ 25 ..... \$ 1,125

1.2 標本抽出準備

10 mm × \$ 25 ..... \$ 250

1.3 事前調査 ( 調査票, 実査要員構成, 調査方法 및 評価를 위한 檢討 ) ..... \$ 1,000

1.4 訓練과 調査票類 ( 印刷等 ) 準備

1,000m·d × \$ 1 ..... \$ 1,000

1.5 事務用品 ..... \$ 1,000

準備를 위한 豫算小計 \$ 4,375

2. 訓 練

2.1 教官訓練

30 名 × 5 日 × \$ 12.50 ..... \$ 1,875

2.2 実査要員訓練

575 名 × 4 日 × \$ 8 ..... \$ 18,400

訓練을 위한 豫算小計 \$ 20,275

이러한 豫算計劃은 事業이 推進될 수 있도록 하기위해 50,000弗의 資金을 割當 받을 수 있게 하고 있다. 이때 調査의 各分野에 對한 事實上의 經費는 매우 注意깊게 管理되어야 한다. 調査遂行期間동안 어떠한 編成豫算이 豫想을 超過할 것이 分명한 경우 (例로서 準備를 위한 豫算 4,375弗, 資料處理를 위한 豫算 5,337弗이 너무 낮다고 判明될 경우) 事業을 完成하기 위해서는 修正豫算이 마련되어야 하며 追加資金이 調達되어야 한다.

### 3. 実 査

3.1 調査票類 発送 ..... \$ 500

### 3.2 実 査

5,572 家口 × \$ 1.50 ..... \$ 8,358

### 3.3 実査指導

5,572 家口 × \$ 0.50 ..... \$ 2,786

### 3.4 内容検討

8 × \$ 190 ..... \$ 1,520

実査豫算 小計 ..... \$ 14,164

### 4. 資料處理

### 4.1 審 査

5,572 枚 × \$ 0.50 ..... \$ 2,786

4.2 電算處理 ..... \$ 5,337

小 計 ..... \$ 8,123

5. 雜 費 ..... \$ 3,063

總 計 \$ 50,000

資料處理 豫算

1. 마그네틱 테이프

1.1 1,600BPI - 600feet : 4 個 × \$ 16.50 = \$ 66

1.2 1,600BPI - 2,400feet : 10 個 × \$ 25 = \$ 255

\$ 321

2. 製表紙

2.1  $14\frac{7}{8} \times 11'$  : 1 枚 1 部 × 2,000 ( 箱子当 )

8 상자 × \$ 22 ..... \$ 176

2.2  $17\frac{25}{32} \times 11'$  : 1 枚 1 部 × 2,000 ( 箱子当 )

2 상자 × \$ 24 ..... \$ 48

2.3  $17\frac{25}{32} \times 11'$  : 1 枚 3 部 × 1,000 ( 箱子当 )

12 상자 × \$ 66 ..... \$ 792

\$ 1,016

3. 穿孔 2,500md × \$ 1 ..... \$ 2,500

4. 分析 및 프로그래밍 ..... \$ 1,500

總豫算 \$ 5,337

2.4.2(d) 資 源

새로운 統計調査에 必要한 資源은 調査規模와 統計局에서의 資料利用度에 따라 매우 多樣하다. 언제나 그러하듯이 必要한 資源은 여러 階層의 人力, 施設, 設備 및 資料(供給物)가 될 것이다. 여러 경우에 있어 高位의 企劃, 設計 및 監督要員과 같



은 약간의 人力이 利用될 것이나 다른 部門의 要員도 確保되어야 한다. 마찬가지로 施設도 利用될 것이며 大規模 事業의 경우 새로운 施設, 備品, 裝備, 訓練場 등의 賃貸를 必要로 할 것이다. 또한 裝備와 資料는 量的으로 매우 多様하며 企劃段階에서 이에 대한 注意깊은 考慮를 해야만 한다. 예를들어 30,000 枚의 調査票가 必要한 人口 센서스에서는 調査票를 印刷하기 前에 必要한 量의 종이를 確保하는데 상당한 조회시간이 所要될 것이다. 이에는 印刷費, 包裝費, 發送費 등이 關係되고 있으며 아마도 包裝記錄 및 管理要員 등이 所要될 것이다. 이와 類似한 事例는 必要한 準備를 하는데 適用된다. 海外注文이 必要한 경우 조회시간은 더욱 길어진다.

다음 統計表는 한 調査의 所要資源이 어떠한가를 말해주고 있다. 이表는 調査에 있어 人的資源을 어떻게 活用할 것인가 하는 細部事項을 보여주고 있는 것으로 생각된다. 同統計表와 위의 社会經濟調査豫算을 比較한다면, 標本으로 抽出된 5,572 家口와 雇傭된 575 名의 実査要員과 같은 統計表의 여러가지 集計數字를 豫算準備에 있어 어떻게 活用될 것인가를 알아보는 것은 흥미로운 일이다.

標本家口・調査区 및 実査要員数

( X地域의 社会經濟調査 )

| 地 域 | 標 本 家 口 数 |       |       | 調 査 区 数 |     |     | 実 査 要 員 数 |       |     | 監 督 者 |
|-----|-----------|-------|-------|---------|-----|-----|-----------|-------|-----|-------|
|     | 都 市       | 郡 部   | 計     | 都 市     | 郡 部 | 計   | 調 査 員     | 指 導 員 | 計   |       |
| X1  | 169       | 194   | 363   | 17      | 39  | 56  | 28        | 10    | 38  | 1     |
| X2  | 74        | 77    | 151   | 8       | 16  | 24  | 12        | 4     | 16  | 1     |
| X3  | 70        | 56    | 126   | 7       | 12  | 19  | 10        | 4     | 14  | 1     |
| X4  | 173       | 145   | 318   | 18      | 29  | 47  | 24        | 8     | 32  | 1     |
| X5  | 581       | -     | 581   | 58      | -   | 58  | 29        | 10    | 39  | 1     |
| X6  | 409       | 577   | 986   | 41      | 115 | 156 | 79        | 26    | 105 | 4     |
| X7  | 326       | 542   | 868   | 33      | 108 | 141 | 71        | 24    | 95  | 3     |
| X8  | 54        | 55    | 109   | 6       | 11  | 17  | 9         | 3     | 12  | 1     |
| X9  | 541       | 639   | 1,180 | 54      | 128 | 182 | 91        | 31    | 122 | 4     |
| X10 | 31        | 59    | 90    | 3       | 12  | 15  | 8         | 3     | 11  | 1     |
| X11 | 32        | 73    | 105   | 3       | 15  | 18  | 10        | 4     | 14  | 1     |
| X12 | 19        | 63    | 82    | 2       | 13  | 15  | 8         | 3     | 11  | 1     |
| X13 | 41        | 62    | 103   | 4       | 13  | 17  | 9         | 3     | 12  | 1     |
| X14 | 109       | 50    | 159   | 11      | 10  | 21  | 11        | 4     | 15  | 1     |
| X15 | 48        | 57    | 105   | 5       | 12  | 17  | 9         | 3     | 12  | 1     |
| X16 | 123       | 123   | 246   | 13      | 25  | 38  | 20        | 7     | 27  | 1     |
| 合 計 | 2,800     | 2,772 | 5,572 | 283     | 558 | 841 | 428       | 147   | 575 | 24    |

#### 2.4.2(e) 調査體系

決定해야 할 가장 重要な 問題의 하나는 全數調査 (Complete census) 로 할 것인가 標本調査 (Sample survey) 로 할 것인가의 與否이다. 이것은 要求되는 結果值의 信憑性 (標本調査 資料는 標本誤差가 發生하기 쉬운 反面, 全數調査資料는 그렇지 않기 때문임) 調査時期 (普通 調査結果值는 標本調査에서 빨리 얻을 수 있기 때문임), 費用 및 蒐集될 資料의 量에 對한 考慮를 한 것이다.

標本調査와 全數調査를 選定함에 있어 考慮해야 할 諸要素에 對해서는 第3章 4節에서 보다 具體的으로 討議할 것이다.

#### 2.4.2(f) 調査時期

調査가 언제 遂行되어야 하는가를 明確하게 決定함에 있어 考慮해야 할 많은 要素들이 있다. 例를들어 農業調査에 있어 道路事情이 좋지 않다면 調査가 지연되기 때문에 季節風이 부는 季節을 避하는 것이 바람직 할 것이다. 即 家口支出額 調査에 있어 라마단의 斷食時期나 크리스마스와 같은 特別한 時期를 避하는 것이 아마 必要할 것이다. 이때의 支出形態는 非正常的이다. 또한 學校教師와 같은 一時的으로 採用된 特殊그룹의 職員이 活用되는 경우 學校放學이라든가 週末等의 期間동안은 調査를 制限하는 것이 必要할 것이다. 몇명 안되는 面接調査員으로 長期間에 걸쳐 調査하거나 많은 面接調査員으로 短期間에 걸쳐 調査하는 경우의 相對的인 長點이 研究되어야 할 것이다. 다른 統計調査와 調査時期上의 對立을 避하는 것은 같은 調査要員이 活用되거나 같은 사람에 對

해 調査하는 경우에 特別히 바람직하다.

調査時期에 影響을 주는 또다른 要因들이 있으며 各調査는 各其의 特異한 問題를 가지고 있을 것이다. 時期는 경솔하지 않고 注意깊게 決定되어야 하는데 이러한 모든 要因들은 統計를 願하는 利用者の 需要에 對해 均衡을 이루어야 한다. 만일 資料가 急히 要求된다면 適切치 않다고 考慮되더라도 一時에 調査를 遂行하여야 할 것이다.

#### 2.4.2(g) 日程表

調査의 諸分野를 網羅하는 日程表는 統制를 維持하고 諸般事項이 時期에 맞게 達成되도록 하기위한 매우 有用한 方案이다. 이것은 初期에 作成되어야 하고 区分된 活動別로 始期와 終期를 나타내 주어야 한다.

公表日자가 記入되어야 하며 이것은 調査를 마치는 豫定日자가 된다. 이러한 日程表는 類似한 調査에 經驗을 갖고 있는 職員이 作成해야 한다. 이들은 活動別로 얼마의 期間이 걸릴지를 評價함에 있어 自己의 知識과 判斷을 利用할 수 있고 問題가 發生될 地域과 所要될 時間을 알수 있을 것이다.

特別히 言及해야 할것이 있는데 이는 計劃評價와 檢討技術 (PERT: Programme Evaluation and Review Technique) 과 같은 網形分析 (Network analysis) 과 크리티칼 패스 시스템 (Critical path system) 이다. 이 名稱이 意味하는 바와같이 이러한 技術은 計劃을 達成하기 위해 遂行되어야 할 모든 機能을 說明해주

는 網 ( Network ) 의 準備를 要求하고 있다. 이들은 各 單位作業이 다른 單位作業의 終結과 各 單位作業을 위해 割當된 時間에 依存하고 있는 그들 相互間의 關係를 說明하기 위해 相互連結되어 있는 것이다. 이것이 行해질때 "크리티칼 패스" ( Critical path ) 는 構成網을 따라서 추적될 수 있다. "크리티칼 패스" 에 對한 모든 單位作業들은 그들을 遂行함에 있어서의 어떤 지연이 全體計劃을 遲延시킨다는 意味에서 볼때 매우 重要하다. "크리티칼 패스" 以外의 單位作業은 그 時期에 있어 신속성의 程度가 크다. 이러한 制度는 신속성 即 各 單位作業의 空白期間 ( Floating time ) 을 測定하기 위한 여러 規定과 各 單位作業의 "始作되는 날짜와 끝나는 날짜" ( Earlist and latest dates ) 를 利用하고 있다.

重要한 狀況을 解決하기 위한 單位業務를 構成하는 網의 調整方法이 있는데 例를 들어 主要活動 ( 調査票 印刷等 ) 에 있어서의 遲延은 어떤것이 遂行되지 않는 限 全體의 調査를 遲延시킬 것이다. 이때 한 例로 調査員 訓練을 위해 職員을 더 割當하는 것을 決定해야 하는데 이는 印刷作業의 遲延을 상쇄하기 위하여 어떠한 單位作業이 보다 짧은 期間內에 끝날 수 있도록 하기 위해서 이다. 萬一 作業이 遂行된다면 單位作業의 構成網을 變更시킬 必要性이 있게 되며 이에따라 새로운 크리티칼 패스 ( Critical path ) 를 定할 必要性이 있게 된다.

大部分의 이러한 制度는 새로운 크리티칼 패스는 勿論이고 새롭고, 修正 調整된 資料로서 繼續 最近의 것으로 設計되고 있

다. 이러한 制度는 컴퓨터에 依해 運營될 수 있도록 設計된다. 그러나 이러한 程度의 細部的인 管理가 必要하지 않다 하더라도 最初 構成網을 갖추는데 있어 眞價가 있는 것이다. 이것은 設計者와 企劃樹立家에게 그들 計劃을 깊이 理解하도록 해주고 있다.

우리의 社会經濟調查를 위한 時間日程表는 다음과 같은 圖表로 說明될 수 있다.

이 圖表에서 우리는 調查管理에 對한 많은 값진 情報를 얻을 수 있으며 어떤 作業이 同時に 推進되고 있는가를 알 수 있다. 예를 들면 1.6, 1.7, 1.8의 活動들은 모두가 7月 初半에 進行되도록 計劃되어 있으며 일단의 高位企劃樹立家들은 調查期間동안 利用될 일관성 檢討를 위한 規定을 設計해 놓은 反面 일단의 프로그래머들은 컴퓨터 프로그램을 作成하는데에 參與하고 있으며 또다른 그룹은 必要한 訓練資料를 準備하고 있다. 이러한 作業들은 相互間에 独立的으로 進行될 수 있으며 이들중 어느 하나의 遲延은 다른 두 活動을 制止하지 못할 것이다.

그러나 其他 많은 活動은 繼統적으로 遂行되어야 하며 하나의 活動은 다음 活動이 始作되기 前에 끝나야 한다. 例로서 訓練資料의 準備事項인 1.8 活動은 實查教官 訓練事項인 2.1 活動이 開始되기 前에 끝나야 한다. 卽 이 活動은 그 調查의 事實上 實查인 3.2 活動以前에 끝나야만 한다. 그러므로 우리는 各各 이러한 活動은 이를 뒤따르는 活動에 對해 매우 重要함을 알 수 있다.

社会經濟調査 日程表 ( X 地域 )

| 活 動 別                              | 1980 |   |   |   |   |   |   |   |   |    |    |    | 1981 |
|------------------------------------|------|---|---|---|---|---|---|---|---|----|----|----|------|
|                                    | 1    | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 1    |
| 1. 準備                              |      |   |   |   |   |   |   |   |   |    |    |    |      |
| 1.1 調査目的 決定                        | —    |   |   |   |   |   |   |   |   |    |    |    |      |
| 1.2 調査票 및 指針書 作成                   |      | — |   |   |   |   |   |   |   |    |    |    |      |
| 1.3 標本設計 및 抽出                      |      | — |   |   |   |   |   |   |   |    |    |    |      |
| 1.4 事前檢討 : 訓練 · 実査 · 評価            |      |   |   | — |   |   |   |   |   |    |    |    |      |
| 1.5 調査票 및 指針書 確定<br>( 資料処理指針書 包含 ) |      |   |   | — |   |   |   |   |   |    |    |    |      |
| 1.6  일관성 檢討를 위한 規定                 |      |   |   |   | — |   |   |   |   |    |    |    |      |
| 1.7  컴퓨터 프로그램과 테스트方法<br>作成         |      |   |   |   |   | — |   |   |   |    |    |    |      |
| 1.8 訓練과 調査票類 ( 印刷等 ) 의 準備          |      |   |   |   |   | — |   |   |   |    |    |    |      |

| 活 動 別                          | 1980 |   |   |   |   |   |   |   |   |    |    |    | 1981 |  |
|--------------------------------|------|---|---|---|---|---|---|---|---|----|----|----|------|--|
|                                | 1    | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 1    |  |
| 2. 訓 練                         |      |   |   |   |   |   |   |   |   |    |    |    |      |  |
| 2.1 実査教官要員 訓練                  |      |   |   |   |   | — |   |   |   |    |    |    |      |  |
| 2.2 実査要員 訓練                    |      |   |   |   |   |   | — |   |   |    |    |    |      |  |
| 3. 実 査                         |      |   |   |   |   |   |   |   |   |    |    |    |      |  |
| 3.1 調査票類 発送                    |      |   |   |   |   | — |   |   |   |    |    |    |      |  |
| 3.2 実 査                        |      |   |   |   |   |   | — |   |   |    |    |    |      |  |
| 3.3 調査票 蒐集                     |      |   |   |   |   |   |   | — |   |    |    |    |      |  |
| 4. 資料処理                        |      |   |   |   |   |   |   |   |   |    |    |    |      |  |
| 4.1 에디팅 및 코딩                   |      |   |   |   |   |   |   |   | — |    |    |    |      |  |
| 4.2 資料 転換                      |      |   |   |   |   |   |   |   |   | —  |    |    |      |  |
| 4.3 컴퓨터 処理<br>( 일관성 檢討 · 製表等 ) |      |   |   |   |   |   |   |   |   |    | —  |    |      |  |
| 5. 公 表                         |      |   |   |   |   |   |   |   |   |    |    |    | —    |  |



한 예로 訓練資料를 準備하는데 遲延이 發生하여 7月 中旬까지 完了되지 않았다면 뒤따르는 各各의 活動, 나아가 全體의 活動은 한달 뒤로 미루어야 할 것이다. 이와같이 調査를 統制하는 指導員들은 訓練資料가 豫定날 짜까지 準備되어야 함을 이 圖表의 研究를 통해 알수 있다.

다른 한편 어떤 活動은 그렇게 重要하지 않음을 알수 있는데 예를 들어 컴퓨터 프로그램의 作成인 1.7 에만 依存하는 活動은 컴퓨터 處理活動인 4.3 뿐이다. 資料취합 (Data capture) 인 4.2 에 直接 關聯되는 다른 모든 活動들은 프로그래머들이 그들의 任務를 遂行하는 동안 繼續 進行될수 있을 것이다. 더우 기나 時間豫定表가 9月末 까지는 終了될 것으로 豫定된 計劃下 에서 두개의 單位作業間에는 한달간의 間隔이 있어야 할 것이나 事實上 11月以前 까지는 調査의 進行上 반드시 必要한 것은 아니다.

그러므로 프로그램을 作成하는데 있어 1個月의 遲延은 그 調査를 전혀 遲延시키지 않을 것이다. 이러한 性質을 把握하는 것은 매우 重要하며 잘 管理된 統計調査는 이와같이 잘 設計된 時間豫定表에 크게 依存하게 될 것이다.

#### 2.4.2(h) 公表 및 弘報

調査結果公表와 弘報活動은 統計人에게 아주 重要하다. 結局 大部分의 統計調査는 要求되는 資料를 提供해주는 사람에 依存하고 있다. 이러한 사람들의 積極的인 協力이나 도움없이 는 調査가 成

功될 수 없는 것이다. 資料가 法律上으로 要求되고 職員이나 機關이 資料를 提供하지 못하여 非難을 받을 경우라 할지라도 調査를 成功的으로 遂行하려면 좋은 意見을 反映시켜야 한다.

사람들은 어디에서나 問題를 要求하는 公務員에 對한 疑心을 가지고 있으며 그들은 本能的으로 資料를 提供하지 않는다. 그러므로 이러한 本能的인 反抗心을 克服하기 위해서는 公表와 弘報活動에 對해 많은 時間과 수고를 아끼지 말아야 할 必要性이 있는 것이다. 그래서 資料를 蒐集하기 前에 왜 調査가 遂行되며 그 資料가 무엇을 위해 利用되는지 等 그 調査에 關해 應答者에게 알려 주어야 한다. 設計段階에서 調査에 對한 反발이나 非協調關係가 커질 것으로 豫想되는 地域이나 集團을 確認하는데 모든 努力을 기울여야 하고 公表活動을 위한 努力이 主로 이러한 集團을 위해 기울여야 할 것이다.

이러한 모든 資料는 쉽게 理解할 수 있는말로 要求되어야만 한다. 그들은 統計의 潛在的 이로움은 무엇이며 어떻게 利用될 것인가를 말하게 될 것이다. 그들이 提供하는 資料는 秘密이 維持되고 그들에 對해 不利하게 利用되지 않음을 確認시켜 주어야 한다. 그러나 質問에 對答하고 資料를 提供해 주는 代價로 一定한 報酬나 惠沢에 對해 統計人은 간혹 장담하지만 約束하지 않는 것이 重要하다. 경우에 따라 要求되는 資料를 提供하는 것은 社会的 義務인 것이다.

適切한 公表方法에는 다음과 같은 것이 있다.

(i) 應答者에 對한 私送

(ii) 新聞과 雜誌記事

(iii) 關心있는 者의 그룹, 政治家, 마을指導者等과의 對話

(iv) 라디오나 텔레비전 放送

企業設備調査에 있어 公表와 弘報活動은 家口調査에서 要求되는 公表와는 아마도 다를 것이며 設備調査를 企劃할때 보다 큰 關心을 기울여야 한다. 첫째 企業體는 要求되는 資料를 提供하기 위해 分明 그들의 記錄事項에 依存해야 하기 때문에 調査員의 訪問을 미리 알려주는 것은 반드시 必要할 것이다. 또한 調査員이 訪問할때 責任있는 職員이 応答할 수 있도록 特別한 時間을 約束하는 것도 必要할 것이다. 이러한 事前準備가 없으면 貴重한 時間이 낭비될 것이며 이러한 것은 統計調査에 費用을 더 加算해주게 된다. 둘째, 企業體는 그들 資料의 秘密性에 대해 매우 敏感하며 그들의 答辯內容이 税金賦課나 다른 公務員 심지어는 그들의 競争者들에 까지 전해지지 않을까 念慮하게 될 것이다.

그러므로 提供된 모든 資料는 秘密이 分明 維持되고 統計目的에만 利用될 것임을 確信시켜 주는데 特別한 關心을 가져야 할 것이다. 大部分의 國家에서 統計法은 이러한 것을 열거하고 있으며, 個人의 資料를 다른 사람이나 企業 또는 政府当局에 알려주는 統計從事者에게는 重罰에 処하도록 하고 있다. 秘密保障手段인 이러한 法律規定에 대해 이야기 해주는 것은 事業體를 最初 訪問하는데 있어 이로울 것이다.

### 2.4.3 技術的 準備

#### 2.4.3(a) 概念 및 定義

새로운 統計調査를 設計하고 企劃함에 있어 概念을 명백히 하는것은 반드시 必要하다.

概念은 統計人과 利用者가 滿足하고 分명한 理解와 同意를 할 때 確定된다고 말할 수 있다. 어떤경우 이것은 統計人과 利用者が 國際機關이나 다른 나라와 같은 他機關에 依해 이미 定해진 概念을 高수하기로 合意본 것임을 意味할 것이다.

前述했던 한 例로서 住宅担当 部처가 特히 都市에서의 住宅不足 統計를 願하고 있음을 생각해 보자. "住宅不足"이라는 概念을 明確히 하기 위해서는 資料가 要求되는 理由를 分明明히 알아야 할 것이다. 前에 言及한 바와 같이 同 部처의 目的은 다음과 같을 것이다.

- 乞人이나 住宅以外의 居處에서 살고 있는 者에게 最小限의 住宅을 마련 해주고
- 모든 都市居住者에게 最小限의 標準住宅을 供給해 주며
- 住宅 確保用 資金을 가지고 있는 者가 住宅을 建築하거나 共同住宅 또는 아파트를 購入할 수 있도록 建設産業問題를 解決하는 것이다.

住宅不足에 對한 이러한 各各의 概念들은 分明 需要에 對處하기 위한 여러가지의 統計를 必要로 할 것이다.

첫째 : 乞人數와 其他 極貧者數를 把握하고 住宅以外의 居處에서

잡자는 理由를 把握하는 것이 必要하다. 이것은 매우 어려운 調査이다.

둘째 : 첫번째 事項이 包含되나 既存住宅에 처한 다른 事項을 追加 調査하고 不良住宅 代替計劃의 必要性을 評價한다.

셋째 : 아직 求하지는 않았으나 앞으로 購入할 사람에 關한 調査 即 家口調査, 建設当局이나 建築者가 保有하고 있는 待期名簿에 關한 資料蒐集 및 지난 2年間 建築承認을 받은 사람에 처한 調査等を 必要로 하게 된다.

統計作業이 意味있고 正確하게 遂行되기 위해서는 用語定義가 必要하다. 우리는 "住宅不足"과 같은 用語는 여러가지 意味를 가질 수 있고 定義없는 用語를 使用하면 混亂을 일으킨다는 것을 알았다. 예를 들면 "住宅不足"이란 말은 住宅以外的 居處에서 잠을 자거나 겨울에 最小 一週日中 4日間을 住宅以外的 建物에서 잠을 자는 사람의 數(家口 또는 家族等)로 定義할수 있다. 위의 例에서 본바와 같이 우리는 "最小限의 마음에 드는" 또는 "都市(境界)" 등과 같은 말을 定義해야 할 必要가 있다. 그러한 定義問題를 說明하고자 다음과 같은 몇가지 例를 考慮해 볼수 있다.

一 生産(例·自轉車)

이는 自轉車 組立, 生産되었으나 組立이 안된것, 顧客의 組立用으로 包裝된것, 장난감自轉車等を 包含하고 있는지

一 年度

이는 曆年, 會計年度 또는 企業會計年度(普通 設立日字와

關係됨)等 어느것을 意味하는지

一 都市

이 都市의 境界가 어디인지 特殊한 地方政府인지 卽 周圍의 地方政府 地域이 包含되는지 또는 이 都市는 特殊性을 지닌 地域만을 包含하는 것인지 等이다.

單純한 概念이라도 定義하기에는 꽤 어렵다는 것을 알수 있다. 보다 複雜하고 매우 重要的 概念들을 定義하는데는 極히 어려울 때가 있다. 貧困, 삶의 質, 福祉, 基本需要, 作業向上等과 같은 概念들은 좋은 例이다.

UN과 같은 國際機構는 統計調査에 使用할 基準概念 및 定義를 發展시키는데 많은 일을 해왔다. 이러한 機構들은 國家的인 次元의 調査目的에 対応하는데 있어 이러한 概念 및 定義를 利用하는 것이 바람직한 일인데 이는 調査結果들 다른 나라 結果와 比較할 수 있는 附隨的인 이로움이 있기 때문이다.

마찬가지로 서로 다른 調査를 위해 같은 定義가 適用된다면 이는 그 結果가 서로 比較될수 있음을 뜻한다. 우리는 12章에서 國際標準分類 및 定義의 一部를 考察하게 될 것이다.

한 國家의 여러地域에 關한 資料를 蒐集할때 行政地域이라 볼수 있는 標準地理明細를 利用하는 것이 가장 좋다. 區, 道, 選挙區等은 이러한 例이다. 이러한 一貫性있는 方法으로 地域을 定義 해둠으로써 同一地域의 여러 統計를 比較할 수 있고 事實上 各地域에 對한 有用한 데이터 베이스를 만들수 있을 것이다.

### 2.4.3(b) 調査形態, 統計活動 및 資料蒐集 方法

本章에서 우리는 새로운 統計調査의 設計와 企劃에 關心을 두어왔다. 그러나 設計段階에서 接近方法에 關聯되는 여러가지 代案을 밝혀두는 것은 有用할 것이다. 이들은 다음과 같은 統計活動에서의 考慮해야 할 事項을 包含하고 있다.

- 既存記錄으로 부터 拔萃
- 家口, 農耕地, 事業體等에 對한 標本調査
- 家口, 農耕地, 事業體等에 對한 센서스, 資料蒐集方法 또는 利用할 統計目的
- 家口, 事業體等에 對한 現地面接
- 郵便에 依한 自計式調査(또는 現地에서의 調査票配付 및 蒐集에 依한)
- 現地에서의 測定(특히 農業의 경우)
- 其他(電話調査)

우리의 目的上 이는 家口를 對象으로 한 標本調査로서 現地面接方法으로 해야 할 것이다.

여러 調査方法은 各各의 長短点(3章에서 討議되는 바와 같이)을 가지고 있으며 統計人은 特別調査를 위해 어느 方法이 가장 좋은가를 決定해야 한다. 여러가지 決定方法을 가지고 어느것이 事實上 가장 適切한가를 試驗해 보는 것이 간혹 必要할 경우가 있다. 여러 調査方法의 結合 例로서 現地面接과 郵便에 依한 資料蒐集方法이 結合되어 利用되는 경우가 있는 것이다.

#### 2.4.3(c) 応答者の 範囲・対象 및 選定

企劃過程에서 調査範圍가 어떤가, 範圍内에서의 單位의 總和는 어떻게 얻어질 수 있는가를 決定하는 것이 必要하다. 範圍로 보아 우리는 調査에 包含되어야 할 母集團에서의 個體 또는 單位를 構成하는 対象母集團에 關係 說明하고자 한다. 이것은 單純하거나 簡單한 問題는 아니며 決定하기가 종종 어려울 때가 있다. 예를 들어 家畜統計를 얻기위한 調査를 企劃한다면 적어도 어떠한 家畜을 所有하고 있는 家口와 農場은 모두 調査의 範圍内에 包含되어야 하는것은 一見 必然的인 것이다. 이것은 예를 들어 家屋내에 몇마리의 家禽과 한마리의 염소를 가지고 있는 家口도 모두 包含해야 함을 意味하는데 이렇게 行하는 努力은 禁止되어야 할 것이다. 그러므로 이러한 意味없는 家口를 除外하도록 範圍를 設定할 必要가 있게된다. 事實上 이는 調査에 採用된 調査員이 어떤 調査單位가 範圍内에 있고, 範圍밖에 있는가를 알 수 있도록 限界線을 그리거나 이것을 分明하게 限定하는 方法이 모색되어야 함을 意味한다. 大部分의 統計調査에서 이와 類似한 問題가 提起될 것이다. 예를 들어 農家住宅産業이나 小規模 修繕事業이 工場센서스 範圍内에 包含되어야 할 것인가, 貧困에 關係한 調査에 모든 家口가 包含되어야 할 것인가, 또는 高所得 家口를 除外시키는 것이 좋은가, 等이다. 事實上の 実査活動에서 어떻게 調査하는가를 決定하기 보다는 範圍를 設定해 두는것이 보다 쉬운 경우가 종종 있다.



範圍가 일단 決定되면 対象을 設定하는것 다시 말하면 調査의 範圍內에 있는 모든 調査單位가 事實上 나타내 지도록 確實하게 方法을 찾아 내는것이 必要하다. 包含되어야 할 어떠한 調査單位도 漏落되지 않는것이 重要한데 漏落하는 경우에는 그 結果는 分明 낮게 나타날 것이다. 例로서 小売業 센서스와 같은 어떤 主要調査에서 그 나라의 單一 商店과 市場들을 모두 찾아 낸다는 것은 매우 어려운 것으로 생각될 것이다.

#### 2.4.3(d) 標本의 抽出

여러 統計調査에서 調査單位(即 面接할 家口)를 抽出하고, 確認할 수 있도록 하기 위해서는 完全하고도 最新화된 單位들이 必要하게 될 것이다. 事業體, 家口, 마을 또는 農村地域의 地圖, 都市開發計劃 및 地籍圖等은 單位들로서 利用될 수 있다. 萬一 單位들이 利用될 수 없다면 記錄해둔 것이나 實查를 遂行함으로써 單位들을 만들 수 있을 것이다. 標本選定 方法이 採択되면 標本單位의 選定은 採択된 標本選定方法에 따라 틀로부터 만들어지며 標本單位名簿가 作成된다. 간혹 틀을 準備하는 것이 不可能할 때가 있는데 이 경우에는 特別한 選定節次가 規定되어야 할 것이다.

아주 完全한 事業體名簿를 準備하는데에는 莫大한 量의 業務가 뒤따르 듯이 適切한 選定틀을 準備하는 것은 事業體調査에서 아주 커다란 問題가 되고 있다. 더우거나 企業體는 普通 家口 보다도 그 形態가 빨리 變化하고 있기 때문에 滿足할만한 틀이 準備되

있다 하더라도 이를 最近의 것으로 維持하는 데에는 繼續 問題가 뒤따른다. 이러한 作業을 위한 充分한 資金과 人員의 確保를 보장하기 위하여 企劃過程에 特別한 關心을 기울여야 할 것이다.

#### 2.4.3(e) 指針書 및 調査票類

統計調査의 設計와 企劃에 있어 가장 重要한 事項의 하나는 어떤 指針書 및 調査票類가 要求되는가를 決定하는 것이다. 例로서 한가지 가장 重要한 것은 指針書를 準備하는 일이 될 것이다. 計劃은 各 活動別로 樹立되어야 하고 參與한 人員은 進行되고 있는 各作業을 어떻게 遂行하는 가에 對하여 指示를 받아야만 할 것이다. 指示는 要求되는 背後知識, 採択되어야 할 技術, 器具와 다른 設備를 利用하는 方法, 取할 豫防措置 및 이에 따르는 行政規則等에 對하여 完全하여야 한다. 別途의 指針書가 實査, 記錄事項의 編制, 資料의 檢討와 에디팅, 코딩, 電算, 製表, 函表作成 및 公表와 같이 여러 다른 活動을 위해 準備되어야 할 것이다. 指針書에는 概念, 定義, 分類, 코드名簿, 實例의 行政書式, 亂數表, 換算表等과 같이 參考資料를 追加시킬 수 있다.

準備해야 할 以外의 것으로는 調査票(아래에서 討議됨)와 調査員이 그들의 作業狀況을 記錄할 수 있는 作業用紙 및 要函等を 들수 있다. 매우 바람직하나 잘 使用되지 않는 그밖의 것으로는 調査經驗 그 自體를 記錄한 것이다. 여러 作業活動에 必要한 人-時間(Man-hours), 費用, 問題點 및 解決事項等에 關한

아주 重要한 情報가 記錄될 수 있다. 이것은 未來의 調查設計  
를 위한 매우 重要한 情報源과 보다 큰 效率을 기할 수 있는  
土台를 마련해 주고 있다.

結果(普通은 事後調查나 獨自의인 結果分析에 依해서 얻어짐)  
의 信憑性 評價를 包含한 調查業務의 모든 事項에 對한 評價를  
할수 있도록 情報를 提供해 주는것은 亦是 重要하다. 細部資料  
가 많이 提供되면 될수록 앞으로의 利用이나 類似한 事業을 着  
手하고 있는 다른 나라의 利用을 위해서 더욱 價值있는일이 될  
것이다.

勿論 이러한 記錄事項은 아주 값비싼 것이며 아주 有能한 高  
位職員에 依해 遂行되어야 하는데 이는 無視될 수 없다. 그러  
므로 이러한 記錄事項의 內容, 妥當性 및 目的이 事前에 注意깊  
게 決定되어야 하며 利用할 수 있는 資料를 網羅한 가장 重要  
한 事項에 만 局限해야 한다.

#### 2.4.3(f) 데이터 시트(Data sheet)의 設計

데이터 시트는 統計調查가 設計되는 경우 情報를 記錄하  
는데 利用된다. 데이터 시트는 2가지의 主要形態가 있는데 이는 調  
査票와 計劃表 또는 一種의 書式이다. 前者는 對答할 空間을 가  
지고 있는 一連의 指示的인 質問書이다. 또한 後者는 普通 一  
覽表 形態이며 情報는 表題에 따라 記錄될 것이다. 그러나 이  
러한 差異에도 불구하고 大部分의 사람들은 本教材에서 使用되는  
바와 같이 데이터 시트, 調查票, 計劃表 및 蒐集表 또는 單純表

라는 말을 쓰고 있다. 데이터 슈트는 여러 목적을 위해設計되어야 한다. 卽 一部는 訓練받은 調査員에 依해 記錄되고, 一部는 応答者가 도움받지 않고 記錄하며, 또한 一部는 觀測과 物理的 測定結果를 記錄하기 위해 設計될 것이다.

데이터 슈트의 設計는 分明 그것의 利用目的에 依存하고 있다. 普通 잘 設計된 調査票나 計劃表는 다음의 要求條件을 充足해야 한다.

- (i) 可能한 限 簡單하여야 한다.
- (ii) 主題와 調査項目 또는 說問事項은 質問에 對한 答辯을 함에 있어서 自然스러운 생각을 해내는데 關心을 갖고 論理的인 順序에 따라 整理되어야 한다.
- (iii) 各 項目 또는 說問事項은 分明히 한가지 意味만을 가져야 한다.
- (iv) 調査項目과 說問事項은 쉽게 理解될수 있어야 한다.
- (v) 說問事項은 答辯에 影響을 주도록 해서는 안된다.
- (vi) 공격적인 質問을 삼가해야 한다.
- (vii) 答辯을 記錄하거나 코딩 또는 計算과 같은 關聯業務處理를 위하여 充分한 餘白이 있어야 한다.
- (viii) 데이터 슈트의 設計는 分類, 穿孔 및 製表等과 같은 活動을 容易하게 할수 있도록 되어야 한다.
- (ix) 応答內容의 信賴性 檢討를 위해 미리 準備해 놓은 照見表를 利用 可能한 곳에 備置하여야 한다.

全體調査의 成功与否가 데이터 슈트의 設計에 달려 있음은 明白

하다. 따라서 모든 觀點에서 事前檢討 및 諸設計에 모든 努力을 기울여야 할 것이다.

#### 2.4.3(g) 에디팅, 資料處理 및 公表

에디팅(內容檢討, 問題點 및 是正事項 包含), 資料處理(코딩 및 資料轉換) 및 公表(結果表, 函表, 記事內容 및 分析報告書 形態로 의 調查結果의 準備 包含) 등에 關聯하여 設計過程에서 내려야 할 많은 決定事項이 있다. 設計過程에서 우리는 무엇이 어떻게 行해져야 하는지에 대해 關心을 가지게 되며 企劃過程에서는 누가 언제 어디에서 그것을 行해야 할 것인가에 대해 關心을 갖게 될 것이다.

內容檢討, 問題點 및 是正事項에 대한 例를 들어보면 어떠한 檢討를 해야하며 調查員이나 応答者에 대한 再照會를 할 것인지의 与否 및 어떻게 是正해야 할 것인가를 決定하는 것은 分明 高位職의 責任인 것이다.

이러한 問題에 關한 決定은 統計의 質, 結果生産에 있어서의 遲延, 調查費用 등에 대해 상당한 影響을 주게 되며 資料處理要員에 대한 指示文의 形態로 文書化 되어야 한다. 企劃過程에서 누가 資料處理를 遂行하고 어느 地域에서 어떻게 그들을 募集하고 訓練시키며 그들이 有能한 것을 어떻게 保障할 것인가 등을 決定하는 것이 必要하게 될 것이다.

類似한 設計 및 企劃에 關한 考慮는 亦是 코딩, 資料轉換, 製表 및 公表 등에 適用된다.

이러한 調査의 一部인 設計事項은 最初 過程에서 遂行되어야 하며 後에 別途 注意해야 함을 잊지 말아야 하는데 이 亦是 強調되어야 할 事項이다. 우리는 오직 이러한 方法으로 調査의 各 部門이 統合되는 것을 알수 있다. 即 例를 들어 調査票設計는 調査事項을 處理하는데에, 그리고 實查活動은 調査事項을 是正하는데에 重點을 두어야 할 것이다.

#### 2.4.3(h) 募集 및 訓練

人口센서스나 大規模調査와 같은 主要事業에서 調査員과 事務處理要員으로서 行動할 數千名の 一時被雇傭者를 募集하는 것이 必要하게 될 것이다. 必要한 人員의 形態 및 그들의 活用可能性에 對한 決定을 하기 위해서는 많은 考慮가 必要하며 募集節次는 注意깊게 計劃되어야 할 必要가 있다.

調査設計過程에서 所要人員數, 最小限의 教育程度, 其他屬性( 그들이 現在 雇傭되었는가 또는 풀타임으로 利用할 수 있는가의 與否) 및 過去經驗等에 關한 指針이 樹立되어야 한다. 企劃에는 弘報日程表, 面接, 要員選拔 및 募集, 再確保할 待期要員數, 即 充分한 要員이 確保되지 않을 경우 臨時措置等을 必要로 하는 事項들이 包含될 것이다.

勿論 訓練은 調査에 雇傭된 모든 要員에게 매우 重要하다. 一部要員 特히 上位職級의 要員은 職業에 關한 教育을 받게 될 것이나 모든 事務處理要員, 實查調査員 및 指導員은 그들이 이러한 特別調査에 參與할 수 있도록 하기 위한 最小限의 補充訓

練을 必要로 할 것이다.

萬一 아주 새로운 要員이 募集될 경우에는 보다 徹底한 訓練이 提供되어야 한다. 訓練活動의 設計에는 各 要員들에게 傳達되어야 할 技術과, 이것이 어떻게 遂行되어야 할 것인가 即 講義, 示範, 指針書, 學習計劃, 口頭訓練方法, 實習, 實查經驗, 演出役割等 어느 方法으로 해야 할 것인가를 決定하는 內容이 內包 되어야 한다.

訓練計劃은 各種 指針書作成, 訓練資料 및 補助資料, 教官의 訓練, 適切한 訓練場 確保, 學課에 參席할 被訓練生에 對한 忠告, 訓練, 參席者評價等を 包含하고 있다.

이러한 計劃은 訓練指針書(單純한 것 일지라도)의 作成과 같은 面에서 볼때 指針書의 內容과 形態에 關해 決定하고 이를 作成, 設計하며 또는 要求되는 部數를 만들어 이들이 利用될 中心 地域에 配布하기 위해서는 상당한 照會時間이 必要하기 때문에 아주 細部的으로 注意 깊게 樹立되어야 할 必要가 있다. 이것은 다른 分野에도 마찬가지로 適用된다.

#### 2.4.3(i) 事前檢討

새로운 統計調査를 設計할 때에는 우리가 이에對한 아무리 많은 經驗과 知識을 갖고 있다 하더라도 發生할 모든 問題를 豫想하는 것은 不可能하다. 이러한 理由때문에 可能的 限 많은 調査節次를 檢討해 보는것은 매우 重要하다.

事前檢討 또는 豫備調査의 設計는 特히 調査結果가 目的에 合

당한가를 決定하는데 重点을 두어야 할 것이다. 設計者는 檢討하는데에 어떤일이 요구되며 그 檢討에 기여하는것이 무엇인가를 決定해야만 한다. 例를 들어 正確한 応答를 얻는데 있어 質問의 正確性を 檢討해 보는 것을 잘 決定해야 한다. 이러한 것을 行하는 한가지 方法은 普通 実査調査員을 利用하여 事實上의 調査狀態에서 面接을 試圖해 보는 것이다. 이때 同一家口와의 잇따른 面接은 質問內容이 事實上 理解되었는지, 応答者가 信賴하고 있는지의 与否를 알아보기 위해 說得과 具體的인 調査方法으로 上位職級의 統計人에 依해 行해질 수 있는 것이다.

以外的 다른 檢討方法으로서는 実査地域에서 調査員의 業務觀察 完了된 調査票의 檢討, 調査員과 其他要員으로 부터의 報告事項 即 現地点檢 및 再面接, 等を 들수 있을 것이다. 大規模 豫備調査에 있어서도 設計, 企劃 및 履行은 이와같은 接近方法을 必要로 할 것이다.

이것은 統計人의 側面에서 볼때 모든 節次가 効率的으로 履行되며 좋은 結果가 얻어지는 가를 確認하기 위한 過程에서 重要한 決定을 要求하는 統計業務의 한 分野이다. 그것은 亦是 効率的인 檢討를 計劃하고 遂行함에 있어 보다 큰 思考力과 創案力을 要求하고 있으며 이 過程에서의 欠點을 事전에 發見, 處理함으로써 앞으로의 심각한 問題를 豫防할지 모르기 때문에 全體事業의 아주 重要한 分野이다.

事前檢討와 豫備調査는 第7章 "誤差의 管理"에서 더 討議될 것이다.



## 第 3 章 資料蒐集方法

### 3.1 概 要

統計目的을 위한 資料를 蒐集함에 있어 統計人은 그들의 目的에 가장 適合한 것을 選擇해야 하는 接近方法上的 수많은 變化에 直面하게 된다. 여러가지 考慮를 해보면 可能한 結合 또는 接近方法의 數는 더욱 많아진다. 例를들어 最適의 接近方法에 對한 決定은 다음과 같은 事項을 考慮해야 할 것이다.

- 統計人이 直接 蒐集할 것인가 또는 現存 行政記錄으로 부터 拔萃할 것인가.
- 센서스(모든 單位로 부터 蒐集) 또는 標本調査中 어느 方法을 利用할 것인가.
- 面接(調査票 利用) 또는 自計式調査(应答者 自身이 調査票를 記錄함)中 어느 方法을 利用할 것인가
- 統計單位(例: 사람, 家口, 關聯事業體) 또는 別途의 蒐集單位(例: 家口主, 里長, 事業體)中 어느것으로 부터 資料를 蒐集할 것인가.

이러한 여러가지 事項에 對한 考慮를 容易하게 하기 위하여 本章은 다음 問題를 다루게 된다.

- (a) 資料出處
- (b) 統計觀察을 行하는 方法
- (c) 統計活動이나 調査의 種類

### 3.2 資料出處

統計單位 (Statistical unit) 는 언고자 하는 資料의 對象을 말하며 蒐集單位 (Collection unit) 또는 応答者는 統計單位에 關한 資料의 出處를 말한다. 어떤 경우에는 이들은 같은 單位이며 어떤때는 다르다. 例를들면 사람 (統計單位) 의 雇傭狀態에 關한 資料는 그 自身으로 부터, 그가 살고있는 家口의 家口主로 부터, 그 동네의 里長으로 부터, 그의 雇傭主 또는 政府当局으로 부터 얻어지게 될 것이다. 마찬가지로 工場이나 商店 (統計單位) 과 같은 事業體에 關한 資料는 事業體로 부터, 事業體가 所屬하고 있는 어떤 企業體로 부터 (例: 하나의 商店이 하나의 會社에 屬하는 연쇄점인 경우) 또는 政府当局으로 부터 直接 얻어질 것이다.

蒐集單位는 資料의 出處이다. 同一한 資料에 對해 여러個의 出處가 있다는 것은 統計人이 이를 選擇해야함을 뜻한다. 統計人은 目的에 가장 合당한 出處를 確認해야만 하고 各 出處로 부터 利用할 수 있는 資料의 質과 그것을 얻는 活動費用을 念頭해 두어야 한다. 萬一 學校出席에 關한 資料가 要求되는 경우 調查目的에 依存하는 여러 가지 出處가 있는데 例를들면

- 政府教育当局이 利用可能한 資料를 이미 가지고 있음.
- 學校로 부터 蒐集하는 것이 必要함 (例: 欠席에 對한 措置가 要求되는 경우)
- 家口로 부터 蒐集하는 것이 必要함 (欠席事由가 要求되는 경우)

一 學生에 對한 調査가 必要함 (出席態度에 對한 措置가 要求되는 경우)

出處의 選擇은 分明 資料의 要求條件에 달려있다. 選擇에는 調査의 範圍, 規模, 接近方法, 費用, 複雜性等에 關係되는 여러가지가 있다.

普通 統計資料와 實際로 蒐集한 資料의 出處는 政府当局(出生, 死亡, 移住, 保健, 教育, 輸入 및 輸出 등의 行政記錄), 事業體(工場生産, 商店別 販賣事項等), 家口(所得 및 支出, 家口事業, 農業生産等), 私設機關(病院施設, 教育施設, 年老者 保護施設), 사람(意見調査), 保有農地(農作物測定과 生産推計)等이다.

### 3.3 統計的 觀察方法

“統計的 觀察”(Making the statistical observation)이란 統計人이나 權限있는 代表者가 資料를 蒐集하는 行爲를 말한다.

利用되는 主要方法으로서 다음과 같이 分類된다.

面接：實查地域 또는 個人(對面)

電話照會

測定

物理的觀察(또는 目測調査)

自計式調査：郵便調査

家計簿

行政記錄으로 부터 拔萃

面接 (Interview) 은 資料提供者에 對해 直接質問 (統計인이 準備한 調査票를 使用) 하는것을 意味한다. 即 調査員이 家口, 工場, 商店 및 其他 資料提供者에 接觸하여 質問하고 答辯을 記錄하는 것이다.

測定 (Measurement) 이란 全體生産量을 推計하기 위하여 출자나 저울等과 같은 測定器具를 利用하여 標本地域과 生産量을 物理적으로 測定하는 農業에 主로 利用된다. 目測調査 (Inspection) 란 測定器具를 利用하지 않고 推計하는 것을 말한다. 經驗있는 觀察者는 그들의 觀察에 根拠하여 推計한다. 이 方法은 農作物 生産을 推計하는데 利用된다.

自計式調査 (Self enumeration) 란 応答者 (資料提供者) 가 그 나름대로 調査票를 記錄하는 것을 뜻한다. 家計簿記帳方式 (Account book method) 은 一部 國家에서 所得 및 支出 調査對象으로 選定된 家口에 利用되는 方式으로 家口主가 所得 및 支出을 每日 記錄할것을 要求하고 있으며, 이러한 家計簿들은 製表 및 分析을 위해 統計人에게 定期的으로 提出되고 있다.

한편 既存記錄으로 부터 拔萃는 記錄의 複写, 写真複写, 컴퓨터 分析用 마그네틱 테프의 確保, 資料處理와 分析을 함으로써 政府 當局을 支援하는 등 여러가지 記錄으로 부터 資料를 얻는 모든 方法을 網羅하고 있다.

다음 句節에서 이러한 方法들의 長短점에 對해 討議할 것이다. 어떤 方法이 다른 方法보다 훨씬 重要하면 이것이 먼저 討議되며 덜 重要的 方法 (電話對話, 家計簿를 利用하는 自計式調査와 같

은)은 簡單히 取扱될 것이다. 既存記錄으로 부터의 拔萃는 그 것의 特別한 性格 탓으로 最終으로 討議된다. 그러나 事實上 統計人은 重複을 피하기 위해서나 經濟的인 理解關係에 있어 이러한 方法으로 資料를 利用할 것인지의 與否를 먼저 考慮하는 것은 重要的 事이다.

위에서 보여준 여러 方法의 簡單한 分類는 統計業務에 있어서의 事實的인 經驗에 따라 여러가지 考慮事項을 綜合함으로써 이룩된다. 例를들어 面接과 實査는 보통 같이 行해지며 마찬가지로 自計式 調査와 郵便 調査方法도 普通 併行된다 (반드시 그렇지는 않음) 그래서 한가지 方法으로서의 現地面接 (Field interviews)의 長短點은 2가지 事項, 卽 對面 疏通 (自計式 方法이나 測定과는 다름)과 組織, 管理, 費用等에 關한 모든것을 지닌 實査 方法의 利用 (郵便이나 다른 方法을 利用하는 것과는 다름)으로부터 起因된다. 마찬가지로 郵便 調査에 依한 自計式 調査는 自計式 調査 方法과 郵便 調査 方法에서 起因되는 長短點을 가지게 된다.

### 3.3.1 實査: 個人面接

#### 長 點

- (i) 個人面接은 大部分의 사람들이 協調할 意思가 있기 때문에 높은 應答率을 보이고 있다.
- (ii) 面接者는 應答者에게 質問事項을 說明함으로써 不正確한 答辯을 明確하게 만들수 있기 때문에 面接에 依해 確保된 資料는 다른 方法에 依해 確保된 資料보다 더

正確하다.

- (iii) 面接者는 応答者의 個性과 調査된 單位의 代表性을 評價하고 結果를 分析하는데 있어 重要な 狀況에 對한 補充資料를 蒐集할수 있다.
- (iv) 応答者가 對答할수 있도록 分명한 質問이 提示될수 있다.
- (v) 計劃된 대로 調査를 끝내거나 誤診을 是正하기 위한 再訪問은 応答者에게 不便을 주지 않고 行해질수 있다.
- (vi) 面接者는 応答者가 記錄하도록 調査票를 郵送하는 경우 보다는 보다 自發的인 對答을 얻을수 있다.
- (vii) 応答者에 對한 密接한 質問은 面接者에 依해 재치있게 表現될수 있다. 応答者의 反應을 觀察함으로써 面接者는 必要한 境遇 調査上의 問題에 對한 補充說明과 応答者 가 參與하도록 고무시켜 주는 말을 할수 있다. 換言 하면 具體的인 事項은 다른 調査方法보다도 個人面接에 依해 보다 效果的으로 處理될수 있다.
- (ix) 調査用語는 応答者의 能力이나 教育水準에 따라 使用될 수 있다. 따라서 잘못된 說明이나 誤解的인 質問을 避하기가 쉬어지게 된다. 그러나 어떤 調査에서 面接者가 信賴性있는 對答을 얻어야 할 경우 質問用語만 의 使用과 調査票上의 質問順序를 固守하는 것이 必要할 때가 있다는 것은 주목되어야 할 일이다. 이 경우 面接者는 訓練期間동안 이에 關한 教育을 받게

될 것이다.

### 短 点

- (i) 面接者 個人の 影響이 応答을 歪曲시킬지 모른다. 例를 들면 面接者가 어떤 편파성을 가지고 있다면, 그는 自己見解를 確固히 하기 위해 無意識的으로 質問하게 될지 모른다( 이에 对備하기 위해서는 面接者는 반드시 調査票에 記録된대로 質問하도록 指示받아야 한다)
- (ii) 応答者가 面接者를 알고 있는 경우 応答者는 自己의 個人的 事項을 提供해 주는데 不快하게 생각할 것이며, 이러한 것을 私權侵害나 秘密의 漏泄로 看做하게 될 것이다.
- (iii) 広範圍한 地域에 걸쳐 散在해 있는 応答者에 对해 調査하기 위해 要求되는 旅行經費와 調査期間은 個人面接 方法을 非經濟的이고 時間 낭비적으로 만든다.
- (iv) 募集, 訓練 및 実查指導要員을 위한 組織은 다른 方法에 依한 調査보다 複雜하다.
- (v) 調査票当 費用은 電話나 郵便에 依한 調査보다 実查調査員을 雇傭하는 調査가 一般的으로 높다.
- (vi) 面接者가 온종일 집안에서만 調査하게 되는 경우 応答者의 大部分은 家庭主婦가 되는데 萬一 応答이 男子家口員으로부터 要求될 경우에는 大部分의 実查業務를 저녁이나 週末에 해야만 할 것이다. 면접하기 위해 저녁에 利用할수 있는 時間은 한 두時間에 不過하기 때문

에 個人面接方法은 面接対象者の 接觸을 要求하는 調査를 위해서는 많은 要員을 必要로 하고 있다.

### 3.3.2 実査：測定 및 物理的 觀察

“觀察”(Observation)이란 말이 使用되고 있음을 주목해야 할 것이다. 個人面接에 依한 統計的 觀察의 몇가지 長短點은 여기에도 適用된다. 以外の 內容은 다음과 같다.

#### 長 點

- (i) 実査測定은 資料를 얻는 모든 方法中에서 가장 正確하며 어떤 面에서 이것은 事實上 唯一한 實踐性 있는 方法일 것이다. 例를들어 大部分의 農村地域 応答者는 要求되는 資料를 測定하거나 記錄하지는 않을 것이다. 따라서 그들은 面接者나 郵便調査에 明確하게 答辯할수가 없다.
- (ii) 記憶에 依存하지 않으므로 再訪問으로부터 생기는 誤差 問題는 提起되지 않는다.
- (iii) 不応과 非協調를 最小限으로 維持할 수 있다.
- (iv) 結果는 調査者 判斷에 影響받지 않는다.
- (v) 同一한 方法이 항상 使用되기 때문에 期間別 地域別 資料의 比較가 可能하다.



### 短 点

- (i) 訪問時期는 例를들어 農作物이 收穫準備가 다 되어 農業生産을 始作할 경우에는 困難하다. 그러므로 特히 여러가지 다른 農作物의 成長度를 서로 다른 時期에 調査하는 경우 한 地域을 여러번 訪問하는 것이 必要하다.
- (ii) 高價의 複雜한 裝備가 要求된다.
- (iii) 觀察方法 特히 農作物의 目測方法은 不正確하며 調査員의 偏見을 받기 쉽다.

### 3.3.3 自計式調査：郵便調査에 依한 蒐集

### 長 点

- (i) 郵送된 調査票가 利用되는 경우 보다 広範圍한 地域을 対象으로 할수 있고 各 応答者와의 個人面接에 依해 達成할수 있는것 보다도 주어진 資金으로 많은 사람을 相對할수 있다.
- (ii) 郵送費는 実査要員의 交通費나 調査經費에 比하면 比較的 적게든다.
- (iii) 調査員의 募集과 訓練을 위한 많은 經費와 時間浪費가 除去된다.
- (iv) 重要的 것은 無記名으로 調査되기 때문에 郵便에 依해 보다 率直하게 質問에 对答할 것이다. 另一方面, 一部 応答者는 그들의 意圖가 書名을 하지 않더라도 밝혀질

지 모르는 두려움 때문에 생각한 바를 기록하는데 주저하게 될 것이다. 事實은 無記名の 이로움이 不利할지 疑問視 된다.

(v) 質問事項이 標準化 되어 있다. 反面 個人面接에서 調査者는 그가 答辯을 要求해서는 안되는 경우에는 質問事項을 變更하게 될 것이다.

(vi) 質問事項은 応答者の 便宜에 따라 答辯될 수 있다. 이는 応答者가 各 說問内容에 對해 생각할 時間的 餘裕를 주고 있으며 必要한 경우에는 各 項目을 記錄하기 위해 必要한 資料를 찾아볼 수 있다.

#### 短 点

(i) 郵便調査에 答辯하는 사람(即 完了된 說問書를 提出하는 사람)이 応答하지 못한 사람까지를 代表하지는 못한다. 特히 応答率이 低調할 경우에는 위에 열거한 大部分의 長点을 最大限 살리도록 이러한 制限을 充分히 強化해야 한다. 不応者에 對해서나 完全한 對答을 얻기 위한 措置를 取하는데 모든 努力을 기울이지 않는 限 結果는 매우 나빠진다.

(ii) 一般大衆에 보낸 調査票가 回送되는 경우는 一般的으로 많지 않다. 그래서 이러한 方法은 主로 事業體에서 資料를 蒐集하는 경우에 使用된다.

(iii) 応答者가 面接者の 補助없이 調査票를 記錄하기 때문에 說問内容을 잘못 解釈하고 必要한 項目을 省略하거나

製表에 알맞지 않는 形態로 記錄한 資料를 提出하게 되어 調査票를 폐기하거나 값비싼 後統措置를 取해야 할 必要性이 있다.

- (iv) 使用된 說問內容은 이의 意味에 對해서나 記錄하는 方法에 對해서 應答者를 訓練시킬수 없기 때문에 簡單하고 明白하여야 한다( 調査票를 作成하는 方法에 對한 指針書는 普通 提供되고 있으나 經驗으로 볼때 이를 注意깊게 읽는 應答者는 거의 없다 )
- (v) 回收된 資料의 正確性和 信賴性을 檢討하기란 쉽지 않다.
- (vi) 滿足스럽지 못하거나 完全치 못한 資料를 是正하기 위해 應答者에게 反送하는 것은 事實上 不可能하다.
- (vii) 應答하기를 拒絶할 만한 大部分의 質問들을 自計式調査에 包含시킬 수는 없다.

#### 3.3.4 個人面接과 自計式調査方法의 差異

##### (a) 電話問議에 依한 蒐集

이 方法은 事實上 開發途上國家에서는 자주 使用되지 않는다. 이의 主要 價值는 事業體調査의 경우 勸諭나 督促을 할수 있는데 있다. 이것은 資料가 要求될때 小規模 調査를 위해 有用하다.

### 長 点

- (i) 調査方法中에서 가장 迅速하고 經費가 적게 든다.
- (ii) 都市内の 넓은 地域에 대해 追加 경비없이 調査할 수 있다.
- (iii) 面接者를 訓練시키거나 指示하기가 容易하다. 이는 그들이 指導員과 가까운 한 事務室에서 일하고 있기 때문이다.

### 短 点

- (i) 電話応答者는 全人口를 代表하지 못하며, 이 方法을 利用함에 있어 커다란 関心을 기울이지 않는한 좋지 못한 結果를 招來할 위험성이 있다.
- (ii) 応答者가 即時 对答하기 때문에 細部的인 資料를 얻을 수는 없다. 나중에 個人面接에서 細密히 要求될 一部 敏感한 質問은 이러한 形態의 電話問議에는 不適切하다.
- (iii) 農村地域의 電話加入者數는 이러한 方法을 試問하기에는 너무 적기 때문에 都市地域 밖에서는 利用될 수 없다. 都市地域이라 하더라도 電話서비스는 非効率的이다.

### (b) 家計簿 記帳方法

이 方法은 家口가 어떻게 돈을 쓰고 있나를 알아보기 위하여 支出狀態를 分析하고자 할때 家口支出調査 (Household expenditure survey)에서 주로 利用된다.

家口는 指定된 期間에 걸쳐 日計簿形態로 모든 支出品目을 記錄, 維持하도록 要求받고 있다. “家計簿 記帳方法”

은 資料蒐集의 唯一한 手段으로는 잘 사용되지 않으며  
面接과 關聯하여 普通 利用되고 있다.

#### 長 点

- (i) 이는 費用이 낮고 便利한 制度이며 特히 実查指導員이  
調査初期에 家計簿를 配付하고 終期에 이를 蒐集할때  
쉽게 管理될수 있다.
- (ii) 去來行爲가 많을때 이 方法은 応答者의 記憶想起를 要  
求하여 資料를 얻는 方法보다는 正確한 資料를 얻을수  
있다.
- (iii) 応答者는 이러한 調査에 자주 関心을 가지며 完全하  
고도 正確한 資料를 提供하기 위하여 神經을 쓰게 된  
다.

#### 短 点

- (i) 한사람(예:主婦)이 家計簿를 管理함으로써 다른 家口  
員의 支出이 報告되지 않는 경우에는 一部 支出品目이  
누락되는 경우가 있다.
- (ii) 一部 応答者는 그들의 消費버릇을 바꾸게 되는데 「例  
를들면 家計簿에 記錄(또는 記錄忌避)하기 위해서 調  
査期間에 購入(또는 購入忌避)함」 이렇게 하여 記錄된  
資料는 正常的인 消費形態를 나타낼수 없게 된다.
- (iii) 이 方法은 자주 購入하는 主要品目에 對한 支出의 細  
部事項을 얻는데는 적당하지 않다.

### 3.3.5 行政記錄으로 부터 拔萃

政府部 処나 機關의 公式記錄은 때때로 대수롭지 않게 보일 때가 있지만 統計의 主要한 원천임에는 틀림없다.

이러한 원천으로부터 蒐集한 統計는 지금까지 討議한 다른 統計와는 根本적으로 差異가 있는데 이는 아주 다른 目的으로 特別히 樹立된 制度로 부터 基因하기 때문이다.

그러므로 이러한 統計는 하나의 副産物이며 組織이 業務를 遂行할 때에는 豫想되거나 전혀 提供될수 없는 必要物인 것이다. 따라서 이것은 組織으로 부터 생기는 資料에 依해 作成되어야 하며 利用者가 이 資料를 利用하는 경우에는 要求되는 一部를 選擇해야 할 必要가 있게 된다.

이 方法은 統計人이 統計活動을 設計함에 있어서 취하는 接近方法과는 다르며 統計人은 利用者の 需要에 對處할 자료를 作成하기 위한 制度를 設計하는데 모든 努力을 기울이고 있다. 아래 列挙되는 大部分의 長短点은 制度의 根本적인 差異에 對하여 說明해 주고 있다.

#### 長 点

- (i) 資料가 이미 蒐集되었기 때문에 상당한 費用이 節約된다.
- (ii) 이러한 統計는 다른 方法으로 蒐集된 資料보다 早期公

표를 위해 迅速히 準備되며 利用될 것으로 豫想된다.

- iii) 統計局과 다른 政府機關이 作成한 統計數字間의 對立危險性이 줄어든다. 統計局은 他部處가 作成한 統計와 類似的한 事項에 關해 아주 빈번하게 資料를 蒐集하고 統計를 公表한다 (例: 統計局에서 遂行한 勞動力調査에서 蒐集한 失業統計와 雇傭機關의 登錄事項에 依해 勞動部가 作成한 統計). 兩機關이 失業에 關係하고 있더라도 調査對象과 定義는 다르며, 이에 따라 統計利用者間의 混亂과 不信을 조장케 한다. 모든 要求條件을 갖춘 統計가 他部處에 依해 蒐集된 資料로 부터 直接作成될 경우에는 이러한 問題는 야기되지 않는다. 이에 例를 들면 出生, 死亡, 結婚, 移住, 教育, 保健, 輸入 및 輸出, 租稅等과 같은 것이 있다.

#### 短 点

- (i) 既存記錄은 要求되는 全體人口를 對象으로 하지 못하며 完全한 것이 못된다. 例로서 出生과 死亡登錄으로부터 作成된 統計는 登錄되지 않는 경우가 많을 때에는 거의 意味가 없을 것이다.
- (ii) 使用된 概念과 定義가 統計目的上 不適切하다.
- (iii) 既存記錄은 統計目的上 要求되는 모든 具體的인 情報를 提供하기 위해서는 不充分的인 資料를 包含할 수도 있다. 例로, 出生記錄은 人種区分이 除外되는 경우가 있다.
- (iv) 一部資料는 關係部處의 直接的인 關心이 아니기 때문에

不正確하고 부주의하게 作成된다. 一例로 輸入 積荷目録이 海外貿易統計에 利用되는 경우 輸入商品價值가 正確히 記錄됨을 알수 있다. 이러한 數字는 붙입할 세금을 算定하는데 關稅当局이 利用할 것이기 때문이다. 그러나 積荷目録에 記錄된 다른 項目(例, 선적량, 선적국等)은 關稅目的에 利用되지 않은 것이므로 全部 檢討할 必要는 없는 것이다.

- (v) 그러나 記錄에 對해서는 統計人의 接近을 除限하는 法的規制나 特殊秘密規定이 있게 되는데 例를들면 税金記錄과 같은 경우이다.

#### 3.4 統計利用의 諸形態

統計를 直接蒐集할 必要가 있는 경우(即, 既存記錄으로부터 얻을수 없는 경우) 統計人은 모든 蒐集單位에 對한 센서스를 實施할 것인가 또는 標本調査를 實施할 것인가를 決定해야 한다. 標本調査의 利用에 有利한 要因中에는 다음과 같은 것이 있다.

- (i) 規模가 크고 複雜한 調査에 있어 全體調査費用은 너무 높을 것이다. 例를들면 食糧生産調査의 경우 모든 庭園, 栽培地 또는 農場의 生産을 測定하기에는 너무 큰 費用이 들것이다.
- (ii) 標本調査의 資料處理에 對한 要求條件은 센서스보다 훨씬 적을 것이다. 소수의 人員과 施設만이 必要하며 結果도 빨리 利用될 수 있다.



事實資料處理뿐 아니라 活用の 모든 過程은 母集團의 一部만을 對象으로 하기 때문에 신속히 處理될 수 있으며 結果를 신속하게 利用할 수 있는 것은 標本調査의 主要 長点의 하나이다.

- (iii) 合理的으로 設計된 標本調査에서 正確한 推計가 可能하며 規模가 작아 非標本誤差(Non-sampling errors)를 줄일 수 있기 때문에 그 活動을 效果的으로 管理할 수 있다. 이와같이 標本調査는 全數調査보다 낮은 費用으로 보다 正確한 結果를 提供할 수 있다. “標本誤差”(Sampling error)는 標本調査方法을 利用하는 結果로서 發生하고 “非標本誤差”(Non sampling error)는 以外の 誤差(完全치 못한 調査對象, 不応, 不正確한 對答, 虛偽, 資料處理上의 誤差等)를 말하며 標本調査나 센서스에서 共히 發生한다.
- (iv) 複雜한 測定이나 어려운 質問을 하는 경우 全數調査를 遂行하기 위한 時間이나 訓練된 人員이 充分하지 않다.
- (v) 全數調査에서는 簡單한 質問만을 할 수 있는데 標本誤差에서는 보다 具體적이고 有用한 資料를 얻기 위하여 特別히 訓練받은 調査員을 雇傭할 수가 있다.
- (vi) 標本調査에서는 母集團의 一部에게만 質問에 答辯하는 부담을 주게 되는데 이것은 標本으로 選定된 對象者를 위해서가 아니고 全體로서의 母集團에 對한 부담을 경감시켜 주는 것이다.

다른한편 全數調査는 경우에 따라 매우 重要한 獨特한 長點을 가지고 있는데 이는 다음과 같다.

- (i) 標本誤差가 없으므로 非標本誤差를 管理할수 있는 適切한 措置가 취해질수 있다면 보다 큰 正確性을 얻을수 있다.
- (ii) 小規模地域( 調査區와 같은 )이나 分類( 例, 個人職業 또는 産業 )를 위한 正確한 統計를 얻을수 있다. 이것은 標本이 크지 않는 限 이로부터 얻을 수 없는 것이다.
- (iii) 節次가 普通 單純하고 보다 簡單하다. 例를들면 標本들과 標本單位를 選定할 必要性이 없으며 合計值를 구하는 것이 바로 標本으로부터의 單純한 推計值인 것이다. 또한 實查와 資料處理要員의 技術과 訓練이 덜 要求된다.
- (iv) 統計利用者는 센서스結果는 믿으나 標本으로부터 얻은 統計는 不信할지 모른다. 이것은 考慮할만한 重要한 事實이다. 統計人은 標本調査의 信賴性을 認定한다 하더라도 利用者가 이를 받아들이지 않으면 그 統計는 利用되지 않을 것이다. 그러나 그들이 標本調査結果를 받아들이도록 그리고 資料蒐集의 妥當한 方法을 利用者에게 教育시키는 것은 統計人의 責任인 것이다.

全數調査와 標本調査는 서로 다른 長短點을 가지고 있기 때문에 目的에 가장 合當한 方法을 選擇해야 하며 兩者에 이로움을 주는 制度를 모색하여야 한다.

센서스나 標本調査方法이 同時 利用되는 경우는 다음과 같다.

- (a) 標本選定을 위한 標本들을 만들기 위하여 센서스結果를 利用함.
- (b) 센서스資料를 最近것으로 維持하기 위하여 간이調査資料로써 補完한 定期센서스(每 5年 또는 10年)
- (c) 基本資料를 얻기 위하여 모든 單位를 調査하고 標本單位로부터 追加資料 蒐集
- (d) 大單位(例, 大規模工場, 小都市)에 對한 센서스와 費用을 줄이기 위한 小單位(例, 小規模工場, 마을)의 標本調査
- (e) 센서스結果를 미리 얻기 위하여 回送된 調査票中에서 選定한 標本을 對象으로 한 우선적인 資料處理
- (f) 小規模標本을 利用하여 센서스節次나 調査票等を 檢討하기 위하여 遂行되는 센서스事前檢討(Census pilot test)
- (g) 센서스에서 얻은 資料의 質, 包括範圍等を 評價하기 위하여 센서스 後에 遂行되는 標本調査인 事後調査(Post enumeration surveys)

## 第 4 章 資 料 蒐 集

### 4.1 概 要

前章의 3.2 절에서 資料蒐集源과 3 가지 主要資料蒐集方法에 대해서 說明하였다. 즉

- 現地에서의 面接, 測定 (Measurement) 觀察 (Observation)
- 郵便 또는 其他手段에 依한 自計式調査
- 行政機關의 保存記錄에 依한 資料蒐集

本章에서는 이러한 方法의 実行과 關聯된 事項에 대해서 說明하고자 한다.

아시아의 大部分의 開發途上国에서는 現地에서의 面接, 測定, 관찰方法과 行政記錄에 依한 資料蒐集方法이 가장 代表的인 케이스이다. 郵便 또는 電話에 依한 自計式調査 方式은 훨씬 使用빈도가 낮다. 여기서는 現地面接과 測定에 대해서 詳細하게 說明하고 相對적으로 重要도가 낮은 郵便調査에 대해서는 간단히 取扱키로 한다.

비록 資料蒐集에 對한 基本的인 원칙이 있다하더라도 現地狀況과 實施될 調査의 性格을 勘案하여 實際로 使用되는 實施方案이 考慮된다는 점을 認識해야 한다. 따라서 이것은 國家에 따라서 統計調査에 따라서 다르게 마련이다. 現地 從事員과 地方事務所 職員들을 訓練시킬 때에는 포괄적인 訓練(本 冊子에서 다루는 內容)과 個別 센서스 또는 統計調査에 對한 特別訓練을 包含하여 綜合的인 訓練을 實施해야 한다.

#### 4.2 現地에서의 資料蒐集(面接, 測定 및 觀察)

現地에서의 資料蒐集을 위해서는 訓練을 받은 사람이 調査對象의 所在地를 訪問하여 必要한 觀察을 하게 된다. 어떤 形態의 調査表나 觀察方法도 使用될 수가 있다. 調査單位는 사람, 家口, 事業體의 生産製品, 動物, 나무, 市場등이 對象이 될 수 있다. 觀察은 面接 또는 測定, 計算, 記錄을 위한 特殊한 도구 또는 다른 方法에 依해서 할 수 있다.

現地從事員은 本部와는 거의 独立的으로 自己일을 遂行해야 한다. 그들은 개념이나 정의 調査節次와 必要한 裝備의 使用方法에 대해서 詳細한 部分까지 理解해야 한다.

정해진 期間內에 正確한 資料를 蒐集하기 위해서는 現地從事員의 個人的인 知識과 能力이 問題가 된다. 그들은 모든 事項을 分明하게 理解해야 하며 正確성을 기하고 오차나 편견을 없애겠다는 情熱을 가져야 한다. 또한 필적이 깨끗해야 하며 융통성, 재치, 참을성과 正直性을 가져야 한다. 健康하고 민첩하고 能動的이어야 한다는 점도 마찬가지로 重要하다. 그들은 公務員으로서의 強한 責任感을 認識해야 하고 특히 얻어진 資料에 대한 秘密保護에 대한 認識을 깊이해야 한다. 이미 앞에서 調査한 資料의 보안유지의 必要性을 언급하였거니와 現地 從事員들은 다른 統計關係 職員들보다 더 密接하게 応答者와 接觸하고 있기 때문에 그들이 蒐集한 資料의 秘密을 그들 스스로 지켜야 한다는 것은 特히 重要한 意味가 있다.

#### 4.2.1 現地調査 節次

現地에서의 業務는 面接, 測定 또는 觀察에 의해서 단 순히 資料를 蒐集하는 것은 아니다. 거기에는 現地從事員이 遂行 해야 할 많은 種類의 課題가 있기 마련인데 이들을 集約하면 다음과 같다.

( 7 가지 과제 )

- (a) 單位의 所在地 確認 및 識別
- (b) 單位의 目錄作成
- (c) 單位의 選定
- (d) 測定
- (e) 面接
- (f) 蒐集된 資料와 內容檢査
- (g) 完成된 調査表의 提出

(a) 單位의 所在地 確認 및 識別

現地從事員은 現地에서 單位의 所在地를 確認하고 이를 識別해 내야 한다. 이것은 다음과 같은 目的때문에 必要하다.

- (i) 표본추출에 앞서 單位의 수를 把握하고 目錄을 作成하기 위해서
- (ii) 現地에서의 표본추출을 위해서
- (iii) 單位(예컨대 가구)에 대한 統計的觀察(예컨대 면접)을 위해서

目的이 어떠한 現地調査員은 必要한 單位의 所在를 確認하고 이를 識別해내는 要領이나 指針을 完全하게 터득하는 것이 重要

하다 対象外의 單位에 對해서 觀察을 한다면 또는 定해진 單位의 所在把握이 不可能하여 觀察을 할 수 없다면 標本調査인 境遇에는 代表性을 喪失하게 되고 또한 센서스인 境遇에는 아무리 全國을 對象으로 한다면이라도 그 以前에 세워놓은 計劃이나 設計와는 다른 結果가 나타날지도 모른다. 이와 같은 事項은 統計調査에 따라 다르게 되므로 調査員은 調査에 따라서 그때 그때 特殊한 訓練이 必要하게 된다. 그러나 要領이나 指針의 種類에도 不拘하고 共通的으로 使用되는 一般的인 事項을 整理하여 보면,

- 地圖. 이것은 地域의 境界와 위치를 알려주는 重要한 事項이다. 地圖는 明確하고 細密해야 되는 것이 理想的이다. 境界, 標識, 도로등이 쉽게 現地에서 識別될 수 있어야 한다. 거리와 方向이 分明하게 表示되어야 한다. 그러나 實際에 있어서 地圖는 統計調査의 遂行에 適合하도록 特別한 作業이 가해지지 않는한 活用性이 크게 떨어진다. 時間이 지나면 새로운 地圖가 作成되고 과거의 地圖는 없어지게 마련이다. 거리이름이 없거나 住宅의 番地가 없을 수도 있다.

이러한 欠點은 現地從事員으로 하여금 첫단계로 單位의 所在를 確認하고 後에 特別히 選定되는 單位의 所在를 재확인하는데 심각한 問題에 부딪치게 한다. 아마도 繼續調査인 境遇에는 調査時마다 그러한 問題가 定期的으로 發生할 것이다.

이러한 問題解決을 위해서 現地從事員은 다음과 같은 事項을 完全히 터득해야 한다.

- 地圖를 읽는법 특히 境界, 標識, 方向, 距離

- 現地에서 實際와 地圖을 어떻게 対応시킬 것인가. 特히 地圖의 方向을 現地에서의 方向과 어떻게 일치시킬 것인가, 現地에서의 距離를 어떻게 測定할 것인가

- 地圖製作 機關에서 製作된 地圖를 使用하여 略式地圖을 어떻게 作成할 것인가

- 새로운 事項을 追加하고 變動된 部分을 修正하여 地圖를 補完하는 方法

境遇에 따라서는 現地從事員들은 콤팩스, 줄자, 색연필(地圖上的 變動을 쉽게 識別할 수 있도록), 나무판(移動式 책상으로 使用하기 위한) 照明機具(야간 作業時 使用을 위한) 등을 準備한다.

- 정의. 必要한 單位의 所在確認이나 識別은 그들 單位가 統計目的으로 使用된 定義와 一致되는가의 檢討와 評價에 의해서만 可能한 境遇도 있다. 이 境遇 現地從事員은 適用할 개념이나 定義에 對해서 완벽하게 터득하고 있지 않으면 안된다.

예컨대 小売價格調査를 위하여 店舖의 所在를 確認하고 이를 識別해내야 한다면 現地從事員으로서의 調査對象과 비대상단위를 各各 識別해 낼 수 있도록 訓練되어 있지 않으면 안된다. 대개의 商店들은 識別하기가 용이하지만 修理店, 카페, 노점상, 市場 등은 特別한 판단기준이 없이는 識別이 곤란하다. 이러한 定義는 신중하게 統一的으로 適用해야 한다. 住宅, 家口, 農村등과 같은 用語는 만족스러운 統計調査를 期待한다면 철저하게 定義되어야 한다.



- 識別表示, 觀察對象이 되는 單位가 番地가 없는 住宅이거나 其他 建物이어서 거리이름도 없는 곳에 위치해 있다면 現地從事員은 다음에 그 單位의 位置를 다시 訪問할때 區別할 수 있도록 어떤 識別表示를 붙여 둘 必要가 있다. 그런 다음 表示된 單位의 거리, 方向, 距離 및 개략적인 位置를 判別하는데 必要한 略式地圖를 作成해야 한다. 單位의 識別表示는 다음 訪問時까지 充分히 견딜 수 있는 材料(예컨대 금속판 또는 番號)를 使用해야 한다. 住宅이나 建物小有主로 하여금 識別表示를 제 位置에 保存해 준다면 약간의 謝禮를 하겠다는 제의를 할 必要가 있다. (어떤 國家에서는 調査員이 調査를 위하여 訪問했을 때까지 識別表示가 제자리에 붙어 있는 住宅에 대해서 相當한 現金이 걸려 있는 복권을 謝禮로 준 境遇도 있다)

- 姓 名

특히 農村地域의 境遇 略式地圖나 位置에 대한 說明書를 보조적인 手段으로 使用하여 家口主이 이름에 의해서 住宅이나 牧場의 位置를 識別할 수도 있다.

現地從事者들은 수시로 바뀌기 때문에 여기에 對備하여 어느 從事員이 配置되더라도 所在地를 識別할 수 있는 形態로 所在地 識別 方法이 文書化 되어야 한다는 점을 명심해야 한다. 調査員은 언제라도 現地業務를 그만 둘 수 있다는 점을 勘案하여 調査員 個人이 現地事情에 밝다는 事實에 너무 많은 期待를 거는 것은 바람직하지 못하다.

(b) 目録作成

센서스나 一般統計調査에 있어서 家口, 事業體 같은 統計 또는 標本單位는 標本틀 (Sampling frame) 이나 實査에 利用될 目的으로 자주 必要하게 된다. 이와 같은 目録은 記錄이나 과거의 센서스결과로 부터는 얻어질 수 없으며 實査 入手가 可能하더라도 오래되었거나 수정을 요한다. 이러한 狀況에서는 必要한 目録은 實査前 또는 實査와 병행하여 現地 調査를 通하여 作成할 必要가 있다.

目録의 作成에 있어서 모든 單位가 包含될 수 있도록 特別한 주의가 必要하다. 單位의 所在確認이나 識別에 있어서와 마찬가지로 目録에 包含된 地域의 精密地圖와 單位에 대한 分명한 實査 指針이 반드시 必要하다. 또한 目録에 包含된 各 單位로부터 蒐集 되어야 할 項目을 明示한 目録樣式과 調査의 進行에 關한 詳細한 指針이 必要하다.

調査地域의 境界는 調査目的에 適合하게 作成된 地圖에 의해서 明確하게 区分되어야 한다. 目録作成 전에 特히 다른 地域이 잘못 包含되지 않도록 地域境界의 正確性を 期하기 위해서 全地域을 調査하는 것이 重要하다. 이러한 作業過程에서 現地에서 蒐集된 知識을 바탕으로 모든 疑問점을 해소시켜야 한다. 이시점에서 現地事情, 예컨대 建設이나 開發같은 要因을 地圖에 表示함으로써 모든 變動狀況이 補完될 수 있다. 現地に 대한 잠정적인 調査가 끝나면 目録作成을 시작할 수 있는 최적지점 (예컨대 西北쪽 끝)의 선정이 可能하다. 地圖上에 그 지점의 表示를 明確하게 한다.

이러한 作業이 끝난 후에는 이 地点에서부터 일정한 方向으로 체계적으로 目錄이 作成되어야 한다. 예컨대 時計方向으로 길을 따라 作成한다. 일련번호는 처음 시작되는 單位부터 순서대로 부여한다. 이러한 일련번호는 後에 어떤 特定單位의 所在를 識別할 수 있도록 地圖上에 表示하는 것이 정상적이다.

하나의 建物内に 있는 여러개의 單位에 대해서 目錄을 作成함에 있어서는 다음 建物로 넘어가기 전에 그 建物内の 모든 單位를 包含시키는 것이 重要하다. 사람이 거주하는 지하층의 누락을 防止하고 보조통로나 초인중 같은 것은 다른 表示가 되도록 特別한 주의가 기울여져야 한다. 事業體나 다른 經濟的인 單位의 目錄作成에 있어서는 新規事業體를 包含시키고 休業, 廢業事業體 및 他地域으로 移轉한 事業體를 除外시키는데 관심을 집중시키는 것이 重要하다. 이미 "所在確認 및 識別"에서 言及한 바 있듯이 目錄作成에 있어서도 対象單位만 包含시켜야 하며 비대상 單位는 除外시켜야 한다는 점에 주의를 기울여야 한다. 여기에서는 目錄作成의 一般的인 節次와 代表的인 問題點에 대해서만 言及하였다.

詳細한 內容은 센서스나 統計調査의 實查에서 具體的으로 言及될 必要가 있다.

### (c) 實查를 위한 單位選定

전수조사에 있어서는 統計調査의 對象에 包含되는 모든 單位가 所在確認, 識別, 目錄이 作成되어 自動的으로 選定된다. 표본조사에 따라서는 표본틀(Sampling frame)과 調査해야 할 單位選定 作業이 中央機關의 設計段階에서 이루어지는데 예컨대 연

간 製造事業體調査 또는 小売物価調査 등의 境遇에 그러하다. 그러나 다른 統計調査 특히 多段표본에 있어서는 調査員 자신이 現地에 도착한 후에 統計單位를 선정해야 한다. 예를 들면 곡물의生産量調査를 위하여 어떤 地域의 일 部分을 선정하거나 부락 또는 도시의 block 單位로 家口를 선정하는 境遇가 이에 속한다. 이러한 境遇에는 調査員이 정해진 作業順序에 의해서 統一的으로 그리고 임의 (random) 로 표본을 추출하기 위해서 모든 주의를 기울이는 것이 重要하다. 예컨대 표본틀 (Sampling frame) 에 있는 單位를 적절히 目錄化하고 第5章에서 說明한 亂수표로부터 random number 를 뽑아내는 作業등이 그 것이다. 個人的인 판단이 意圖的으로 介入되어서는 안된다. 또한 指示事項을 적절히 따르지 않음으로 因한 부주의에 의한 誤謬가 있어서도 안된다.

표본이 임의로 추출되었다 하더라도 調査結果의 질은 選定된 單位의 어느 하나라도 調査事項의 全部 또는 一部를 提供할 수 없다면 떨어지게 마련이다. 이러한 境遇가 發生되면 調査員은 모든 努力을 기울여 처음에 選定된 單位로부터 必要的인 情報을 蒐集해야 할 必要가 있다. 다른 單位로 교체하는 것은 최후의 手段임을 명심해야 하며 전출, 부재 또는 其他 由로 추출된 單位에서의 調査가 現實적으로 不可能할 때에 어쩔 수 없이 托록 한다. 無回答인 境遇에도 回答과 無回答單位의 相互比較의 基礎를 마련하기 위하여 이웃사람부터 라도 回答을 얻어내는 努力이 必要하다. 그러나 이것도 定해진 기본원칙에 따라야 한다. 이러한 無回答에 의한 표본의 대체가 적절하다는 확신이 선다면 여기에서 얻어지는 情報 또한 유용할 것으로 볼 수 있

을 것이다. 아무튼 無応答으로 표본교체가 不可避한 境遇에도 調査員에 의한 추출작업절차는 이러한 問題로부터 파생되는 影響을 최소화하기 위하여 엄격하게 지켜져야 한다.

#### (a) 測定

調査單位에 대한 觀察은 現地 종사원에 의한 測定の 形態로 遂行되기도 하고(測定에는 개수, 세기, 걸음, 目測 또는 個人的인 판단등이 包含될 수 있다) 또 다른 方法은 面接에 의한 応答의 記錄으로 觀察하는 것을 들 수 있다. 때로는 特히 農業 統計調査에서 面接과 測定이 同時에 이루어지기도 한다. 本章의 마지막部分에서 測定手段의 種類나 面接技術등에 關해서 言及코자하며 여기서는 一般的인 면만을 다루고자 한다.

觀察의 手段으로 測定이 동원되는 境遇에 指定된 測定單位나 정확성이 유지되어야 한다. 地方單位로 測定이 이루어진다면 적절한 段階에서 標準單位로 換算시켜야 한다. 저울과 저울추 같은 地方에서 使用되는 測定裝備는 피하는 것이 原則이다. 現地종사원은 機具나 其他 장비가 包含된 Kit를 供給받아야 한다. 中央에서 供給되는 測定用 Kit를 使用하면 정확성을 확신할 수 있는 큰 이점이 있는 반면 地方裝備는 이에 비하면 아주 부정확하다. 測定은 정확성의 観点에서 보면 가장 소망스러운 形態의 觀察方法이 된다. 그러나 觀察은 때로는 目測이나 個人的인 판단(예컨대, 양떼를 보고 마리수를 추정한다든지 재배면적비율의 추정하는 境遇)과 같은 정확성이 낮은 手段을 利用해야 되는 境遇도 있다.

(e) 面 接

觀察이 面接을 통하여 이루어 지는 境遇 現地종사원은 応答者와 接觸하여야 하는데 이것은 어떤 公式的인 권한의 行使가 아니다. 오히려 신분을 밝혀야 하고 (예컨대 조사원증의 제시) 応答者가 이해할 수 있는 말로 訪問目的을 說明해야 하며 面接을 통하여 얻어지는 情報은 秘密保護가 된다는 점을 인식시켜야 한다. 地方의 狀況이나 관습을 이해하고 있으면 올바른 분위기 조성을 위해서 크게 도움이 될 것이다. 面接時間의 선택 또한 마찬가지로 重要하다. 왜냐하면 家口나 事業體에서 必要한 情報을 정확하게 提供할 수 있는 사람은 대개 한두사람 밖에는 없기 때문이다. 따라서 이러한 応答者가 부재중인 時間을 피할 必要가 있으며, 너무 바빠서 面接에 応할 수 없는 時間을 피하는 것이 좋다. 境遇에 따라서는 저녁에 面接해야 되는 境遇도 있을 것이다. 同一한 応答者를 여러번 訪問해야 하는 調査에 있어서는 처음 訪問時間을 가장 편리한 時間으로 設定하고 계속되는 訪問은 처음방문시 정한 時間에 따라 하는 것이 效果的이다. 이러한 情報은 영구보존용으로 現地 종사원에게 배부된 作業일지 또는 其他 書式에 記錄해 두어야 한다.

調査表에 觀察結果를 記載할 때는 예 - 아니오形 応答, Kg, cm, hr 같은 略語, 연필로 記載할 부분과 잉크로 記載할 部分, 使用할 색갈등에 대한 기준에 관해서 特別한 주의를 기울여 全国的으로 記載 形式이 統一되도록 해야 한다. 또한 미리 코드가 부여된 설문 項目이 있다면 精確한 코드가 선택되었는가를 確認하

는 特別한 주의가 요망된다.

(f) 記載된 情報의 檢査

응답자에 대한 面接이 完了되는 즉시 조사원은 調査表를 檢査해야 한다. 調査員은 기억이 생생할 동안에 한번더 応答內容을 檢査하여 모든 설문에 応答內容이 빠지지 않고 記載되었는지 그리고 応答이 關聯項目과 일관성이 유지되는가 또는 合理的인가를 確認해야 한다. 測定單位를 재검토하여 單位가 정확하게 적용되었는가를 확인하여야 하고 모든 계산을 신속하게 재검토하여야 한다. 모든 記載事項이 읽어서 識別이 안되는 部分이 없는가를 確認하고 調査表作成을 完了한다. 檢査하는 동안에 의심나는 記載事項이 있으면 応答者를 통하여 재확인하고 必要한 곳에 주기를 달아준다.

調査表에 따라서는 關聯項目間의 비교로 檢査가 可能하도록 되어 있다. 이것은 記載된 情報의 조직적인 檢증을 可能하게 한다. 예를 들면, 토지소유규모, 토지이용과 가축수는 相互間에 일관성이 유지되어야 하며 성, 연령 및 직업간에는 논리적인 關係가 있다는 점을 들 수 있다. 때에 따라서는 調査表作成이 完了되기 전에 먼저 必要한 주거나 圖表가 作成되는 境遇도 있다. 이러한 주거나 圖表 또는 必要한 計算을 記載한 종이등도 재검토해야 하며 이들도 調査表와 함께 본부에 송부해야 한다. 정해진 原則을 고수하거나 調査表作成上에 야기된 問題點도 역시 中央에 보고되어야 한다.

調査員은 調査結果를 指導員에게 提出하기전에 센서스 혹은 統

計調査에 관한 모든 유인물을 재확인 해야하며 必要한 境遇에는 그때 까지도 분명하지 못한 部分을 応答者에게 재확인 해야 한다. 이러한 檢討는 現地중사원이 現地に 있을때 이루어져야 한다는 점이 重要하다. 이 段階에서 어떤 必要한 情報가 더 있다면 나중까지 기다리는 것보다는 応答者와 接觸하여 얻는 것이 더 용이하고 신속하다. 통신시설이 여의치 못한 지역에서는 특히 그러하다.

#### (g) 完成된 調査表의 提出

調査表가 完成되고 재확인이 끝나면 날짜 記載와 서명 을 한 후 業務遂行時 얻은 参考事項, 地圖 또는 計算에 使用한 종이, 調査 및 出張일지등을 添附하여 調査表를 監督機關에 提出하여야 한다. 監督機關에서 發見된 誤謬는 될수 있는대로 신속하게 수정하기 위해서 現地指導員에게 보내져야 한다.

#### 4.2.2 指導員의 역할

많은 國家에서 現地業務 담당 職員을 地方事務所에서 관장하고 있으며 사무소 職員들은 現地業務를 만족하게 수행 되도록 감시적인 역할을 한다. 장비나 材料의 供給, 必要한 指針, 諮問 및 協助, 俸給 支給과 같은 면에서 지도원의 支援을 받지 않는 한 현지중사원이 소기의 성과를 거둘 수 있다고 기대하기 어렵다. 실제적으로 現地에서 수행되는 業務와 關聯하여 地方事務所에서는 많은 일을 담당하고 있다. 이러한 業務는 대략 다음과 같이 8가지로 요약할 수 있다.



- (a) 본부와의 연락업무
- (b) 現地종사원의 업무할당 및 裝備의 支援業務
- (c) 訓練 및 關聯業務
- (d) 応答거절 및 其他 發生되는 問題点解決을 위한 協助
- (e) 面接實施와 關聯된 管理業務
- (f) 完全한 資料의 蒐集業務
- (g) 조사원이 提出한 結果의 檢討
- (h) 行政 및 재정상의 管理業務

모든 國家統計調查의 全般的인 책임은 中央統計機關에 속하기 때문에 指導員이 본부와 연락하는 일은 가장 핵심이 되는 것중의 하나이다. 면적이 넓은 國家에 있어서는 하나의 中央統計機關이 全國에 산재한 많은 현지종사원의 活動을 직접 管理한다는 것은 실행불가능하므로 地方事務所가 重要的 명령전달의 역할을 담당하게 된다. 이중에는 본부로부터 調查遂行 절차에 관한 지시를 받아서 이를 다시 現地業務 담당직원에게 전달하는 것과 같은 業務도 包含된다. 그러나 現地業務 담당직원으로부터 詳細한 調查進行狀況, 發生된 問題点, 수정을 必要로 하는 불만족스러운 調查事項을 전달받는 것도 또한 重要的 일중의 하나이다. 이러한 形態의 情報를 본부에 전달하는 일(통상 Feedback 이라 稱함)은 매우 重要하다. 왜냐하면 調查의 기획과 調整을 책임지고 있는 고위직 통계담당자들은 조사체계상의 약점을 수정하고 必要하다면 調查要領을 변경시킬 수 있기 때문이다.

調査員에게 業務를 할당함에 있어서 地方事務所가 할 일은 정

해진, 調査期間內에 모든 調査員의 業務가 完了될 수 있도록 業務를 할당하거나 可能하면 調査員이 그의 거주지를 벗어나는 出張期間이 가급적 짧게 조정하는 등 그 地域의 全體 實查計劃을 作成해야 한다. 하나의 인접된 넓은 地域을 調査員에게 할당하는 것이 몇개의 떨어져 있는 地域을 맡기는 것보다 분명히 훨씬 效果的이다. 할당된 詳細한 業務內容과 계속적으로 完了된 業務를 記錄하는 관리대장을 作成하는 것도 역시 重要的 일이다.

調査員에게 支給되는 것에는 적절한 양의 調査表, 調査要領書와 作業記錄인지 등에서부터 其他裝備와 測定에 必要的인 도구 같은 物件까지 包含된다. 測定の 신빙성은 測定道具의 우열에 따라 左右 된다. 따라서 그들 도구가 支給되기 전에 精確하게 작동하는 가를 檢査해야 한다.

현지종사원에 대한 訓練이 완벽한가를 확인하는 것은 重要的 일인데, 특히 重要的인 新구조사를 실시함에 있어서는 더욱 그러하다. 全體訓練計劃중 보다 포괄적인 부분만 본부에서 담당하고 개별 統計調査에 대한 實查訓練 책임은 地方事務所에 위임되는 것이 보통이다. 이러한 訓練은 예상되는 어떤 狀況에서도 그들이 취할 行動을 精確하게 알 수 있도록 세밀하고 구체적인 것이어야 한다. 調査要領書에는 이러한 境遇에 必要的인 詳細한 情報를 提供하여 調査員으로 하여금 要領書를 参照하여 그가 처한 의 문점에 대처하고 指導員에게 자주 問題解決의 諮問을 구하지 않더라도 要領書에 의해서 처리될 수 있도록 해야 한다. 地方事務所의 現地業務支援은 調査指針이 充分히 이해되었는가 조사

설차가 정확하게 지켜지고 있는가를 確認하는 일이라 할 수 있다.

현지담당직원이 봉착하는 보다 一般的인 問題點에 대해서는 다음 節에서 說明될 것이다. 때로는 이와 같은 問題들 중에는 地方事務所의 支援을 必要로 하는 것이 있는데, 이 境遇에는 만족스러운 해결책은 調査員이나 応答者의 기억이 생생할 동안이 훨씬 쉽기 때문에 신속하게 措置가 강구되어야 한다. 때로는 地方事務所의 담당직원 線에서도 応答을 거절한 사람으로부터 재차 필요한 情報의 蒐集이 可能한 境遇가 있다.

조사원의 面接을 監督하는 內容에 대해서는 7장에서 言及코자 한다. 조사원의 面接 실시 여부를 確認하는 方法은 여러가지 要素에 左右된다. 즉 이를 確認하는 담당직원의 能力, 현지종사원의 技術과 經驗 그리고 調査의 性格과 복잡성에 달려있다. 그러나 일정한 방문계획에 의하여 현지종사원을 만나는 것은 바람직하지 못하고 예고없는 불시 점검방법이 調査員의 의무수행 여부를 점검하는 가장 적절한 것으로 판단된다. 불시점검이 實施되고 있다는 점을 인식시키는 것으로도 현지종사원들이 양심적으로 業務를 수행할 수 있게 하는 方法이 된다.

현지종사원들로부터 그들이 完了한 業務를 提出받아 이를 조정하고 記錄하는 체계를 완벽하게 갖추는 것은 作業量을 할당하는데 必要한 조직을 갖추는 것만큼 重要하다. 이러한 조직을 갖춘 후에는 이것에 의해서 作業完了 比率을 산정해야 하고 全體 調査完了에 대한 심각한 지연요인을 가려 내도록 해야 한다. 이

러한 체계속에는 각 調査員에 대한 관리대장과 숫자 또는 그래프의 形態로 作成되는 指導員의 全體作業 진척도를 包含하는 것이 바람직하다.

完了된 作業結果에 대한 檢討에 대해서는 第7章에서 詳細히 說明하고자 한다.

行政 및 재정적인 管理는 특정조사가 必要로 하는 內容을 만족시키기 위하여 作成한 일련의 記錄을 통하여 이루어진다. 재정적인 記錄은 장비와 資料, 調査員의 수당과 사후적으로 支給되는 비용(예컨대 교통비와 타지역 출장시의 숙박비)과 같이 項目別 지출내역을 包含시켜 作成되어야 한다. 調査員은 일정한 週給, 업무수행상 소요된 時間에 따른 시간급 또는 성공적으로 완수한 면접건수에 의해서 給料를 받는다. 地方事務所는 給料算出 방식에 따라서 적절한 면접업무의 記錄을 作成保存할 책임이 있으며 수당이나 費用이 지체없이 支給되도록 해야 한다.

지금까지는 어떤 統計調査의 現地業務를 監督하고 조정하는데 있어서 地方事務所의 책임에 대해서 간략히 요약하여 보았다. 실제로 현지업무수행에는 여러가지 면이 있으나 이는 실시에 정인 센서스나 統計調査의 性格에 따라 다르게 된다.

#### 4.2.3 現地業務에 있어서의 기본문제점

현지종사원들은 現地에서 필연적으로 크고 작은 問題點에 봉착하게 되는데 이들 중에는 심각하고 어려운 問題도 있기 마련이며 이 境遇에는 담당 고위 統計專門家の 도움이 必要하게

된다. 大部分의 問題點은 성격상 現場에서 즉각적으로 發生되고 신속한 해결책이 必要하게 되므로 地方事務所의 담당자에 의해서 해결되어야만 한다. 그러나 어떠한 重要問題도(특히 같은 問題가 여러번 반복해서 發生되어 調査의 설계상에 어떤 근본적인 약점이 있는 境遇) 본부에 報告하여 그들의 參考資料로 活用되게 하여야 한다.

本 指針書에서 事業體調査, 家口調査 또는 소매물가조사와 같이 다양한 調査에서 發生되는 모든 一般的인 問題를 取扱하는 것은 不可能하다. 그러나 現地業務의 大部分이 家口調査와 關聯된 것이므로 이러한 形態의 調査에서 공통적으로 發生되는 問題點에 대해서 간략하게 說明하고자 한다. 여기서는 地方事務所의 지도원의 관점에서 問題點을 다루고자 하는 바, 조사원 자신이 직접 봉착한 주요 問題點과 地方事務所에서의 심사과정에서 發生된 問題點을 함께 說明할 예정이다. 여기서 간략하게 取扱한 몇가지 問題點에 대해서는 第7章에서 詳細하게 說明될 것이다.

(a) 미완성 面接: 調査表의 면밀한 심사단계에서 일부 項目의 回答이 누락된 境遇가 發見된다. 이러한 事例는 조사원이 調査表에 있는 어떤 項目을 부주의로 누락시켰거나 回答者가 어떤 질문에 대해서 回答을 거부한 境遇에 發生할 수 있다. 미완성 調査表는 可能하면 面接者인 조사원에게 되 돌려져서 그가 할 수 있다면 조사원 자신이 누락된 情報를 蒐集해 넣도록 해야 한다. 만일 그것이 不可能하다면 조사원으로 하여금 回答者를 재방문시킬 것인가 아닌가를 결정해야 한다. 때로는 그것이 간단치 않은

境遇도 있는데 특히 원거리에 있는 地域은 왕복에 素要되는 時間과 費用으로 볼때 正當化되기가 어렵다. 그러나 대체로 미완성 面接의 境遇에는 應答者를 재방문하여야 하는 것으로 되어있다. 비록 몇몇 사소한 項目일 境遇에는 재방문을 생략하기도 하는데 이때는 관련 記載事項으로부터 생략된 部分의 추리가 可能해야만 한다.

(b) 住所에 의해서 가구가 發見되지 않은 境遇: 이러한 問題는 여러가지 이유로 인해서 發生될 수 있다. 즉 家口의 소재확인인 잘못된 單位目錄 또는 地圖를 使用함으로써 어렵게 되거나 不可能한 境遇, 주소가 주거와는 무관한 商店, 事務室 또는 다른 建物인 境遇, 住宅이 파괴되었거나 空가인 境遇등에 이러한 問題가 發生될 수 있다. 여기에 대한 措置方法은 狀況에 따라 다르며 이와 같은 모든 境遇가 지침서에 명시되어야 한다. 家口의 대체추출이 必要할 것인데 狀況에 따라서는 地域을 넓혀가며 실제로 추출된 家口를 밝히는 努力이 必要하다. 조사원은 家口가 확인될 때까지 이러한 問題에 봉착하면서 계속적으로 業務를 수행하지 않으면 안되는데 단순히 家口의 소재확인이 어려워 無應答이 나올 수 있다는 것을 인정해 주어서는 안된다.

(c) 應答者와의 면담불능: 조사원이 마땅한 應答者를 찾지 못했다면 (특히 情報가 특정인을 상대로 면접해야만 얻어지는 境遇) 그는 적어도 세번은 그 家口를 다시 訪問해야 한다는 指針에 따라야 한다. 應答者가 귀가하는 시기를 이웃사람들로부터 蒐集할

수는 있어도, 선정된 應答者를 除外한 다른 사람으로부터 調査에 必要한 情報를 얻는 일은 그러한 지시가 따로 없는한 있어서는 안된다. 재방문을 용이하게 하기위하여 다음 訪問약속을 해두는 것이 必要하다. 재방문은 주어진 양식에 따라 적절히 記錄해야 한다. 재방문은 설사 調査를 할 수가 없는 경우라도 너무 많은 時間을 허비하지 않은 범위내에서 調査員에 의해서 수행되어야 한다.

(d) 應答拒絶 : 현지종사원들은 여러가지 이유에 의해서 應答拒絶 事例를 경험하게 된다. 어떤 應答者들은 調査가 징세목적으로 수행된다고 생각하거나 번거로움을 피하려고 생각한다. 만일 어떤 應答者가 面接을 거절한다면 조사원은 指導員과 상의하도록 되어 있는 것이 보통이다. 應答을 거절한다고 해서 그것으로 모든것이 끝났다고 생각하지 않는 것이 重要한 일이며 應答者가 必要한 情報의 提供할 법적인 의무가 있음을 기억해야 한다. 應答拒絶의 유형은 다양한데 그 狀況에 적절히 대처함으로써 사태가 극복되는 境遇도 있다. 提供되는 情報에 대해서는 엄격하게 秘密이 보장된다는 점을 應答者에게 믿도록 하는 事例는 아주 흔히 있다. 앞에서 지적한 바와 같이 地方事務所의 고위 통계담당자 수준에서는 面接者에게 協助하지 않으려는 應答者로 하여금 必要한 情報를 얻어내는 境遇도 있다. 또한 어떤 境遇에는 그 地域의 최고지위자의 影響力을 동원하기도 한다. 마지막 手段으로서 統計法에 의해서 그 調査에 강제규정이 있는 境遇에는 法的인 措置가해진다는 점을 利用할 수도 있다. 그러나 그러한 행동은 역효과가 發生할 소지가 다분하다. 가령 應答者로부터 원한이나 적개

심을 자극할 수도 있고 장차 그러한 위험이 가해졌다는 事實이 알려진다면 그의 친구나 이웃으로부터도 그러한 取扱을 받을 可能性이 있다. 一般的으로 조사원이든 그의 指導員이든 応答者에게 위협적인 行動을 취하여 강제로 답변을 받아내려는 의도는 応答者들을 크게 실망시키는 것이 大部分이다.

만일 応答拒絶 事例에 대한 解決에 실패하는 境遇에는 指針에 따라 다른 家口를 대체하여 추출할 必要가 있다.

(e) 情報를 提供할 能力이 없는 応答者: 한 家口의 구성원 중에는 질문에 대해서 만족스러운 応答를 할 수 없는 사람도 있다는 것을 알 수 있다. 이러한 境遇는 특히 応答者 自身이 調査表를 作成해야 할 必要가 있을 때 發生한다. 예를들면 第3章에서 家口小費支出調査의 예에서 지적한 것과 같이, 가구주로 하여금 일정기간(일주일 또는 한달간)에 所得을 支出한 모든 項目을 記錄하게 하는 것이 公同적인 方法中の 하나인데, 가구주에 따라서는 이러한 記錄을 할 수 없는 사람도 있다. (아마도 글을 쓸 수 없는 사람인 境遇가 될 것이다)

이 問題에 대한 解決策이 모색되려면 가령 父母가 문맹자일지라도 學校에 다닐 정도의 어린이들로 하여금 家計式 調査表의 作成을 맡길 수도 있을 것이다. 그러나 현실적으로 解決方法을 찾을 수가 없는 境遇에는 그러한 가구를 除外시키고 다른 가구로 대체시키지 않으면 안된다. (調査指針에 그러한 지시가 있는 境遇에 한함) 설사 資料가 提供되더라도 부정확하거나 不完全할 것이 명백한 境遇에도 応答를 얻으려는 努力을 기울이는 것은



의미가 없다. 이 境遇에는 초기에 그러한 問題를 인식하고 그에 따른 적절한 措置를 취하는 것이 바람직하다.

(f) 実査期間中の 調査員 사임 및 철수 : 일정한 期間中에 調査를 成功的으로 完了한다는 것은 調査員이 그에게 할당된 일을 完了하지 못할 때에 發生되는 어려움때문에 순조로운 것만은 아니다. 이러한 事例는 조사원의 사임이나 건강악화 같은 돌발요인에 의한 조사철수 때문에 자주 發生한다. 이런 境遇에 조사원은 現地에서 사용한 用品이나 調査表 또는 其他 材料 (使用했던 안팎든)를 반납해야 한다는 것을 명심해야 한다. 해당 指導員은 그 調査員이 맡은 全體 일중에서 어느 部分을 完了했는지를 정확하게 確認해야 하며 재방문에 대한 詳細한 事項을 얻어내야 하고 그때까지 해결안된 問題點에 대처해야 한다. 또한 作業表와 其他 管理台帳등이 제대로 記載되었는지 그리고 최근 상태로 보완되었는지를 確認해야 한다. 이러한 인계가 끝날 때까지는 部分的으로 完了한 作業分에 대한 수당이나 費用의 支給은 일체 보류시켜야 한다. 새로 교체된 調査員은 나머지 일의 完成을 위하여 세심한 作業處理設計를 해야한다. 그리고 지금까지 完了된 作業에 대해서 적절한 정리가 이루어지지 않거나 부지런히 作業수행을 하지 않는다면 지금까지 그에게 남아 있는 일에대한 확신을 가질 수 없게됨으로써 作業을 복잡하게 만들 우려가 있다. 이러한 作業은 기존의 調査員에게 그에게 부여된 본래의 일 외에 부가하여 할당하거나 可能하다면 보조조사원에게 맡김으로써 完了시키게 된다. 이러한 形態의 대체에 있어서는 신중한 판단을

요한다. 우선 기존 調査員에게 맡길 境遇 業務에 과중한 부담을 주게 되어 할당 받은 業務의 完了를 의식하여 종전보다 무성의하게 일에 임하게 될 위험이 따르고 보조조사원에게 새로이 맡기는 境遇에는 적절한 訓練이 되어 있는가 그리고 調査에 대한 원칙이나 진행절차에 대해서 익숙한가를 指導員이 확신한 후에 결정해야 하기 때문이다.

(g) 대상이 되지 않는 가구나 事業體 : 조사대상이 되지 않는 가구나 事業體에 관한 問題는 특정 센서스나 統計調査에 있어서 주어진 指針에 따라 사례별로 處理해야 한다. 복합건물( 하나의 건물이 여러개의 독립된 單位로 갈라지는 境遇)의 境遇와 도로 건설 노동자의 캠프와 같은 임시건물 같은 境遇에는 特別한 問題에 빠지게 된다. 調査指針書에는 이러한 비정상적인 狀況이 發生할 수도 있다는 것을 考慮하지 않는 境遇가 가끔있다. 이때에는 指導員이 現地에서 必要한 措置를 취하여 해결하든가 아니면 본부의 결정에 따라서 解決하지 않으면 안된다.

(h) 指針, 개념 및 조사진행과 關聯한 問題點 : 指導員이나 조사원은 모든 指針, 개념이나 調査進行등에 대해서 일일이 기억하지는 못한다. 만일 發生된 問題가 특수한 狀況에서 적용되어야 하는 개념과 關聯된 것이라면 반드시 調査指針을 参照하여야 한다. 實查業務의 效率성을 提供하기 위해서는 이러한 問題가 신중히 檢討되어야 하고 調査指針, 개념 및 조사진행을 명확하게 하기위해서 이러한 問題를 최대한 利用할 수 있어야 한다.

#### 4.2.4 면접기술

前節에서는 지도원의 立場에서 問題点과 解決方法을 알아 보았다. 지금부터는 調査員의 立場으로 돌아가서 가능한 한 最良質의 応答를 얻기 위한 면접기술을 설명하기로 한다.

일반적으로 면접자와 응답자는 서로 초면이고 면접자의 基本的인 目標은 응답자로 하여금 그에게 協助할 의사가 생길 수 있도록 호의적인 분위기를 조성하자는 데 있다. 이것을 일반적으로 동기 (Motivation)라고 하는데, 이것에 의해서 면접에 응해야 되겠다는 생각을 불러 일으킬 뿐만 아니라 나아가서는 진실한 열의를 가지고 正確한 응답을 추구할 수 있도록 關心을 자극하게 된다.

면접시에 가장 重要視 되는 것은 正確性이라는 점을 잊어서는 안되며 만일 応答者가 그의 대답이 正確하고 完全하다는 것을 確信하는 정도가 아닌 質問에 대답을 한다는 정도로만 협조한다면 응답자의 동기를 불러 일으키려는 努力은 失敗하는 것이 되고 만다. 첫인상이 가장 重要하기 때문에 調査員은 그의 태도나 외모가 응답자에게 있어서 받아들일 만한 것이어야 한다는 점을 명심해야 한다. 이것은 면접수행 場所에 어울리는 복장이나 몸가짐을 해야 한다는 뜻이다. 調査員은 可能하다면 그 지방의 관습에 따라야 한다는 점도 명심해야 한다. 예컨대 어느 社會에서는 방에 들어 가기 전에 신을 벗는 것이 습관으로 되어 있다. 어떤 종교집단에서는 반드시 저녁에만 면접에 응하는 경우도 있다.

어떤 地域에서는 그 地方 거주자들이 외부인, 예컨대 세무서원, 商人등과 같은 사람을 경계하는 경우가 있으므로 統計調査員들은 그러한 사람으로 認識되지 않도록 努力해야 한다. 따라서 면접에 들어가기 전에 미리 응답자로 하여금 自身이 누구이고 어디서 일하고 있으며 무슨자료를 提供해야 하며 수집되는 모든 자료가 嚴格하게 秘密保障이 되며 重要的 통계 목적에만 利用된다는 점을 充分히 이해하고 있는가를 確認해야 한다. 신분증을 제시하거나 案内文을 제공하는 것이 좋은 方法이 된다.

대개의 경우에는 면접을 위한 기초적인 정보를 提供함으로써 充分하지만 때에 따라서는 이것으로 불충분할 때도 있다. 응답자가 이웃집은 調査對象이 되지 않고 자기집이 調査對象이 되는 이유를 알고자 한다면 표본추출방법에 대해서 간명하게 說明을 해야 한다. 이것은 運이나 “惡運”에 속하므로 “運” 운운하는 말은 피해야 한다. 그러나 표본추출은 完全히 임의로 된 것이며 선정될 확률은 누구나 동일하다는 점을 強調해야 한다.

만일 실시되는 調査가 표본이 일정하게 교체되는 계속조사라면 표본으로 선정된 응답자의 住宅이 언제까지 표본으로 남아 있게 되고 응답부담을 可能的 한 公평하게 分散시키기 위하여 표본이 바뀌게 된다는 점을 응답자에게 充分히 說明할 수 있는 機會가 된다. 調査에 대한 보다 상세한 내용을 要求한다면(즉 무엇때문에 調査하며 結果가 어떻게 이용되는가) 一般的인 事項에 대해서 개인식시키는 方法으로 친절하게 說明해야 한다. 그러나 調査員은 그 調査가 실시할 가치가 있는가 없는가에 대한 討論

을 벌어서는 안된다.

면접을 시작함에 있어서는 물론 어떤면에서는 대화가 아닐지 모르나 아뭏든 일상적인 對話形式으로 進行하는 것이 최선이다. 대화와의 重要的 差異點은 한 사람은 묻기만하고 또 한 사람은 대답만 함으로써 생각이나 의견교환이 없다는 事實이다. 그러나, 면접자가 대화에서 처럼 자연스럽게 면접의 흐름을 유도할 수 있다면 응답자로 하여금 이것이 試驗이나 심문이라는 인상을 갖지 않도록 할 수 있겠고 응답자의 信賴를 얻을 수 있는 가장 바람직한 경우가 될 것이다.

일단 이러한 信賴를 確保하면 성공적으로 시작된 면접을 계속 하여 성공리에 마칠 수 있도록 주의를 게을리하지 말아야 한다. 면접과정에서 고려해야 할 重要的 두가지 요소가 있는데 흔히 이것을 접근( access )과 이해( understanding )라고 부른다. 처음의 access는 분명하다. 즉 응답자로 하여금 必要的 모든 情報에 접근하도록 하는 것인데 실제로 자주 그렇게 되지 않는 것은 놀라운 일이다. ( 모든 情報에 접근하지 못하는 境遇가 자주 發生한다 ) 여기에는 많은 原因이 있을 수가 있는데, 예컨대 (a) 응답자가 必要的 정보를 測定하지 않았거나 계산해 내지 않았을 경우 (이것은 농업통계조사에서 특히 그러하다), (b) 정보가 어디엔가 記錄되어 있으나 응답자 자신에게 알려지지 않은 경우, (c) 응답자가 평상시에는 대답을 알고 있었으나 지금은 잊어버린 경우, (d) 질문하는 形態가 응답자에게 完全하게 適用되지 못하는 경우를 들 수 있다.

면접자는 이러한 問題에 對해서 항상 注意를 集을리하지 말아야 하며 특히 보다 正確한 回答이 나올 可能性이 보일 때마다 이미 기록된 內容을 檢討해 보지 않으면 안된다. 그리고 설사 이로 인해 면접시간이 지연되더라도 인내로서 이를 극복해야 한다. 記憶의 착오로 發生되는 不正確한 回答이나 臆作에 의한 答변등을 주의하지 않으면 안된다. (통계전문가들은 이를 記憶誤差 (recall error) 라고 부른다) 왜냐하면 이러한 형태의 오차는 많은 통계조사에서 가장 심각한 오차의 하나로 나타난다. 質問이 조사표의 설문과 동일하게 행하여지지 않았을 때의 조치에 대해서는 實施되고 있는 調査의 형태나 지시된 調査原則에 따라 다르기 때문에 一律적으로 말하기 困難하다. 그러나 때에 따라서는 만족스러운 回答을 얻어내기 위하여 質問을 다른 표현으로 바꾸어 하는 것도 許容될 수 있을 것이다. 한편 조사표 內容을 보충설명하기 위해서 조사표상에 適當한 內容을 써 넣을 수도 있을 것이다.

두번째 고려사항은 回答자에게 質問한 사항에 대한 이해가 어떨까 하는 問題이다. 예를 든다면, 각 質問에 적절한 情報가 무엇이며, 또는 回答자가 어느정도로 完全하게 答변하여야 하는가, 사용된 技術的인 용어가 正確하게 어떤 의미인가를 판단할 수 있어야 한다. 回答자에게 조사의 目的에 대해서 回答초기에 完全하게 說明하고 回答자의 이해가 부족하여 答변內容이 不完全하거나 齟齬가 빚나간 경우에 상세한 설명을 부가함으로써 回答자를 이해시키는 일은 면접자의 의무이다.

면접에서 사용되는 언어와 關聯하여 發生되는 문제가 종종 있기 마련인데 이것은 면접 초기에 해결되어야 할 事項이다. 공용어로서 들내지 세계 이상의 言語를 使用하지 않는 국가나 지역에서 모든 사람이 어느 하나의 언어는 말하고 이해할 수 있을 것이므로 이것은 전혀 심각한 문제가 될 수 없다. 조사표가 各各의 言語로 인쇄된다면 면접에는 어느 言語가 적합한가를 決定하기만 하면 원만하게 면접이 進行될 수 있다. ( 言語에 따라 조사표를 選擇하여 사용하면 되기 때문이다 ) 그러나 어떤 지역에서는 사용언어나 지방사투리가 다양하여 각 言語나 사투리별로 조사표를 인쇄하기가 불가능하다. 면접자와 응답자가 자유로이 對話할 수 있는 언어로 조사표가 작성될 수 없을 때 어려움이 發生한다. 이러한 문제가 발생하면 調査員은 면접에 必要한 言語로 질문할 수 있도록 조사표의 질문내용을 번역해야 하고 응답을 기록하기 전에 조사표상의 言語로 응답을 번역하여야 한다. 질문이나 응답을 正確하게 번역하기 위하여 많은 주의를 하지 않는 한 오류가 발생할 수 있는 危險이 있다는 것은 분명하다. 이러한 문제를 극복하는데는 調査員이 그가 조사할 地域에서 사용되는 言語나 사투리를 미리 파악하여 그 地域에서의 조사착수에 앞서 모든 技術的인 用語나 어려운 표현을 그러한 言語나 사투리로 번역하는데 익숙해지는 것이 바람직한 방법이 될 것이다.

면접자체가 어떻게 遂行되어야 하는가에 대해서는 여러가지 原則이 제시될 수 있겠으나 여기서는 다음과 같은 몇가지 重要的

事項에 대해서 說明하기로 한다.

(a) 될 수 있는대로 질문은 조사표에 있는 그대로 문도록 해야 한다. 조사표는 利用前에 신중한 實驗을 거치도록 되어 있으며 모든 질문의 의미가 질문별로 응답자에게 가장 분명하도록 表現되어야 한다.

따라서 인쇄된 표현에서 조금 벗어나는 표현을 사용하는 경우에도 다른 方向으로 이해되어 일관성 없는 응답을 얻어내게 된다. 그러나 응답자가 질문의 의미를 이해하지 못하거나 적합한 답변을 하지 못할 때는 다른 표현으로 質問하거나 보충설명을 해야 한다. 이 경우에 본래의 질문의 의미가 변질되지 않고 따라서 수집된 情報가 인쇄된 질문에 대한 적합한 응답이 될 수 있도록 주의가 要望된다.

(b) 조사표상의 모든 항목을 면접해야 한다. 가구조사에서는 一般的으로 모든 質問을 모든 사람에게 대해서 해야 한다. 물론 지침에 의해서 특정질문은 特定人에게 하고 또 특수한 경우에 질문을 특정인에게서 省略하는 경우도 있다. 그렇다고 해서 調査項目의 어느 하나에 대해서 調査하지 않아도 되는 것이 있다는 가정을 해서는 안된다.

(c) 유도질문은 결코 용납되지 않는다. 유도질문이란 면접자의 생각에 소망스럽다거나 기대할 수 있는 응답을 유도하는 것이다. 면접자는 유도질문을 하려는 경향이 있다. 왜냐하면 그렇게 하는 것이 調査를 쉽게 만들 수 있다는 생각에서이다.



그러나 이러한 種類의 質問은 유도질문에 의해서 제공된 응답을 응답자가 사실이라고 믿고 있는 것으로 가정함으로써 사실을 심 각하게 왜곡할 가능성이 있다. 유도질문의 예를 보면,

(i) 당신은 지난주 어떠한 일도 하지 않았지요? 그렇지요?

(“일하지 않았다”는 대답을 예상하는 경우임)

(ii) 당신은 가축을 소유하고 있지요? 그렇지요?

(“소유하고 있다”는 대답을 예상하는 경우임)

(iii) 당신의 어머니는 아프지요? 그렇지요?

(“아프다”는 대답을 예상하는 경우임)

(d) 면접자는 應答者의 말을 방해해서는 안된다. 응답자로 하여금 그의 말을 마치도록 하여 끝난 다음에 응답을 記錄해야 한다. 단호하게 必要性을 인정할 때까지는 대화를 중단시켜서는 안된다.

만약 응답자가 어느정도 느린 속도로 應答한다면 계속적으로 생각할 수 있도록 방해하지 말고 질문에 대하여 생각할 수 있는 충분한 시간을 주어야 한다.

(e) 면접자는 다음 질문을 묻기 전에 응답내용의 기재를 完了해야 한다. 면접자의 주의가 분산됨으로 인해서 조사표 기재 사항이 실제 응답내용과 다를 수 있다.

(f) 응답을 가정해서는 안된다. 어떤 질문에 대한 응답을 분명하게 어떻게 되리라는 것이 명백하여 굳이 질문하는 수고를 하지 않아도 좋을 것 같은 경우가 나타난다. 예컨대 어떤 부

락의 농가는 어떤 特定作物을 재배하지 않는다는 것이 이미 確  
認되어서 調査員은 면접자에 대해서 다른 狀況이 나타나리라고  
기대하지 않을지도 모른다.

그러나 꼭 그러리라고 確信할 수는 없는 일이다. 따라서 一  
般的인 대화형식으로 조사표의 설문에 따라 질문하는 것이 중요  
하다.

(g) 變化가 일어나지 않았다는 가정을 해서는 안된다: 동일한  
가구를 두번째 방문하여 면접을 계속하는 경우 지난 방문이래  
變動이 없다는 가정을 해서는 안되며 그렇다고 해서 면접시에  
지금까지 어떤 변동이 있었는가 없었는가를 질문해서는 안된다.  
이렇게 하는 것은 마치 응답자로 하여금 “없었다”는 대답을  
유도하는 것 같은 소위 “유도질문”을 던지는 것과 유사한 영  
향을 미칠 수 있다.

(h) 選定된 回答者만이 응답해야 한다. 선정된 응답자가 잠  
깐 부재중인 경우에는 언제 귀가할지를 確認하여 그때 재방문해  
야 한다. 어린이를 면접대상으로 해서는 안되며 설혹 응답이  
可能할 것으로 생각되더라도 10대 소년소녀를 면접대상으로 할  
수 없다. 앞에서 언급한 바와 같이 어린이들이 調査에 協助할  
수 있는 경우가 있다. 특히 그들의 부모가 言語上的 문제가  
있거나 문맹일 경우가 그러하다. 그러나 그들로 하여금 부모의  
승인없이 부모의 일에 關聯된 情報를 提供하도록 요구해서는 안  
된다.

표본으로 선정된 가구에 아무도 없을 때는 이웃사람들에게 인제 어디서 면접이 가능한가를 물을 수는 있으나 그들로부터 調査에 必要한 情報을 얻으려고 해서는 안된다.

(i) 타인이 같이 있는 곳에서 면접을 해서는 안된다. 이런 경우 응답자는 신경과민이 될지도 모르며 결국 秘密保護에 역행한다. 비록 그 地域의 최고 공직에 있는 사람이 면접자와 응답자의 접촉을 가능케하여 응답자의 協助를 구했다 하더라도 그 사람이 동석한 자리에서 면접을 해서는 안된다. 동시에 여러 응답자를 対象으로 하는 집단면접은 도저히 용납될 수 없다.

지금까지는 주로 가구를 대상으로 하는 면접기술에 대해서 설명하였다. 事業體를 対象으로 하는 調査에 있어서도 같은 基本原則이 적용될 수 있으나 면접 접근방식에 있어서 차이가 있을 수 있다. 예컨대, 보다 공식적인 基準을 必要로 할 것 같다. 또한 調査員은 가구조사에 비하여 지속성을 훨씬 必要로 할 것이다. 事業體調査에서 正確한 응답을 얻기 위해서는 그 事業體의 고위직에 있는 직원에 의해서만 자료수집이 가능한 것이 대부분이기 때문이다.

#### 4.2.5 현지측정 및 관찰의 기술

現地從事員의 業務를 용이하게 하고 수집되는 자료의 信賴度를 높일 수 있는 면접기술이 있는 것과 마찬가지로 보다 좋고 正確한 測定과 관찰을 가능케 하는 技術이 있다. 여기에 포함되는 현지활동에는 여러가지가 있는데 그중에는 다음과 같은

것이 포함될 수 있다: 작물의 目測; 면적계 산시 결음수에 의한 概略的인 측정에서부터 측량용 체인이나 콤파스를 사용한 複雜한 方法에 이르기까지 다양한 계산방법; 가축수 측정; 수확량측정을 위한 표본구역의 選定 및 경지내에서의 표본구역의 위치확인; 표본구역에서 자란 작물을 베어서 타작하고 말려서 중량계산; 소매물가 자료를 상점이나 市場에서 수집하거나 기타의 여러가지 活動, 대부분의 技術은 國家에 따라서 (경우에 따라서는 한 국가내의 지역별로) 적용방식이 다르며 적용방식은 그 국가 또는 地方의 사정이나 必要에 따라 적절히 조정되기 마련이다. 이러한 상세한 내용까지는 본 教材에서 다룰 수가 없다. 이유는 이러한 내용이 각 통계사업별로 作成되어 있는 구체적인 指針書나 教育教材에 훨씬 잘 나타나 있기 때문에 그러하다.

그러나 제 6장에서 이러한 技術中 몇 가지에 대해서 一般的인 事項을 다루고자 한다. 제 6 장에서는 통계자료의 信賴性和 오류가 發生할 수 있는 경우에 대해서 설명할 것이다. 농업통계자료 수집에 使用되는 大部分의 技術은 주관적인 판단을 必要로 하거나 아주 不正確한 경향이 있는데 이는 正確한 技術의 活用에는 資源의 범위내에서 選択하기에는 무척 비싸기 때문이다. 따라서 이러한 기술을 사용함으로써 正確한 통계자료가 수집될 가능성은 높아질 것으로 생각된다.

#### 4.2.6 지도의 사용과 작성

앞에서 重要한 現地活動을 다룰 때 설명한 바와 같이

지도의 사용은 현지에서의 統計活動 중에서 아주 중요한 事項이다. 지도는 여러가지 方法으로 使用되고 있는데 이들은 大略 다음과 같은 세가지 유형으로 사용된다.

(a) 현지조사자의 責任調査區域을 正確하게 명시해 주는데 사용된다. 특히 인구센서스 같은 大規模 調査에서는 수만명의 조사원이 동원되는 관계로 각 調査員이 그들이 担当해야 할 正確한 구역을 알지 못한다면 어떤 지역(예컨대 street block 또는 마을전체)이 누락될 심각한 危險이 따른다. 반대로 두 調査員이 같은 지역을 자기의 담당구역으로 생각한다면 중복될 危險이 있다. 각 조사구의 경계를 분명하고 正確하게 명시한 지도를 사용한다면 중복이나 누락으로 인한 不正確한 調査는 되지 않을 것이라는 確信을 가져도 무방할 것이다.

(b) 住宅 또는 다른 추출단위의 소재를 確認하는데 사용된다. 표본조사에 있어서는 精確하고 상세한 주기를 包含하는 지도를 사용하여 住宅의 調査對象 여부를 분명하게 구분해 낼 수 있을 것이다.

(c) 표본틀(Sampling frame) 즉 표본추출의 기초로 이용된다. 여기에 대해서는 제 5장의 Area(또는 Map) Frames에서 설명하고자 한다.

지도의 형태는 用途에 따라 달라진다. 여기에는 조사구의 方向과 位置確認에 必要한 약식 지도, 조사원들이 그들의 맡은 지역을 지도상에 正確하게 포함하고 있는가를 確認할 수 있도록 지도원이 사용하는 대지역지도(즉 지역지도로서 많은 조사구가 표

시 되어 있음) 그리고 正確하게 축적이 표시 되어 개별 단위의 식별이 可能한 소지역지도가 있다.

새로운 지도를 작성하고 모든 변동사항을 綜合적으로 修正補完하는 작업은 현지중사원의 重要한 임무이다. 지도작성시에 지켜야 할 原則의 大部分이 국가에 따라, 조사에 따라 다르긴 하지만 통계목적의 지도작성에 있어서는 여러가지의 공통적인 사항이 있다. 여기에는 다음과 같은 내용이 包含된다.

(a) 모든 공통적인 물리적 外形(즉, 강, 도로, 철도, 산, 건물, 상점등)과 행정구역경계(즉 시도, 구시군, 동읍면, 부락, 센서스조사구등)를 나타낼 수 있는 통일된 方式이나 기호를 설정해야 하고 이러한 표준표기법이 모든 種類의 統計目的 지도의 작성에 利用되어야 한다. 특별한 모양을 표시하거나 색깔(또는 양자의 결합)을 사용하는 것도 이러한 目的에 活用할 수 있다.

(b) 축적과 四位가 모든 지도에 분명하게 표시되어야 한다.

(c) 조사지역의 경계선을 定하는 경우 현지에 나갔을 때 분명히 식별될 수 있는 물리적 外形物이 使用되어야 하며 이들 외형물은 앞으로 지도가 사용되는 경우에 언제든지 식별이 용이할 수 있도록 영구적인 것이어야 한다. 따라서 강, 도로, 철로와 전력선등은 경계선으로 적합하다. 또한 두개의 고정되어 있는 분명히 눈으로 볼 수 있는 영구목적물(라디오 송신탑같은) 사이를 잇는 假想線도 역시 경계선으로 적합하다. 반면에 눈으로 식별되지 않는 행정구역경계나 土地는 경계선으로 不適當하다. 이들은 장차 지도를 使用해야 하는 현지중사원들이 위치를 달리

인식할 수 있기 때문인데 울타리 같은 구조물은 그들이 長期的으로 現在位置에 존속할 수 있는 경우에만 경계선으로 이용할 수 있다.

(d) 경계선은 경계를 中心으로 어느쪽에 건물이나 기타형태의 單位가 위치하는가를 판단하는데 애매한 部分이 없을 정도로 精密해야 한다. 따라서 강이 경계선으로 使用되는 경우 “강의 북쪽둑” 또는 “강의 중심”과 같이 주기를 할 必要가 있다. 만일 막연히 “강”이라고만 표시한다면 水上住宅과 기타 船舶들을 어느 調査區로 할당해야 할 것인지가 불분명하여 오류와 혼동이 야기될 수 있을 것이다.

(e) 지도와 목록이 조사구와 이 조사구내의 단위의 식별에 使用된다면 수정보완시에 모든 變動事項은 양쪽에 記錄된다. 이렇게 되지 않으면 지도와 목록이 일치될 수가 없으므로 오류의 발생은 거의 피할 수가 없을 것이다.

#### 4.3 우편조사

##### 4.3.1 우편조사의 체계

우편조사가 성공적으로 遂行되기 위해서는 적절한 목록과 効率的인 우편제도와 응답자 자신이 조사표를 작성할 수 있는(자제식조사) 能力이 전제되어야 한다. 조사표나 기타 必要한 資料를 準備, 발송하고 또한 최신접수의 관리방법과 무응답자의 추적방법 등이 설계되어야 한다. 실제의 作業準備는 조사의 형태, 이용가능한 자료에 따라 事例別로 다르게 된다.

물론 우편조사가 현지조사를 必要로 하는 調査에 비해서 비용이 훨씬 적게드는 것은 사실이다. 자료처리를 除外하면 大部分의 費用은 조사표작성과 우표 값이다. 그러나 많은 국가에서는 우편망이 전국을 包含하지 못하고 주요 도시지역에 한정되어 있기 때문에 우편조사가 제약을 받고 있다. 나아가서는 人口의 대다수가 읽고 쓸 수 없거나 설사 문맹은 아니더라도 複雜한 통계표 양식에 따라 正確하게 資料를 提供할 能力이 없다. 우편조사는 一般的으로 企業經營調査나 기타 도시지역에 국한되는 調査에 적용될 수 있을 것 같다.

우송해야 할 여러가지 자료(예컨대, 서신, 응답자의 보관용을 包含한 2매의 조사표, 기재요령서, 우표가 붙은 회신용 봉투)가 묶음으로 準備되어야 한다. 봉투를 봉투속의 일부가 들여다 보이는 것으로 準備하면 편지와 봉투에 각각 주소를 기입하는 수고를 덜 수 있을 것이다.

봉투를 봉하기 전에 봉투와 봉투속의 내용물을 점검할 必要가 있다. 이들 봉투는 적당히 分類하여(가령 地域的으로 分類하여) 배부일정에 맞추어 우송하면 될 것이다. 이러한 작업과정에는 주소작성기계, 번호작성기계 등을 包含하여 다양한 장비가 동원될 수 있다.

#### 4.3.2 資料蒐集의 管理

우편조사에서의 자료수집 관리는 다음과 같은 사항을 包含한다. 즉,



- (a) 계속적으로 접수되는 회답의 取扱節次
- (b) 일정 期日까지 회신이 없는 사람에게 적절한 독촉서한 송부
- (c) 응답자에게 조사표와 案内文 배부
- (d) 資料의 수집과정에서 정기적인 보고서를 作成하여 최종적으로 더 以上の 회신을 기대할 수 없는 마감일을 결정
- (e) 一般的인 독촉서한으로서는 별다른 効果を 기대할 수 없을 때 전화, 서신 또는 可能하다면 직접방문등에 의해서 후속수단을 강구
- (f) 資料蒐集이 법적인 구속력이 있을 때 마지막 수단으로서의 법적조치의 活用

회신접수 이후의 작업으로서,

- (i) 회신의 개봉, 접수일자 記錄 및 分類
- (ii) 관리대장에 접수狀況 記錄
- (iii) 資料處理를 위한 內容檢討 및 分類
- (iv) 의문시 되는 내용의 精密檢査 및 질의
- (v) 관리대장에 질의내용 記錄
- (vi) 회신진행상황의 지속적인 추적
- (vii) 독촉서신 발송과
- (viii) 집중적인 후속조치등이 있다.

이러한 모든 作業을 적절히 遂行하여야만 調查의 성공을 기대할 수 있다. 가장 중요시 되는 관리의 수단은 응답자별로 하나의 특별 “관리대장”을 작성하는 것이다. 이 대장에는 조사

표의 발송일자, 회신접수일자와 후속조치 및 의문점에 관한 조치 사항등이 記錄된다.

調査의 進行狀況을 効率的으로 점검하기 위해서는 일주일 단위로 응답수를 記錄하는 것이 바람직하다. 이러한 기록은 통계표로 作成한다. 또한 지금까지 접수된 총회신수를 한눈에 볼 수 있도록 그래프를 作成할 수 있다. 이러한 형태의 그래프는 소위 “ogive”라고 하는데 여기에 대해서는 제10장에서 설명될 것이다. 일정한 간격으로 實施되는 계속조사에 있어서는 조사별로 이러한 그래프를 작성비교함으로써 아주 유용한 分析結果를 얻을 수 있는데 예컨대 応答率이 호전되는가 아니면 갈수록 떨어지는가를 한눈에 볼 수 있기 때문이다. 경우에 따라서는 실제 応答率과 기대 応答率을 比較하여 調査가 예상대로 進行되는가를 確認할 수 있다.

이미 회신을 보낸 응답자가 錯誤로 독촉서신을 받는 일이 없도록 특히 주의할 必要가 있다. 독촉서신을 발송할 때는 발송 직전까지 도착한 최후의 회신까지 確認해야 한다. 만약 응답을 발송한 후에 독촉서신이 도착되었다면 독촉서신은 무시하라는 주기를 添加하는 것도 좋은 方法이 될 것이다.

이미 言及한 바와 같이 모든 우편조사의 조사표는 2매를 보내는 것이 바람직하다 - 하나는 회송용으로 다른 하나는 응답자의 보관용으로 이용될 수 있다. 여기에는 두가지 목적이 있다 - (a) 회신내용에 대한 의문사항을 電話나 郵便으로 질의 했을 때 응답을 위하여 자기가 이미 보낸 資料를 参照할 수 있고

(b) 정기적으로 實施되는 調査에 있어서는 그가 이미 提供한 資料의 사본을 가지고 있기 때문에 이것을 이용하여 현재의 조사표작성을 단순화할 수 있다. (예를 들면, 前期末의 在庫는 今期初의 재고가 되므로 이것은 그가 보관하고 있는 資料에서 直接 조사표를 作成할 수 있는 항목이 된다) 또한 제공되는 자료의 질을 높이는데 도움이 되며 조사표작성이 용이해 질수록 보다 많은 회신을 기대할 수 있기 때문에 応答率을 높일 수도 있다.

#### 4.3.3 목록의 작성과 비치

목록이란 특정조사의 응답자 姓名과 住所의 記錄(또는 일련의 카드)이다. 여기에 추가하여 업종이나, 농업조사에 있어서의 영농형태와 같은 기타 情報의 記錄이 包含될 수 있다. 동 목록에는 기록된 情報의 출처나 작성일자등을 기재하는 것이 일반적이다.

현지조사든 우편조사든 간에 대개 목록은 必要로 한다. 표본조사의 경우에는 목록은 표본틀(sampling frame)로 사용되기도 한다.

目錄에 包含되는 資料의 可能한 출처 :

- (a) 잠정적인 現地調査
- (b) 인구센서스결과
- (c) 과세대장, 자격증발급대장, 등기대장, 전화번호부와 같은 행정 기록
- (d) 기타조사의 조사표

(e) 貿易協會등과 같은 특수한 출처

목록의 수정보완을 위하여 주기적으로 記錄된 모든 사람에게 簡略한 수정보완용 양식을 발송하면 좋을 것이다. 이러한 方法에 의해서 가장 最近의 情報만 사용할 수 있으며 調查가 大規模일 때는 더욱 效率的이다.

여러 형태의 기관에서 目錄을 保有하고 있다. 만일 같은 목록이 행정목적과 통계목적 양쪽으로 사용될 수 있다면 努力이 절약될 수 있으나 통계정보는 秘密이 보호되어야 하므로 이것이 不可能한 경우도 있다.

최근에는 많은 나라에서 모든 통계목적에 사용할 수 있도록 하나의 中央 기록소를 아주 유용하게 운영하고 있다. 전통적인 예로서 사업체등록을 들 수 있다. 모든 事業體 關聯調查는 이러한 記錄을 기초로 실시할 수 있다. 이러한 기록을 이용하면 努力의 중복을 피할 수 있고 企業體의 입장에서는 여러 政府機關으로부터 같은 種類의 정보에 대한 요구를 받지 않을 수 있다.

등록자료는 유용성이 높은 반면 적절한 유지보전이 必要하고 이에 따르는 주기적인 비용을 必要로 한다. 이러한 면에서 볼 때 規模가 큰 선진국에 비하여 작은 나라에서는 문제점이 적게 발생하고 어려움이 거의 없으므로 훨씬 유리한 입장에 있는 것이 분명하다.

#### 4.4 기존기록으로부터의 통계 산출

#### 4.4.1 기록의 형태

기록에 의하여 수집되는 資料는 간행물, 행정기록 또는 경상통계조사결과로부터 얻어진 資料이다. 이러한 자료는 政府機關이나 기타 공공기관이 작성하는 大部分의 경상통계의 기초를 형성하고 있다.

이러한 記錄은 다음과 같은 3가지 유형으로 分類된다.

(a) 行政規則이나 法律에 의해서 유지되거나 접수, 발표되는 문서: 예컨대 출생신고, 혼인, 사망신고, 사업등록, 자동차등록, 노동 및 세금보고등

(b) 수입지출명세, 企業의 經營活動, 연간보고서등과 같이 어떤 組織의 내부활동과 관련한 기록

(c) 상부기관에 제출된 통계보고, 각 事業體로부터 本社에 보고하는 자료, 지방관서에서 중앙관서로 보고되는 실적보고, 각국 政府機關으로부터 蒐集되는 국제기구자료등이 여기에 속한다.

대개의 경우 기록으로부터 추출된 資料는 必要한 통계로 작성되기 전에 조사표에 적절한 형태로 이기해야 된다. 이러한 형태의 자료수집방법은 별도설계에 의한 통계조사와는 다른 문제점을 안고 있다. 이러한 차이점은 주로 資料의 범위나 対象이 記錄에 包含된 정보이상을 넘어 설 수 없다는데 기인한다.

다시 말하면, 통계전문가의 입장에서는 기초자료의 수집에 대해서 전혀 조정역할을 할 수 없다는 것이다. 기록을 관장하는 기관으로 하여금 통계적인 必要性에 보다 접근될 수 있도록 체계를 수정보완하도록 설득하는 것이 불가능하다면 현실적으로 가능

한 資料以外에는 蒐集이 不可能해진다. 그러나 자료처리방법이나 결과의 제공형태에는 큰 차이가 없다.

行政遂行 과정에서 發生되는 記錄, 등록이나 기타 보고에 의해서 얻어지는 資料를 利用하여 다양한 통계를 작성할 수 있다. 政府機關, 民間機關, 公共機關 및 企業體로부터 生産되는 통계중 많은 부분이 내부 또는 외부적인 활동기록에 의한 資料를 기초로 하고 있다. 이들 기초자료는 정상적인 업무로서 작성되며 여러 가지 목적으로 제공된다.

기록에 따라서는 기초자료로서 뿐만아니라 특별조사의 Sampling fram으로 사용되거나 다른 통계조사결과의 確認에도 利用된다. 이러한 자료수집 方法은 일반적으로 비용이 절감된다는 理由때문에 가능한한 이것을 이용할 수 있도록 努力을 기울여야 한다. 무역통계, 인구동태통계와 기타 여러가지 통계가 그러하듯이 통계에 따라서는 자료수집체계를 행정조직에 依存하지 않으면 正確性 보장이 어려운 경우도 있다.

정부기관, 공공기관 및 많은 企業體에서는 아주 貴重한 통계 자료를 包含하는 여러 種類의 기록을 보존하고 있다. 적은 비용으로 活用도가 높은 통계를 作成하기 위하여 이러한 기록을 점차적으로 많이 이용할 可能性은 충분히 있을 것 같다. 또한 현행 行政組織을 活用하여 많은 통계를 작성하고 질을 높일 수 있는 가능성도 있다.

地方統計事務所 職員 특히 고위직에 있는 사람들은 행정기록을 통계 작성에 必要한 기초자료로 더욱 広範하게 이용하는데 있어서

큰 역할을 할 수 있다. 이러한 기록은 大部分 地方의 중심이 되는 행정기관으로 모여질 것이다. 통계조직의 전반적인 發展計劃을 관장하는 本部의 통계담당자는 이러한 기록에서 유용한 통계작성이 可能하다는 事實을 전혀 認識하지 못하고 있을 가능성은 充分히 있다. 따라서 地方에 있는 통계담당자가 이러한 기록의 利用可能性을 確認할 수 있도록 해당부처와 密接하게 연락을 계속할 수 있고 중앙통계기관에 적절한 提案을 할 수 있다면 中央統計機關으로 하여금 이용가능한 자료를 최대한 活用할 수 있도록 協力할 수 있다.

이것은 처음 제 1장에서 言及한 바와 같이 통계조직내에서 상호협조하는 또 다른 事例가 될 것이다.

#### 4.4.2 資料의 제약과 질적개선

統計作成에 사용되는 大部分의 기록들이 基本的으로는 통계이외의 다른 目的으로 사용되기 때문에 이들 資料에는 제약이 따른다. 예를들면, 그와 같은 기록作成에 사용되는 概念과 定義는 다른 통계계열에서 사용되는 것과 다를 경우가 있다. 따라서 다른 資料를 기초로 작성된 통계와의 비교를 대단히 어렵게 만든다. 자료의 대상범위가 不明確하거나 不完全한 경우도 있을 것이다. 그러나 이러한 기록체계를 관장하는 責任者를 설득하여 처리방식을 바꿀 수는 있을 것이다. 기록이 통계목적에 사용되는 것을 반대하는 사람은 없을 것이나 이러한 가능성을 신중하게 검토해 본 적은 없는 것 같다. 따라서 통계관계 담

당자들은 이들 기록관계 담당자들과 業務上 密接한 관계를 정립  
해야 할 특별한 의무가 있다.

통계담당자는 記錄을 개선하여 보다 향상된 통계정보의 生産이  
可能的 方法을 모색하고 認識시키도록 努力해야 한다. 同時에  
이러한 改善이 관계당국에 의해서 쉽게 받아들여질수 있기 때  
문에 이로인하여 行政目的으로 이용하는데는 하등의 영향을 주지  
않는다는 점을 確信시켜주는 것이 重要하다. 이를 위해서는 기  
록자체와 수집의 법적근거, 자료의 출처, 概念과 定義, 양식의 설  
계, 기록의 完成, 다른 출처로부터 수집되는 유사정보와의 관계등  
에 대해서 신중하게 검토해 볼 必要가 있다.

調査對象이 完全한가를 確認하기 위하여 모든 단위와 개인에  
대한 定義가 分明한가 누락된 것은 없는가를 檢討하는 것이 効  
果的이다. 여기에서 使用되는 概念과 定義는 통일성이 없거나  
수정보완을 必要로 하는 경우도 있을 것이다. 만약 이러한 事  
項이 여의치 못하면 일련의 정의를 다른 것으로 바꾸는 方法을  
만들어 두어야 한다. 특성이나 단위에 대한 정의방법이 다른면  
이를 묶어서 보다 포괄적인 집단으로 만들거나 資料를 再分類하  
면 이렇게 하여 얻어진 자료를 기초로 作成되는 통계는 이용도  
가 보다 높아질 것이다. 기록을 수정보완하고, 자료의 作成對象  
期間이 겹치거나 중간에 빠지는 期間이 없도록 분명히 하고, 記  
錄을 신속하게 이용가능하게 하는데 努力을 기울인다면 통계작성  
의 기초자료로서의 질이 높아질 것이다.

기록양식을 재설계할 때 作成責任者로 하여금 하등의 불편을



주지 않는 범위내에서 추가로 새로운 특성을 記錄對象에 包含시킬 수 있을 것이다. 境遇에 따라서는 “秘密保護”로 分類된 資料이기 때문에 자료의 수집이 不可能할 때도 있다. 유용한 資料가 모두 使用될 수 있는가를 주기적으로 檢討해야 한다.(秘密保護가 해제되는 경우등이 있기 때문에) 또한 記錄을 유지관리하는 사람들은 이러한 記錄들이 重要的 통계에 必要的 기초자료라는 점을 認識할 수 있어야 한다. 이들 기록담당자들에게는 기초통계의 必要性을 認識시켜야 하며 통일성, 연속성, 완전성, 正確性和 적시성이 어떻게 자료의 질을 改善하는지에 대해서 이해시켜야 한다. 이러한 면에서 보면 기록관련 책임자에 대한 訓練이 매우 効果的일 것 같다.

#### 4.3.3 편집계획

編輯計劃을 樹立하는데 있어서 發生되는 重要的 문제점은 주어진 記錄으로부터 얻어질 수 있는 최종결과를 미리 把握하여 가능한한 신속하게 이용가능한 情報를 이용할 수 있게 하는데 있다. 기록을 생산한 機關이나 정부부처 자신이 간행물발간을 위하여 통계를 作成하거나 다른 機關에 통계를 보고하기 위하여 편집하는 경우도 있다. 또한 基本的인 기초자료의 사본이 통계분석이나 제표를 위하여 통계기관으로 보내지는 경우도 있다. 편집계획을 設計함에 있어서는 별도의 現地調査의 경우와 유사한 주의가 要望된다. 그리고 자료수집, 자료처리, 자료의 작성과 간행물발간에 대한 행정적인, 기술적인 면은 이러한 설계속

에 포함될 수 있다. 때에 따라서는 한 機關으로부터 다른 機關으로 記錄資料를 보낼 때 법적인 規定을 必要로 한다.

편집계획설계의 첫단계는 基本的인 記錄의 내용을 면밀히 把握하여 이것으로부터 어떠한 통계제열이 얻어질 수 있는가를 決定하는 것이다. 이렇게 하기 위해서는 記錄의 生産者와 統計의 利用者의 의견을 반영해야 할 것이다. 基本的인 기록을 개선하면 이것으로부터 作成되는 통계의 질을 改善할 수 있는 경우도 있다.

記錄上の 資料가 가령 人口에 관한 것이라면 인구의 특성에 관한 연구과제의 対象으로 적합해야 한다. 때로는 어떤 제약을 피할 수 없는 경우가 있겠으나 人口에 관한 定義를 약간 수정하여 이를 克服할 수 있을 것이다. 아니면 結果를 공표할 때 주거나 다른 제약내용을 명시하여 이용시 참작토록 하는 것이 바람직한 것이다. 때에 따라서는 표본추출 같은 특별한 작업을 통하여 보충자료를 제공함으로써 이러한 문제점을 解決하기 위한 대책을 제시하는 것이 유용할 수도 있다.

記錄을 檢討하고 이것으로부터 어떤 자료가 도출될 수 있을 것인가를 결정하고 나면 조사표를 設計하고 편집계획의 다음단계를 고려해야 한다.

통계분석을 위해서는 기록에 있는 資料가 조사표에 移記되어야 할 것이다. 대개의 경우 조사표형태가 단순한 것이 보통이나 조사표작성과 조사표기입요령에 必要한 一般的인 原則이 적용되어야 한다. 資料의 移記란 조사표의 正確한 항목에 기록으로부터

나온 수치를 채워 넣는 과정을 말한다. 따라서 조사표의 設計 때에는 資料의 移記와 그 후의 작업을 용이하게 할 수 있도록 항목별 기재란의 위치와 크기가 정해져야 한다.

경우에 따라서는 자료가 두가지 이상의 기록으로부터 같은 조사표에 移記해야 할 때가 있을 것이다. 이와 같은 狀況에서는 다른 記錄에서 얻어지는 資料가 移記되는 난은 小計와 合計를 쉽게 산출할 수 있도록 設計되어야 한다. 기입란의 위치가 잘못되었다는 것은 合計의 계산시 항목을 계속하여 直接 합하도록 되어 있는 것이 아니라 항목을 몇개씩 골라 가면서 合計해야 하는 경우를 말한다. 말하자면 계산요원은 合計의 계산을 위하여 긴 칼럼에 계속하여 적혀 있는 각 항목의 수자 중에서 단지 몇개의 해당항목의 수자만 골라내어야 하기 때문에 많은 時間이 소요될 뿐 아니라 칼럼상의 모든 수자를 단순히 合計하여 나가는 것보다 더 많은 오류를 유발하게 되는 것이다. 이 밖에도 必要한 어떤 계산이나 수정이 可能하도록 기입란이 설계되어야 한다. 예컨대, 수입량이 品目別로 각각 다른 단위로 記錄되어 있다면, 같은 단위로 변환시켜야 할 것이다.

分析이나 제표가 펀치카드방식으로 이루어진다면 자료는 기록으로부터 直接 카드로 전환될 수도 있다. 그러나 경우에 따라서 조사표를 거쳐서 資料를 수집하는 것이 문제를 단순화시키고 오류를 감소시킬 수도 있다.

資料를 조사표에 기록하거나 카드에 入力시키기 전에 資料를 포함하고 있는 기록이 기록내의 항목간 일치성이 있는가 正確

성과 완전성이 있는가를 검토해야 하며 필요하다면 수정해야 한다. 사본은 작성 즉시 정확한가를 검토해야 한다.

통계작성기관의 공통적인 기능중의 하나는 개인이나 기관이 작성한 공식보고서에서 자료를 추출하여 편집하는 것이다. 통계전문가들은 여건이 허락하는 한 기록작성 양식이 통계작성기관의 조사자료로 활용될 수 있도록 설계되어야 하며 따라서 자료편집에 직접 사용될 수 있어야 한다는 점에 대해서 기록을 관장하는 기관과 접촉을 계속해야 한다.

## 第 5 章 標本抽出의 原理

### 5.1 概 要

統計人 特히 統計를 企劃하는 사람이 資料를 蒐集하고자 할때 그것을 全數調査에 依해서 蒐集해야 할지 혹은 標本調査에 의해 蒐集해야 될지를 결정해야 될 경우에 考慮해야 할 要素들에 關하여 이미 3.4의 標本抽出의 概念에서 소개하였다. 그때 政府統計에 있어서 標本抽出方法이 중대한 역할을 한다는 것을 보았지만 그러나 우리는 이것에 關하여 關連된 基本原理에 對한 要約으로 시작해서 標本抽出에 關하여 좀더 자세히 討論하게 될 것이다.

우리는 모두 어떤 項目에 關하여 적은 情報를 갖고 있을때 그 項目에 關하여 어떤 判斷을 내려야만 될 경우를 자주 접하여 왔다. 만일 우리가 한 푸대의 쌀의 品質을 알고자 한다면 아마도 몇줌의 쌀만을 보게 될것이며 그 푸대 全體의 쌀을 전부 調査하지는 않을 것이다. 마찬가지로 船積된 말린 야자열매의 等級을 調査하고자 할때 그 一部分만을 調査하고 그 나머지는 모두 調査된 그 一部分과 같다고 추측하게 된다. 만약 맥주공장에서 監督이 生産되는 맥주의 品質을 檢査하고자 한다면 아마도 그는 몇개의 甕통을 골라서 맛을 보거나 分析하고자 하고 生産되는 맥주를 모두 檢査하려고는 하지 않을 것이다.

위의 세가지 경우의 예들은 모두 標本抽出이라고 하는 方法을

使用한다. 統計學者들과 數學者들이 標本抽出에 관하여 상세히 研究를 하여 그것은 이제 統計業務에 있어서 매우 重要한 한 分野가 되었다. 確實히 우리가 對象이 되는 集團의 一部分만을 調查하여 合理的인 結果를 얻을수 있다면 그것은 전부를 調查하는 것보다 迅速하고도 費用이 적게 들 것이다.

(a) 變異性

위와 같은 세가지 예의 경우에 있어서 共通되는 要素는 調查對象의 變異性이 매우 적다는 事實이다.

우리는 푸대 속의 쌀이 잘 섞여져 있음으로 그 푸대 속의 한줌의 쌀이 그 속에서 꺼낸 다른 어떤 한줌의 쌀과도 거의 같다는 事實을 알고 있다. 만일 우리가 한줌의 쌀이나 혹은 열줌의 쌀을 調查하더라도 결코 다른 結果를 얻게 되지는 않을 것이다. 이와 마찬가지로 의사가 환자의 피를 檢査하고자 한다면 身體의 어느 部分에서 採血하든지 간에 그는 거의 같은 結果를 얻게 된다는 것을 알고 있다. 만일 몇 개의 깡통맥주가 만족할 만하다면 나머지 모두도 만족스러운 品質이라고 確信할 수 있는 것등이다.

全體의 變異性이 매우 적다면 標本을 어떻게 抽出하든지 간에 혹은 그 크기가 얼마이든지 간에 염려할 必要가 없다. 極端的인 경우 어떤 대상이 모두 同一한 경우 그 集團에 관한 調查를 하고자 한다면 그 集團에 관하여 모든것을 알기 위하여 한 개의 對象만을 調查하여도 된다. 그 한 개의 對象을 어떻게 選擇하든지 간에 關係가 없다.

그러나 政府統計에 있어서 直面하는 問題는 주로 사람들과 그들의 行動에 관련된 것들이다. 人間의 行動에 있어서의 重要한 特性의 하나는 그들은 매우 커다란 變異性을 나타내고 그 變異性은 위에서 들은 예보다도 標本抽出을 한층 더 어렵게 하고 많은 問題를 야기시킨다.

#### (b) 確 率

그러한 問題들은 어떻게 標本을 抽出할 것인가, 모집단에 關한 推定을 어떻게 할 것인가, 추정치의 正確度를 어떻게 測定할 것인가등을 다루는 標本論이라 불리우는 統計學의 一分野의 發展을 招來하게 되었다. 標本抽出理論의 大部分은 數學의 一部分인 確率理論에 基礎를 두고 있다. 우리는 이 理論에 關하여 상세히 研究하지는 않겠으나 確率理論에 關聯된 基本原則과 政府統計에 있어서의 그 應用에 關하여 집중적으로 研究하게 될 것이다.

#### (c) 標本誤差

標本調査에 의하여 調査單位로 부터 蒐集한 情報은 모집단 全體를 調査하는 것과 같지 못하다는 것을 우리는 確信할 수 있다. 가령 어느 마을의 3살짜리 어린이의 平均體重을 알고 싶다고 가정해보자. 우리는 그 마을의 모든 3세 어린이의 體重을 달 수도 있고 또한 標本을 뽑을 수도 있다. 그러나 그 標本調査에 있어서는 서로 다른 어린이의 體重에는 確實히 많은 體重의 變異性이 있게 될 것이다. 어떤 애는 살이 찼을 것이고 어떤애는 여위었을 것이며, 어떤애는 영양상태가 양호할 것이

나 어떤 때는 영양부족 일수도 있다. 비록 우리가 體重을 正確히 측정할 能力이 있고 또한 어떠한 실수도 하지 않는다고 할지라도 우리가 標本으로 부터 얻은 結果가 모든 어린이의 실제 平均體重과 같다고 할수는 없다. 우리는 나중에 이러한 差異가 標本誤差라는 것을 알게 될것이다. 모든 어린이의 體重을 달아보지 않고 標本の 結果를 信賴하는데 대하여 치루어야 할 代價가 바로 標本誤差이다.

그러나 標本理論의 實際적인 結果는 매우 잘 設定된 數學的 法則에 따라 抽出된 標本은 그 標本誤差의 크기를 統制할 수 있고 그의 크기를 計算할 수 있다는 것이다. 이것이 統計業務에 있어서 標本抽出이 重要하다는 것을 나타낸다.

본 장의 나머지 部分에서는 우리의 討論은 거의 전부가 標本 調査에 대해 집중적으로 실시될 것이다. 이것은 즉 住民全部를 調査하는 것보다는 標本으로 抽出된 住民들로부터의 資料蒐集이 될 것이다. 이것은 政府統計業務에 종사하는 직원들이 자주 직면하게 되는 標本抽出 方法의 使用이다. 그러나 標本抽出의 技巧은 다른 많은 경우에도 使用될 수 있기 때문에 이에 관련된 몇가지 예를 들어 보겠다.

- 一 統計從事者の 業務의 質도 또한 그 一部分을 点檢함으로써 評價할 수 있고 또 統制할 수 있다. 이렇게 해서 일선 監督者(Field supervisor)가 어느 調査員의 面接調査를 몇번 觀察함으로써, 그의 業務遂行狀態를 点檢할수 있고, 事務室 監督者(Office supervisor)도 또한 業務의 一部分



을 點檢함으로써 事務處理의 質을 統制할 수 있다.

— 資料蒐集過程의 適正与否, 調査票設計, 그리고 統計業務의 다른 觀點들도 調査가 實施되기 이전에 住民의 一部를 標本으로 抽出하여 試驗調査(Trial run)를 實施하여 評價할 수 있다. 우리는 이러한 것들을 第7章의 試驗調査와 파이롯트 씨베이(Pretests and Pilot Surveys)의 題目下에서 討議하게 될 것이다.

— 調査의 暫定的인 結果는 調査票 一部를 資料處理함으로써 얻어질 수도 있다. 이것은 이용할 수 있는 最終結果가 確定되기 이전에 미리 推定値가 만들어 질 수 있게 한다.

이 위의 예에 있어서의 처음의 두가지는 統計業務에 있어서 同一한 觀點의 質 또는 正確度를 決定하는 標本抽出方法의 使用을 內包하고 있다는 것을 發見하게 될 것이다. 이것은 바로 標本抽出技巧의 使用으로서 一般的으로 質的管理라고 불리우는 것이다.

마지막으로 어떤 對象을 使用해보거나 그것을 破壞해봄으로써 만 그에 대한 情報를 蒐集할 수 있게 되는 狀況에 부딪히게 된다는 것을 言及해 둔다. 앞에서 들은 예에서 맥주의 檢査에 있어서는 標本으로 抽出된 몇개의 甕통을 열어서 그 內容物을 檢査함으로써 그 質을 檢査하였다. 그 工場의 모든 맥주를 이러한 方法으로 檢査할 수는 없고 그렇게 한다면 전혀 아무것도 판매할 수가 없게 될 것이다. 이와 마찬가지로 의사는 환자의 피의 一部分만을 檢査를 위하여 채혈할 수 있을 뿐이다. 그러

므로 이러한 경우에는 標本抽出方法外에는 選擇의 여지가 없다.

## 5.2 概念과 定義

統計學의 다른 部分에서 처럼 標本抽出에 관하여 더 상세히 論하기 이전에 特別한 意味를 가지고 있는 單語나 用語에 관하여 定義를 내릴 必要가 있다.

### (a) 母集團

第1章에서 統計單位의 特性의 “觀察”에 關聯하여 “觀察(Observation)”의 意味가 소개되었다. 調查單位라는 用語는 觀察이 되어질 對象을 지칭하기 위하여 一般的인 用語보다는 더 넓은 意味로 使用되었다. 그것은 가구, 工場, 사람등이 될 수 있다. 우리에게 關心이 있는 蒐集해야 될 모든 統計單位(Statistical unit)를 모집단(Population)이라 한다. 이것은 우리가 보통 使用하는 사람의 集團을 인구(Population)라고 하는 意味보다 더 넓은 정의이다. 그러나 統計學에서는 모든 統計集團을 包含하도록 定義를 擴張시킨다. 예를 들면 어느 마을에 있는 모든 가구, 어느 地域에 있는 모든 소매점 혹은 어느 섬의 모든 코코아농장이 모집단에 속한다. 第10章에서 보게 되겠지만 有限母集團과 無限母集團을 區別할 必要가 있다. 有限母集團이란 앞에서 言及한바와 같이 그 크기에 限定이 있는 것을 말한다. 無限母集團이란 그 限界가 없을 정도로 매우 크기 때문에 限界를 정하지 못하는 母集團을 말한다. 예를 들면 바다에 있는 고기, 해변의 모래알등이다.

(b) 標本抽出單位

個別單位나 單位의 集團等 標本抽出目的을 위한 편리한 單位를 標本抽出單位라 한다. 예를들면 농작물통계에서 資料를 蒐集할 농장을 抽出하는 것이 편리할 것이다. 이때 農장이 抽出單位이다.

여기서 우리는 標本抽出單位와 統計單位와의 用語를 區別하는데 注意를 要한다. 이들은 서로 같을수도 있으나 반드시 그렇지 않는다. 가령 勞動統計에 있어서 우리가 關心을 가지고 있는 것은 個人的 特性인 僱傭關係이며 이때 사람은 統計單位이다. 그러나 조사상 가구를 抽出하는 것이 편리하고 抽出된 가구내의 모든 사람에 關한 情報를 얻게 된다. 이때 가구는 標本抽出單位이다.

우리가 情報를 蒐集하고자 하는 單位를 統計單位라 정의하고 標本틀(Sampling frame)을 構成할 수 있는 單位나 單位의 集團을 標本抽出單位라 한다.

(c) 標本틀(Sampling frame)

資料를 蒐集할 때 使用하는 標本抽出方法에 있어서 研究할 母集團에 속하는 모든 標本抽出單位의 틀을 갖는 것이 필수적이다. 이때 우리는 이것을 標本틀이라 한다.

넓게 말해서 우리는 두 種類의 標本틀로 区分할 수 있다.

(1) 명부틀(List frame) - 母集團에 속한 모든 抽出單位가 열거된 명부, 一般的으로 重要調査에 있어서 모든 抽出單位의 完

소한 명부를 만드는 것은 어렵고 또한 費用도 많이 들게 되지만, 그러나 이러한 目的으로 이미 만들어져 있는 명부를 使用할 수도 있다. 예를 들면 가장 最近의 인구센서스에서 調査된 가구 명부 혹은 選舉管理事務所에 登錄된 유권자명부, 상공회의소에 登錄된 事業體 명부등은 명부틀로서 使用될수 있다. 때때로 그 명부는 어떤 行政目的에 맞도록 作成되었기 때문에 우리가 調査하고자 하는 모든 母集團을 正確히 모두 包含하고 있지 못한 경우도 있다. 그것은 統計目的을 위한 標本틀로서 使用하기 위한 적합한 形態로 變型해야만 할 때도 있다.

(2) 地域틀 혹은 지도틀(Area frame or map frame) — 地域의 境界를 나타내는 지도이다. 예를 들면 都市地域을 많은 區劃(Street block)으로 나누거나, 郡部地域을 작은 區域이나 마을로 나눌수도 있다. 各 區劃이나 다른 地域抽出單位에 標本抽出을 위하여 일련부호가 부여될 것이다.

본 장의 후반에서 보게 되겠지만 우리는 자주 2 단추출 혹은 그 이상의 標本抽出을 하게 된다. 예를들면 농작물통계에서 우리는 全國에 있는 모든 농장의 명부를 갖고 있지는 못할것이다. 그러나 地域틀( 아마도 지도의 형태이거나 또는 마을단위의 명부의 형태 )이 第1次 抽出單位인 마을을 抽出하기 위하여 使用할 수 있을 것이다. 그리고 다음단계로 最終적으로 농장을 抽出하기 위하여 抽出된 마을의 농장명부를 작성하는 것이 필요하게 될 것이다. 이 경우 마을을 Primary 또는 第1次 抽出單位라 하고 농장을 Secondary 또는 第2次 抽出單位라 한다.

(d) 標本의 크기

標本의 크기란 標本으로 抽出된 單位의 總數를 말한다. 분명한 것은 標本의 크기는 母集團의 標本抽出單位의 總數보다는 작아야한다는 것이다.

적당한 規模의 標本의 크기를 選擇하는 것은 몇가지 要因들에 의하여 決定된다. — 가장 重要하게 요구되는 두가지는 正確性和 調査에 소요되는 費用이다. 높은 正確度를 얻기 위하여는 많은 標本의 抽出을 요하기 때문에 많은 費用의 支出을 하여야 한다. 다시 말하면 적은 規模의 標本은 費用은 적게 들지만 正確度에 있어서는 그 質이 떨어진다. 그러므로 적당한 크기의 標本의 規模는 이러한 두 요소의 均衡을 이루게 되어, 그것은 즉 주어진 費用에서 더 높은 正確度 혹은 정해진 正確度를 위하여 더 적은 費用으로 結果를 얻을 수 있게 된다.

(e) 抽出率

抽出率이란 抽出된 標本의 母集團에 對한 比率을 말한다. 一般的으로 母集團을  $N$ 으로 表示하고 標本을  $n$ 이라 表示한다. 이 경우 抽出率  $n/N$ 으로 表示된다.

(f) 퍼라미터 (Parameter)

觀察이 되는 母集團의 特性의 참값 (True value)을 의미하는 것으로 “퍼라미터” (혹은 모수)라는 용어를 使用한다. 이러한 예로서는 어느 國家에 對한 쌀生産量 調査에 있어서 母數는 어떤 特定對象期間 동안에 쌀의 總生産量을 意味한다.

(g) 推定公式 (Estimator)

標本調査를 實施할때 母數를 推定하기 위하여 추출된 標本 抽出單位를 觀察한 特性值를 使用한다. 推定公式 (Estimator) 이란 用語는 그것에 의하여 母數의 추정이 行하여지는 統計公式을 指칭하는데 使用된다. 어떤 特定標本에 推定公式을 適用하는데 따라 각각 다른 母數의 推定值를 얻게 된다.

(h) 標本誤差

앞에서 觀察한 바와같이 標本誤差란 母集團의 모든 單位를 調査하는 대신에 그 一部の 觀察을 信賴하는데 대한 代價로 看做된다. 一般的으로 標本으로 부터 계산된 推定值는 비록 아무런 실수를 하지 않는다 하더라도 母集團의 참값이나 母數와는 完全히 一致하지 않는다. 標本으로부터 얻어진 推定值와 母集團의 참값과의 差異가 바로 標本誤差인데 그것이 크면 클수록 標本으로 부터의 推定值는 不正確해진다. 보통 우리는 母集團의 참값을 알수 없기 때문에 標本誤差를 正確하게 계산할 수는 없다. 그러나 數學的理論에 근거하여 추출된 標本에 대하여 그 誤差의 크기가 어느 정도인가를 推定할 수가 있다.

나중에 보게 되겠지만 標本誤差는 서로 다른 몇가지 要素들에 의하여 決定되기 때문에 實際적으로는 可能한 한 誤差가 적게 되도록 가장 効果的인 方法으로 標本을 設計한다.

第10章에서 標本誤差의 計算과 單純標本 (단순임의 추출에 의한)에 適用하는 공식에 대하여 소개될 것이다.

(i) 偏倚(Bias)

만약 우리가 얻은 推定値가 母集團의 참값과 같게 되었다면 그 標本은 合理的인 結果를 나타내었다고 할 수 있다.

만일 標本抽出을 반복 施行한다면 그 각각 다른 標本에 따라 推定値가 달라질 것이다. 그것은 또한 각기 다른 標本誤差를 招來하게 될 것이며 母集團의 값과는 다르게 될 것이다. 만약 이렇게 抽出을 계속하여 많은 標本을 抽出한다면 그리고 각각의 標本の 推定値가 母集團의 참값주위에 集重된다면 그 標本抽出은 合理的이라고 생각할 수 있다. 어떤 標本の 값들은 母集團의 참값보다 높을 것이며 어떤 것들은 낮게 될 것이나 大部分의 많은 標本推定値의 誤差들은 서로 상쇄되는 경향이 있다. 이러한 상쇄가 되지 않는 경우 — 예컨대 모든 推定値가 참값보다 높거나 혹은 낮은 경우 — 에는 標本誤差는 서로 상쇄되지 않는데 이때 推定公式이 偏倚(Bias)되어 있다고 말하게 된다.

예를 들면 어떤 學生集團의 平均身長을 측정하고자 標本調査를 실시한다고 假定해보자. 만약 抽出方法이 그 集團으로부터 큰 學生들만을 뽑기로 한다면 偏倚가 있는 平均身長을 얻게 될 것이다. 이것은 標本抽出을 몇번이나 반복하든 관계없이 正確한 推定値를 결코 얻게 되지 못할 것이다.

방금의 예는 명백하나 이것을 認識하지 못한채 標本設計에 偏倚를 介入시키는 것이 자주 쉽게 일어난다. 다음의 예를 생각해 보자. 政府의 어떤 統計人이 어느 都市에 살고있는 고용이 되지 않은 15세부터 60세 사이의 사람의 수를 알고 싶어한다.

이때 사람을 抽出하여 그들이 職業을 갖고 있는지 아닌지를 面接을 통하여 標本調査를 하기로 한다. 標本으로서의 応答者를 구하기 위하여 어느 월요일 아침 10 시에 어느 시장에 가서 정확하게 나이를 물어 100 사람을 調査하였다. 그중에서 65名이 職業을 갖고 있지 않았다. 이것은 非雇傭者 比率에 偏倚가 있는 推定値인데 그것은 월요일 아침 10 시에는 雇傭者의 大部分이 職場에서 勤務하고 있기 때문에 市場에 갈수 없기 때문이다. 이것은 市場에서의 非雇傭者의 比率이 都市 全體의 比率보다도 매우 높게 나타나게 될것이다.

偏倚는 標本에 對하여 또하나의 결점을 介入시키게 된다. 一般적으로 標本の 크기를 增加시켜 誤差를 줄임으로써 推定値를 改善시킨다. 다시 말하면 더 많은 觀察을 하는 費用으로써 더 正確한 結果를 얻게 된다. 만약 偏倚가 있게 되면 標本の 크기가 얼마나 크든지 간에 關係 없이 여전히 偏倚가 存在한다.

第6章과 第7章에서 資料의 質에 어떠한 영향을 偏倚가 미치는지 討論하게 될것이다.

#### (j) 非標本誤差

統計調査活動에서 發生하는 모든 誤差를 非標本誤差라고 하며 우리가 檢討한 標本誤差와 區別하기 위하여 非標本誤差라는 用語를 使用한다. 標本誤差란 標本抽出過程에서 發生한다는 것과 母集團의 一部分을 觀察하는데 대하여 치루어야만 되는 대가라는 것을 기억해 보자. 다시 말하면 非標本誤差란 偏倚가 있는 標本抽出過程을 包含하며, 不完全한 調査( 조사구내의 모든 대상을



調査하지 않은 調査), 調査員의 실수, 計算이나 資料處理過程에서 發生하는 誤差등을 말한다. 이러한 誤差는 調査過程의 어느 段階에서든지 發生할수 있으며 실제로 그 誤差의 大部分은 標本 調査나 全數調査에서 똑같이 發生한다.

非標本誤差는 매우 중대하며 그 크기는 標本誤差의 몇 배가 될 것이다. 문제는 그것을 統制할수 없고 그 영향을 계산할수 있는 고정된 법칙이 없다는데 있다. 이에 대하여 후에 자세히 어떤 경우에 誤差와 錯誤가 發生하며(第6章), 그것을 제거하거나 統制할 수 있는 方法(第7章)에 관하여 研究하게 될 것이다.

그러나 측정기구나 計算등에 있어서의 誤差는 全數調査보다는 標本 調査에 있어서 더욱 중대하다는 것이 강조되어야 한다. 이것은 標本으로 抽出된 各 單位는 母集團의 어느 部分을 代表하게 되며, 母集團 全體를 推定하기 위하여 標本 調査結果를 곱한다면 誤差 역시 곱해진다.

### 5.3 標本抽出方法

먼저 우리는 任意抽出과 非任意抽出을 區別해야 한다. 任意抽出方法은 모집단의 各 單位에 정해져 있거나 계산할수 있는(0이 아닌) 抽出될 確率을 부여한다. 어떤 교과서에는 任意抽出 대신에 確率抽出이라는 용어를 사용하고 있는데 이는 추천할만 하다.

그 외의 다른 抽出方法은 非任意抽出 혹은 非確率抽出이라 한다. 一般的으로 이러한 方法으로 抽出된 標本은 진실로 全體母

集團을 代表하지는 않는다. 이에 대한 예로는 다음과 같은 것들이 있다.

(a) 어느 市의 市民을 나타내는 電話番號簿로부터 추출된 사람의 이름들, 이때 電話를 갖고 있지 않은 사람들은 抽出될 機會가 전혀 주어지지 않는다. 電話를 가진 사람의 특성은 갖지 못한 사람의 특성과는 매우 다를 것으로 期待되며(예컨대 電話를 갖고 있는 사람은 높은 所得階層을 나타내는 경향이 있음), 이러한 標本은 만족스럽지 못하며 그로부터 얻어진 結果는 重大한 偏倚를 招來하게 될 것이다.

(b) 有意 標本抽出(Purposive sampling)이란 標本을 抽出하기 편리하도록 抽出하는 것이다. 가장 잘 알려진 이러한 抽出方法의 하나는 농작물의 수확을 調査하기 위하여 도로주변에 있는 경작지를 抽出하는 方法이다. 抽出에 있어서 部分的으로는 도로주변을 따라가면서 일정한 거리(난수표에 의해 정해진 거리)마다 抽出함으로써 任意抽出方法을 使用하며, 일정한 거리에 멈출때마다 그 도로에서 가장 가까운 경작지에서 調査를 한다. 그러나 道路와 인접하지 않는 경작지는 결코 抽出될 機會를 갖지 못하므로 結果적으로 抽出된 標本은 실제로 任意抽出이 되지 못한다. 만약 道路에 인접한 경작지의 수확량이 다른 경작지의 수확량과 같을 때에는 상관이 없으나 반드시 그렇지 않을 것이다. 예를 든다면 道路周邊의 경작지들은 통행하는 차량의 매연이나 먼지 때문에 영향을 받게 되어 수확량이 적거나 혹은 道路에서 먼 경

사진 곳의 경작지보다 저지대에 위치하여 수확량이 다르게 될 것이다.

(c) 判断抽出 (Judgement sampling) 이란 일선조사원 (혹은 監督者)의 直觀에 의해 母集團을 잘 代表한다고 判断되는 標本을 抽出하는 方法이다. 이 方法에 對한 重大한 問題點은 偏倚를 招來하기 쉬운 傾向이 있다는 점이다. 예컨대 만약 어떤 地方 官吏들에게 代表的인 표본가구를 選출하라고 한다면 그들은 그 地域의 이미지를 살리려고 많은 영향력 있는 가구를 추출하게 될 것이다. 만약 調查結果가 조세를 增加시킨다거나 원조를 청하기 위한 것이라면 그들은 平均值보다도 더욱 낮게 評價하려고 할 것이다. 비록 標本에 影響을 미치지 않으려고 意識적인 노력을 할지라도 심사숙고하여 代表的인 標本을 抽出한다는 것은 不可能하다. 사람이 달라질 경우에는 代表的이라는 意味도 또한 달라지게 마련이고 달리 抽出된 標本은 다른 結果를 招來하게 될 것이다.

불행히도 이러한 非確率抽出은 매우 重大한 결점때문에 敬시되고 있다. 標本誤差를 계산할수 있는 數學的方法이 存在하지 않기 때문이다. 다시 말하면 “機會法則” (Laws of chance)에 근거한 標本에 있어서는 推定値가 어느정도 正確한 것인가를 밝힐 수 있다. 우리는 이러한 調查結果의 評價를 위한 客觀的인 手段을 갖고 있다. 이것이 바로 可能하면 任意抽出 혹은 確率抽出方法을 使用하는 傾向이 생기게 된 重要한 特徵이고 또한 가장 큰 理由이다. 본장의 나머지 部分에서 우리는 任意抽出에 関

하여 集重的으로 討論하게 될것이다.

任意抽出方法에는 여러가지가 있다. 어떤 方法은 母集團의 各 單位가 모두 같은 抽出確率을 갖고 있으나 어떤 다른 方法은 一部單位가 다른 것보다도 더 많은 機會가 주어지는 것도 있다. 그러나 모든 任意抽出方法에 공통적인 것은 모든 單位에 抽出될 機會가 주어지며 그것을 계산할수 있다는 것이다. 그러나 有意 標本抽出 (Purposive sampling)에 있어서는 그러하지 못하다. 標本틀 (Sampling frame)로부터 標本抽出을 하고자 한다면 그러한 方法中에서 몇가지를 使用하여야 한다.

우리가 任意的이라고 생각하는 숫자를 기재하는 것은 任意抽出이 아니다. 우리가 지금 쉽게 연습할수 있는 것처럼 任意的인 順序로 숫자들을 主觀적으로 나열한다는 것은 不可能하다.

다음과 같은 任意的인 結果를 만드는 方法이 必要하게 된다.

- i) 동전던지기
- ii) 주사위 굴리기
- iii) 모자에서 숫자꺼내기

위의 세가지 方法이 올바르게 使用된다면 任意的인 結果를 가져올수 있지만 매우 規模가 큰 調査를 위한 標本을 抽出하기 위하여 使用한다면 時間을 浪費하게 되고 실용적이 되지 못한다.

이 問題를 극복하기 위하여 오늘날의 統計人들은 컴퓨터에 의해 만들어진 난수표를 使用한다. 이 난수표에는 숫자들이 수록되어 있는데 결코 서로 一定한 順序가 있거나 關聯이 없다. 그 표의 어떤 方向으로든 숫자를 읽어서 임의의 順序를 만들수 있다. 본장의

後半에 작은 規模의 난수표가 있는데 그 사용에 대하여 後에 討論하게 될 것이다.

#### 5.4 任意標本抽出의 種類

가장 좋은 標本이란 이 標本이 抽出이된 母集團을 가장 잘 代表할수 있는 것이다. 지금까지 여러가지의 任意抽出方法이 開發되었는데 統計人들은 항상 여러가지 경우에 使用될 수 있는 標本抽出方法을 發見하려고 努力한다 - 즉 어떤 경우에도 標本으로 抽出된 單位가 全體集團을 잘 代表하도록 標本을 設計하려고 努力하는 것이다. 여기서는 一般的으로 使用되는 간단한 형태에 관하여 言及하겠다.

##### 5.4.1 單純任意抽出

標本을 抽出하여 모든 單位가 標本으로 抽出될 機會가 同一한 경우에 그 抽出方法을 單純任意抽出이라 한다. 예컨대 지금 이방에서 單純任意抽出로 5名을 뽑는다면 모든 사람이 뽑힐수 있는 機會가 均等하다. 이것은 任意抽出의 一般的이고 도 單純한 形態이다.

單純任意抽出은 同一한 確率로 하나씩 차례로 抽出할 경우

i) 다음 單位를 抽出하기 전에 일단 抽出한 單位를 다시 집어넣고 抽出하는 경우

ii) 일단 抽出된 것은 다시 집어넣지 않고 나머지 중에서만 그대로 抽出하는 경우로 分類할 수 있다. 전자는 代替 (With replacement) 單純任意抽出이라 하고 후자는 非代替 (Without

replacement) 單純任意抽出이라 한다.

그러나 單純任意抽出은 사실상 널리 利用되지 않고 있는데, 이는 一部資料만이 母集團의 모든 單位에 適用될 수 있고, 一般的으로 標本設計의 效率性を 높이기 위한 抽出計劃에만 이용할 수 있기 때문이다.

#### 5.4.2 系統抽出

系統抽出이란 모집 단명부로부터 미리 정해져 있는 體系的인 順序에 따라 抽出하는 方法을 말한다.

아마도 가장 빈번히 使用되는 方法으로서 명부로부터 一定한 間격으로 標本을 抽出하는 것이다. 예를들면 10%의 標本을 抽出한다면 1부터 10 사이의 숫자사이에서 맨처음에 抽出될 單位에 부여할 숫자를 선택해서 그로부터 매 10 번째의 單位를 抽出한다. 만약 처음 뽑히는 單位의 숫자가 5 라면 5 번째, 15 번째, 25 번째, 35 번째 등의 순서로 抽出한다. 이 方法에 의해서는 각각의 單位의 組合으로부터 抽出될 機會 - 말하자면 명부상의 첫 번째와 두 번째 單位가 같이 標本으로 抽出되는 것은 不可能하다 - 는 單純任意 抽出方法으로써는 可能하나 이 方法에 의하면 전혀 주어지지 않는 경우도 있다.

이러한 사실은 한편으로는 장점이 될 수도 있는데, 이 方法에 의하여 單純任意 抽出로서 얻을 수 있는 것보다 더 效果的이고 代表的인 標本을 마련해 줄수도 있다. 이것은 어떤 적당한 순서의 배열에 의하여 이루어진다. 예컨대 어느 마을에서 사람을

標本으로 抽出하고자 한다면, 이러한 사람들을 연령순으로 명부를 작성한다. 이에 대한 계통추출은 標本에 있어서 연령을 골고루 拈散시키는 것을 보장한다. 單純任意抽出에는 항상 存在하는 可能性· 즉, 실제모집단의 상태보다도 나이가 많은 사람의 比率이 더 크게 되는 可能性이 우연히 일어나게 되는 것을 排除한다. 그러므로 作成된 명부가 만족하게 배열이 된다면 우리는 더 代表的인 標本을 抽出할 수 있다는 것을 확신할 수 있다. 그러나 그 명부는 일정하게 반복되는 형태를 취하면 안된다. 예를 들면 어떠한 이유로 인하여 男女가 교대로 명부에 記載되었다면, 10개의 標本中에서 하나는 모두 남자들만을 혹은 모두 여자들만이 뽑히게 될 것이며 그것은 명백히 全體母集團을 잘못 代表하게 된다. 이러한 명부를 계통추출에 使用하려면 일정하게 반복되는 형태를 고쳐야만 한다. 매우 效果的인 것으로서 계통추출의 또 하나는 均衡된 系統抽出(Balanced systematic sampling)이 있다. 이 方法은 母集團을 적당한 순서에 따라 배열하고 그 명부의 처음과 끝으로부터 같은 간격으로 抽出하는 方法이다. 예를 들면 100개의 調査單位가 있는 명부에서 任意의 方法으로 첫 번째가 뽑혔다면 역시 끝에서 첫 번째인 100 번째의 單位를 뽑고 만약 12 번째가 뽑혔다면 끝으로부터 12 번째인 89 번째 單位를 抽出하여 均衡을 이루게 되는 것이다.

一般的으로 系統抽出은 大部分의 경우 偏倚가 없고, 理解하고 実行하는 것이 용이하며, 效果的인 標本을 抽出케 해주는 좋은 方法이라 할 수 있다. 이 방법은 뒤에 說明될 方法과 結合되어

실제로 広範圍하게 使用되는 方法이다.

### 5.4.3 層化抽出

우리는 母集團의 變異性を 줄이기 위하여 層化를 使用하는데 이 方法으로 標本の 크기를 늘리지 않고도 標本誤差를 줄일수 있다. 가끔 母集團은 자기 独立的으로 標本을 抽出할수 있는 副母集團의 系列로 이루어진 것으로 생각되고 있으며 이 각부모집단을 層이라 한다. 우리의 目的은 各層을 層内에서 可能的한 한 同質的으로 만드는데 있으며, 또한 뽑히는 調査單位가 어떤것이든지 간에 各層의 다른 調査單位들을 代表할 수 있게 하는데 있다. 예를 들자면 어느 學校에서 모든 學生들의 平均身長을 알고 싶다고 가정해 보자. 우리는 平均身長이 다른 무엇보다도 연령과 性에 關係된다는 것을 알고 있다. 그러므로 우리들은 學生으로 構成된 모집단을 각각 다른 부모집단으로 나누거나 혹은 연령별 그룹이나 혹은 남성이나 여성으로 된 자기 다른層으로 나눌수 있다. 그리고는 各層으로부터 單純任意抽出에 의하여 標本을 抽出할 수 있다. 만약 各層이 注意깊게 抽出이 되었다면 層化抽出된 100 名의 學生들은 같은 크기의 單純任意抽出에 의한 것보다도 더 正確할 것이다. 즉 그것은 期待되는 標本誤差가 적다는 것이다.

우리가 層化를 하는 目的은 層内에서의 各調査單位가 可能的한 그층의 다른 調査單位와 性質이 갖게 하는데 있으며, 그렇게 해서 어떤 任意의 숫자를 抽出하든지 간에 그層으로부터 抽出된



標本이 나머지 다른것들과 매우 유사하도록 하는데 있다. 그러나 서로 다른 층끼리는 서로 차이가 크게할수 있는데 그것이 바로 長点이기도 하다. 이것은 벼를 재배하는 농장과 채소를 재배하는 농장으로 構成되는 國家의 경우처럼 매우 單純化된 예로서 說明될수 있다. 만약 이것을, 한개의 層은 모두 벼농사를 하는 농장과 다른 하나는 모두 채소를 재배하는 농장등 2個의 層으로 나눌수 있다면, 두개의 층은 서로 별개의 층이 될것이나 各層內에서의 各 調査單位는 서로 유사하게 될것이다. 그리고는 各層에서 各各 10%의 標本을 抽出한다면 各各 다른형태의 농장을 正確한 比率로 抽出할수 있다고 장담할수 있다. 그러나 만약 層化抽出에 의하지 않고 單純任意抽出에 의하여 10%의 標本을 抽出한다면 우연히도 벼를 재배하는 농장을 더 높은 比率(혹은 더 낮은 比率)로 抽出하게 될수 있다. 이렇게 層化는 標本으로 하여금 代表性을 높일 수 있다. 統計學者들은 이것을 “層別로는 더큰 變異性을” 그리고 “層內에서는 낮은 變異性을” 이라고 하는데, 이것이 標本設計에 있어서 가장 중요한 特徵의 하나이다.

層化抽出을 使用하는 또하나의 경우로서는 각기 다른 副母集團으로부터의 結果가 要求되는 때이다. 우리는 전국뿐만 아니라 各 地域別로 코코넛 농장의 總面積을 알고자 한다. 이때에는 全國을 地域으로 層化해서 各 地域으로부터 別個의 標本을 抽出해야만 한다.

#### 5.4.4 集落抽出

集落抽出은 各 個別的인 標本抽出單位 대신에 標本으로서 調査單位의 그룹이나 集落을 抽出하는 것인데 集落到 속해 있는 모든 調査單位를 包含한다. 集落抽出은 一般的으로 地域을 基準으로 適用되기 때문에 地域抽出로도 알려져 있다. 集落抽出의 예로서 人口調査를 위하여 郡部地域의 가구를 標本으로 抽出한다고 가정해 보자. 單純任意抽出로서 각 가구를 抽出하는 대신에 먼저 마을을 抽出하고는 그 抽出된 마을내의 모든 가구를 調査하면 된다. 層化抽出과는 대조적으로 集落抽出의 目的은 가능한한 集落内の 調査單位들을 서로 異質的이게 하여 各 集落을 가능한한 서로 同質的으로 하는데 있다. 이러한 形態의 標本은 보통 같은 크기의 單純任意抽出보다는 期待되는 標本誤差가 더 크다는 短點을 갖고 있다. 그러나 標本抽出單位가 集落으로 모여 있기 때문에 各 調査單位間의 조사시간이 단축되고 또한 調査費用이 현저하게 節減된다는 長點을 갖고 있다. 다른 또 하나의 長點은 單純任意抽出에 必要한 完全한 명부(모든 가구에 대한 명부) 대신에 매우 간단한 標本들(마을에 대한 명부)을 必要로 한다는 것이다.

#### 5.4.5 二段抽出

二段抽出은 처음 抽出되는 것이 調査單位의 그룹이나 集落이라는 點에서 集落抽出과 매우 유사하다. 그러나 集落抽出과는 달리 二次抽出이 第一次抽出에서 抽出된 그룹내에서 이루어

진다. 예로서 앞의 예를 사용한다면 먼저와 마찬가지로 우선 마을을 抽出한 다음 抽出된 마을내의 모든 가구를 標本으로 하는 대신에 抽出된 마을내에서 다시 2 차로 가구를 抽出하는 것이다. 마참가지 方法으로 第一次抽出로 먼저 区域을 抽出하고 第二次抽出로 다시 그 区域内에서 소매상을 抽出하게 될 것이다.

集落抽出과 같이 二段抽出은 全體母集團을 위한 標本틀로서 第一次抽出에 쓰일 그룹이나 集落到 관한 것이라는 장점을 지니고 있다. 전에 地域틀(Area frame)을 討論했을때 이점을 강조하였는데, 第一次 抽出에서 마을을 抽出함으로써 어떻게 標本地域이 구해지며, 第二次抽出에서 抽出된 마을로부터 어떻게 농장이 정해진다는 것을 알게 되었다. 그러나 이것이 二段抽出의 標本設計를 使用하는 단하나의 이유만은 아니다. 비록 우리가 전체농장의 적절한 명부틀(List frame)을 가지고 있다 할지라도 아마도 二段抽出(혹은 集落抽出)을 사용하기를 좋아할지도 모르는데, 그 이유는 全國(혹은 어느도)의 모든 농가를 單純任意抽出에 의하여 全國에 広範圍하게 散在해 있는 농장으로부터 資料를 蒐集한다는 것은 매우 많은 費用이 소요되기 때문이다. 統計人は 항상 資料蒐集에 소요되는 費用과 必要한 資料의 正確度を 유지하기 위하여 費用을 최소로하는 과정을 채택하여야 한다는 것을 念頭에 두는것이 重要하다.

#### 5.4.6 多段抽出

이것은 二段抽出의 연장으로서 다음단계의 標本抽出은 항상 바로 그 전 단계에서 抽出된 그룹내에서 이루어지는 것이다.

에컨대, 처음에 道 또는 地域을 抽出하고 第二次抽出에서는 抽出된 道나 地域에서 마을을 抽出하고, 第三次抽出에서는 抽出된 마을에서 다시 家口를 抽出한다. 大部分의 地域이 広範圍한 國家에서 全國的인 家口調査는 多段抽出로 行하여 진다. 위에서 본 二段抽出의 경우와 마찬가지로 第二次抽出單位나 마지막단계의 抽出單位에 대한 完全한 명부를 必要로 하지 않는다는 것이 長点이다.

#### 5.4.7 크기에 比例한 確率抽出 (Sampling with probability proportional to size)

우리과정에서는 매우 複雜한 標本設計에 대하여는 言及하지 않겠으나 크기에 比例한 確率抽出 (Sampling with probability proportional to size, 혹은 P.P.S. Sampling)의 한가지 特別한 기법에 대하여 言及하겠는데, 이 方法은 매우 広範圍하게 使用되고 있으며 特히 多段抽出의 경우 第一次抽出에 많이 使用되고 있기 때문이다. P.P.S 標本抽出은 각 抽出單位에 抽出될수 있는 均等한 機會를 주는 대신에 큰 抽出單位에는 적은 抽出單位보다 더 많은 機會를 부여하는 절차를 適用한다.

P.P.S. 標本抽出을 適用하기 위하여 亂수표를 사용하는 예로서, 이 기교에 대한 論評이 본장의 조금 뒤에서 다루어질 것이다. 이 책에서는 이 問題에 관하여는 더 이상의 상세한 說明을 피하려고 한다. 그것은 理論이 매우 複雜하기 때문이다. 그러나 P.P.S. 抽出方法은 全體母集團에 걸쳐 골고루 標本을 抽出하는데 實際적으로 淸快할만한 長点을 지니고 있다는 것을 간과해서는 아니된다.

## 5.5 標本誤差

이미 본바와 같이 標本誤差란 우리가 단지 標本을 抽出함으로써 치루어야만 되는 代價로 생각될 수 있다. 그것은 바로 母集團의 참값과 標本으로부터 얻어진 推定値와의 差異이다. 우리가 標本을 어떻게 抽出하든지간에 誤差는 發生하는데 단지 任意抽出法을 使用한다면 誤差의 期待되는 크기를 계산할 수 있다는 것 뿐이다.

標本을 抽出하기 위하여 任意抽出 方法을 使用하기 때문에 可能的 모든 標本中에서 어떤 標本 하나를 抽出하는 機會를 알 수 있다. 또한 標本の 抽出과 推定에 있어서 偏倚가 없는 方法이 있다면 여러가지 標本으로부터의 推定値의 平均은 母集團의 참값과 같게 된다는 것을 우리는 알고 있다.

우리가 할 수 있는 것은 모든 可能的 標本에 있어서 標本推定値가 어떻게 分布되어 있는지를 計算하는 것이다. 만약 標本推定値가 母集團의 참값과 매우 근사하다면 標本抽出計劃이 잘 되어 있고 標本誤差가 적다고 期待할 수 있다. 그러나 만약 각각 다른 標本으로부터의 推定値가 참값으로부터 넓게 分布되어 있다면 標本誤差는 매우 크게 된다. 같은 標本을 반복추출함으로써 얻은 推定値의 變異性의 分布가 正確性의 척도이거나 혹은 標本抽出過程의 正確性의 尺度를 나타낸다. 確率理論을 使用함으로써 계산할 수 있다는 것은 이러한 分布이다. 第10章에서 이 標本誤差를 계산하는 公式에 관하여 간단히 取扱하게 될 것이며, 推定値의 信賴性을 評價하기 위하여 이것이 어떤 의미를 갖고 있는

가를 다루게 될 것이다.

標本抽出의 正確性은 다음과 같은 事實에 依存한다.

- (a) 標本이 抽出되는 方法
- (b) 標本の 크기
- (c) 母集團의 變異性

매우 變異性이 큰 母集團에 대하여 正確性을 높이기 위하여는 매우 規模가 큰 標本이 必要하게 된다. 그러나 (이미 본바와 같이) 副母集團의 變異性을 감소시키는데 도움이 되는 層化抽出과 같은 方法을 使用하여 各 副母集團으로 부터 獨立된 標本을 抽出할수 있다.

#### 5.6 標本調査에 있어서 實際적으로 特別히 考慮되어야 할 事項

본장에서 標本抽出의 基本概念에 대하여 討論해 왔다. 그러나 더욱 중요한 몇가지 要因들을 標本設計에 있어서나 그것의 信賴性을 評價하는데에 考慮해야만 한다. 그러나 地方事務所나 立선조사원의 立場에서는 實際적인 問題로서 標本을 抽出하고, 標本을 維持하고, 資料를 蒐集하는 등등의 標本設計에 대해 크게 염려할 必要는 없다. 후속절차중 大部分이 各 調査에 따라 特別히 지시를 받게될 것이나, 자주 접하게 될 몇가지 問題에 대하여 言及 하겠다.

##### 5.6.1 亂수표의 使用

統計業務에 있어서 亂수표는 매우 重要한 도구이다. 亂수표의 使用은 標本の 抽出에 있어서 客觀性을 보장하고 個人

判斷으로 부터 發生하는 偏倚를 排除한다. 난수표는 여러가지 방법으로 만들수 있으나 오늘날에는 컴퓨터를 사용하는 것이 一般的이다. 어떠한 방법으로 만들어졌든지 간에, 만들어진 숫자들이 실제로 任意的인지를 確認해야 하는데 이것은 각 숫자가 正確하게 同一한 抽出의 機會를 갖고 있는지, 뿐만 아니라, 非任意的인 순서가 피해졌는지를 確認해야 된다는 것을 意味한다. 앞선에서 使用하기 위하여 準備된 난수표는 만족스럽다고 確信할수 있어야 한다.

그러나 비록 숫자들이 自體的으로 任意的이라 하더라도 그것들은 正確하게 미리 정하여진 規定에 따라 使用되어야 하는데 그것은 숫자들의 非任意的인 使用은 非任意的인 抽出로 結果되기 때문이다. 500 個의 숫자로 된 난수표(25 個 行과 20 個의 列로 構成된)가 아래에 마련되었는데, 이 난수표를 參考로 하여 난수표의 올바른 使用에 대하여 예를 들어보겠다.

몇개의 마을을 包含한 몇개의 層으로 어느 國家의 地方을 나누어, 二段抽出의 第一段階 抽出로서 各層으로부터 한개의 마을을 抽出한다고 가정해 보자. 만약 첫번째 層에 19 個 마을이 있다면, 마을들을 1 번부터 19 번까지 일련 번호를 부여하여 抽出될 마을을 결정하기 위하여 01 부터 19 까지의 숫자중에서 任意的 숫자를 한개 選擇한다. 이것은 2 자리 숫자이기 때문에 우리는 난수표의 앞의 2 행을 使用하여 그 숫자들을 찾아내려다가 01 과 19 사이의 숫자를 選擇한다. (任意的으로 결정된 난수표의 2 자리 숫자를 어떤 행의 위에서 아래로 찾거나 혹은 줄의 왼쪽에서

|    | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | (行) |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|
| 1  | 5 | 0 | 4 | 4 | 6 | 6 | 4 | 4 | 2 | 1  | 6  | 6  | 0  | 6  | 5  | 8  | 0  | 5  | 6  | 2  | 6  | 8  | 1  | 5  | 5  |     |
| 2  | 3 | 2 | 6 | 6 | 2 | 2 | 1 | 5 | 8 | 6  | 2  | 6  | 6  | 3  | 7  | 5  | 4  | 1  | 9  | 9  | 5  | 8  | 4  | 2  | 3  |     |
| 3  | 9 | 6 | 2 | 4 | 4 | 0 | 1 | 4 | 5 | 1  | 2  | 3  | 2  | 2  | 3  | 0  | 8  | 8  | 5  | 7  | 9  | 5  | 6  | 7  | 4  |     |
| 4  | 3 | 1 | 7 | 3 | 9 | 1 | 6 | 1 | 1 | 9  | 6  | 0  | 2  | 0  | 7  | 2  | 9  | 3  | 4  | 8  | 9  | 8  | 5  | 7  | 0  |     |
| 5  | 7 | 8 | 6 | 0 | 7 | 3 | 9 | 9 | 8 | 4  | 4  | 3  | 8  | 9  | 9  | 4  | 3  | 6  | 4  | 5  | 5  | 6  | 6  | 9  | 4  |     |
| 6  | 8 | 4 | 3 | 7 | 9 | 0 | 6 | 1 | 5 | 6  | 7  | 0  | 1  | 0  | 1  | 0  | 2  | 3  | 9  | 8  | 0  | 5  | 8  | 5  | 1  |     |
| 7  | 3 | 6 | 6 | 7 | 1 | 0 | 0 | 8 | 2 | 3  | 9  | 8  | 9  | 3  | 3  | 5  | 0  | 8  | 8  | 6  | 9  | 9  | 2  | 9  | 7  |     |
| 8  | 0 | 7 | 2 | 8 | 5 | 9 | 0 | 7 | 4 | 8  | 8  | 9  | 6  | 4  | 5  | 8  | 8  | 9  | 7  | 5  | 8  | 3  | 8  | 5  | 6  |     |
| 9  | 1 | 0 | 1 | 5 | 8 | 3 | 8 | 7 | 6 | 0  | 7  | 9  | 2  | 4  | 3  | 1  | 6  | 6  | 5  | 6  | 2  | 1  | 4  | 8  | 2  |     |
| 10 | 5 | 5 | 1 | 9 | 6 | 8 | 9 | 7 | 6 | 5  | 0  | 3  | 7  | 3  | 5  | 2  | 1  | 6  | 5  | 6  | 0  | 0  | 5  | 3  | 5  |     |
| 11 | 5 | 3 | 8 | 1 | 2 | 9 | 1 | 3 | 3 | 9  | 3  | 5  | 0  | 1  | 2  | 0  | 7  | 1  | 3  | 4  | 6  | 2  | 3  | 3  | 7  |     |
| 12 | 5 | 1 | 8 | 6 | 3 | 2 | 6 | 8 | 9 | 2  | 3  | 3  | 9  | 8  | 7  | 4  | 6  | 6  | 9  | 9  | 4  | 0  | 1  | 4  | 7  |     |
| 13 | 3 | 5 | 9 | 1 | 7 | 0 | 2 | 9 | 1 | 3  | 8  | 0  | 0  | 3  | 5  | 4  | 0  | 7  | 2  | 7  | 9  | 6  | 9  | 4  | 7  |     |
| 14 | 3 | 7 | 7 | 1 | 6 | 7 | 9 | 5 | 1 | 3  | 2  | 0  | 0  | 2  | 4  | 4  | 9  | 5  | 9  | 4  | 6  | 4  | 8  | 5  | 0  |     |
| 15 | 9 | 3 | 6 | 6 | 1 | 3 | 8 | 3 | 2 | 7  | 9  | 2  | 7  | 9  | 6  | 4  | 6  | 4  | 7  | 8  | 2  | 8  | 5  | 4  | 9  |     |
| 16 | 0 | 2 | 9 | 6 | 0 | 8 | 4 | 5 | 6 | 5  | 1  | 3  | 0  | 5  | 0  | 0  | 4  | 1  | 8  | 4  | 9  | 3  | 0  | 7  | 5  |     |
| 17 | 4 | 9 | 8 | 3 | 4 | 3 | 4 | 8 | 3 | 5  | 3  | 2  | 8  | 8  | 3  | 3  | 6  | 9  | 9  | 6  | 7  | 2  | 3  | 6  | 0  |     |
| 18 | 8 | 4 | 6 | 0 | 7 | 1 | 6 | 2 | 4 | 6  | 4  | 0  | 8  | 0  | 8  | 1  | 3  | 0  | 3  | 7  | 3  | 4  | 3  | 9  | 2  |     |
| 19 | 1 | 8 | 1 | 7 | 3 | 0 | 8 | 8 | 7 | 1  | 4  | 4  | 9  | 1  | 1  | 4  | 8  | 8  | 4  | 7  | 8  | 9  | 2  | 3  | 3  |     |
| 20 | 7 | 9 | 6 | 9 | 1 | 0 | 6 | 1 | 7 | 8  | 7  | 1  | 3  | 2  | 7  | 6  | 9  | 5  | 6  | 2  | 8  | 7  | 0  | 0  | 2  |     |

(列)



시작하여 오른쪽으로 찾아가든 어떤 方法이든지 무방하다. 실제로는 調査員이 어디서 시작해야 될지, 이를테면 5 번째 줄과 9 번째 행에서와 같이 特別한 指示가 주어진다) 만약 처음의 두개 행의 위에서 부터 시작한다면 50, 32, 96, 31, 78, 84, 36 등은 크기 때문에 차례로 제외되어야 하고, 우리의 명부에서 정해진 範圍에 해당하는 처음의 두자리 숫자인 07 을 抽出한다. 그리고는 07 의 숫자밑에 줄을 치거나 혹은 表示를 한다(혹은 만약 난수표를 表示없이 保存해야 된다면 따로 抽出된 숫자를 記錄한다), 7 번째의 마을이 첫번째 層에서 그 層을 代表하는 任意의 方法으로 抽出된 마을이다.

명부상의 07 까지의 모든 숫자들은 이제는 使用된 것으로 간주되어야 하고 다음 抽出은 그 행의 다음 숫자로부터 시작되어야 한다. 그 행의 처음부터 다시 시작하여서는 안되며(혹은 처음에 시작할데서 부터 다시 시작해서는 안되며), 이미 使用된 숫자들에 대해 다시 抽出機會를 부여해서는 안된다. 예를 들면, 만약 두째 층에 34 個 마을이 있다면, 위로부터 다시 시작하여 50 은 크기때문에 제외하고 抽出番號로서 32 를 취하는 것은 매우 잘못된 方法이다. 올바른 절차는 07 바로 직후부터 숫자를 점검하여 다음 숫자는 10 으로서 이것은 1 부터 34 사이의 範圍에 속하는 것으로서, 우리가 요구하는 範圍內이기 때문에 두번째 층을 代表하는 마을로 10 번 마을을 抽出하고 10 의 숫자에 表示를 하거나 따로 記錄을 해둔다. 다시 되돌아가서 앞의 숫자들을 使用하는 것이 어찌서 잘못된가 하고 의문을 提起할지 모르나, 그

것은 浪費的인 절차로서 必要이상으로 빠르게 난수표를 使用해 버리기 때문인데 조금만 생각해보면 그 이유를 알게 될 것이다. 처음의 7개 숫자들은 크기 때문에 제외된 것들이다. 즉 19보다 크기 때문이다. 그러므로 만약 우리가 이러한 숫자들을 다시 찾는다면 20부터 34까지의 숫자는 매번 抽出機會를 갖게 되나, 1부터 19까지의 숫자는 전혀 抽出의 機會를 갖지 못한다.

이것은 偏倚가 있는 方法으로 높은 숫자가 뽑히게 될 경향이 있으나(실제로 32가 抽出되는 것처럼), 반면에 최후로 抽出된 번호의 다음부터 시작한다면 01부터 34까지의 숫자는 同一한 抽出機會를 갖게되는 偏倚가 생기지 않는 抽出이 된다.

이러한 方法은 큰 問題가 되지 않는 것으로 생각되고 있다. 만약 마을들이 任意의 순서대로(이를테면 가나순으로) 記載되어 있다면, 그것은 그다지 중요한 問題가 되지 않는다. 그러나 만약 그 마을들이 어떤 體系的인 方法으로 作成되었다면 그 方法은 커다란 差異를 나타내게 될 것이다. 예를들면 편리하게 어떤 都市에 인접된 마을들을 먼저, 그리고 멀리 떨어진 마을들은 뒤에 羅列했다고 가정해 보자. 앞에서와 같은 난수표의 不正確한 使用은 中央에 가까운 마을보다는 먼 마을이 抽出될 機會가 더 많게 되어, 그러한 먼 마을에 對하여 調査된 特性에 差異가 있게 되기 때문에 事實上 그러한 抽出은 代表性을 상실하여 그 結果는 심각한 偏倚(Bias)를 招來하게 될 것이다. 난수표를 使用하는데 있어서는 정해진 規則을 따르는 것이 매우 중요하다.

이미 본바와 같이, 꽤 많은 숫자들을 제외하는 方法으로 난수

표를 使用하는 것은 非效果的으로서 그 숫자들은 “浪費” 되었다. 그러한 結果를 피하기 위하여는 節次가 좀더 複雜해진다. (예컨대 01 부터 19 까지의 숫자를 抽出하기 위하여 21 부터 40 까지의 숫자에서는 20 을 控除하고, 41 부터 60 까지의 숫자에 대하여는 40 을 控除하는 등의 方法을 使用할수 있다. 우리의 난수표에서는 처음숫자가 50 이므로 10 번 마을을 抽出하게 된다. 20, 40, 60, 80, 00 의 숫자를 除外하고는 모든 숫자가 마을을 抽出하게 되며 棄却되는 숫자의 수가 매우 적게 된다.)

오늘날에는 매우 큰 規模의 난수표를 만들어 낼수있는 컴퓨터를 使用하게 되어 난수표가 使用하는데 不足되는 일은 결코 없다. 節次는 가능한한 간단해야 할것이 매우 強하게 要求되며, 난수표의 浪費를 없애기 위하여 이러한 特別한 手段이 日常적으로 使用되어서는 안된다.

이러한 節次들이 얼마나 단순하게 만들어 졌든지간에 統計一線에서 從事하는 사람들에게 난수표를 使用하는 올바른 指針이 주어지지 않는다면 그들은 여전히 혼동을 일으키게 되고 또한 不確實해지게 된다. 예를들면 앞의 예에서 첫번째 행을 다 使用해 내려갔다면 그 다음은 어디서 시작해야 될것인가? 아마도 가장 명백한 答은 3 째와 4 째행의 위에서(다음 숫자들은 44, 66 등) 시작될 것이나 마찬가지로 2 째와 3 째행의 04, 26, 62 등을 使用하더라도 任意的이고 덜 浪費的이 될것이다. 난수표는 수직대신에 수평 - 예컨대 50, 44, 66 등 - 으로도 使用할수 있는데 역시 長점이 있다. 어떠한 方法으로 使用하든지간에 關係는

없으나, 單純하고 任意的이어야 하며 혼동을 피하기 위하여 명백하고 適切한 成文化된 指針을 選定해야만 한다.

P.P.S 抽出을 위하여 난수표를 使用하는 앞의 예로 돌아가자. 어느 군에서 多段抽出로 5 個의 마을중에서 하나를 抽出하고 다음단계로 抽出된 마을에서 다시 어떤 個人들을 抽出한다고 가정해보자. 抽出方法은 이미 앞에서 言及되었으므로 우리는 단지 마을을 抽出하기 위하여 1부터 5까지의 任意的 숫자중에서 하나를 選択하기로 한다. P.P.S 抽出에 있어서는 각마을의 人口 (혹은 調査單位로서 使用하고자 하는 單位에 따라서 家口數等) 를 알아야만 하는데 다음과 같이 마을을 抽出하여야 될 것이다.

| 마을 | 人口  | 累積人口  | 抽出範圍          |
|----|-----|-------|---------------|
| 가  | 226 | 226   | 1 ~ 226       |
| 나  | 705 | 931   | 227 ~ 931     |
| 다  | 339 | 1,270 | 932 ~ 1,270   |
| 라  | 104 | 1,374 | 1,271 ~ 1,374 |
| 마  | 295 | 1,669 | 1,375 ~ 1,669 |

다섯 마을의 總人口는 1,669 이므로 우리는 0001부터 1,669 사이의 任意的 數를 選択한다. 001부터 226 사이의 어떤 숫자는 “가” 마을을 抽出하고 위의 抽出範圍欄에 해당하는 숫자에 따라서 각마을을 抽出하게 된다. 이 方法에 의하면 가장 작은 “마” 마을보다는 가장 큰 마을인 “나” 마을을 抽出하는 것이 有力해질 것이다. 왜냐하면 “나” 마을을 抽出하게 될 任意的 숫

자가 705 個나 되나 ( 226 부터 931 까지 ), “ 마 ” 마을을 뽑게 될 任意의 숫자는 104 個 ( 1,271 부터 1,374 까지 ) 뿐이기 때문이다. 실제로 抽出될 機會는 正確하게 마을의 크기에 比例하게 된다. 만약 우리의 난수표에서 처음의 4 자리 숫자를 찾는다면 ( 왜냐하면 1,669 는 4 자리수이기 때문에 ), 5044, 3266 등은 크기 때문에 제외되고 마침내 “ 나 ” 마을에 해당되는 숫자 0728 에 도달하게 될 것이다.

비록 지금의 예에 있어서는 마을을 抽出하는 確率이 同一하지는 않지만 다음단계에서는 마을에서 각 個人을 抽出할때 그 確率이 均等하도록 할수 있다.

#### 5.6.2 標本의 交替

勞動力調査와 같은 繼續調査에 있어서는 每 調査때마다 標本으로서 각 個人을 계속해서 調査하는 것이 상당한 長點이 있다. 일단 標本이 確定되면 抽出된 單位를 交替하지 않는 것이 費用이 적게 所要되고, 또한 応答者들이 調査員과 質問内容에 親密해지기 때문에 調査가 迅速해지고 쉽게 될수 있다. 더우기 같은 사람들이 調査때마다 同一하게 維持된다면 그 사람들에게 發生한 變化 ( 고용과 같은 ) 는 실제적인 變化로 나타난다, 그것은 우연히 그들에게 고용특성이 달라진 것은 아니다. 이러한 사실은 調査에 있어서 推定値의 傾向을 잘 나타낸다. 그러나 標本으로서 応答者를 오랫동안 維持하는데 있어서도 위험이 따르게 된다. 그들은 피조사자로 선정되었기 때문에 - 즉 그들은 調査의 応答者로서 어떤 特別한 行動을 해야 되기 때문에 - 그들이

提供하는 情報을 약간 변형시키기도 하여 더 이상 母集團을 대표할 수 없게 되는 경우도 있다. 혹은 상당히 오랫동안 調査에 応하게 되어 마침내는 応答를 거부하거나, 不正確하게 对答하는 경우도 있게 된다. 이것을 “ 応答者 기피 ” (Respondent fatigue) 라고 한다. 어떻든 어떤 사람들에게 계속적으로 調査에 応하도록 期待하거나, 처음부터 調査에서 除外시켜 장래에도 그렇게 되도록 하는 것은 公평하지 못하다. 그러므로 이러한 要因들 때문에 한편으로는 応答者를 계속 維持하려고 하고 다른 한편으로는 応答者를 交替하려고 한다.

어떤 調査에 있어서 어떤 個人 또는 家口가 包含되는 回數를 제한함으로써 위의 要素들을 均衡이 되게 할 수 있다. 만약 이 回數를 10 회로 制限한다면 各 家口는 연속하여 10 회동안 標本으로 調査되고는 交替된다. 家口를 交替하는데 所要되는 費用과 時間은 그렇게 해서 節減되고 期間에 따라 分散되며 大部分의 家口들이 어떤 調査期間에서 다음 調査로 이어진다. 단지 每調査마다 10 分の 1 의 家口가 交替되거나 交替되어 除外된다. 이 方法이 10 회동안 모든 家口를 標本으로 維持하고 一時에 모두를 交替하는 것보다 좋은 方法이다.

이 선정된 單位의 점진적인 交替가 바로 標本の 交替라 불리며 定期的인 期間으로 반복되는 調査에 널리 使用되는 有益한 기교이다. 交替될 標本的一部分은 各 國家의 必要와 條件에 따라 다르지만 一般的으로 標本の  $1/8$  이나  $1/10$  정도이다.

### 5.6.3 標本の 代替

標本이 처음에 抽出되면, 嚴格한 任意的 節次를 使用하여 母集團의 各 單位가 미리 定해진 抽出機會를 갖도록 한다. 바로 이렇게 선정된 單位로 부터 応答를 얻고자하며 이에 偏倚(Bias)가 생기지 않도록 한다. 그러나 실제로 이러한 理想的인 條件이 갖추어지기는 어렵다. 때로는 応答를 거부하는 사람도 있고, 精確한 对答를 하지 못하는(문맹때문에) 応答者도 있게 된다. 抽出된 單位로 부터 応答를 얻기 위하여 모든 노력을 하여도 応答를 얻을수 없다면 그 單位는 調査에서 除外시켜야 된다. 이러한 경우가 發生할 경우 취할수 있는 方法이 2가지 있는데 하나를 택하여야 한다.

(i) 처음에 抽出된 水準으로 標本の 규모를 유지하기 위하여 無応答者를 代替하는 方法

(ii) 처음에 抽出된 水準以下로 되더라도 抽出率을 變更시키고 無応答者를 代替하지 않는 方法. 이 경우 抽出率에 따라 使用된 推定方法을 變更시켜야만 한다.

어떠한 方法이 使用되든간에 偏倚(Bias)가 없어지지 않는 다. 어떻게 代替되든 代替된 것은 원래 抽出되었던 無応答者와 는 特性이 다르기 마련이다. 예컨대 만일 無応答이 応答者의 문맹으로 인한 것이고 이런 사람들의 所得이 낮은 경향이였다면 그러한 標本을 除外시키면 偏倚(Bias)가 생기게 되어 이러한 偏倚를 除去할 方法이 거의 없다. 이 경우에 代替할 單位를 無応答者와 可能的한 유사하도록 주의깊게 선출한다면(같은 크기

의 代替할 住宅을 選定하거나 혹은 代替될 住宅과 같은 材料를 使用한 住宅을 선정한다면) 비록 偏倚가 完全히 除去되지는 않는다 하더라도 그것을 줄일수는 있다. 만일 일선에서 調査員의 一時的인 기분에 따라 代替가 이루어진다면 그것은 더 많은 費用이 들게 되며 偏倚를 增加시키게 될 것이다. 그러므로 無応 答者가 생기는 경우 양자택일할 수 있도록 미리 標本을 抽出하여 놓거나 必要한 경우 代替하도록 監督者에게 위임하는 方法이 좋으나 이러한 方法도 역시 費用이 많이 들게 될 것이다. 代替를 할것인가, 안할것인가하는 결정은 이러한 要因의 均衡에 依存하게 될것이다.

#### 5.6.4 標本の 維持

母集團은 時間의 經過에 따라 變化되는데 母集團에 새로운 單位가 追加될 경우에 標本 역시 이러한 單位를 반영하도록 擴大해야 될 경우가 있다. 만약 標本の 單位가 住宅이라면 標本에 新축되는 住宅의 一定한 比率을 반영하는 方法을 마련하는 것이 必要하게 된다.

이러한 것을 達成하는 方法은 標本設計에 依存하게 되며 標本維持를 위한 어떤 嚴格한 規則을 따로 세울수는 없다. 그러나 이것은 매우 중요한 고려이며, 몇몇 調査에 있어서는 일선 調査活動에 있어서 가장 많은 時間을 消費하는 것중의 하나가 될것이다.



## 第 6 章 誤差의 源泉

### 6.1 概 要

우리는 이미 統計가 正確해야 되고 信賴할 수 있어야 한다는 것에 대하여 이미 몇번이나 언급하였는데 이제는 매우 상세히 誤差에 대하여 토론하고자 한다. 이것에 대하여 본장과 다음 장에서 論議하게 될 것이다. 본장에서는 統計資料의 處理過程에서 發生되는 重要한 誤差의 形態와 그것이 어떻게 발생하는가에 대하여 볼 것이나 제 7 장에서는 어떻게 오차의 발생을 防止하고 發生된 오차를 어떻게 발견하고 또 그것을 어떻게 正正할 것인가에 대하여 고찰하게 될 것이다.

통계는 日常生活의 모든 分野에 사용이 되고 그것을 기초로 하여 重要한 결정을 내려야 되기 때문에 公表된 자료는 可能한限 完全하고 正確해야 한다. 公표된 자료에 비교적 重大하지 않은 것같은 숫자의 착오가 包含되었다면 그것은 豫見하지 못한 影響을 미치게 될 수도 있다.

예를들어 人口增加率의 계산에 重大한 착오가 있었다고 가정해보자 : 즉 실제로는 연 2.2%인데 이것을 3.2%로 발표하였다고 가정해보자. 만약 그 틀린 숫자를 가령 1억인구의 10년후에 기대되는 인구증가를 계산하는데 사용하였다면 천 3백만이라는 인구가 지나치게 추정될 것이다. 차례로 이것은 실제로는 必要치 않은 特別教育施設이나 医療施設등과 같은 여러 분야에 國家로 하여금 資源을 낭비하게 만들 것이다.

정부로 하여금 공표된 통계에 오차가 전혀 없다고 확신할 수 있도록 統計人은 그 能力의 최대한도로 역할을 다하여야 한다. 資料의 蒐集, 處理, 公表등 여러 분야에서 일하고 있는 모든 사람들이 그들이 하고 있는 일로부터 可能한 限 많은 부정확성을 없애려고 노력하는 것이 그 직업의 일부분이다.

물론 공표될 자료의 모든 오차를 제거한다는 것은 不可能하다는 것을 우리는 알게 될 것이다. 만약 우리가 표본조사를 실시한다면 우리가 얻은 모집단의 추정치는 표본오차를 包含하고 있다는 것을 안다. 이것에 관하여 이미 전장에서 論議한 바 있다. 표본오차를 없앨 수 있는 오직 한가지 방법은 전수조사를 실시하는 것인데 그것은 必要한 資源이 없기 때문에 不可能한 때가 있다.

그러나 전수조사에서 조차도 보통 근절 시키기가 어려운 어떤 오차들이 있게 된다. 예컨대 인구센서스에 있어서 누락된 사람과 중복된 사람이 있기 마련이다. 비록 資源이 豊富하게 있다하더라도 이러한 일이 일어나지 않는다고 확신하는 것은 不可能하다. 우리가 할 수 있는 것은 이러한 오차-비표본오차라고 부르는-들을 될 수 있는 한 적게 하여 최종결과에 影響을 미치지 않게 하는 것 뿐이다.

그러므로 經驗이 많은 통계인은 매우 完全한 정확성을 획득한다는 것이 지극히 힘들다는 것을 認識하게 될 것이다. 그는 수집되는 情報에 대하여 거의 統制를 가할 수 없는 分野가 있고 完全한 절차를 使用하는데 대해 時間과 資源의 制約이 있게 되

는 어떤 상황도 있다는 것을 認識하게 될 것이다. 그러나 그  
는 훌륭한 設計, 注意깊은 기획, 숙련된 統計技術의 使用에 의해  
그러한 오차들의 影響을 最小化할 수 있는 方法을 찾아내게 될  
것이다. 그가 충분히 管理할 수 있는 분야에 대해서는 어떠한  
착오도 일어나지 않았고 또한 그것을 完全히 없앴다는 것을 확  
신해야 한다.

어떤 種類의 오차들은 다른 오차들보다 더 一貫的으로 최종결  
과에 影響을 미친다는 것을 우리는 알게 될것인데 統計를 좀더  
正確하고 信賴할 수 있게 하기 위하여 우리는 이러한 중요한 오  
차들에 대하여 주의를 기울여야 한다.

## 6.2 非標本誤差와 偏倚 (Bias)

우리가 비표본오차라고 말할 때에는 표본추출방법을 사용하  
는 直接的인 結果로 일어나는 이외의 모든 오차를 말한다. 그  
러므로 만약 일선조사원이 표본조사에 있어서 경작지가 실제로는  
4헥타인데 이것을 3헥타라고 기록을 하였다면 오차가 발생하였  
는데 이 오차는 조사원의 착오로 인한 결과(추출방법의 실시로  
일어난 조사결과가 아니기 때문에)이기 때문에 非標本誤差로 묘  
사된다. 標本誤差는 단지 표본추출과정에서만 발생되지만 비표본  
오차는 全數調査에서든 或은 標本調査에서든 어떤 統計調査에서도  
발생한다는 것을 명백히 이해하여야 한다.

### 6.2.1 応答誤差

비표본오차의 가장 일반적인 형태중의 하나가 바로 応

答誤差라 불리운다. 調査에 包含되는 어떤 單位에 대한 어떤 質問에는 단하나의 올바른 대답이 있는데 이것을 개별적인 참값 (ITV, Individual True Value)이라고 한다. 應答者가 이 ITV가 무엇인지 알지 못하는 경우 (예컨대 그의 나이가 몇살인지를 알지 못하는 경우)가 있을 수 있으나 그러나 거기에도 질문에 대한 正確한 답변은 하나밖에 없으며, 이것이 바로 調査에서 把握되어야 하는 것이다. 만약 조사한 것과 기록된 것에 어떤 차이가 있게 되면 이것이 바로 오차이며 이것을 개별적인 應答誤差 (IRE, Individual Response Error)라고 한다. 이러한 IRE 들은 양이거나 음을 취하는데 이것은 참값보다 높거나 또는 낮게 調査된 값이다.

물론 統計에 있어서 우리가 관심을 가지고 있는것은 個人에 대한 어떤 수치보다는 總計이며, 우리가 얻은 個人의 應答을 모았을때 어떤 결과가 나타나는지 보아야 한다. 만약 여러가지 센서스나 調査활동들이 정해진 규정에 따라 嚴格하게 행하여진다면 모든 ITV의 합계는 正確한 단하나의 값으로 나타나는데 그것을 참값 (True Value)이라 한다. 우리는 참값과 조사된 값 (調査로 부터 얻어진 결과)의 차이를 應答誤差라 한다.

또한 다른 형태로서 不完全한 調査範圍 (Incomplete coverage) 그릇된 抽出方法, 推定節次의 不適切한 適用등이 있는데 이러한 것들은 다음 부분에서 비표본오차의 원천이란 제목하에 상세히 검토될 것이다.

### 6.2.2 偏倚(Bias)

系統的으로 발생하는 오차와 그렇지 않은 오차를 구별하는 것이 매우 중요하다. 우리가 자주 접하게 되는 応答誤差의 특별한 예를 고찰해봄으로써 이러한 구별을 하고자 한다.

예 1 : 応答者가 사실대로 35세라고 応答을 했으나 조사원이 부주의로 25세라고 기재한 경우

예 2 : 応答者가 실제로는 일주일 동안 麥酒를 10병을 마시는데 調査에서 6병만을 마신다고 応答한 경우

이 두가지 誤差가 暗示하는바를 考察해 보자. 만약 調査員이 부주의로 応答을 잘못 기록하였다면 応答誤差가 발생한다. 이러한 誤差는 매우 계통적으로 發生하지는 않는다. 때로는 調査員이 나이를 적게 記載할 것이며 때로는 많게도 記載할 것이다. 各 調査員에 있어서나 모든 標本에 있어서 어떤 方向의 오차는 반대 방향의 다른 오차와 크기가 같게 될 것이다. 그러한 誤差들은 全体的으로 상쇄되어 순수한 오차(예컨대 平均年齡에 있어서의)는 매우 작게 될 것이다. 그러나 두번째의 예처럼 그렇게 낮게 대답하는 경우는 分離된 誤差가 아니라 一般的인 傾向의 예이다. 맥주소비에 대하여 大部分의 応答者들이 낮게 대답해 줄 것이다. 왜냐하면 사람들은 그들이 얼마나 많이 마시는지 밝히기를 꺼려하기 때문이다. (이것은 応答時에 부인이 동석해 있다면 매우 쉽게 나타나게 될 것이다. 그것은 자기 부인으로 하여금 자기의 주량을 알게 하고 싶지 않기 때문이다. 또한 조사원이 부인에게 그러한 質問을 한다면 그 부인은 남편의 正確

한 주량을 알지 못하기 때문에도 이런 현상이 쉽게 일어날 것이다) 이때에는 첫번째 예와같은 誤差의 상쇄경향을 기대하는것은 어렵다. 이때 우리는 系統的인 誤差가 발생하였다고 하거나 応答의 偏倚(Response Bias)가 있다고 하는데 이때 각 個人의 값을 合計하면 참값보다 調査된 값이 낮게 될 것으로 期待하게 된다.

이와 마찬가지로 만약 統計가 租稅当局에서 使用하여 각 개인이 세금을 얼마나 支払하여야 될지를 評價할 수 있는 경우에 각 個人의 所得에 관한 資料를 蒐集한다면 아마도 系統的인 誤差가 發生하게 될 것이다. 이 境遇에는 応答者들이 실제로 얻는 所得보다도 낮게 알려주어 그릇된 情報를 얻게 될 것이다. 그렇게 되는 偏倚(Bias)가 있는 全体的으로 實際보다 낮게 계산된 所得水準을 얻게 될 것이다.

이 책의 곳곳에서 統計單位에 관련된 資料에 대하여 秘密이 維持되어야 한다는 重要性을 강조했다. 앞의 예들은 그러한 것을 초월했다는 것을 보여준다: 資料는 秘密이 유지되어야 할뿐만 아니라 応答者도 또한 그렇다는 것을 信賴해야만 한다. 그렇지 않으면 系統的인 誤差가 생기게 되고 심각한 偏倚(Bias)가 있는 결과를 얻게 될 위험이 매우 크게 된다.

偏倚(Bias)의 概念은 이미 第5章에서 소개되었다. 그것은 統計學에서 매우 중요한 概念이고 偏倚가 介在된 그릇된 統計가 되지 않도록 어떠한 대가를 치루더라도 그러한 偏倚는 除去되어야만 한다. 이것은 偏倚가 없는 節次로부터 發生된 오차를 무시

해도 된다는 것과 調査員의 不注意가 重要하지 않다는 것을 意味하는 것은 아니다. 可能한 限 모든 誤差는 除去되어야 하는데 때로는 偏倚(Bias)는 存在하지 않으나 실제로는 오차가 상쇄되지 않는 경우도 있게 된다. 예컨대 만약 調査員이 어느 耕作地를 실제로는 1,000 헥타인 것을 10,000 헥타라고 잘못 記載하였다면, 이 중대한 착오는 매우 큰 것이기 때문에 실령 그가 다른 착오를 하거나 어떤 작은 偏倚가 系統的으로 存在한다고 해도 그 착오를 상쇄시키에는 불충분할 수도 있다. 모든 誤差는 重要하며 다른 모든 것을 除去하기 위하여 偏倚(Bias)있는 節次에 우리의 관심을 집중시킬 수도 없다. 한편 偏倚가 없는 節次로부터 誤差들은 서로 상쇄되는 傾向이 있으며, 偏倚가 있는 節次로부터 발생한 誤差는 서로 상쇄되지 않는다는 것과 그러므로 편의를 제거하고 확인하는데 대하여 特別한 주의가 要請된다는 것을 우리는 결코 잊어서는 아니된다. 다음에서 보게 되겠지만 거기에서는 좀더 상세하게 비표본오차가 어떻게 발생되며 대부분의 발생하는 偏倚가 그릇된 計劃이나 그릇된 組織設計에 기인한다는 것이 紹介되어 있다. 이로부터 실제로 調査를 實施하기 이전에 適當한 주의와 豫測을 함으로써 統計의 正確性이 매우 개선될 것이라는 것이 이해될 것이다.

第2章에서 指摘한 바와 같이 훌륭한 調査企劃에 대한 責任은 전적으로 中央에 있는 사람한테만 있는것이 아니다, 왜냐하면 調査의 모든 部分에 대하여 專門家로서 期待할 수가 없기 때문이다. 地方事務所의 모든 職員들도 調査体系의 設計를 改善하고 地

方事情에 밝은 知識과 經驗으로써 調査를 遂行하는데 情報의 回饋 (feedback)의 原理에 따라 중요한 役割을 할 수 있다.

(예를들면 調査票 用語의 不正確한 表現때문에 發生되는 어려움, 부적당한 時期等の 問題에 接하였을 때)

### 6.3 非標本誤差의 源泉

非標本誤差는 調査의 企劃, 實查, 資料處理等 그 어느 段階에서나 發生할 수 있다. 여기서 우리는 誤差를 3 가지로 分類할 것이다.

(a) 企劃段階에서의 誤差

(b) 資料蒐集過程의 誤差

(c) 資料處理過程의 誤差

이것은 發生하는 誤差를 分類하는데 편리하기는 하나 서로 중복되기 때문에 理想的인 方法은 되지 못한다. 예컨대 대부분의 応答者名簿가 中央에서 調査以前에 作成되기 때문에 여기서는 명부작성에서 發生하는 誤差를 (a)의 部類에 包含하였으나, 때로는 일선 調査過程에서 調査員에 의해 現實에 따라 수정되기 때문에 실제로 發生하는 誤差는 (a)와 (b) 모두에 該當된다.

#### 6.3.1 企劃段階에서의 誤差

기획 단계에서의 오차는 偏倚가 있는 道具나 節次로 分類될 수 있다. 때로는 두 部類를 分간하기가 용이하지 않지만 (예컨대 만약 어떤 結점이 있는 亂수표를 가지고 있다면 그것을 偏倚가 있는 도구라 할 수 있고, 만약 亂수표를 使用하는 指



針이 完全하지 못하였기 때문에 偏倚가 있는 추출을 하게 되었다면 이것은 偏倚가 있는 절차라고 할 수 있다), 一般的으로 이것은 企劃段階에서의 誤差를 分類하는 훌륭한 方法이다. 偏倚가 있는 節次의 예로서는 數學的으로 正確하고 任意的인 方法 대신에 個人的 判斷에 따라 標本을 추출한 경우: 應答者 명부의 不完全으로 調査範圍가 좁혀진 경우: 경사지에 있는 耕作地의 不正確한 調査(예컨대 農業統計에 있어서 경사지의 경우 수평으로 들출된 부분이 必要한데 경사를 무시한 耕作地의 길이의 測定) 등이 있다.

本章의 끝에서 農業統計調査에 있어서 발생하는 一般的인 誤差에 대한 설명이 있을 것이다. 이러한 것들의 大部分이 企劃段階에서 발생하는 것으로서 偏倚가 있는 節次이다.

偏倚가 있는 도구는 저울, 자, 사슬, 실등 測定道具와 조사표, 일선직원에 대한 指針書(실제 測定하는 경우), 標本틀등을 일컫는다. 이러한 모든 도구들은 결점이 있을 수 있으며, 그것들로부터 일어나는 誤差의 特徵은 系統的이라는 것이며, 즉 그것은 一貫性있게 너무 높거나, 일관성있게 너무 낮다는 것이다. 편의를 초래하는 設計의 착오는 다음과 같다.

(a) 그릇된 概念과 定義

부적당하거나 一致되지 않는 定義를 使用함으로써 企劃段階에서 誤差를 발생하게 되는 경우가 있다. 第2章에서 調査에 包含되어야 하는 모든 變數에 대하여 定義를 내려야 한다는 것을 論議하였다. 우리가 명백하다고 생각하고 있는 單純한 概念

조차도 실제로는 定義를 내릴 必要가 있는 것이다.

發生할 수 있는 오차의 형태는 다음과 같다.

(i) 어느 읍의 業務統計에 있어서 境界가 正確하게 定義되어 있지 않다면 邑境界內에 속하지 않는 地域까지도 包含 調査하게 될 것이며, 結果値는 매우 높거나 과장이 될 것이다.

(ii) 家口調査에 있어서 가구의 概念은 訪問客等과 같은 부류의 사람을 누락시켜 人口를 적게 말하게 될 것이다.

(iii) “生産과 蒐集”과 같은 생산의 定義는 만약 商品이 한 応答者에 의해 生産되고 다른 応答者가 이를 蒐集하였다면 계속해서 중복조사가 되게 할 것이다.

(iv) 賃金労働者에게 關聯된 生活費에 사치품을 包含시킨다면 그 지수는 실제를 正確하게 반영하지 못할 것이다.

(b) 잘못 設計된 조사표

資料蒐集에 있어서 조사표는 아마도 가장 重要한 도구일 것이다. 그것은 応答者가 要求되는 情報를 提供하는 手段이 될 뿐만 아니라 調査의 結果를 기록하는 수단이기도 하다. 그것은 教育程度, 理解關係, 社会的 背景등이 어떠한 모든 応答者가 사용하기에 適合해야 한다. 조사표의 사용은 다양한 행동, 견해, 要求되는 情報의 의미를 해석하는 手段이 되기도 한다. 그러나 조사표의 目的은 同一한 解釈을 確保하는데 있다. 그러므로 조사표는 상당한 주의를 기울여 서로 다른 여러 條件下에서 試驗을 해보아야만 한다.

조사표에 있어서 誤差가 發生하는 重要한 源泉은 바로 不正當

한 質問을 하는 말씨이며 이러한 誤差는 統計에 系統的인 誤差를 발생시킨다. 어떤 國家에서 地域에 따라 意味가 다르거나 어떤 사람의 集團에 따라 意味가 달라지는 言語의 使用은 피해야만 한다. 質問을 표현하는 순서는 論理的이어야 한다. 어떤 對답을 기대하는 유도질문은 때로는 피해야 한다.

(c) 그릇된 指示

資料를 蒐集할 責任이 부여된 調査員에 대한 지시는 조사표 자체 만큼이나 重要하다. 그 지시는 質問의 意味, 概念 資料蒐集에 使用되는 節次등을 同一하게 번역하도록 설계되었다. 지시는 自然的으로 調査마다 다르게 될 것이고, 教育程度, 訓練, 經驗等에 따라 해석이 될 것이다. 지시는 단순한 용어를 사용해야 되고 調査員이 다른 해석을 하지 않도록 可能的 限 完全해야 된다. 예컨대 第5章에서 본 바와같이 난수표 사용에 대한 그릇된 지시가 내려지면 不正確한 使用이 調査의 結果에 偏倚(Bias)를 초래하게 될 것이다.

(d) 잘못 작성된 명부 또는 표본들

또 하나의 매우 중요한 統計道具는 前章에서 이미 討議한바 있는 調査된 모집단의 명부나 登錄簿로서 센서스에서 모든 統計單位의 所在把握을 위하여 使用되거나 標本抽出을 위한 표본들로서 使用될 것이다. 명부나 틀은 여러가지 原因때문에 不充分할 것이며, 가장 一般的으로 發生하는 명부작성의 착오라고 불리는 過失은 다음과 같이 要約될 수 있다.

- (i) 包含되어야 할 單位의 누락
- (ii) 母集團에서 除外되어야 할 單位의 包含
- (iii) 이름이나 住所의 不正確한 記載
- (iv) 중복된 명부작성
- (v) 死亡, 不在, 轉出등의 單位의 包含

그러한 誤差의 예로서는

- 오래된 업무명부 - 즉, 더이상 遂行되지 않는 業務를 包含시키게 될 것이나 새로운 業務는 除外될 것이다.
- 선거인명부에 있어서 同一한 사람이 다른 이름으로 중복된 경우
- 勞動部에서 保管하고 있는 事業體 名簿가 5 인미만의 事業體는 除外되었기 때문에 不完全한 경우

등이 있다.

가끔 規模가 큰 調査에 있어서 다른 目的으로 作成된 표본들이나 명부를 使用하여야 될 경우가 있다. 이 경우에 명부를 작성하기 위하여 使用된 定義가 調査와 다른 경우도 있고, 調査된 모집단이 對象母集團과 다른 경우도 있다.

명부작성에 있어서의 대부분의 誤差가 境界線의 確定이 不明確하기 때문에 발생한다. - 예컨대 공장의 명부작성에 있어서 어떤 事業體가 일부는 製造業에 종사하고 일부는 小売業을 하는 경우에 이를 包含시킬 것인가 또는 除外시켜야 될 것인지 분명하지 못한 경우가 있다. 또 다른 重要한 要因은 모집단이 끊임없이 변하기 때문에 명부나 틀이 낙후되어 비록 그것이 처음에

는 매우 잘 作成되었다 하더라도 正確하게 維持하는 것은 매우 어렵다.

우리가 부디치게 되는 중대한 不正確性은 말할 必要도 없이 명부로부터의 누락인데, 만약 누락이 많으면 모집단 全體의 推定値는 매우 下向된 偏倚를 초래하게 될 것이다. 調査範圍의 誤差라는 用語는 對象母集團 全部를 調査하지 않으므로써 발생하는 誤差를 말하며, 잘못 작성된 명부나 다른 類似한 原因때문에 발생한다.

### 6.3.2 資料蒐集過程의 誤差

資料蒐集過程에서 발생하는 誤差의 種類와 그들이 어떻게 극복될 수 있는가에 대해 考察해 보겠다. 여기에서는 그릇된 設計로 인한 資料蒐集過程에서 발생하는 誤差에 대해서는 論議하지 않겠다. 資料蒐集에 責任있는 者에 의한 調査業務에 있어서의 착오에 대해서만 強調할 것이다.

#### (a) 調査途中 發生하는 誤差

이러한 오차는 一線調査員이나 實查指導를 하는 사람들에 의한 부적절한 調査業務에 기인하는 것이다. 오차가 발생하는 중요한 원인은 다음과 같다.

(i) 不適當한 調査技術: 조사원이 취해야 될 節次는 企劃段階에서 이미 결정되며, 嚴格한 시험조사를 거쳐야 한다. 그것은 指針書나 教育을 통하여 조사원에게 傳達된다. 만약 조사원이 要求되는 절차를 따르지 않는다면 표본의 代表性을 깨뜨리게 되거나 応答에 影響을 미치므로써 統計의 質을 低下시키게 될 것이다.

(ii) 抽出單位나 個人의 所在錯誤

제 5 장에서 본 바와 같이 統計單位의 所在把握이 調查員에게 賦與된 主要業務中の 하나이며 이것은 주위 환경에 依存하게 되어 실제 把握하는데 상당한 어려움이 따르게 된다. 一般的으로 調查員에게는 調査에 包含시켜야 할 統計單位나 個人에 관한 명부가 주어지는데, 지도에 의해 명부나 일련의 定義를 일일이 확인해야 한다. 만약 調查員이 지도를 解讀할 能力이 모자라거나 그 태도를 決定할 能力이 不足하거나, 주어진 명부나 定義를 올바르게 사용할 能力이 欠如되었다면 必然的으로 다른 單位를 選擇하여 調査를 하므로써 統計의 질에 중대한 影響을 미치게 될 것이다.

(iii) 副次單位 (Sub-unit) 名簿의 誤差

2 단추출이나 집락추출에 있어서 제 1 차추출에서 추출된 單位의 부차단위 (Sub-unit) 명부를 작성할 必要가 있다. 예컨대 어떤 마을에 사는 모든 가구의 명부를 작성하거나 어떤 特定地域의 모든 農家名簿를 작성할 必要가 있게 된다. 만약 이러한 명부작성이 제대로 되지 않으면 誤差가 발생할 수 있다. 위의 (ii)와 같이 이러한 種類의 오차는 지도를 사용하는 能力이 不足하거나 名簿作成에 組織的으로 接近하지 못한데에 기인한다.

(iv) 測定誤差

調查員들은 때때로 어떤 特定變數를 測定하여 資料를 蒐集하여야 되는 때가 있다. 이러한 測定은 때로는 무리로 몰려있는 가축을 세거나, 住宅의 면적을 測量하거나, 標本區域에서 어떤

곡식의 무게를 달거나 하는 등이 있다. 이때 두가지 오차가 발생할 수 있다. - 실제 測定이나 결과의 기록에 대한 誤差와 測定機具가 正確하지 못하여 발생하는 誤差이다.

(v) 質問이나 面接때의 誤差

잘못 기재된 質問을 떠나서 調査員이 正確하게 질문을 하지않거나, 나쁜 方法으로 面接을 행할 때 誤差나 偏倚가 발생할 수 있다. 예컨대 “당신은 지난 일주일 동안 수입있는 일을 하 셴습니까?” 라고 질문한다고 가정해보자. 만약 “당신은 지난 일주일동안 수입을 위한 아무일도 하지 않았지요, 그렇죠?” 하고 묻거나, “지난 일주일동안 일하셴습니까?”, “당신은 수입을 위해 일주일동안 내내 일하셴습니까?” 라고 물었을때 그에 대한 각각의 反應을 생각해 보자. 분명히 어떠한 特別한 資料를 얻기 위한 질문은 그것이 쓰여진 대로 질문하여야 한다. 모든 面接에 있어서 질문을 하는 사람과 應答을 하는 사람사이에는 親切하고 協力的인 關係를 갖는 것이 必要한데 만약 調査員이 거칠다거나, 통명스럽거나, 협박적이면 그런 분위기는 만들어 질수가 없다. 사람들은 非友好的이거나 친절하지 않게 대하게 되면 협력을 하지 않게 될 것이다. 그들은 應答을 거부하거나 성의없이 대답해 버리게 될 것이다.

질문이 어떤 言語로 쓰여 있는데 다른 言語로 묻는다면 오차가 발생하게 될 것이다. 번역된 言語가 元來言語의 意味대로 表現되었는지 確信하기는 쉽지 않다.

제 4 장에서 面接要領에 대한 지표를 들었는데 그러한 지표를 무

시한다면 오차가 발생하게 될 것이다. 이런 성질의 특수한 問題는 어떤 調査에서 調査員이 지난번 調査以後 變動이 없었다는 応答를 기대하는 것이나 推測하는 것과 같으며 正確한 답변을 얻는데 실패하게 된다.

(vi) 調査員의 알아볼 수 없는 기록

이것은 글씨를 잘못써서 읽을수가 없는 조사표에서 발생하는 문제이다. 숫자가 혼동을 일으킬 수 있다. - 이에 대한 예를 드는것은 매우 용이한데 6을 잘못써서 0으로, 5를 잘못 기재해서 3으로 읽는 경우등이다. 주의를 기울여 깨끗이 쓸 것이 必須的으로 요구된다.

(vii) 時間의 不足

조사원이 그의 業務를 遂行하는데 충분한 시간이 주어지지 못하는 때가 있다. 만약 그가 時間에 쫓기게 되면 모든 節次를 올바르게 이행하지 못하게 되고 착오를 일으키게 되거나, 質問에 대한 답을 모두 記錄하지 못할때도 있게 된다. 언제 발생할지도 모를 예견하지 못한 問題에 대해서도 時間이 주어져야 한다.

(viii) 無応答

이것은 單位에 대한 情報를 얻지 못하는 것을 의미한다 이것은 單位나 応答者가 부재중이거나 応答를 거절함으로써 발생한다. 無応答의 정도가 심각할 경우에는 調査의 結果에 重大한 影響을 미치거나 或은 結果를 無効化하게 될 것이다. 때로는 応答者와 無応答者와는 어느 정도 차이가 있다. 이것은 全体 모



집단을 추정하기 위하여 応答을 한 單位들로 부터 얻어진 情報을 使用하게 되면 偏倚가 있는 결과를 초래할 수 있다는 것을 의미한다. 예를들면 가구의 所得水準을 알고자 하는 調査에 있어서 所得이 많은 가구는 調査結果가 그들의 세금을 增加시킬지도 모른다는 憂慮때문에 応答을 거부하게 될지도 모른다. 그러므로 얻어진 답변을 근거로한 平均所得의 추정치는 偏倚가 있게 되어 그 水準은 매우 낮아질 것이다.

(ix) 調査員에 의한 虛偽報告

調査員은 業務를 빨리 끝내려고 조사구의 일부를 끊어내버리려는 유혹에 빠질 수도 있다. 예를들면 그는 주변의 농부들이 어떤 곡물은 심지를 얇기 때문에 어떤 농부에게는 그 곡물의 경작여부를 묻지 않기도 한다. 또는 情報를 蒐集하기 위하여 먼 旅行을 하기보다는 멀리 떨어져 있는 応答者의 답변을 조작할 유혹에 빠지기도 한다. 만약 調査員들이 그들의 義務를—양심적으로 그리고 完全하게 수행하지 않는다면 統計의 질에 상당한 影響을 미치게 될 수도 있다.

(b) 郵便調査에 있어서의 誤差

우편조사에서 발생하는 誤差도 지금 우리가 考察한 것과 유사한 것이다. 応答者가 訓練을 받은 調査員의 도움이 없이 情報를 提供하기 때문에 오차가 발생할 可能性이 높다. 誤差가 發生하는 源泉은 다음과 같다.

(i) 不正確한 住所錄

調査될 單位名簿나 일선에서의 調査管理를 위하여 必要한데

특히 郵便調査에 있어서는 住所錄이 必須的이다. 만약 住所錄이 不正確하다면 調査結果에는 誤差가 생기게 될 것이다. 이 경우 住所錄이 正確하게 作成되고, 가장 最近것으로 補完되어야 하며 우편을 보내는 각 단위의 住所가 올바르게 기록되어 있다는 것을 確信할 수 있는 것이 매우 중요하다.

(ii) 質問의 誤解

만약 질문이 명백하지 못하고 사용된 概念과 定義가 明確히 설명되지 않으면 應答者들은 質問을 오해해서 그릇된 대답을 보내줄 것이다. 우편조사에 있어서 주의해야 할 것은 應答者를 도와줄 수 있는 調査員이 그자리에 없기 때문에 應答者가 임의대로 질문을 해석해 버린다는 사실이다. 더우기 전에 論議한 바와같이 大部分의 應答者가 調査에 관심이 있어서 指針을 주의깊게 읽어보려는 노력을 하는것이 아니기 때문에 질문이 그릇 해석되었다 할지라도 놀랄일은 아니다.

(iii) 不正確한 答辯

우편조사에 있어서 질문사항이 어렵거나 많은것을 요구할 때에는 아마도 우편으로 받은 답변은 正確하지 못할 것이다. 사람들은 착오를 일으키거나 혹은 正確한 답변을 하는데 必要한 措置를 취하지 않게 될 것이다.

(iv) 읽기 어려운 글씨

일선조사와 마찬가지로 오차를 발생하게 된다. 應答者와 直接的인 接觸이 없기 때문에 우편조사의 경우 이것을 극복하는것은 매우 어렵다.

(v) 無応答

情報를 提供해야 된다는 法的根拠가 있더라도 우편조사에 있어서 이 것이 問題가 된다. 이것은 바로 無応答者를 추가 조사하는 절차를 마련해야 되는 이유이기도 하다.

(vi) 不完全한 様式

大部分의 応答者들이 쉬운 질문만을 답변하고 질문이 어려운 것은 공란으로 남겨두는 경향이 있다. 어떤 사람들은 처음에 시작할 때는 기재를 잘하나 나중에는 興味를 잃게 되어 조사표의 後半部分은 完成하지 못하는 경우도 있다. 이것은 無応答의 또 다른 形態로서 별도 措置가 必要하게 된다.

(c) 報告統計에서 發生하는 誤差

行政報告는 政府統計의 중요한 源천이나, 그 성질상 特別한 問題가 야기되어 誤差나 착오를 일으키게 된다. 이러한 報告는 非統計的인 目的으로 작성되기 때문에 위와 같은 問題가 야기된다. 숫자로 된 정보의 수집은 일반적으로 부수적인 重要性이 있을 뿐이며 일단 작성된 후에는 주의깊게 確認하지 않는다. 더구나 統計人은 資料蒐集過程에서는 거의 關여하지 못하거나 전혀 關여하지 못한다. 그가 할 수 있는 것은 후에 情報를 確認해 보는 것뿐이다.

誤差가 發生하는 源泉

(i) 動機의 不足

기록을 完成해서 보내는 立場이나 그것을 받는 立場 모두가 단지 統計的 目的만으로 必要한 답변을 確認해 보려는 열의

가 부족하다. 예컨대 관세를 목적으로 하는 輸入登錄은 한 國家의 國際貿易에 관한 중요한 資料이다. 船積때 支払해야 하는 稅金이 그 당시의 價額에 依存한다면 이것은 正確하게 登錄이 되고 세관원에 의하여 確認이 될 것이다. 선적량이 統計 目的을 위하여 把握될 必要가 있다면 그것은 세관원이나 輸入業者에 의하여 把握되기는 힘이 들것이다. 때로는 資料가 전혀 무시될때도 있다.

(ii) 부적합하거나 矛盾된 定義의 사용

報告時에 使用된 定義는 아마도 行政目的에 부합되는 것이지만 統計에 필연적으로 적합한 것은 아닐 것이다. 이것은 報告로부터 얻어진 資料가 正確히 대상 모집단을 설명하지 못한다는 것을 의미할 수 있다. 資料가 한개 이상의 報告에 의해 統合된 것이라면 정의가 完全히 一致할 수는 없다.

(iii) 無応答

單位에 대한 情報가 전혀 없다는 의미에서 어떤 特別한 질문에 대한 답변이 없다는 문제가 발생한다. 위에서言及한 바와 같이 行政目的에 直接的으로 必要한 資料는 그 기록이 完全하고 正確할지 모르나 단지 統計目的에만 必要한 資料는 一部事項이 누락되어 있거나 正確하지 못할 것이다.

### 6.3.3 資料處理過程의 誤差

資料處理에 使用될 수 있는 設備에 따라 (즉 使用되는 技術과 使用할 수 있는 도구에 따라) 資料處理에서 發生하는 오차를 몇가지로 분류할 수 있다. 부호기입, 內容檢査, 천공, 集計,

計算, 對照, 統合 등에서 誤差가 발생할 수 있다. 誤差는 어떤 資料에서 다른 資料로 移記하면서 發生할 수 있다. - 예컨대 집계용지에서 타이프를 치기 위하여 정리하는 最終的인 표로 옮기는데서 발생할 수 있다.

많은 조사표나 천공카드를 취급할때나 資料處理를 할 때에 매우 잘 訓練된 專門家조차도 錯誤를 일으키게 된다. 여기서 資料處理의 4 단계 - 內容檢査, 符號記入, 分類, 편집 (Compilation) - 에서 일반적으로 발생하는 몇가지 誤差에 대하여 考察해 본다.

#### (a) 內容檢査

다음 章에서 보게 되겠지만 이 問題가 더 상세히 討論된다면 資料의 질을 개선하려고 하는 蒐集된 자료의 內容檢査는 統計資料의 處理에 있어서 重要的인 연결의 역할을 한다. 이것은 때로는 缺어진 資料의 推定을 수반하거나 틀린 것으로 명백히 드러난 資料의 變更을 包含한다. 비록 이러한 過程들이 상세히 기록된 指針에 따라 內容檢査要員이 定規的으로 행하고 있지만 이러한 指針들은 새로운 資料나 수정된 資料를 삽입하는데 대한 客觀的인 判斷基準을 提供하기는 하나 부가적으로 誤差를 발생시킬 여지가 되기도 한다. 正確한 숫자를 變更시키거나 內容檢査要員의 그릇된 판단으로 正確한 것을 틀리게 만들거나 혹은 指針을 잘못 해석할 위험도 있다. 그러나 內容檢査過程에서 발생하는 誤差의 主要原因은 해야될 內容檢査를 하지 않거나 수정을 하지 않았기 때문이다.

(b) 부호기입

부호기입 과정에서 가장 보편적으로 발생하는 誤差는 두가지 種類가 있다.

(i) 変数나 屬性에 대한 그릇된 부호의 사용 : 부호기입은 조사표가 完成되었을때 행하여 지거나 혹은 그 후에 행하여 질수도 있다. ( 地方에서 혹은 모든 調査票가 中央에 도착된 후에 ) 어떤 경우이거나 부호를 잘못 기입 함으로써 誤差가 발생할 수 있다. 存在하지 않는 부호를 使用할 경우도 있고 分類表에 있기는 하나 잘못 使用할 경우도 있다.

(ii) 完全히 누락된 부호 : 질문에 대한 답변이 完全히 누락되어 있는 때 조차도 분류번호가 기입되는 것에 주의하여야 한다. 예를 들면 부호기입요령서는 기록이 되지 않은것을 處理하기 위하여 99 로 부호기입하는 마지막 부류를 마련하는데 適合한 때에는 이러한 부호의 挿入은 必要하게 될 것이다.

(c) 分類

일단 부호기입이 完了되면 集計를 위하여 조사표를 分類하여야 한다. 分類는 손으로 행하여 지거나, 어떤 機械的 手段에 의하거나 혹은 必要한 조작을 遂行할 컴퓨터에 資料를 입력할 수도 있다. 이러한 과정에서 發生할 수 있는 誤差는 다음과 같다.

(i) 그릇된 分類 : 單位나 기록을 다른 部類에 包含시키는 것, 이것은 손으로 分類할 때 일어나기 쉽다.

(ii) 機械分類를 위한 欠工이나 準備中の 誤差 : 機械에 의한 分

類나 컴퓨터가 사용된다면 材料가 必要하게 된다. 보통 이것은 천공된 카드, 종이 테이프, 디스크등이다. 분명히 천공도중에 誤差가 발생할 수 있는데 그것을 발견하고 수정할 特別한 장치가 必要하다.

(ii) 記錄의 紛失 혹은 削除: 매우 規模가 큰 調査에 있어서 손으로 分類할 경우에는 기록을 분실하거나 잘못두는 境遇가 흔한데 이것을 防止하기 위하여 注意를 하여야 한다.

(d) 편집 (Compilation)

편집이란 分類된 節次에 따라 統計票를 작성하고 必要한 計算을 행하며 標本으로부터 모집단값의 추정치를 準備하는 조작을 의미한다. 過程의 複雜性은 어떤 特別한 調査에 따라 다르나 간단한 경우에도 誤差를 除去하기 위하여 세심한 주의를 기울여야 한다. 이 段階에서 發生하는 착오는 매우 重大하기 때문에 總計나 母集團의 값에 影響을 미치게 된다. 그러므로 모든 계산과정이나 조작을 確認하는 것이 重要하다. 만약 편집이 컴퓨터에 의해 행하여진다면 機械로 하여금 計算을 시키고 集計를 위한 프로그램을 주의깊게 確認해야 한다. 機械自体는 數學的인 錯誤는 일으키지 않지만 그릇된 것을 행하라고 한다면 그릇된 結果를 만들어 낼 것이다. 어떤 國家에서는 統計表의 작성을 손으로 행하기 때문에 각 과정에서 發生할 수 있는 誤差들에 관해서 集中的으로 討論하게 될 것이다.

오늘날에는 大部分의 事務室에 각각의 業務遂行을 補助하는데 必要한 전자계산기가 있게 마련인데 대부분이 複雜하다. 그러나 아

직도 계산기를 使用하는데 착오를 일으키기 쉽기 때문에 다음과 같은 問題들이 관련된다.

다음과 같은 題目下에 發生할 수 있는 誤差에 대해서 考察할 수 있을 것이다.

(i) 計算誤差 : 두가지 형태가 있는데 첫째 數學的인 착오 - 더하기, 빼기등과 같은 數學的 조작을 遂行하는데서 發生하는 誤差이다. 둘째는 그릇된 조작때문에 야기된 誤差인데 예를 들면 어떤 單位에서 다른 單位로 轉換시킬때 틀린 숫자로 나눌때 등이다.

(ii) 移記誤差 : 이것은 어떤 資料로 부터 다른 資料로, 어떤 종이에서 다른 종이로, 어떤 표에서 계산기로 옮길때 일어나는 오차이거나 혹은 계산결과를 記載할 때 잘못 써서 發生하는 誤差이다. 사사오입의 誤差(혹은 근사오차)라고 불리우는 또 다른 오차가 있다. 이것은 一般的으로 資料處理가 끝난후에 공표하려고 準備하는 資料에서 발견된다. 이것은 숫자를 끊어내거나 계산도중에 근사치로 만들었을때 發生한다. 사사오입의 誤差(Rounding error)란 우리가 토의한 非標本誤差의 形態와는 전혀 다르며 資料處理要員에 의해 야기되는 錯誤가 아니라 다음과 같은 지시의 結果로 發生된다. 第8章에서는 사사오입에 관하여 전반적인 討議가 있게 될 것이다.

#### 6.4 農業統計에 있어서의 特別한 問題

일선조사원들의 業務中 상당한 部分이 農業統計調査와 關聯된 것일것이다. 이러한 調査는 蒐集된 資料의 질과 관련하여 獨



특한 問題를 안고 있는데 重要한 몇가지 觀點에 대해서 討論하게 될 것이다. 어떤 책은 全体가 이 問題에 관하여 쓰여진 것도 있으나 우리는 모든 問題를 모두 상세히 檢討할 수 없기 때문에 우리가 자주 대하게 되는 一般的인 狀況에 대하여 檢討하는데 集中하게 될 것이며 使用된 節次때문에 야기되는 結果의 不正確性에 대하여 集中적으로 論議하게 될 것이다.

올바른 技巧를 使用하여 農業統計資料를 蒐集하는데 많은 費用과 時間을 消費하게 되는데 따라서 根本적으로 發生하는 問題들에 대해서 檢討하게 될 것이다. 그러므로 使用할 수 있는 豫算에 따라 調査를 할 수 있는 節次를 選定하는데 어떤 確定되어 있는 方法은 없다.

農業統計에 關한 9가지의 特別한 觀點을 열거하게 되겠는데 그중 앞의 5가지는 地域의 特定에 대한 不正確性에 關聯된 것이고 마지막 4가지는 穀物生産量의 추정의 어려움에 關한 것이다.

(a) 눈짐작 : 이것은 地域이나 生産量 혹은 두가지 모두를 추정하는데 使用되는데 平均生産量을 추정하는데 一般的으로 使用되고 있다. 눈짐작은 確實히 主觀的이며 資料의 正確性은 각 調査員의 經驗과 感覺에 상당히 依存하게 되며, 可能한 限 버려야 할 方法이다. 그러나 그것은 費用이 적게 들고 신속하다는 장점을 지니고 있기 때문에 結果가 매우 絶박하게 必要한 경우 계속 使用될 것임에는 의심의 여지가 없다.

눈짐작은 所謂 '관리' 測定과 결합하여 使用되는 것이 바람직하다. 예를들면 조세 또는 다른 行政報告가 어떤 地域의 耕作地나 곡물의 總生産量을 나타내는데 使用될 수 있다면 각 調査員의 눈

집작의 結果를 더하여 報告資料와 比較를 해보아 그 結果가 일치하지 않을 때에는 調整이 될 것이다. 이와 마찬가지로 平均 生産量의 눈짐작은 어떤 전형적인 農耕地의 客觀的인 測定과 결합되어 행하여지는 것이 바람직하다. 農耕地로 부터 生産量의 주관적인 평가가 실제 測定된 生産量과 약간의 차이만 있다면 어떤 관리가 전혀없이 추정된 生産量보다는 더 信賴할 수 있게 된다.

(b) 徒步測定 : 이것은 農業統計調査에 있어서 地域測定の 한방법으로 자주 사용되는데 장방형의 農耕地를 測定하는 데에만 그 사용을 한정하여야 한다. 이 方法을 사용하기 위하여는 어떤 標準化된 걸음걸이로 이를테면 75 cm정도로 걸수 있도록 訓練이 되어야 한다. 地域測定은 耕作地의 가장자리를 따라 도보로 測定하여 각 가장자리의 걸음걸이수를 기록하면 된다. 徒步測定에 의하여 얻어진 결과는 어떻게 걸었는가에 따라 다르게 된다.

아프리카에서 徒步測定에 의하여 地域을 推定한 것과 正確하게 사슬이나 컴 파스를 가지고 測定한 實驗은 (i) 調査員에 따라 추정치의 正確성에 있어서 현저한 차이를 보이고 있으며 (ii) 經驗이 많은 調査員일수록 正確성에 있어서 현저히 좋은 결과를 나타내고 있다.

地域資料가 徒步測定에 의하여 定期的으로 蒐集되고 있을 때에는 監督者로 하여금 가끔 正確한 測定道具를 사용하여 불시점검을 행하여 資料의 確認을 함으로써 許容할 수 있는 水準의 正確성을 항상 維持하도록 하여야 한다.

(c) 경사진 地域의 測定 : 다음의 예는 경사진 地域의 測定에 관련된 것이다. 調査員이 경사진 耕作地의 여러 가장자리의 길이를 잴 때 傾斜를 무시한다면 도구가 아무리 正確하다 할지라도 偏倚(Bias)가 생기게 될 것이다. 農業統計에 있어서는 耕作地가 水平으로 돌출된 部分이 必要하게 된다. 그러므로 많은 耕作地가 언덕에 있다면 測定은 현저하게 偏倚(Bias)가 發生하게 될 것이다.

(d) 쇠사슬과 컴파스의 不一致 : 일반적으로 쇠사슬과 컴파스를 사용하여 面積을 計算하는 것은 正確할 것이다. 이 方法에 있어서는 耕作地를 실물에 比例하도록 그래프用紙에 그리게 되는데 耕作地의 한모퉁이에서 시작하여 컴파스를 사용하여 다음 모퉁이로 이어지고 그 점으로부터 水平으로 돌출된 거리를 測定하게 된다. 이 過程이 反復되어 처음 시작한 지점으로 돌아가게 되는데 이 지점은 이 과정의 처음과 끝에서 겹쳐지게 된다. 이 출발점이 경우에 따라서는 서로 正確하게 一致되지 않게 되는데 특히 不規則的인 形態의 耕作地에서나 언덕 또는 까다로운 地形에 있어서 현저한데 이 地点의 차이를 不一致(Misclose)라고 한다. 許容할만한 不一致를 미리 정할 必要가 있는데 不一致가 許容限度를 초과할 경우에는 다시 작성하여야 한다. 그렇지 않으면 要求되는 水準의 正確성은 달성되지 못할 것이다.

(e) 순회 : 어떠한 이유로 말미암아 任意抽出原則의 嚴格한 適用이 실제로 困難할 경우에 이 方法이 單純抽出로서 使用된다. 어떤 適合한 地域單位에서 이 方法으로 標本(이른데면 마을)이

抽出된다. 調査員이 이 마을에 가까운 地域을 가로질러 運輸을 하고 간다. 같은 간격으로 혹은 임의로 選擇한 地點에서 차를 세우고 그차에서 가장 가까운 그리고 指定된 穀物을 경작한 農耕地를 標本으로 추출한다. 이 方法을 使用한다면 명백히 대상 곡물을 경작하는 경작지가 동등한 추출기회를 갖지 못한다. 道路로부터 멀리 떨어진 경작지는 전혀 추출될 기회를 갖지 못하는데 이러한 耕作地가 도로에 가까운 경작지와 特性이 같지 않다면 추정치는 偏倚(Bias)가 있게 될 것이다.

이런 方法으로 얻은 穀物生産量의 추정치에 偏倚가 생긴다는 것은 누구든지 쉽게 推測할 수 있다. 예를 들자면 도로에 가까운 耕作地의 生産量은 먼지나 자동차의 매연때문에 影響을 받게 될 것이다. 다른 한편으로 도로는 낮은 地帶에 建設되기 때문에 產出量이 많을 수 있으며, 均衡을 잃을 정도로 많은 耕作地가 標本으로 추출될 수 있으며 그것은 도로에 가까운 耕作地에 는 물 등을 供給하기가 더 쉽기 때문이다.

순회는 매우 큰 偏倚를 내포한 推定値를 만들 수 있는 危險을 잠재적으로 갖게 되기 쉽기 때문에 可能的 限 이 方法은 피하여야 한다.

(f) 境界線偏倚(Border bias): 穀物の 平均生産量의 推定値는 任意的인 手段에 의하여 추출된 경작지에서 嚴格한 틀(Rigid frame)을 設定하여 이 틀의 標本區域으로 곡식을 추수하고 탈곡하여 건조하고 무게를 달아봄으로써 얻어진다. 이것은 正確히 標準化된 틀을 사용하여 쉽게 헥타당 平均生産量으로 轉換시킬 수

있다.

그러나 만약 틀로 추출된 위치가 경작지의 境界線과 겹쳐지게 되면 混動이 일어나게 되고 틀내에서 추수할 수 있는 地域이 標準보다 작게 된다. 그러므로 실제에 있어서는 境界線이 重複되어 추출되면 기각하거나, 혹은(一般的인 形態의 耕作地로) 이러한 추출이 되지 않도록 特別한 기교를 使用하여야 한다. 이 경우에 一般的인 추출방법은 耕作地의 넓이에서 틀의 넓이를 뺀 나머지 임의의 수를 더하여 標本區域의 한모퉁이의 좌표를 추출하는 것이다. 만약  $3\text{ m}^2$ 의 틀이 사용된다면 길이가  $100\text{ m}$  이고 너비가  $60\text{ m}$ 인 경작지에서 1과 97 사이의 임의의 수와 1과 57 사이의 임의의 수를 취하여 좌표가 정해지게 될 것이다. 이러한 方法으로 추출된 區域은 서로 境界線이 重複되는 일은 없게 될 것이다. 이 方法의 問題點은 境界線쪽의 耕作地 部分이 경작지의 中心部分보다 추출기회가 적게 되고 耕作地의 모퉁이부분이 가장 적은 추출기회를 갖게 되는 것이다. 이 抽出方法은 의심할 여지도 없이 偏倚가 생기게 되나, 이 偏倚가 중대한 결과를 초래하는가에 대하여는 論爭의 여지가 있다. 만약 境界線을 따라서 얻어진 生産量이 中心地의 生産量과 別 차이가 없다면 結果에 影響이 있더라도 偏倚는 별문제가 되지 않으나, 서로 差異가 많다면 境界線편의(Border bias)가 問題가 된다.

아프리카와 인도에서의 어떤 研究는 境界線偏倚를 무시할만하나 인도에서의 다른 어떤 研究는 어떤 작물(황마)에 있어서는 境界線보다는 中心地의 生産量이 현저하게 많다는 것을 보여 주었

다. 生産量의 추정치에 影響을 미치는 境界線偏倚의 위험이 인식되어져야 하고, 耕作地의 가장자리로 부터 거리에 따라 生産量의 變化가 있는지를 결정하기 위하여는 實驗을 해야 한다는 것을 건의하고 싶다.

(g) 줄로 심어진 作物의 生産量推定: 穀物生産량이 正方形이나 長方形의 틀내의 標本区域으로부터 추수에 의하여 추정이 될 때 곡물이 줄로 심어졌을 때에는 正確한 틀의 위치를 결정하는데 있어서 注意를 기울어야만 한다. 만약 틀이  $1\text{ m}^2$  라면, 줄이 18 cm간격으로 놓여 있을 때에는 그 틀을 어떻게 놓는가에 따라 틀 내에는 5 줄 혹은 6 줄이 包含되게 될 것이다. 어떤 生産量의 推定値가 어떤 경우에는 20%가 높기 때문에 調查員으로 하여금 난수표를 使用하여 대충 標本区域을 抽出하도록 許容하거나 調查員 자신의 裁量으로 틀의 正確한 位置를 決定하게 한다면 심각한 偏倚를 초래하게 될 것이다. 틀의 위치선정에 있어서 調查員의 自由裁量을 박탈하는 명백한 指針을 마련하는 것이 必須的이다.

분명히 틀의 크기가 크면 클수록 그 内部에 包含되는 줄의 수는 많게 될 것이고 百分率로 나타낼 때에는 잠재적인 偏倚는 작게 될 것이다. 잠재적인 편倚는 틀을 耕作地에 平行하게 놓는 대신에 대각선으로 놓게 되면 작게 될 것이다.

(h) 生産量의 調查時期: 生産量 調查를 위한 理想的인 方法은 추수직전에 작물을 베어내는 것이다. 調查와 추수사이에 時間的인 空白이 있게 되면 平均生産量의 추정치는 하향하는 편倚가 생기

게 될 것이다. 이것은 작물에 따라 다르게 되겠는데 추수전에 매일 3 퍼센트씩 건조의 成長率을 測定한 스웨덴의 한 研究에 의하여 關聯된 偏倚(Bias)의 觀念이 評價될 수 있다. 그러므로 調査와 추수 사이의 며칠 간의 간격은 結果에 重大한 影響을 미치게 된다. 調査員이 調査地域에 살거나 가까이 살 때에는 農夫와 같이 확인할 수 있고 또한 正確한 時期를 맞춰 耕作地를 訪問할 수 있기 때문에 時期는 別 문제가 되지 않는다. 그러나 調査期間동안 訪問해야 될 마을이나 地域이 많다면 적절한 時期에 正確한 장소를 訪問한다는 것은 不可能하게 되어 추수가 되기도 전에 어떤 區域에서는 生産量을 調査해야 되는 경우도 있게 될 것이다. 環境에 따라 行動을 취해야 되며, (최소한 推測으로라도) 그의 業務日誌에 추수전 어느날에 調査를 해야될지 또는 그가 취해야 될 接近方法을 決定해야 되며 事務室의 資料處理過程에서 作成해야 될 적당한 수정계수를 마련해야 한다.

(i) 잠재적인 生産量과 實際生産量 : 어떤 作物 특히 채소에 있어서서는 作物의 一部分이 밭에 남게 되어 추수되지 않거나 혹은 非效果的인 추수방법 때문에 실제에 있어서 損失이 생기는 것이 보통이다. 이렇게 되면 잠재적인 生産量(損失이 없었다면 추수가 되었을 量 - 이것은 때때로 生物学的인 生産量이라고 불리운다)과 經濟的인 生産量이라고 불리우는 實際로 추수된 生産量과는 一般的으로 상당한 차이가 있게 된다. 일상적인 調査方法은 잠재적인 生産量을 測定하는 傾向이 있는 반면 대부분의 統計的 目的에 必要한 資料는 實際 生産量을 要求한다. 그러므로 이러한

이유때문에 必要한 경우에는 平均生産量の 過大評価를 防止하기 위한 調整이 必要하게 된다. 몇년전에 어느 國家의 農業統計에 있어서 잠재적인 生産량과 經濟的 生産量사이의 차이를 없애도록 許容하였으나 실패하였다. 그것은 經濟가 정원에서 경작하는 자경 자급농업에 상당히 依存하였기 때문에 이 國家의 總生産량이 상당히 과대평가 되었기 때문인 것으로 여겨진다.

위와같은 問題들은 農業統計에 있어서 여러가지로 偏倚가 생길 수 있다는 것을 보여주기엔 充分할 것이다. 調査가 어떻게 되었는가에 따라 생길 수 있는 重要한 不正確性에 대해서만 言及했을 뿐이다. 다른 製造業이나 鎔業의 生産량의 調査보다 農業에 있어서는 그 生産량의 測定에 있어서 더 複雜한(그리고 誤差가 더 發生하기 쉬운) 業務로서의 特徵을 지니고 있다. 분명한 하나의 요소는 生産者가 많기 때문에 대부분의 경우 실제적 必要때문에 標本을 추출한다는 것이다. 또 하나는 全体生産량의 많은 부분이 팔리지 않고(鎔業과 2次産業의 産出量の 大部分처럼) 生産者에 의해 消費되기 때문에 얼마나 生産되었는지 保存된 기록이 전혀없다. 農業統計를 원래 어렵게 만드는 이러한 모든 要因들에 관하여 상세히 討議하지는 않겠으나 단지 特別한 問題를 야기시키는 混合耕作과 계속경작에 대해서만 言及을 할 것이다. 混合耕作이라 함은 사이갈이나 같은 耕作地에 서로 다른 作物을 동시에 재배하는 것을 말한다. 混合耕作地에 있어서 각 作物의 재배면적을 어떻게 正確하게 分配시킬 것인가에 대한 지역통계에 있어서는 어려운 問題가 發生하게 된다. 이 問題에 대



하여 모든 環境에 만족할만한 一般的인 해결책은 없으나, 연습에 의하여 發生할 수 있는 어려움에 대하여 이해를 시키고자 이러한 경우 地域統計를 作成하는데 있어서 接하게 되는 5 가지 경우를 言及하려고 한다.

(i) 全体地域을 가장 중요하거나 우세한 작물에 割當한다. 이 방법은 適用하기가 쉽고 우세한 작물에 대하여 믿을만한 評價를 할 수가 있다. 그러나 가장 큰 단점은 어떤 작물이 다른 것보다 더 頻繁하게 耕作된다면 이 방법은 어떤 作物에 대하여는 過大評價를 하게 되는 반면 다른 작물에 대하여는 過小評價를 하게 된다는 것이다.

(ii) 全体地域을 混合된 모든 作物에 할당하는 방법인데 이것은 가끔 使用되기는 하나 모든 작물에 대한 報告된 總面積이 실제 耕作된 面積을 초과하게 되기 때문에 一般的으로 장려되지 않는다.

(iii) 경작지 全体를 混合된 作物의 各各에 均等하게 配分하는 방법

(iv) 各 作物이 占有하고 있는 面積을 눈짐작으로 추정하는 방법인데 이것은 매우 主觀的이기 때문에 統計에 심각한 偏倚를 發生시킨다.

(v) 각 作物에 따른 地域分配 : 混合되어 있는 각 作物에 地域을 配分하는 여러 方法이 考案되었는데 보통 作物의 密度나 混合耕作된 상태의 生産量을 그 作物만 심어져 있는 狀態의 生産量에 対応하여 配分한다. 예를 들자면 단지 한작물만 심어져 있는 一般的인 密度에 대한 추출이 된 混合된 區域의 單位區域當

植物의 密度의 比率을 總面積에 곱해주면 각 작물에 대한 할당 면적이 얻어진다. 이 방법은 적어도 理論적으로는 長點을 가지고 있으나 두가지 短點이 있는데, 하나는 小區域의 植物의 密度에 관한 情報과 한가지 作物만을 심었을 때의 密度에 대한 情報가 必要하게 되는데 이것은 아마도 많은 費用을 들여야만 얻을 수 있을 것이다. 또 한가지는 耕作地의 모든 作物에 配分되는 總面積은 耕作地의 실제 耕作면적보다 크거나 작다는 것이다.

이로부터 混合耕作의 경우 面積配分이 매우 어렵다는 것과 실제로 適用되는 節次가 全國에 똑같이 適用될 수 없다는 것 그리고 각 地域의 環境에 맞도록 위의 方法들을 混合하여 使用하여야 된다는 것을 깨달았을 것이다. 세계농업센서스(World Census of Agriculture)를 위한 勸告에는 混合耕作地域을 어떻게 處理해야 된다는 것을 밝혀 놓았는데 거기에는 “만약 各 作物이 그 作物 하나만 재배되었다면 어느 정도의 面積이 되겠는가를 추정하고자 하는 努力을 可能한 限 하여야 한다”고 서술되어 있다. 그 原則은 각각의 作物에 적당한 比率로 地域을 配分하여야 한다는 것인데 그러므로 (i), (ii) 그리고 (iii)의 方法은 이 原則에 따른다면 기각을 하여야 하나 때때로 실제로 있어서는 不可能한 것이다. 混合耕作에 있어서의 다른 분류는 만약 여러 作物이 성장속도가 다르다면 이러한 區域으로부터 生産量의 추정은 더욱 複雜해 지는데 그때에는 調査員이 한번의 訪問으로 모든 作物의 生産量을 同時에 調査할 수 없다는 것이다. 이러한 問題

는 調査員이 그가 거주하는 마을로부터 멀리 떨어진 地域을 할당받아 再訪問이 困難한 경우에 特히 發生되는 問題이다.

계속경작은 한해에 同一한 耕作地에 같은 作物이나 혹은 서로 다른 作物을 한번이상 연속적으로 심어서 추수를 하는 慣習을 말한다. 이 관습은 또한 問題를 야기하게 되는데 예컨대 어떤 作物이 자랄때마다 耕作地의 總面積을 한번씩 계산할 것인지 아니면 耕作面積의 重複計算을 防止하기 위하여 作物사이에 할당을 할 것인지를 決定하여야 한다. 그것은 일년동안의 總生産량을 調査하기 위하여는 再訪問을 해야 한다.

混合耕作이나 繼續耕作과 같은 관습에서 생기는 어려움과는 달리 農業生産의 調査의 正確性은 기후의 惡條件, 洪水, 가뭄 등과 같은 예기치 못한 發生에 影響을 받게 된다. 비록 적은 地域만이 影響을 받았다 하더라도 이러한 것들은 統計에 重大한 影響을 미칠수가 있다 - 예를들면 그들은 이미 實施된 測定을 모든 環境에 代表할 수 없게 한다. 이러한 사건들은 必要한 事後措置가 행해질 수 있도록 確認이 되어야 하고 蒐集된 資料의 正確性에 影響을 미치는 非正常的인 特徵에 대하여 確認하고 報告하는 것이 調査員의 責任이 될 것이다. 이러한 特殊한 環境의 주시를 계속하여 일선조사업무에서 偏倚가 발생하지 않도록 주지시키는 監督業務에 責任이 있는 地方事務所의 管理職員도 또한 任務를 充實히 遂行하여야 한다.

## 第 7 章 誤差의 管理

### 7.1 概 要

前章에서는 統計調査에 있어서 오차가 發生할 수 있는 여러가지 경우에 대해서 考察하였다. 이장에서는 誤差의 防止와 發生한 誤差의 發見이나 수정에 대하여 考察해 보고자 한다. 취해야 될 여러 단계가 있는데 여기서는 다섯가지 부류로 分類하고자 한다.

- (a) 改善된 設計와 企劃
- (b) 改善된 資料의 蒐集
- (c) 資料의 確認 및 內容檢査
- (d) 事後調査
- (e) 改善된 資料處理 方法

等 重要한 觀點들이 本章에서 論議될 것이다.

### 7.2 改善된 設計와 企劃

#### 7.2.1 定義와 分類

調査의 設計段階에 있어서는 前章에서 觀察한바와 같이 誤差가 여러가지로 發生할수 있는데 이러한 誤差들은 系統的으로 作用하기 때문에 偏倚를 일으키게 된다. 이러한 設計誤差는 각기 다른 概念과 變數에 對한 그릇된 定義의 使用때문에 發生한다. 이러한 問題들 극복하기 위한 重要한 方法의 하나로서 國際적으로 承認되어 作成된 定義를 使用하는 것이 있다. UN을 비롯한 여러기관에서는 農業統計로부터 雇傭統計에 이르기까지 各

種統計調査를 위한 定義問題에 關聯하여 많은 연구를 해 놓았다.

물론 이러한 定義들이 國家가 추구하는 目的에 부합되어야 할 것이 要求된다. 그러나 定義가 國家目的에 부합되지 않는다 하더라도 必要한대로 변형을 시키는데 있어서 國際的인 基準을 使用하는 것은 長점이 있기 마련이다. 때때로 기교적인 변형으로써 國家的인 要求事項을 만족시키고 國際的인 比較나 改造가 可能하게 된다. 또한 많은 問題에 關하여 資料가 수집되고 처리되는 경우에 적당한 부호기입을 위하여 定義를 包含하여 있고 기초를 제공해 주는 상세한 분류들도 있다. 國際的인 分類로서 一般的인 것들은 다음과 같다.

貿易統計를 위한 標準國際貿易分類(SITC, Standard Trade International Classification)

企業과 産業統計(모든 經濟統計)를 위한 標準 國際産業分類(ISIC, International Standard Industrial Classification)

雇傭統計를 爲한 標準國際職業分類(ISCO, Industrial Standard Classification of Occupations)

각종 통계에 國際的인 標準體系를 使用하는데 있어 또 하나의 長점은 各國別로 情報를 比較해 볼 수 있고 또한 각 統計別로 比較해 볼 수 있다. 이러한 標準國際分類에 對하여는 第12章에서 討議될 것이다. 確實히 새로운 統計調査를 企劃하는 段階에서는 各種誤差가 發生할 여지가 많다. 부적합한 定義나 分類에 더 나아가서 모집단 確認의 결점, 標本設計와 표본들, 調査表를 통한 正確한 의사소통, 調査員에 對한 그릇된 指針이나 道具, 事務

실에서 資料處理와 그에 對한 감독등 모든 과정에서 系統的인 오차를 發生시킬 수 있다.

### 7.2.2 事前調査와 試驗調査( 파일로트 씨베이 )

이러한 誤差가 發生하는 原因을 確認하는 方法으로서 事前調査와 試驗調査( Pilot survey )를 本調査以前에 實施하는 것이 있다. 이런 方法에 依하여 오차가 發生하는 原因이 발견되고 수정될 수 있으며 時間, 費用, 人力等에 關하여 必要한 情報을 얻을 수가 있다. 이책의 앞부분에서 사전조사와 파일로트 씨베이에 關한 言及이 있었다. 本章에서는 이들에 關한 상세한 說明이 주어질 것이며 이것의 主要목적은 計劃段階에서 어떻게 오차가 除去되어야 하는가를 보여주는데 있다. 그러나 다른 目的도 또한 있는데 그에 對한 說明도 있게 될 것이다. 計劃段階에서 새로운 調査表를 가지고 試驗的으로 面接을 實施해 보거나 몇몇 応答者에게 편지를 보내어 그 效果를 試驗해 보는 것이 一般的이고 또한 바람직하다. 이것을 事前調査라고 한다. 그러나 때로는 그 以上으로 바람직한 것은 組織的으로 調査의 모든 過程을 試驗해 보는 것이다. 이 方法은 특별히 調査가 새로 시작되어 計劃樹立에 必要한 아무런 情報도 없을때에 더욱 必要하다. 이런 경우에는 本調査를 모방한 작은 規模로 調査를 實施하여 그 效果를 測定해 보는 것이 도움이 된다. 이것을 파일로트 씨베이라고 하는데 충분한 時間의 여유가 있고 이것을 實施하여 어떤 變경을 하고자 할때 本調査 이전에 實施된다. 특히 다음과 같은 觀點에 對한 지표들 마련해 줄 것이다.

(a) 調査에 分配될 要素

예를 들면 計劃의 각 段階別로 所要되는 時間은 모든 活動을 遂行하는데 必要한 사람의 數를 算出하는 根拠가 되며, 出張, 장비, 임금등 必要한 總豫算의 윤곽을 파악할 수 있다. 이것이 파일로트 씨베이의 가장 重要한 點의 하나이다.

(b) ( 센서스에 있어서 ) 応答者의 명부나 ( 標本調査를 위한 ) 표본들이 만족스러운지 或은 부적당한 명부가 調査에 편의를 초래하는지에 대해서 점검해 볼 수 있다.

(c) 資料를 蒐集하기 위한 最적의 方法

예를 들면 面接調査나 우편조사중에서 어떤것이 바람직한 것인지 판단을 내리기가 어려운 경우가 있다. 파일로트 씨베이는 양자택일할 수 있는 方法中에서 選擇할 수 있도록 相對的인 費用, 正確性, 応答率등을 나타내줄수 있다. 또한 必要한 情報가 応答者에게 어느정도 知曉되어 있는지를 把握할 수 없는 경우 事前調査는 모든 調査가 面接調査로서 實施되어야 하는지 或은 調査員이 보다 많은 비용이 소모되는 實地조사에 의하여야 할 必要가 있는지에 관하여 알려준다.

(d) 調査될 모집단의 變이성은 標本誤差의 크기가 얼마나 될지를 암시해 준다. 실제로 모집단의 變이성에 對한 知識은 調査에 必要한 標本の 規模를 決定하는데 使用된다.

(e) 期待되는 無応答率 즉 調査에 協助를 거부하는 사람의 比率 或은 必要한 資料를 提供할 能力이 없는 사람이나 부재중이어서 만날수 없는 사람의 比率 등에 對한 情報는 調査設計에 어

면 변경을 일으킬 수도 있다. 예컨대 미리 대체를 決定하여 無応答者를 다른 사람으로 대체시키든지 或은 大部分의 사람들이 집에 있을 時間인 저녁이나 주말에 調査를 하든지 할수 있다.

(f) 아마도 가장 重要한 것은 파일로트 씨베이는 調査表가 만족할만한지 或은 변경을 해야될지를 나타내 줄수 있다는 것이다. 調査表가 너무 긴지 或은 質問順序가 論理的으로 되어 있는지, 表現이 명백한지, 어떤 質問이 어색한지, 분개를 살만한지 或은 대답을 거절할 것인지를 決定해 줄수 있다.

몇년전 인구센서스에서 용어사용의 잘못으로 특수한 사례가 발생하였다. 이민은 사람에게 대하여 얼마동안(즉 몇년동안) 그 國家에서 거주하였는지를 알고자 하였다. 실제의 質問은 "length of residence" 이라고 물어 보았다. 그러나 大部分의 사람들이 住宅의 길이로 대답하였다. 만약에 적절한 파일로트서베이가 실시되었더라면 그러한 용어사용의 잘못을 저지르지 않았을 것이며, 統計人들로 하여금 보다 나은 용어 즉 "거주기간(Period of residence)" 를 使用토록 하였을 것이다.

(g) 面接調査를 위하여 作成된 指針이나 訓練計劃이 만족할만한지를 알게 해 준다.

물론 파일로트 씨베이를 實施하는데는 費用이 많이 들게 되는데 그 실시의 必要性与否는 所要되는 費用과 均衡이 되어야 하거나 或은 단지 한번만 實施되는 特히 重要한 調査나 (人口센서스와 같이) 매우 간격이 큰 調査를 除外하고는 그 長点이 所要되는 費用과 均衡이 맞지 아니한다. 만약 파일로트 씨베이가



實施되지 않으면 너무 늦어서야 問題가 明白하게 드러나게 된다.

앞의 論議에서 事前調査란 本調査에 앞서서 적은 規模로 實驗을 해보기 위하여 實施되고, 파일로트 씨베이는 全體調査의 完全한 形態를 갖춘 豫行演習을 意味한다. 그러나 實際로는 두가지의 한계를 짓는것이 어려우며 많은 統計學者들이 두 用語를 병행하여 使用한다.

### 7.3 改善된 資料蒐集

#### 7.3.1 訓練

資料蒐集方法은 第4章에서 이미 說明되었기 때문에 여기서는 더이상 論議하지 않겠다. 거기서 서술된 方法 특히 面接技術과 關聯하여 사려깊은 研究는 調査員으로 하여금 誤差의 水準을 줄일 수 있게 한다. 여기서는 다른 몇가지 蒐集에 關한 重要한 觀點에 對하여 論議 하겠다.

일선 調査業務에서 蒐集된 資料의 誤差를 줄이기 위하여는 適切한 指導下에서 잘 訓練되고 業務에 積極性을 갖고 있는 調査員이 必要하다. 訓練은 調査員이 擔當할 모든 業務에 對하여 行하여져야 하며 특별한 問題點과 特別한 狀況에 부딪혔을때 處理하는 方法에 對하여 實施되어야 한다. 調査員이 擔當하고 있는 統計調査에 關하여 취하여야 할 節次에 따라 可能한 限 분리하여 實際的인 訓練이 되어야 한다. 이러한 特殊教育과 병행하여 一般的인 教育을 強化하는 것이 가장 重要하다.

### 7.3.2 監 督

調査員에 對한 監督은 일선 監督者의 核心的인 業務이다. 調査가 效果的으로 實施되기 위해서는 調査員의 그릇된 절차등 業務가 評價되고 수정이 되어야 하고 調査의 技術을 改善하는데 도움이 되어야 할 것이 必須的이다.

調査員의 業務는 다음에 의하여 評價될 수 있다.

(a) 面接의 觀察 : 監督者는 調査員을 訪問하여 일선에서 實際로 面接을 觀察하여야 한다. 특히 이것은 新規調査員이나 業務의 処理能力에 익숙치 않는 調査員의 경우에 重要하다.

(b) 不時点檢 : 이것은 誤差를 点檢하는 效果的인 方法의 하나이며 不時点檢은 다음에 討議될 一線과 事務室의 内容檢査에 보충적인 것으로 간주되어야 한다. 더우기 이것은 不誠實한 調査員들이 있어 그들을 대체할만한 時間的 餘裕가 있을때 그러한 調査員을 確認하는 重要한 手段이다. 만약 調査員들이 철저히 業務를 遂行하지 못하거나 명령된 지시나 節次를 따르지 않는다면 보수의 一部가 지급되지 않거나 해임된다는 것을 充分히 理解시켜야 된다. 또한 그들의 業務에 對한 不時点檢이 어느때든지 實施된다는 것과 不誠實하거나 부주의한 業務를 防止하기 위하여 實施된다는 것이 傳達되어야만 한다.

調査員 監督의 한가지 方法으로서 어느 國家(인도)에서 使用한 方法은 監督者가 面接調査를 녹음하는 것이다. 테이프는 나중에 再生되어 調査員의 面接技術과 節次에 對한 理解度を 分析하는 것과 아울러 規定된대로 모든 답변을 올바르게 記錄했는지

도 確認할 수 있다. 이 方法은 調査員이 後에 그의 監督者와 함께 테이프의 內容을 檢査하여 그의 業務遂行을 改善하는데 크게 도움이 될 수 있다. 반면에 그 方法이 廣範圍하게 使用되 기에는 너무나도 많은 費用이 들게 된다는 단점이 있다. 어떤 応答者는 이런 節次를 싫어 할지도 모르므로 許諾을 事前에 얻 지 않고서는 어떠한 面接도 녹음되어서는 아니된다. 또한 応答 者가 그들의 답변이 녹음된다는 것을 알게 되면 소심하게 되거나 意識的이 되어 그들의 답변의 正確性에 影響을 미치게 될 危險 도 또한 存在한다. 아마도 이 方法은 實際調査 보다는 試驗的으 로 하는 面接을 分析하는데 使用하는 것이 適合할 것이나 경 우에 따라서는 利用價值가 있을 수도 있다.

調査員들은 그들의 業務遂行에 關한 일지를 每日作成하여 그 사 본을 定期的으로 本部로 送付하여야 한다. 이 制度는 一線 調 査員들의 活動을 督促하는 役割뿐만 아니라 뒤따라야 할 資料蒐集 의 進行도 또한 可能하게 한다.

#### 7.4 資料의 確認과 內容檢査

##### 7.4.1 內容檢査의 種類

內容檢査는 必要한 경우에는 미리 定해진 規則에 따른 資料의 檢査, 質疑 수정하는 것을 의미한다. 때때로 그것은 단지 報告 書의 檢査와 質疑를 意味하는 것으로 간주되는 경향이 있다. 그 러나 內容檢査過程에서 必須的인 것은 內容檢査要員에 의한 報告 된 情報의 변경이다. 때때로 調査表의 어느 部分에는 어떠한 質

問에 關하여 어떤것이 正確한 答변이 된다는 것을 암시하는 단서가 있기 마련이다. 그러나 어떠한 변경이든지 成文化된 內容檢査 要領에 따라 嚴格하게 行하여져야 한다. 어떠한 理由로도 報告된 情報의 변경이 內容檢査要員의 個人的인 判斷에 맡겨져서는 아니된다.

資料의 內容檢査는 事務所의 內容檢査, 부호기입, 或은 센서스나 調査의 다른 段階에서도 行하여진다. 이틀테면 一線의 조사도중, 機械에서의 資料處理途中, 或은 結果의 공표이전, 技術的인 再確認等.

統計資料處理의 各段階에서 行하여지는 5가지 內容檢査形態는 다음과 같다.

(a) 調査員의 內容檢査: 調査의 마지막에 記錄의 再確認을 하는 것도 內容檢査이다. 어떠한 指針이 없을수도 있으며 指針이 段階的으로 이루어졌을 수도 있다.

(b) 一線 內容檢査: 어떤 組織的인 方法으로 一線監督者가 調査員으로 부터 調査票를 받는다. 그가 調査票를 事務所로 보내기전에 內容檢査를 해야 한다. 이 內容檢査는 임의적으로 確認하는것 以上이 되지 못할수도 있다. 한편 그것은 精密한 節次를 內包할수도 있다. 또는 重要的 項目에 對하여 檢査를 하거나 經驗한 바에 의하여 자주 틀리는 어려운 項目에 對하여 檢査를 하는 중간형도 있을수 있다. 이러한 것들은 一線內容檢査라고 불리우며 一般的으로 눈에 띄는 明白한 誤差를 除去하기 위하여 高안되었다( 수정을 위하여 調査票를 調査員에게 反送하거나

나중까지 기다리기보다 이 段階에서는 必要한 情報을 얻기 위하여 원래의 应答者를 만나는 것이 쉽다. 특히 이것은 통신망이 제대로 發達되지 못한 地域에 適合하다)

(c) 事務室에서의 임시 內容檢査: 調査票가 中央이나 地方事務所に 接受되면 때때로 대충 內容檢査가 되고 있다. 이때의 內容檢査는 事務室 臨時內容 檢査라고 하는데 만약 重要質問이 만족스럽게 答변되지 않았다면 이것을 다시 調査員에게 보낼 것인지를 決定하게 된다. 이 審査는 不完全한 調査票가 더 以上 處理되는 것을 防止하고 同時に 처음 面接과 그뒤에 行하여질 面接과의 時間의 경과를 短縮시킨다.

어떠한 調査나 센서스의 個人面接의 形態는 의심할 여지 없이 調査員이 面接을 끝낸 직후 즉시 調査票의 內容檢査를 하는 것이 가장 좋은 方法이다. 繼續되는 一線內容檢査, 事務室에서의 臨時內容檢査 或은 審査의 양과 種類는 時間, 人力, 豫算, 調査의 性質, 正確度 및 可能的 다른 要素들에 依存한다.

(d) 事務室에서 손으로 하는 內容檢査: 內容檢査過程의 다음 段階는 事務室에서 손으로 하는 內容檢査이다. 이것은 正確하고 상세한 內容檢査指針에 따라서 精密한 條件下에 實施되어야 한다. 事務室 內容檢査는 应答者와 接觸하면서 수정을 行하는 調査員의 內容檢査와는 根本적으로 다르다. 事務室 內容檢査에 있어서는 一般的으로 어떤 事項간에 가장 개연성이 크다는 것을 根拠로 決定이 내려진다. 複雜한 內容檢査를 實施하기 위하여는 內容檢査要員은 內容檢査要領을 철저히 이해해야 될 뿐만 아니라 調査票,

概念, 調査員에 대한 指針書도 또한 잘 이해해야 한다. 事務室에서 손으로 하는 內容檢査 역시 때로는 質問의 性質이나 使用할 수 있는 資料處理裝備의 種類에 따라 縮小되거나 배제될 수 있다.

(e) 機構內容檢査: 根本적으로 機械內容檢査는 컴퓨터나 통상적인 集計機械로 行하여지는데 손으로 하는 內容檢査方法和 類似하나 調査票 代身에 통상적인 집계기구에 입력되는 穿孔카드를 使用한다. 컴퓨터에서는 内部的으로 變更되어 종이의 취급과 재천공과 같은 問題들은 排除된다. 그러므로 可能하면 機械에 의한 內容檢査를 增加시키는 것이 바람직하다. 機械에 의한 內容檢査는 손에 의한 內容檢査보다 正確性和 統一性이 높다. 機械內容檢査는 카드를 穿孔한 뒤에 行하여지는데 調査員의 錯誤( 그리고 內容檢査要員과 審査要員의 손으로 行하는데서 發生하는 )뿐만 아니라 穿孔착오도 또한 수정된다. 機械內容檢査는 착오를 發見하는 手段으로 使用되는데 착오가 發見되면 손으로 수정하게 된다. 이미 지적한바와 같이 어떤 狀況에서는 機械內容檢査가 손으로 하는 內容檢査를 최소화 하거나 完全히 배제한다.

(f) 컴퓨터에 의한 內容檢査: 이것은 손으로 하는 內容檢査와 通常적인 機械에 의한 內容檢査의 長點을 結合한 것이다. 同時에 양자가 갖고 있지 않은 長點들을 많이 갖고 있다. 컴퓨터內容檢査에 있어서는 記錄의 正確性與否의 確認에 있어서 많은 條件의 一致與否를 同時에 確認할 수 있다. 더 나아가서 通常적인 機械에 의한 分類나 재천공에 의지함이 없이 自動적으로 記錄이

변경될 수 있다. 잘못된 특성치의 변경이 比較할 수 있는 그 以前의 資料에 根拠하여 行하여 질 수도 있다.

컴퓨터의 용량을 전부사용 하여야 되는 複雜한 內容檢査 프로그램은 그것을 試驗하고 정착시키는데 엄청난 조작시간을 必要로 한다. 그것이 올바르게 試驗되지 않으면 資料의 수정과 아울러 誤差도 또한 發生시키게 된다. 보통 갖가지의 모든 誤差를 수정하도록 期待하는 것은 무리다. 프로그램이 指示에 따라서 변경이 가능하도록 짜여져 있지 않으면 사무기구나 通常的 機械에 依한 內容檢査에 있어서와 같은 彈力性を 잃게될 것이다. 컴퓨터는 절충된 解決策으로 比較的 單純한 착오를 發見하는데에만 使用하여야 한다. 그리고는 손으로 수정하는 절차를 利用함으로써 彈力성이 維持된다. 이것은 특히 計劃된 時間內에 複雜한 컴퓨터內容檢査의 프로그램을 作成할 수 있는 프로그래머의 確保가 되어 있지 않은 國家에 그가치가 있게 된다.

#### 7.4.2 內容檢査의 原則

調査節次는 複雜한 活動이어서 때로는 一致되지 않거나, 不完全한 資料의 記錄을 초래하기도 하고, 명백히 잘못된 資料를 만들어 놓기도 한다. 때로는 센서스나 調査에 必要한 事項의 記錄을 漏落시키기도 한다. 調査票의 內容檢査要員은 집계를 위하여 調査票를 準備하거나 公表와 分析을 위하여 集計된 資料를 準備하는 여러가지 機能을 遂行한다. 調査되는 모든 狀況으로부터 모든 項目에 對한 完全하고 正確한 資料를 얻는 것은 不可能하다. 一般的으로 調査員이나 応答者의 고의에 의한 虛偽는 比較

的 적으나, 어떤 項目에 對한 系統的인 偏倚( 아마도 人間の 속성이나 文化的 要素의 結果에 의한 )가 發生한다는 것을 確信할 수 있다. 使用할 수 있는 資源의 한도내에서 不一致와 未詳의 것을 除去함으로써 一線에서 調査되어 接受된 資料를 可能的 限 正確하게 만드는 것이 問題이다. 調査票의 內容檢査에 있어서 지표가 될 수 있는 原則은 다음과 같다.

(i) 처음 作成된 資料에 不必要한 變경을 가하지 말것.

(ii) 記載事項中 명백한 모순과 不一致를 除去할것.

(iii) 參考할 수 있는 다른 情報을 利用하여 어떤 漏落된 項目에 對한 추가기재, 이 情報은 同一한 報告書의 다른 답변으로부터 얻어질 수도 있고, 과거의 調査에서 같은 應答者에 의해 提供된 답변에서 얻어질 수도 있으며, 類似한 답변을 해줄 수 있을 것으로 期待되는 인근의 應答者의 답변에 의해서도 얻어질 수도 있다. 어떠한 變경을 가하든지간에 미리 성문화된 內容檢査要領에 의거 變경하여야 한다.

本來의 記載事項을 變경할 때에는 처음 使用한 잉크와는 다른 색깔을 使用하여야 한다. 모든 內容檢査表記에 있어서 特別한 색깔( 이를테면 녹색 )을 지정하는 것이 좋다. 調査員은 초록색 잉크로 조사표를 作成하는 것을 금지시켜야 되며, 內容檢査要員은 다른 어떤 색깔을 使用하도록 許容해서는 안된다. 그러면 內容檢査過程에서 發生하는 變경에 對하여 監督者가 어떤 段階에서든 混同을 일으키는 일은 없게 될 것이다. 本來의 記載事項은 직선이나 원으로 表示하고 수정이 되었거나 內容檢査된 事項은 그



옆에 쓰여져야 한다. 문자나 숫자를 重複하여 기재하거나 변경시키지 말아야 한다. 이런것들은 一般的으로 읽어 볼수가 없다.

~~Female~~ 69 68                      Fmmale 69  
 Male or (69) 68

올바른 内容檢査方法

그릇된 内容檢査方法

#### 7.4.3 統計報告書의 内容檢査節次

여기서는 誤差를 네가지 큰부류로 分類하여 어떻게 하면 内容檢査를 通하여 各 境遇에 따라 資料의 질을 改善할 것인가에 對해서 考察하게 될 것이다.

(a) 漏落 : 이것은 調査員이 記載하는 것을 漏落함으로써 생기는 단순한 경우이다. 可能的 답변이 “그렇다-아니다” 或은 “男-女”와 같은 質問처럼 制限되어 있다면 올바른 대답이 때로는 조사표의 다른 事項으로부터 決定될 수 있는 경우도 있다. 예컨대 가구주와의 關係가 “처”로 나타나 있다면 性別은 女性이다. 그러나 때때로 正確한 답변은 명백하게 決定될 수 없고 단지 推測될 수 밖에 없는 경우도 있다. 예를 들면 어느부인의 나이가 기재되어 있지 않았는데 남편의 나이가 36歲이며 첫애의 나이가 15歲일 경우에 아마도 그녀의 나이는 40歲를 넘지 않는다는 것과(부인의 나이가 남편의 나이보다 많은 경우는 흔하지 않기 때문이다) 30歲는 넘을 것이라는 것을(그렇지 않다면 첫애가 태어났을때 그녀의 나이는 15歲 이하였을 것이다) 推定할 수는 있으나, 그녀의 正確한 나이를 알수는 없다. 그럼에도 불구하고 그녀의 나이를 최종집계에서 “미상”으로 기재하

는 것은 바람직하지 못하며, 内容檢査要領에 대한 적절한 推定節次가 마련되어야만 한다. 예를들자면 “처의 나이가 기재되어 있지 않을 때에는 남편보다 5살 적거나 2살 많은 범위내에서 任意의 나이를 拈할것”과 같이 특정시켜야만 한다. 이경우 31과 38사이의 임의의 수를 選擇하기 위하여 난수표가 使用될수 있을 것이다. 選擇된 이 숫자가 그녀의 나이로 조사표에 기입될 것이다. 모든 各個人의 경우에 正確히 맞을수는 없으나(예컨대 그녀의 실제 나이가 35세인 경우 임의의 수는 33일때), 최소한 올바른 연령계급(이들때면 30~39)에 記錄될 수 있는데 이것은 使用될수 있는 단서를 무시하고 그녀의 나이를 “미상”으로 分類하는 것보다는 낫다. 各個人에 對한 조그마한 착오는 合計를 하는 過程에서 높게 추정된것과 낮게 推定된 것이 서로 상쇄되게 될 것이다.

(b) 不可能한 記載事項: 이것은 부호기입 체계에 의하여 금지된 記入事項이다. 예컨대 부호1이 “그렇다”를 意味하고 부호2는 “아니다”를 나타낸다면 그외의 어떤 다른 부호도 있어서는 안될 記載事項이다. 다시 이착오는 조사표에 記載된 다른 답변을 根拠로 符號記入過程에서 수정될 수 있다.

(c) 不一致되는 事項: 不一致되는 記載事項은 두가지나 그이상의 項目들이 서로 어떤 特別한 關係를 가져야 되는데 그렇지 않은 경우에 나타난다. 예를 들자면 만약 應答者가 “당신 지난 일주일 동안에 임금이나 收入을 目的으로 일을 하셨습니까”라는 質問에 “없다”라고 대답하였다면 “지난 일주일 동안의 봉급이나

賃金は 얼마였습니까”라는 질문에는 “없다”라는 답변이 따라야만 한다. 만약 그렇지 않다면 그의 답변은 一貫性이 없게 되고 두가지 답변중에서 하나는 그릇된 대답이다. 마찬가지로 어느 土地의 總面積이 12헥타였다면 그중관개를 하고 있는 土地의 面積은 12헥타이거나 그보다는 적어야지 달리 어떤 높은 숫자는 답변이 一致되지 않는다.

(d) 不合理的 크기 : 이것은 內容檢査에서 부딪히게 되는 가장 一般的이고 어려운 問題이다. 이것은 맞지 않는것 같은 (아직은 맞는 것)것으로 부터 명백히 잘못된 답변에 이르기까지 広範圍하다. 예를 들자면 어느地域의 관개가 되지 않은 경작지에서 쌀의 平均生産量이 헥타당 4톤이었다면 각 논의 平均生産量은 3~6톤이 되어야 正常的이나, 때로는 8톤으로 보고되기도 한다. 만약 內容檢査要員이 조사표에서 헥타당 8톤으로 보고된 쌀의 平均生産量을 發見하였다면 그는 아마도 수치가 非正常的으로 높기 때문에 잘못 기재한 것으로 의심하게 될 것이나 그것이 틀린 것이라고 確信할 수는 없다. 그러나 만약 헥타당 65톤의 生産量이 보고된 것을 發見하였다면 그것은 完全히 不合理的 것이기 때문에 수정을 하여야 할 것이다. 이러한 狀況은 어떤 調査에서나 끊임없이 부딪히게 될 것이며 內容檢査要員으로 하여금 그러한 問題를 어떻게 処理해야 되는가를 알 수 있을 정도로 重要的 指針의 뼈대를 만드는 것이 重要하다. 위에서 引用한 바와 같이 헥타당 生産量이 7톤까지는 許容하고, 生産量이 7톤이상 10톤未滿이면 질의를 하고, 10톤 以上の 生産量이 보고되었을 때에

는 定해진 基準에 依하여 變更을 加한다는 등의 指針이 있어야  
만 한다. 헥타당 生産量이 8톤처럼 의심스러운 경우의 決定은  
監督者에게 넘겨지게 되겠지만( 그는 아마도 重要하다고 생각된다  
면 調査員에게 질의를 보낼 것이다), 65톤의 生産量에 처하여는  
그것은 명백히 잘못된 것으로 수정되어야 하므로 監督者의 考慮  
가 없이도 決定이 내려질수 있을 것이다.

이러한 性質의 內容檢査確認은 손으로 할수 있고 컴퓨터로도 할  
수 있는데, 事務員은 狀況에 따라 監督者에게 질의를 하든지, 그  
스스로 변경을 가하게 될 것이며, 경우에 따라서는 컴퓨터로 하  
여금 必要한대로 질문서를 프린트하거나 情報를 内部的으로 變更  
하도록 프로그램을 만들수도 있다. 전에 말한바와 같이 內容檢  
査過程에서 可能 限 컴퓨터를 使用하는 것이 훌륭한 생각이며,  
이것은 變更이 統一的으로 될수 있다는 것을 確信케 한다.

어떠한 內容檢査方法이 使用되든지간에 때로는 참값이나 없어진  
값들을 도저히 추정할수 없는 경우도 생기게 된다. 이러한 경  
우에는 關聯된 변수에 "미상"이라는 부류를 使用하여 조사표에  
기재한다. 이런 변수에 관한 統計表가 作成될때에는 이런 부류  
도 또한 표에 나타나게 될 것이며 이러한 記入內容도 總計에 包  
습이 될 것이다. 분명히 "미상"이라는 부류는 될수 있는 한  
적게 나타나야만 한다. 만약 이러한 부류가 2%내지 3%를  
초과하게 된다면 情報의 價値는 상당히 減少될 것이다. 조사표에  
報告된 實際的인 數值뿐만 아니라 그에 의하여 계산될수 있는 比  
率이나 百分率도 또한 內容檢査에 있어서 考慮되어야 하는 것이

雇傭과 賃金統計調査에서 5 個工場에 依하여 報告된 다음의 例를 생각해 보자.

바람직하고 또한 그것을 認識하는 것이 重要하다. 때때로 그것은 명백히 나타난다. 雇傭과 賃金統計調査에서 5 個工場에 依하여 報告된 다음의 例를 생각해 보자.

| 工場 | 雇傭員數 | 年間賃金支払總額(\$) |
|----|------|--------------|
| A  | 32   | 29,588       |
| B  | 20   | 21,305       |
| C  | 67   | 75,565       |
| D  | 54   | 169,173      |
| E  | 159  | 167,120      |

可能的한 명백한 確認은 各工場의 従業員當 年間平均賃金を 計算해 보는 것이다. 이 境遇의 5 個工場의 年間平均賃金は 다음과 같이 計算되었다.

| 工場 | 年間平均賃金(\$) |
|----|------------|
| A  | 925        |
| B  | 1,065      |
| C  | 1,128      |
| D  | 3,133      |
| E  | 1,051      |

비록 다른 어떤 情報가 없더라도 平均賃金이 다른 4 個工場보다 훨씬 많은 D工場에서는 착오가 發生하였다고 結論을 내릴 수 있을 것이다. 그러나 實際로는 内容檢査要員이 D工場의 資料에 對해 質疑를 할 것인지 또는 變更시킬 것인지를 決定하는데 도

움을 줄수 있는 어떤 許容할수 있는 平均賃金水準의 範圍(그國家에서 알려져 있는 支払되는 賃金이나 과거의 調査에서 報告된 資料에 근거한) 등의 形態로 內容檢査의 要領은 다른 어떤 情報을 包含하고 있을 것이다.

一貫性確認(Consistency check)이라는 一般的인 用語는 조사표의 各質問과 다른 応答者에 依하여 作成된 조사표나 過去調査에 있어서 그 応答者가 提供한 다른 源泉의 情報와의 比較를 包含한다. 内部一貫性確認은 그 조사표의 記載事項들간의 比較를 뜻한다.

#### 7.4.4 合計段階에서의 內容檢査

各 調査票의 一貫性確認도 매우 重要하지만 그것은 調査에 있어서 그 자체만으로는 適切な 正確度を 確信하는데는 充分하지 못하다. 任意的 誤差와 같이 偏倚(bias)는 상쇄되지 않는 傾向이 있기 때문에 前章에서 偏倚를 除去하는 것이 必要하다고 強調한 것을 생각해 보자. 불행히도 報告된 資料가 許容할만한 範圍內에 속하는지를 알기 위하여 各 調査票를 確認하는 것은 相對的으로 작은 偏倚(bias)를 찾아내는 것보다 커다란 任意的 誤差를 確認하게 될 것이다. 예를들자면 헥타당 65톤의 쌀생산량은 發見되어져야 하는 커다란 誤差로서 그것은 系統的인 偏倚 때문이 아니고 아마도 어떤 形態의 不注意(6.5 대신에)때문에 發生하는 것이다. 이러한 하나의 착오는 그 자체로서는 큰 것 같으나, 비록 各各의 境遇에는 작지마는 偏倚가 매 조사표마다 같은 方向으로 나타나는 것 보다는 總量統計에 훨씬 더 적은 影響을 미치게 된다.

偏倚가 있는 節次의 단순한 例로서 內容檢査體系에 依해서 發生된 誤差를 어떻게 確認할수 있는가에 대해 考察해 보겠다. 쌀 생산량의 調査에 있어서 4 m<sup>2</sup>의 正確한 틀이 추수될 區域의 畝幅을 나타내기 위하여 使用하려고 하는데 製造過程의 잘못으로 길이가 1.8 m로 만들어진 틀이 調査員에게 支給되었다고 가정해보자. 만약 이러한 틀을 正確한 것이라고 믿고 使用하였다면 각 구역의 報告된 生産量은 實際보다 約 20 %程度가 적게 될 것이다. (나뉘어진 區域은 3.24 m<sup>2</sup>인데 生産量은 4 m<sup>2</sup>라는 가정하에 計算이 될 것이다) 이렇게 헥타당 實際生産량이 5톤일 때에는 4톤으로, 6톤일 경우에는 4.8톤 등으로 보고가 될 것이다. 헥타당 4톤이나 4.8톤 등의 生産量은 許容된 範圍內에 있으므로 지적이 되지 않기 때문에 이러한 各各의 誤差들은 調査票의 內容檢査에서 드러나지 않을 것이다. 그러나 어떤 마을이나 地域의 全體數値는 實際보다 20%가 낮기 때문에 總計에 있어서는 이것은 重大한 誤差가 되며. 이러한 誤差를 發見할 수 있는 確認節次가 必須的으로 要求된다.

물론, 장비나 器具가 標準節次에 의하여 確認되어야 하고 그러한 誤差는 發生해서는 아니되나 만약 일어났다면 事務室의 資料 處理段階에서 發見하는 것도 可能하다. 이러한 것은 마을이나 地域과 같은 각기 다른 水準의 合計에서 一貫性確認을 해봄으로써 可能하게 될 것이다. 비록 각논의 生産量은 단위당 2톤처럼 매우 낮거나 8톤처럼 매우 높더라도 全地域의 平均生産量은 그렇게 많은 격차가 생기지는 않을 것이다. 비록 總生産量은 우량

이나 다른 要素들에 의하여 影響을 받게 되겠지만 期待되는 生産量은 있을 것이다. 예를들어 만약 어떤 地域에서 지난해의 生産량이 헥타량 平均이 4.3 톤이었다면 今年の 기상조건이 얼마나 좋았던지간에 금년의 生産량은 4.5 톤과 5 톤 사이에 있게 될 것이며, 內容檢査는 이러한 限界를 利用하여 各地域水準에서 行하여 질수 있을 것이다. 보고된 生産量에 있어서 그릇된 장비에 의하여 發生한 20%의 誤差는 이러한 限界를 초과하게 될 것이며 만약 그렇다면 이러한 問題를 確認하는 追加調査를 해야할 것이다.

이러한 段階의 內容檢査의 確認에 있어서는 一般的으로 其 期待値로부터 相對誤差의 매우 적은 것만을 다루게 되어 各票 査票를 조사하는 것처럼 명백하고 分명한 誤差의 徵候를 나타내지는 못한다. 오히려 正確性与否가 分明하지 못한 地域을 암시할 뿐이며 追加調査를 必要로 한다. 이러한 確認은 確定的이고 運轉이 뚜렷한 誤差의 증거를 거의 마련해주지 못하기 때문에 때로는 소홀하게 행하여지나, 大部分의 調査에 있어서는 그것은 內容檢査過程의 必須的인 特徵이 되어야 하며, 調査體系의 偏倚를 發見하는 최상의 手段의 하나이다. 이러한 確認에 있어서는 때때로 主觀的인 判斷을 包含하기 때문에 컴퓨터 內容檢査는 適合하지 못하고 事務室의 監督者에 依하여 行하여지는 것이 좋다.

合計過程의 어느 段階에서 正確性을 確認하는 다른 또 하나의 方法은 獨立된 資料를 利用可能한 다른 資料로부터 얻어진 結果와 比較해 보는 것이다. 가계조사의 맥주소비에 관한 資料는 시



람들이 그러한 費用에 관하여 낮게 대답하는 경향이 있기 때문에 결함이 있는 것으로 의심이 가는데, 各 個人의 消費는 매우 다양하기 때문에 各 調査票에 對한 一般的인 內容檢査로써는 確認하기가 어렵다. 그러나 (양조장에서 얻을수 있는) 맥주생산통계와 輸入 또는 輸出된 맥주 (貿易統計로부터 얻을수 있는)는 全國의 실제맥주소비수준을 推定할수 있게 될 것이며, 이것을 調査結果와 比較해 볼수 있을 것이다. 최소한 이것은 全國調査에 있어서 尙 報告된 水準을 測定할 수 있는 좋은 手段이 될 것이며, 이것을 다른 資料로 補完하여 地域이나 地方單位의 合計를 調整할수 있게 한다. 政府의 各部에 依하여 편집된 많은 다른 記錄도 센서스나 調査結果의 有效性與否를 確認하는데 使用될 수 있다. 예를들어 마을이나 地方政府는 租稅目的으로 경작지에 對한 記錄을 갖고 있을 것인데, 이것은 農業統計에 있어서 논밭작에 의한 경작지의 面積의 確認에 귀중하게 쓰일 수 있다. 教育庁에는 年齡別로 學生에 對한 記錄을 가지고 있을 것인데 이것은 人口調査에 있어서 아이들의 年齡分布를 確認하는데 쓰일수 있다.

#### 7.5 사후조사 (事後調査)

파일로트 씨베이 (즉 本調査가 實施되기 以前에 調査母集團의 적은 標本에 對한)가 本調査에서 發生할지도 모를 誤差에 對한 情報를 마련해주는 것처럼 事後調査 (Post-enumeration survey)는 本調査直後에 發生되었거나 發生되었을지도 모를 誤差와 그 크기에 對한 情報를 우리에게 알려준다. 事後調査는 一般的으로

그 결과를 이미 調査되어 있는 것과 比較하기 위하여 모집단의 적은 標本에 對하여 再調査를 하는 것이다. 보통은 事後調査를 本調査와는 完全히 獨立시켜 本調査와는 다른 調査員을 通하여 各 応答者가 변동되기전에 本調査 바로 直後에 實施하는 것이 좋다. 가장 훌륭하고 經驗이 있는 調査員이나 一線監督者가 擔當하는 것이 바람직한데 事後調査를 可能한 限 거의 完全에 가깝게 하기 위하여는 面接回數가 적어야 되고, 特別한 注意를 기울일수 있어야 하며, 正常的인 調査보다는 더 많은 時間을 할애할 수 있어야 한다. 두조사 사이에서 생기는 차이는 本調査에서 생긴 誤差 때문이며, 事後調査에 依한 것이 아니라고 推測될수 있다. 事後調査資料의 正確性에 있어서 完全히 信賴할 수 있는 것이 重要하며 그렇지 않으면 事後調査의 實施는 無意味한 것이다. 事後調査에서 發見하려고 하는 誤差는 調査範圍의 誤差와 內容誤差로 나누어질 수 있다. 第6章에서 본바와같이 特定調査의 規則에 依抛하여 包含되어야 할 応答者를 調査에서 調査를 하지 않았기 때문에 發生된 것이 調査範圍의 誤差이다. 調査範圍의 誤差는 다음과 같다.

(a) 틀린 명부, 조사구경계의 不明確 또는 다른 原因에 依하여 調査에서 빠진 住宅이나 事業體

(b) 경계선의 確認이 그릇되어 잘못 包含된 住宅(같은 調査에서 어느 応答者가 두번 訪問을 받게 되면 調査員에게 그런 事實을 알려주게 되어 센서스에서는 별문제가 되지 않으나, 標本調査에서는 重大한 問題가 될 것이다

(c) 어느주택에서 몇 사람이 빠졌기 때문에 그 住宅의 人口가 잘못 調査된 경우

(d) 訪問者나, 雇傭統計에 있어서 15歲未滿의 아이등 調査對象이 아닌 사람을 추출된 住宅에 包含된 사람들

內容誤差(分類誤差)는 調査範圍에 맞게 包含 調査된 사람으로부터의 맞지 않는 답변을 말한다.

때로는 費用을 줄이기 위하여 事後調査는 調査範圍誤差를 確認하기 위하여 設計되고 內容誤差의 確認은 內容檢査나 다른 方法에 의하도록 미루어지기도 한다. 특히 費用이 關係되는 한에 있어서는 이러한 方法에 對한 훌륭한 理由로서는 모든 調査票의 內容을 반복 調査하는 것은 不必要하며, 올바른 家口가 이러한 家口內에 사는 올바른 사람들이 正確히 調査되었는지를 決定할수 있는 質問에 集中的으로 質問을 할 수 있다는 것이다. 그러나 事後調査는 다른 形態의 誤差(예를 들면 調査員이 유도질문을 하여 답변에 影響을 미친 경우)에 對한 情報를 마련해 줄수 있으며, 事後調査의 調査員은 調査表의 內容을 全部 反復 調査하여 可能하다면 內容誤差와 調査範圍誤差 두가지 모두를 確認하는 것이 바람직하다.

만약 本調査의 重大한 결함이 發見되면 事後調査結果는 關聯된 統計表를 調整하는데 使用이 될 수 있다. 最近에 적어도 2個国 또는 3個国에서 事後調査에 의하여 센서스에서 상당히 적게 調査된 것이 發見되어 인구센서스결과를 調整하였다. 이러한 行動은 事後調査에는 標本誤差가 있기 때문에 경솔하게 취할수 없다.

이러한 경우 統計人은 標本誤差가 包含되어 있음에도 不拘하고 事後調査에 의하여 調査된 人口가 센서스에 依하여 本來 報告되었던 人口보다 더 正確하게 調査되었다는 것으로 滿足하는 것을 意味한다.

#### 7.6 改善된 資料処理方法

內容檢査活動은 資料蒐集過程에서 發生된 誤差를 可能な限 確認하여 除去하려는 목적으로 實施된다. 그러나 誤差는 內容檢査過程에서도 發生될 수 있는데 특히 그것은 손으로 行하여질때 그러하다. 內容檢査要員에 依하여 發見되어야 할 誤差를 發見하지 못하거나 때로는 맞는 답변을 그릇되게 변경시키는 경우도 있다. 이것은 특히 內容檢査指針이 모호하여 잘못 解釋하거나 或은 不完全하여, 부딪히게 되는 모든 狀況에 適用할 수 없을 경우에 일어난다. 게다가 資料処理途中에 다른 誤差도 생겨날 수 있는데 특히 符号記入過程에서 內容檢査要員들이 適切な 期間內에 모든 節次를 完了하지 못하였을 때에는 더큰 困難이 뒤따르게 될 것이다. 그러므로 內容檢査나 符号記入에 있어서는 그 生産과 질에 基準을 세워야하고, 그러한 觀點에 처하여 점검하는 공식적인 體系를 세우도록 중고할만 하다. 이러한 두가지 管理의 形態는

(a) 從事者에 對한 再教育이나 교체

(b) 節次自體에 對한 再點檢

의 必要性에 對한 指標들 마련해 줄수 있다. 그 目的은 최소한의 時間과 費用으로써 許容할수 있는 水準의 正確性을 얻는데

있다.

비록 새로운 業務에 對한 推進実績을 確定하기는 힘이드나 後에 변경될 수 있다는 것을 認識한다 하더라도 推進計劃도 없이 繼續하거나 活動이 어느정도 繼續된 後에 그것을 導入하는 것보다는 처음부터 確定하여 實施하는 것이 더 좋다.

初期段階에서는 “猶豫期間”을 두어 그기간내에는 推進計劃에 해당하지 않더라도 어떤 不利益을 주어서는 안된다. 推進計劃은 事務員들에게 客觀的인 基準을 부여하고 그들에게 期待하는 것에 對한 보다 나은 理解를 提供해 준다. 僱傭者들은 이러한 基準이 전혀 없는것 보다는 一般的으로 일을 더 잘하게 되고 職業에 對한 滿足減을 얻게 된다.

一般的인 信念에 反하여 僱傭員으로 하여금 매일 業務実績의 最低限度를 達成하도록 要求하는 것이 반드시 正確性を 減少시키지는 아니한다. 착오를 일으키는 경향이 있는 사람들은 最低기준의 維持를 要求받더라도 그들이 평상시 하는것처럼 일을 천천히 하기 때문에 역시 착오를 일으키게 될 것이다. 推進計劃의 樹立은 行政을 담당한 사람으로 하여금 그 活動에 對한 實際完了日字를 알 수있게 해준다. 더우기 그들로 하여금 센서스의 企劃에 있어서 이러한 觀點에 對하여 合理的으로 確定할 수 있는 豫算의 推定을 可能하게 한다.

매우 規模가 큰 業務의 初期段階에서는 內容檢査와 符號記入의 業務가 全體적으로 모두 確認된다. 역시 여기에서도 実績에 있어서 誤差를 發生시키는 比率이 許容水準을 超過한다면 不利益을

주기전에 “유예기간”을 준다. 一般的인 節次는 特定된 일의 單位數나 調査票數에 關하여 事務員의 모든 業務에 對하여 確認을 한다. 이기간이 지난 다음에는 만약 착오율이 許容된 水準 內에 있으면 그의 業務中에서 標本을 추출하여 確認을 한다. 만약 착오율이 許容水準을 초과하게 되면 일의 單位의 追加에 對하여도 모든 業務가 確認을 받아야 한다. 만약 이러한 두번째 期間의 經過에도 不拘하고 착오율이 要求되는 水準을 達成하지 못한다면 再教育이나 交替가 고려되어야 한다. 標本으로 實施한 確認에서 資格이 있다고 判定이 되었더라도 후에 또다시 許容限界를 초과하게 되면 모든 業務의 確認을 받아야 하며 標本確認에서 다시 資格이 있는 것으로 判定을 받아야 한다. 規模가 작은 活動에 있어서는 細密한 節次는 實用的이 되지 못할 것이다. 오히려 적은수의 內容檢査要員이나 符號記入要員이 세밀한 監督下에 단순하고 비공식 절차에 의하여 수행하는 것이 效果的인 것이다. 使用된 事務室의 節次에 對한 管理方法이 무엇이든지 간에 各 內容檢査要員으로 하여금 그의 業務가 確認을 받게되고 後에 監督者에 依하여 評價를 받아야 된다는 것을 認識하도록 하는 것이 重要하다. 이러한 認識은 그들을 자극하여 可能한 限 正確히 일을 處理하도록 할 것이다. 또한 이러한 方法으로 職務를 수행하도록 격려함으로써 그가 취한 行動이 쉽게 確認이 될수 있는 것이다. 예를들면 調査員이 기재한 本來의 記錄을 알수 없게 만들거나, 지우거나, 파괴하지 말아야 된다는것과 區別될수 있는 색깔로 調査票에 表示를 하거나 깨끗이 써야된다고 지시를 하

어야 한다. 그렇게 하지 않으면 内容檢査의 確認者가 調査員이 처음에 報告하였던 情報가 무엇이었는가를 正確히 알수 없게 되고, 内容檢査要員이 變更시킨 것에 對하여 内容檢査要員의 業務에 대한 適切한 確認을 不可能하게 한다.

## 第 8 章 資 料 處 理

### 8.1 概 要

資料處理은 調査票上의 情報로부터 統計를 作成하기 까지의 모든 作業을 包含하는 内容이라 할 수 있다. 여기에 包含되는 作業은

- (a) 調査票의 檢討 및 오류修正
- (b) 資料의 符号化(Coding)
- (c) 資料의 變換(Computer가 處理할 수 있는 매체로의 轉換 및 資料處理裝置에 入力시키는 단계)
- (d) 資料의 分類集計(Sorting)
- (e) 資料의 演算等を 들 수 있다.

政府機關에 종사하는 統計專門家와 기타機關에 종사하는 統計專門家가 担当하는 業務의 차이점의 하나가 大量資料를 取扱해야 할 必要性 有無라 할 수 있다. 人口 1億인 나라에서 15개의 調査項目을 가지고 人口센서스를 實施한다면 결국 15億個의 情報를 얻는 結果가 될 것이다. 이와 같은 방대한 資料는 各 項目別로 別途集計가 되어야 함은 물론이고 項目간의 Cross-classification도 作成하여야 한다. 이것은 資料處理에 많은 時間이 所要되고 적절한 資料處理方法을 구상하는 데도 많은 時間이 必要하다는 意味가 된다. 이 章에서는 많은 資料處理作業中 一部에 대해서만 言及하고자 한다.



資料處理過程中 많은 부분이 사람 손으로 처리되는 경우가 있으나 資料處理를 迅速, 正確하기 하기 위해서는 機械的인 도움을 받는 경우가 大部分이다. 이러한 機械的인 도움의 수단에는 간단한 携帶用 計算機에서부터 複雜한 電子計算組織까지 있는데, 전자 계산조직은 技術의 發展과 比例하여 重要性이 增大되고 있으며 價格 또한 下落되는 趨勢이다. 實際로 資料處理方法을 선택함에 있어서는 處理할 資料의 量, 分析形態 및 分析의 複雜性 그리고 資料處理가 수행되는 場所(예컨대 本部, 市道, 其他地方 등) 등을 감안해야 한다. 關聯資料의 量이 많고 고도로 複雜한 分析을 요하거나 한 군데서 集中的으로 資料處理를 必要로 하는 統計調査의 資料處理에 있어서는 컴퓨터處理方式이 바람직 할 것이다. 반대로 말한다면, 小規模 統計調査 또는 地方에 分散하여 資料處理가 要求되는 경우에는 手作業 또는 단순한 계산기로 處理하는 것이 有利할 것이다.

本章에서 說明할 資料處理는 다음의 세 부분으로 나누고자 한다.

- (a) 資料處理組織의 計劃
- (b) 위에서 言及한 資料處理의 다섯단계
- (c) 機械的인 資料處理의 手段 및 處理方法

## 8.2 資料處理組織의 計劃

資料處理의 遂行에 앞서 많은 事前業務의 處理가 完結되어야 하며 計劃이 樹立되어 있어야 한다. 調査票를 接受하고 적

적절한 分類計劃이 先行되어야 한다. 이러한 作業은 小規模調査에서는 많은 問題點을 야기시키지 않으나 大規模調査에서는 적절한 事前措置가 이루어지지 않으면 크나큰 차질을 초래할 수 있다. 調査票를 습기나 곤충 같은 다른 損傷要因으로부터 保護하기 위한 적절한 措置가 이루어지지 않으면 안된다. 一般的으로 各 調査上의 情報은 비밀이 保障되어야 하기 때문에 적절한 保安場所가 마련되어야 함을 명심해야 할 것이다.

人口센서스 같은 大規模의 統計作成에 있어서는 관련 調査票類 등의 管理自体를 하나의 큰 作業으로 볼 수 있다. 예컨대 保管場所는 萬t의 종이무게를 지탱할 수 있을 만큼 튼튼해야 하고 調査票類의 이동이 용이해야 하며 保管場所와 資料處理室間에는 적당한 간격을 유지해 주어야 한다. 獨立된 복도나 계단이 確保되어야만 各 資料處理室間에 資料의 운반에 완벽을 기할 수 있다. 調査票類(대개는 조사구별로 묶어져 있을것임)는 적당한 크기로 安全하게 운반될 수 있도록 묶어주고 식별이 용이하도록 조건표를 붙여야 할 것이다. 매 조사표 다발이 保管場所에 들어오거나 나갈때 또는 資料處理室間에 이동할 때等 모든 移動事項을 빠짐없이 記錄해야 한다. 하나의 調査票다발이 잘못 운반되었거나 다른場所에 保管되고 있다면 大規模調査의 경우 이를 다시 찾아내는데 3個月은 걸릴 것이다.

이 밖에 事前에 計劃樹立을 요하는 事項으로서는 資料處理 段階別 所要人力, 作業處理順序, 所要裝備, 符号目錄(Code lists), 오류검사, 프로그램, 작업표(Working sheets) 및 正確性 檢討등을

들 수 있다.

作業에 따라서는 處理速度가 다를 수 있다는 점을 명심해야 한다. 결론적으로 資料處理部署間에 調査票類가 円滑하게 운반되기 위해서는 段階別 作業量이 달라져야 할 것이다. 더구나 作業內容에 따라서는 訓練이나 技術을 要하는 경우가 있으며 訓練을 받는데 充分한 時間을 미리 確保해 두어야 한다. 또한 資料處理에 所要되는 作業量은 一般的으로 增加하는 경향에 있다.

大規模 統計調査에 있어서는 作業日程表를 準備하는 것이 바람직한데, 作業日程表에는 每 資料處理 段階別로 作業完成日程表를 作成해 넣을 수 있을 것이다. 每 資料處理段階別로 이에 決定된 作業完了日과 實際로 處理되고 있는 作業進行狀況을 比較하여 調整하는 管理카드 또는 管理台帳( Control card, Control ledger )을 使用할 수도 있다. Ogives ( 累積度數曲線 )를 우편 조사( Mail survey)의 回收率( 百分比 )記錄에 活用할 수 있는 것과 마찬가지로 資料處理上의 每 段階別 進度( 百分比 )를 이 Ogives를 利用하여 그려 낼 수 있다.

이런 形態의 記錄은 資料處理過程에 投入된 人力에 대해서 個人別로 맡은 業務의 소화능력을 평가하는데 使用될 수도 있다. 명백하게 소화능력이 없는 사람에 대해서는 다른 種類의 일을 맡기고 다른 사람으로 대체시킬 수 있을 것이다.

統計調査에 따라서는 地方事務所에서 資料處理가 이루어지는 경우가 있으므로 이러한 地方事務所の 責任者가 資料處理에 必要한 作業遂行을 計劃하고 管理하지 않으면 안될 것이다. 이러한 일

은 責任者로서는 가장 重要한 役割중의 하나가 될 것인데 만일 資料處理의 어느 한 過程이 遲延되면 그 다음過程이 연쇄적으로 遲延되어 全体作業의 마무리가 늦어질 것이기 때문이다.

### 8.3 資料處理作業

#### 8.3.1 檢討 및 오류수정

앞에서 우리는 資料의 內容檢討 및 오류수정에 관한 몇가지 問題點을 세밀히 檢討한 바 있다. 資料의 內容檢討와 오류수정은 資料處理의 첫번째 과정이며 다음 과정이 시작되기 전에 完結되어야 할 作業이다.

이 과정이 끝나면 調査票上的의 오류는 完全히 除去된다. 오류는 반드시 修正되어야 하며 누락은 補完되거나 미상으로 分類시켜야 한다.

#### 8.3.2 符號記入(Coding)

우리가 부호기입이라고 말할 때는 흔히 어떤 特性을 코드번호로 대체하는 것을 의미한다. 우리에게는 코드번호가 주어지면 이에 対応하는 특성이 식별될 수 있는 코딩시스템이 必要한 것은 명백한 일이다.

##### 8.3.2(a) 특성의 코딩

대개의 機械的인 계산조직에서는 번호에 의해서만 作業이 處理되므로 이러한 形態의 作業處理方法을 선택하는 경우

에는 모든 특성에 코드가 부여되어야 하는 것은 필수적이다. 다시 말하면, 이것은 문자에 의한 分類를 번호로 번역하는 과정이라 할 수 있다. 여기에는 가령 남자에는 “ 1 ”, 여자에는 “ 2 ”라는 번호를 부여하여 성별을 分類하는 것과 같이 單純한 作業에서부터 수백개의 職種 또는 職業群別로 相応하는 코드번호를 부여해야하는 아주 複雜한 分類도 있다.

### 8.3.2 (b) 變數의 코딩

그러나 가령 觀察値가 이미 數字로 되어 있는 경우에도 變數에 코드를 부여하는 것이 有用한 경우도 간혹 있다. 이를테면, 우리가 달러表示 個人別 年間所得資料를 가지고 있다 하더라도 실제 製表는 所得階層別로만 必要한 경우가 있을 것이다. 이런 경우에 우리는 다음과 같이 코드를 부여할 수 있다. 即,

| 年 間 所 得 (\$)  | 코 드 |
|---------------|-----|
| 500 以 下       | 0   |
| 501 ~ 1,000   | 1   |
| 1,001 ~ 1,500 | 2   |
| 1,501 ~ 2,000 | 3   |
| 2,001 ~ 2,500 | 4   |
| 2,501 ~ 3,000 | 5   |
| 3,001 ~ 3,500 | 6   |
| 3,501 ~ 4,000 | 7   |
| 4,001 以 上     | 8   |
| 未 詳           | 9   |

以上은 所得資料를 10個의 集團으로 나눈 것인데 하나의 Digit로 構成된 코드로서는 10個以上 만들 수 없음을 알 수 있다. 이 경우 만일 所得階層을 더 늘린다고 하면 2 Digit 코드를 使用해야 할 것이다. (예컨대 4,001 달러 이상을 더 세분하는 경우) 지금까지 특성과 變數를 区分해 보았으나 이 경우에는 그 区分이 아주 明確하지가 않다. 따라서 어떤 個人의 所得(變數)이 2,320 달러라고 하는 것은 그가 2,001 ~ 2,500 달러의 所得階層에 속하는 특성을 지녔다는 것과 同一하기 때문이다. 이러한 意味에서 모든 코딩은 특성에 코드번호를 配定하는 것을 意味한다.

### 8.3.2 (c) Hierarchical 코드구조

2개 이상의 자리수로 構成되는 코드를 必要로 하는 아주 세분된 分類가 必要한 경우, Hierarchical 코드構造를 使用하는 것이 바람직하다. 이러한 코드구조를 利用하면 첫번째 자리에는 集團을 分類하고 다음자리에는 小集團 (Subgroup) 또는 個別資料를 分類할 수가 있다. 지금 貿易計統를 國別로 分類하기 위하여 國別 코드를 設計한다고 假定하자. 가장 간단한 方法中の 하나는 國名을 알파벳순으로 정리한 다음 3자리수코드를 使用하여 첫번째 國家인 Abu Dhabi에는 001, 다음 國家인 Abyssinia에는 002 를, 그 다음은 003 ………과 같이 코드를 부여하는 形態이다. 그러나 우리는 흔히 貿易形態를 分析하기 위하여 個別 國別로는 물론 아시아, 中東, 유럽등과

같이 地域別로도 分類하는 경우를 볼 수 있다. 위에서 說明한 단순코드분류를 쓰면 일일이 모든 國家에 코드를 부여하여야 하므로 어렵고 또한 많은 時間이 所要된다. 이런 方法으로는 가령 아시아의 總 貿易資料를 얻기가 아주 어렵다.

이러한 어려움을 解決하기 위하여 첫번째 자리에는 地域을 식별할 수 있는 코드를 부여하면 될 것이다. 즉,

0 아 시 아

1 아프리카

2 中 東

3 유럽 등과 같이 分類하고 두번째와 세번째 자리

에는 그 地域에 속한 國家를 分類한다. 이렇게 하여 中東地域 (코드번호 2)의 첫번째 國家인 Abu Dhabi는 코드 201, 아프리카地域의 첫번째 國家인 Abyssinia는 101로 코드가 부여된다. 이와 같은 方法으로 각 地域의 모든 國家가 몇개의 國家集團(地域)으로 分類될 수 있다. 아시아의 對外去來總額은 첫번째 자리에 0으로 되어 있는 國家를 합하면 될 것이다.

以上の 例에서 하나 더 言及해야 할 事項은 이런 形態의 코드構造를 利用하면 交易量이 아주 작아서 個別, 國別 資料는 별 重要性이 없는 모든 國家를 “其他國家”로 묶을 수가 있다는 점이다. 그러나 단순한 分類方法으로 其他國家群을 만들려고 하면 각 地域에서 國家를 모아야 하므로 어떠한 資料處理方法을 使用하더라도 아시아, 中東 등과 같이 地域別 資料의 集計는 不可能할 것이다. 두번째 方法을 利用하면 각 地域別 其他코드가 있

을 것이므로 地域別 總計는 아무 어려움 없이 計算될 수 있을 것이다.





우리가 위에서 檢討한 分類方式은 단순 2 단계 Hierarchical 分類方式이 있으나 첫번째 자리에 大分類, 다음 자리에 中分類等과 같이 계속적으로 코드를 부여할 수 있는 多段階 Hierarchical 分類法을 택할 수도 있다. 이것은 아주 有用한 方法으로써 資料의 細分程度에 따라 資料를 分類해 낼 수 있는 方法이다. SITC, ISIC 또는 ISCO (12章 参照)와 같은 主要 國際分類시스템은 4 또는 그 以上の 자리수로 構成되어 있는데, 願하는 細分程度에 따라서는 3, 2 또는 자리수로 分類될 수도 있다.

#### 8.3.2 (d) 調査票上에 주어진 코드

지금까지는 通常 応答者の 답변내용을 応答者 또는 調査者が 記載하여 実査担当者에 의해서 코드를 決定하고 分類하는 경우를 전제로 코드를 說明하였다. 이제 또 다른 重要한 코딩시스템에 대해서 알아 보기로 한다. 즉, 調査票上에 주어진 코드가 그것이다. 이 시스템에서는 코드가 調査票上에 미리 주어지며 調査者(郵便調査에서는 応答者)가 해당코드에 “○” 표시, “V” 표시 또는 밑줄을 긋기만 하면 된다. 가장 一般的인 形態는 아래에서 보는 바와 같이 “V” 표시를 할 수 있는 一聯의 사각형의 나열이다.



## 婚 姻 状 態

|       |   |
|-------|---|
| 未 婚   |  1 |
| 有 配 偶 |  2 |
| 離 婚   |  3 |
| 死 別   |  4 |

이러한 코드체계는 말로서 応答된 内容を 코드화하는 方法에 비하여 몇가지 명백한 利点이 있으나 反面 단점도 있다. 장점으로는,

(a) 担当部署의 業務量이 적어지므로 費用이 節減되고, 調査結果를 迅速하게 集計할 수 있다.

(b) 별도의 코드紙에 移記할 必要가 없으므로 移記上의 오류를 피할 수 있다.

(c) 応答者가 답변할 때 發生할 수 있는 애매모호한 内容を 줄일 수 있다.

단점으로는,

(a) 該當 코드에 대한 表示外에는 알 수가 없으므로 調査票의 效果的인 内容檢査가 어렵고 경우에 따라서는 不可能하기도 하다.

(b) 모든 応答는 미리 정해진 범주에 속하도록 되어 있으므로 形편에 따라서는 分類된 범주가 적합하지 않을 수도 있다. 이러한 단점은 마지막 란에 “其他” 또는 “詳述하시오”와 같은 표시를 하여 補完할 수가 있다.

(c) 応答者 자신이 調査票를 作成할 때에는 混同을 일으켜 코드를 잘못 선정할 可能性이 있다.

이러한 코딩시스템 (Precoding)에서 發生할 수 있는 問題點의 한가지 예를 든다면 婚姻狀態의 코드分類에서 본 바와같이 우리는 “別居”를 따로 分類하지 않았다는 점이다. 여기서 우리가 分明히 알 수 있는 것은 結婚은 했으나 離婚이나 死別이 아닌 사람은 그들이 實사 別거하고 있는 경우에도 有配偶로 分類되는 것이 明白하다고 생각할 수 있다.

그러나 이것이 應答者에게 항상 分明한 것이 아니고 質問에 대한 應答에 一관성이 없을 수도 있으므로 調査結果는 信賴度가 낮아질 것이다. 이러한 경우에는 不正確한 應答을 가려낼 方法이 없으므로 우리가 이를 간과하는 한 問題點으로 남을 것이다.

통상 이러한 方法은 應答의 選擇이 제한된 경우에 適合하고 또한 일정 水準以上の 信賴度를 기대할 수 있는 熟達된 調査員을 投入하는 調査에 適合하다. 그러나 應答者 자신이 調査票를 作成해야 하거나 複雜한 質問 혹은 可能한 應答의 범위가 넓은 경우등 應答의 信賴度가 疑問시되는 調査에는 適合하지 못한 코드分類方法이라 할 수 있다.

### 8.3.3 資料의 轉換(Data Transfer)

資料의 轉換이란 手集計, 計算機 또는 EDPS로 資料를 處理하는 데 適合한 形態로 정리하거나 轉換시키는 手段을 意味한다.

코딩을 必要로 하는 모든 調査票上의 應答은 Pre-coding 이 아닌한 반드시 부호기입을 하여야 한다. 부호기입에는 直接 調査票上

에 할 것인가 아니면 전환표( Transfer Sheet )를 사용할 것인가를 決定해야 한다. 대부분의 質問이 Pre-coding되는 경우에 전환표를 사용한다는 것은 부호기입자들이 단순히 부호기입 결과를 전환표에 移記한다는 意味로서 이는 時間浪費와 오류발생을 수반하기 마련이다. 이 경우에는 오히려 調査票上에 応答結果를 分類하는 것이 바람직할 것이다. 따라서 Pre-coding 에서는 앞에서 본 바와 같이 코드번호를 調査票에 미리 印刷해 놓거나 코드번호를 記載할 수 있도록 빈칸을 남겨두는 形態를 취하게 된다. EDPS로 資料處理를 수행하는 경우에는 천공카드를 必要로 하는데 이 方法이 不利한 것은 調査票로부터 직접 카드를 천공해야 하는 점이다. 이것은 상당히 힘든 作業이며 전환표로부터 천공하는 것보다 오히려 오류가 더 많이 發生한다. 調査票 디자인이 複雜하고 調査項目間에는 여러가지 관련성이 있으나 천공결과의 論理的인 一致性이 維持될 수 있도록 調査票가 設計되지 않으면 안된다. 전환표의 사용여부를 決定하는 데는 여러가지 요소가 있다. 一般的으로 大部分의 統計作成機關에서는 可能한 限 전환표의 사용을 기피하는 경향이 있다. 왜냐하면 전환표를 사용하면 資料處理過程이 늘어나게 되며 移記에 따른 오차의 發生可能性을 피할 수 없기 때문이다. 그러나 調査票가 특히 複雜하여 전환표를 사용하여 보다 效率的인 資料處理를 할 수 있는 경우도 있다.

전자계산조직에 사용되는 Transfer Sheet의 設計에 있어서는

担当者가 컴퓨터 管理者의 咨문을 구하는 것이 必須的이다. 카드친공 또는 다른 方法에 의한 컴퓨터입력자료의 準備는 잘 設計된 전환표로 하는 것이 容易할 뿐 아니라 費用 節減效果도 크다. 이러한 部分이 바로 컴퓨터 專門家가 統計業務 종사자보다 經驗이나 技術이 앞선 分野이므로 Data Sheet의 設計時 이들 專門家の 咨문을 받으면 效果的이다.

機械集計方式이 아닌 手集計時의 자료전환에는 몇가지 方法이 있다. 그중 한가지는 여러 欄( Column )으로 構成된 作業표에 資料를 전환하는 것으로 한 줄( line )을 Data Sheet 한장, 한 欄( Column )에 하나의 變數 또는 特性으로 使用할 수 있도록 되어 있다. 이 方法은 적은 量의 資料에는 適合하나 資料가 많을 때는 適合하지 못할 뿐 아니라 每 移記時마다 오류의 發生可能性이 높다.

수집되는 資料의 調査項目이 간단하다면 Data Sheet 한장分の 資料를 카드한장에 입력시켜 資料의 集計를 容易하게 할 수 있다. 여기에서는 특별히 印刷할 必要는 없으나 Blank Card로부터 Index Card까지도 準備해야 한다. 예컨대 入国者 統計調査에서 다음과 같은 特性이 蒐集되었다고 하면,

性 別  
年 齡  
国 籍  
居 住 国  
入国目的

## 滯留期間

카드의 設計는 다음과 같이 될 수 있다.

| 카드 일련 번호 |         |         |
|----------|---------|---------|
| 性 別      | 年 齡     | 国 籍     |
| 居 住 国 家  | 入 国 目 的 | 滯 留 期 間 |

調査結果 얻어진 情報은 카드의 적절한 곳에 들어 갈 것이다. “国籍”, “居住国”, “入国目的”에 各各 적절한 코드를 부여하는 것은 容易할 것이다. 카드는 보통 処理되지 않을 때는 반듯하나 集計用으로 處理될 때는 왼쪽 상단을 잘라서 쓰면 区分이 確實해 진다. 실제 集計時에는 Data Sheet 보다는 이들 카드를 處理하는 것이 훨씬 쉬울 것이다.

資料處理를 機械集計로 하는 경우 機械가 判讀이 可能한 資料의 전환은 여러가지 方法이 있는데 이는 使用되는 機械의 形態에 따라 달라지게 된다. 이러한 作業은 “Data Capture” 혹은 단순히 자료입력(Data Entry)이라고 부른다. Data Capture의 技術에 관해서는 이 章의 마지막 부분(Mechanical Aids to Data Processing)에서 說明될 것이다.

### 8.3.4 資料의 分類集計

#### 8.3.4 (a) 手分類

사람의 손으로 資料가 處理되는 어떠한 統計調査에 있어서도 資料를 적절한 集團으로 分類하는 것은 資料處理

作業中에서 가장 많은 시간이 所要되는 作業中の 하나이다. 그러므로 가장 效率的인 作業過程을 선정하기 위하여 신중하게 計劃을 樹立하여야 한다.

一般的으로 자료분류작업은 각 個別單位를 分類하고 分類別로 數를 計算하는 것을 意味한다. 그러므로 分類作業은 우리가 必要로 하는 모든 集計를 쉽게 할 수 있도록 資料를 再配置하는 것이다. 예컨대, 우리가 위에서 본 入国者の 예를 보면 우리가 알고자 하는 集計結果로서 入国目的 및 滯留期間別 入国者數를 생각할 수 있다. 우리는 아래의 分類를 假定할 수 있다.

| 入 国 目 的   | 滯 留 期 間   |
|-----------|-----------|
| 事 業       | 3 日 未 滿   |
| 就 業       | 3 ~ 5 日   |
| 教 育 및 訓 練 | 6 ~ 8 日   |
| 會 議       | 9 ~ 12 日  |
| 觀 光       | 13 ~ 20 日 |
| 其 他 目 的   | 21 ~ 30 日 |
| 未 詳       | 30 日 以 上  |
|           | 未 詳       |

資料를 正確하게 分類하기 위하여 첫째로 必要한 것은 入国目的의 別集計이다. 入国目的으로 “事業” …… 未詳順으로 資料를 分類한 다음 各 滯留期間別로 다시 細分한다. 끝으로 各 分類

別 사람수를 計算하여 必要한 結果表를 完成할 수 있다. 앞에서 說明한 바와 같이 만일 資料가 카드에 入力되어 있다면 이러한 作業은 더욱 쉬울 것이다.

#### 8.3.4 (b) Tally 表示法

다른 한가지 方法으로는 各 分類內의 數值를 Tally로 만드는 것이다. Tally는 적절한 곳에 表示를 하여 만드는 것인데 그 表示는 후에 모든 數值를 計算하여 合計를 쉽게 낼 수 있도록 5를 하나의 그룹으로 하여 만든다. 이제 性別 入國目的別 入國者數를 나타내는 表를 만들어 예를 들어 다음과 같은 作業표를 作成하였다.

|           | 男 子     | 女 子     |
|-----------|---------|---------|
| 事 業       | /// /   | //      |
| 就 業       | //      | ///     |
| 教 育 · 訓 練 | ///     | //      |
| 会 議       | /       | ///     |
| 觀 光       | /// /// | /// /// |
| 其 他       | //      | /       |
| 未 詳       | /       | //      |

以上과 같이 그룹別로 사람수와 Tally 표시를 집어넣는다.

그러므로 사람의 손으로 資料를 分類하고 集計하는 경우에는 分類別로 資料를 正確하게 区分하여 觀察值 하나하나를 세어서 集

計할 수도 있고 Tally로 묶어서 集計할 수도 있다. 어느方法을 택할 것인가는 여건에 따라 다르겠으나 대개는 分類는 카드로 하는 것이 쉽고 大量資料는 Tally로 묶는 것이 편리하다. 경우에 따라서는 카드와 Tally를 並用할 수 있다.

#### 8.3.4 (c) 手分類의 限界

앞에서 言及한 入国者の 例는 Two-way tabulation 즉, 두가지 特性의 Cross-Classification의 作成과 관련된 것으로 特性별로 각각 7개와 8개의 集團으로 分類했었다. 따라서 이와 같이 단순한 結果표 작성을 하기 위해서도  $7 \times 8 = 56$ 개의 서로 다른 集團(Group 또는 Cell)으로 만드는 作業이 必要하다. 단순한 分類와 集計를 요하는 이러한 정도의 作業量은 사람 손으로도 별다른 어려움 없이 감당할 수 있겠으나 우리가 作成하는 대개의 통계結果표에서는 보다 세분된 Cross-Classification을 요한다. 예컨대, 취업자에 대한 居住地別 職業別 주당임금別 Three-way Tabulation을 하는 경우 地域이 287個 職業이 1,00가지 그리고 임금그룹이 15개로 나누어 진다면 Sorting 또는 Tally로 묶어서 만들어야 하는 集團은  $287 \times 1,100 \times 15$ 개가 된다. 이것은 거의 500만 Cell이 되는데 사람손으로는 도저히 作業遂行이 不可能 할 것이고 실사 편리한 機械的인 方法으로도 處理가 容易하지 않을 것이다. 이와 같은 경우에 컴퓨터處理가 아주 有用하다. 컴퓨터는 大量資料의 分類處理에 있어서 融通性이 높고 따라서 複雜한 分類(Classification) 특히 컴퓨터 登場 以前에는 시



도조차 못했던 小地域分類의 處理를 可能하게 하였다.

機械處理方式에 의한 分類와 統計票 作成에 대해서는 다음 章에서 다시 言及될 것이며 機械處理에 利用되는 基本的인 機械形態에 대해서도 說明될 것이다.

### 8.3.5 演算作業

전형적인 演算處理作業으로는 合計, 단위 환산, 사사오입, 比率 및 構成比計算, 人口推計, 標本調査에 의한 추계시 標本誤差計算, 平均 및 標準偏差計算등이다.

이러한 作業의 어려움에 대해서는 10 章에서 一部 言及하였으며 이러한 일에 종사한 사람들은 보다 단순한 資料處理에 익숙해 있기 때문에 여기서는 상세한 言及을 피했다. 그러나 統計的으로 볼 때는 計算上의 오류가 없다는 것이 아주 重要的 意味를 갖는다는 점을 強調하고 싶다. 사람손을 거친 作業은 반드시 再檢討되어야 하며 경우에 따라서는 원래의 計算方式과 다른 方式으로 再檢討될 수도 있다.

#### 8.3.5 (a) 사사오입

사사오입은 우리가 자주 접하는 計算中の 하나로써 여기서 특별히 言及코자 한다. 이미 第6 章에서 사사오입상의 오류를 言及할 때 이는 다른 오류와는 달라서 바람직한 근사치에 接近될 수 있도록 신중히 사사오입되어야 한다는 점을 지적한 바 있다. 사사오입은 作成되는 統計資料가 信賴할 수 있는 이상의 자리수를 가질 때 利用되며 統計利用者들이 實際로

必要로 하는 水準以上の 正確性이 있다고 믿게 할 必要는 없다. 農業統計調査의 例를 든다면 어느 地域의 集計된 소의 數가 386,294 마리로 나타났다면, 이 數字는 6 자리수 즉 6 자리까지 信憑性이 있는 수라고 말할 수 있는데 이는 소의 마리수가 최종자리수까지 正確하다는 의미이다. 실제에 있어서 소의 마리수를 이렇게까지 正確하게 測定할 必要는 없을 것이며 386,000 마리 정도로 만족할 수 있을 것이다. 즉 천단위 정도로 사사오입해서 세자리까지 信賴性 있는 數字를 만든 경우로서 이러한 것이 實際의 資料의 正確性이라 말할 수 있다. 최종 자리수까지 正確한 統計를 包含하여 여하한 경우에도 全体 자리수로서 作成된 統計를 만들 必要는 없다. (例를 들면 간행물에 收錄하는 경우) 천단위까지의 사사오입이라면 모든 利用者에게 적절한 水準이 될 수 있다고 보며 사사오입한 結果를 使用함으로써 비용 및 공간의 節減과 統計資料利用의 편의를 기할 수 있다.

사사오입이 正確하게 이루어졌다는 것을 確信하기 위해서는 적절한 節次를 취하는 것이 重要하다. 一般的인 原則으로는 사사오입은 演算後에 해야한다. 즉, 처음부터 하면 誤差가 커지게 마련이다. 사사오입시 지켜야할 原則은 다음과 같다.

(a) 信賴性이 있는 마지막자리 다음의 數字가 5보다 크면 마지막자리의 數는 1이 커져야 한다.

(b) 信賴性이 있는 마지막자리 다음의 數字가 5보다 작으면 마지막자리의 數字는 그대로 둔다.

(c) 信賴性이 있는 마지막자리 다음의 數字가 正確하게 5인

경우에는

- (i) 信賴性이 있는 마지막자리의 數字가 짝수이면 그대로이고
- (ii) 홀수이면 1이 커져야 한다.

예컨대 다음 數字는 세자리까지 信賴性이 있도록 사사오입한 예이다.

(i) 27,236 : 27,200

(ii) 13.485 : 13.5

(iii) 0.3725 : 0.372

(iv) 63,750 : 63,800

(i)에서 信賴性 있는 마지막자리 數字는 2이고 다음 數字가 36이므로 2는 변함없다. (原則(b))

(ii)에서 마지막자리 數字가 4, 다음 數字가 85이므로 4는 5로 바뀐다. (原則(a))

(iii)에서 마지막자리 數字가 2, 다음 數字는 5이므로 2는 짝수이므로 변하지 않는다. (原則(c)(i))

(iv)에서 마지막자리 數字는 7, 다음 數字가 5이고 7이 홀수이므로 7은 1이 增加된다. (原則(c)(ii))

#### 8.4 機械를 利用한 資料處理

資料處理를 보다 쉽고, 빠르고, 正確하게 하기 위한 技術과 기기형태는 여러가지가 있다. 그러나 기기에 의한 도움을 속도 와 正確性이라고만 단정해서는 안된다. 특히 전자계산기는 統計에 엄청난 影響을 주고 있는데 그 理由는 規模가 크고 세

분된 統計結果表의 작성을 可能케 하였고 고도로 複雜한 分析을 可能케 하였기 때문이다. 이러한 作業은 사람의 손이나 단순한 機械的인 方法으로는 전혀 불가능했을 것이다.

여기에서는 資料處理上 도움을 다음의 3가지로 区分하여 설명하고자 한다.

(i) 計算機

(ii) 80 칼럼 천공카드를 利用한 機械集計와 Edge-punch card와 같은 其他의 機械的인 Sorting system

(iii) 컴퓨터

처음의 2가지 方法에 대해서는 아주 簡略하게 說明하겠으나 컴퓨터에 대해서는 아주 상세하게 言及코자 하는데 이는 그 重要性이 크게 增大되고 있기 때문이다.

#### 8.4.1 計算機(Calculator)

計算機는 世界의 모든 統計機關에서 利用되고 있고 단순연산과 分析業務를 수행하는데 사용되고 있다. 이들 계산기는 크기와 複雜性에 있어서 가장 기초적인 연산이 可能한 간단한 携帶用에서부터 資料를 축적하고 處理하여 화면에 結果를 나타낼 수 있으며 또한, 高度의 數學的인 演算을 自動的으로 수행할 수 있고 結果를 印刷할 수 있는 기능을 갖추고 있는 프로그램이 可能한 것까지 있다.

모든 계산기는 그 특성이 다르며 계산기 使用要領을 習得하여 특별한 기능을 익히고 어떤 處理가 可能한가를 正確히 把握하는

것이 重要하다. 요즈음은 아주 싸 계산기라도 어느정도 複雑한 계산을 遂行할 수 있다. 여기서는 두가지의 基本的인 기능만을 言及코자 한다. 즉, 모든 計算機 製作会社別로 標準化되고 또한 統計的인 계산에 아주 有用한 機能으로서,

(a) 계산기 내부에 記憶된 일정한 수치를 反復해서 곱하거나 나누는 기능 : 이는 比率計算에 아주 有用하다. 예를 들면 合計에 대한 個別 특성의 構成比 計算을 들 수 있다. 이러한 기능을 活用하면 일정한 數字를 매번 계산기에 집어 넣을 必要가 없으므로 신속하게 結果를 얻을 수 있다.

(b) 계산기의 記憶裝置에서 積의 合計를 계산할 수 있다. 이러한 기능은 예컨대 소매 물가지수의 계산에 아주 편리하다. 一般的으로 지수는 각 항목별로 비교시 가격과 기준이 加重值의 積을 계산한 다음 小分類 및 中分類別로 作成된다. 이러한 形態의 계산은 基本演算機能을 活用하여 수행할 수 있으나 많은 時間이 浪費되고 아마도 가장 效果的인 지름길을 이용하지 못하는 한 오류가 발생될 危險이 增大될 것이다. 따라서 계산기를 가장 效果的으로 利用하기 위해서는 신중한 研究가 必要하며 有用한 기능을 최대한 活用할 수 있겠끔 實驗을 해야할 것이다.

계산기의 가치를 높게 평가하는 것이 重要하지만 소형계산기는 인쇄기능이 없다는 점을 명심해야 하며 우연히 틀린 키를 누를 수 있는 오류의 發生可能性이 남아 있다. 따라서 結果集計에 있어서 오류가 絶대로 발생할 수 없는 오류검증체계의 樹立이 필수적이다.

#### 8.4.2 穿孔카드 시스템

##### 8.4.2(a) 기계집계 (Mechanical Tabulation)

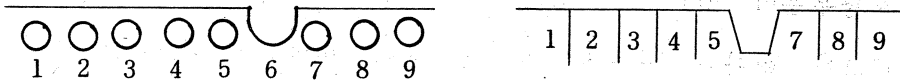
80 칼럼의 표준펀치카드를 이용한 기계집계는 모든 統計業務 종사자들에게 익숙할 것이다. 카드에는 入力資料를 넣을 수 있는 80 칼럼이 있고 每 칼럼마다 적절한 位置에 구멍을 뚫을 수 있도록 되어 있다. 오늘날 이러한 카드는 컴퓨터에 資料를 入力하는 가장 基本的인 수단으로 使用된다. 그러나 한편으로는 특별히 設計된 장치를 이용하여 分類되고 集計된다. 이러한 기계장치는 分類해야 할 칼럼에 맞추어지고 구멍이 그 칼럼의 어느 位置에 있는가에 따라 카드는 自動적으로 区分, 分類된다. 同時に 각 分類別 카드매수가 自動적으로 계산된다.

이러한 裝置는 컴퓨터의 출현에 앞서 大規模 統計調查 結果의 자료처리수단으로 사용되었다. 이러한 장치는 컴퓨터에 비해서 성능이나 融通性面에서 크게 떨어지며 分類, 集計裝置의 使用頻度は 점차 떨어지고 있다. 컴퓨터에 의한 資料處理에는 많은量的 집계프로그램을 작성해야 되고 여기에는 여러달 심지어는 몇년이 걸리는 경우도 있으나 分類·集計裝置에 의해서는 신속하게 資料處理 準備가 完了될 수 있다는 큰 장점을 가지고 있다. 따라서 이러한 裝置는 사람의 힘으로 處理하기에는 量이 너무 많고 컴퓨터處理로는 신속하고 低廉하게 處理할 수 없는 1회한 統計調查의 資料處理에 아주 有用하다. 그러나 人口센서스, 農業센서스등 大規模 資料가 生産되는 調查는 컴퓨터處理가 더욱

적절한 方法이 된다. 또한 分期別 노동력조사와 같은 일정한 주기로 反復되는 調査도 컴퓨터로 處理하는 것이 가장 效果的이다. 그 理由는 프로그램을 한번만 作成해 놓으면 계속해서 같은 프로그램으로 處理가 可能하기 때문이다.

#### 8.4.2 (b) Edge punched cards

Edge-punched card는 사람의 손으로 카드의 辺을 도려내어 資料를 表示하는 方式인데 小規模 資料의 處理에 利用되고 있다. 이것은 전적으로 分類(Sorting) 수단으로 活用되며 두가지 形態의 카드가 있다. 하나는 일련의 작은 구멍이 카드의 辺에 이미 뚫려져 있어서 펀치를 하게 되면 구멍과 카드의 辺사이가 찢어져 나가게 된다. 다른 하나는 구멍이 없는 것인데 카드의 해당부분이 단순히 찢어나가기만 한다. 가령 6 Digit는 두가지 形態에서 각기 다음과 같이 記錄된다.



카드의 辺은 여러개의 부분으로 나누어지며 대개는 미리 印刷되어 있다. 그리고 하나의 Field (혹은 카드의 부분)는 하나의 특성(Characteristic)을 나타내도록 되어있다. 가령 “性”을 나타내는 부분은 두개가 必要할 것이며 “婚姻狀態”를 나타내기 위해서는 5내지 6개의 필드가 必要하게 된다. 모든 資料가 카드에 記錄되고 나면 結果적으로 많은 부분이 카드로부터 떨어져 나가게 되는 것이다.

이 방식에 의한 분류는 카드량이 많지 않다면 신속 간편하다. 辺에 구멍이 뚫려 있는 Edge-punched card 방식으로 분류를 할 때는 긴 강철침을 사용하게 된다. 카드를 구멍뚫린 쪽을 위로 향하게 하여 一列로 정돈한 다음 바늘을 특정구멍(예컨대 “남자”를 나타내는 6번 구멍)으로 집어넣고 서서히 흔들어 준다. 따라서 6번 구멍에 천공된 모든 카드는 바늘에 걸려 있는 그 이외의 카드에서 분리될 수 있다. 카드의 수가 곧 남자가 될 것이다. 바늘 두개를 사용하여 단순 Cross-Classfication을 만들 수 있다. 가령 “既婚男性” 카드를 분리해 낼 수 있다.

다른 Edge-punched card 방식도 유사한 방법으로 処理된다. 카드는 작은 Shacking Machine에 옮겨지고 막대기(bar)가 카드 아래쪽 해당위치에 들어 오며 가령 남자를 記錄한 카드는 해당부분이 잘려져 나갔으므로 공간이 생겨있고 막대기 위로 내려오게 된다. 따라서 막대기 위에 머물러 있는 즉 높게 솟아 있는 카드를 골라내게 되면 나머지 낮은 위치에 있는 카드가 남자가 될 것이다.

이러한 종류의 분류방법은 조작하기 쉽고 費用이 低廉하며, 극히 效率性이 높기 때문에 小規模 地域統計調査의 資料處理에 利用될 수 있을 것이다. 그러나 資料處理는 地方統計機關에서 分散되어 이루어지는 경우에도 全國的인 統計調査는 컴퓨터나 機械的인 방식으로 處理되는 것이 通例인데 이는 통일성을 기하기 위하여 표준화된 全國的인 資料處理方式이 要求되기



때문이다.

나아가서 Edge-punched card 방식이 小規模調査의 資料處理에 아주 效果的이긴 하나 몇 가지 제약이 있다. 가장 큰 問題는 이 방식이 속성 (Attributes)을 記錄하는데는 편리하나 變數, 특히 많은 수의 變數가 관련되는 경우에는 불편하다는 점이다. 때로는 變數의 값을 카드의 표면에 쓰는 경우도 있는데 이러한 項目들은 속성에 대한 分類가 完成된 후에 사람손으로 處理해야 한다.

#### 8.4.3 컴퓨터에 의한 資料處理

##### 8.4.3 (a) 컴퓨터의 크기 및 價格推移

最近 들어 컴퓨터는 資料處理分野에 革新을 가져왔다. 컴퓨터는 複雜한 構造를 가진 전자기기이며 大量의 資料를 신속하게 處理할 수 있는 能力을 갖추고 있다. 컴퓨터가 登場함에 따라 전에는 試圖조차 할 수 없었던 모든 處理가 可能하게 되었다. 政府機關에 종사하는 統計專門家들은 大量의 資料處理와 관련된 業務에 종사하는 關係로 컴퓨터가 아주 重要한 역할을 한다는 것은 분명한 일이다. 그러므로 지금부터는 컴퓨터가 어떻게 處理하고 무엇을 할 수 있는지에 대해서 살펴 보고자 한다.

최초의 현대식 컴퓨터는 美國에서도 1946년까지는 完成되지 못했다. 그러나 그 以後부터 開發은 급속도로 進陞되었다.

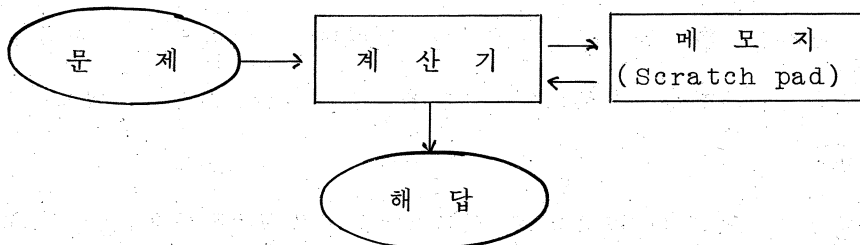
10年前에 컴퓨터는 크고 비쌌다. 단지 規模가 큰機關에서만

컴퓨터를 구입할 여유가 있었고 또한 그들을 利用할 수 있는 充分한 業務處理量을 確保하고 있었다. 그 以後부터 일련의 새로운 발명이 성공하여 機械의 크기와 價格을 크게 줄이게 되었다. 예컨대, 現在 中형정도의 컴퓨터 價格은 大略 US\$80,000에서 \$ 100,000 정도이다. 이것은 10年前 같은 性能의 컴퓨터價格의  $\frac{1}{10}$ 에 불과한 水準이다. 앞으로도 새로운 開發이 이루어짐에 따라서 價格은 계속 下落될 것으로 보인다. 이미 소형 컴퓨터는 US\$ 2,000 以下로 구입할 수 있는 것도 登場했다.

#### 8.4.3 (b) 컴퓨터의 原理

컴퓨터는 數를 다루는 전자기기이다. 컴퓨터는 초고속처리가 可能하고 많은량의 數의 처리가 可能할지라도 基本的으로는 우리가 단순한 계산기를 사용하는 것과 같은 方式으로 作業을 處理한다. 지금 계산기 (Calculator)를 하나의 “System”으로 使用한다고 하면, 이 시스템은 몇개의 서로 다른 部分으로 나누어질 수 있을 것이다.

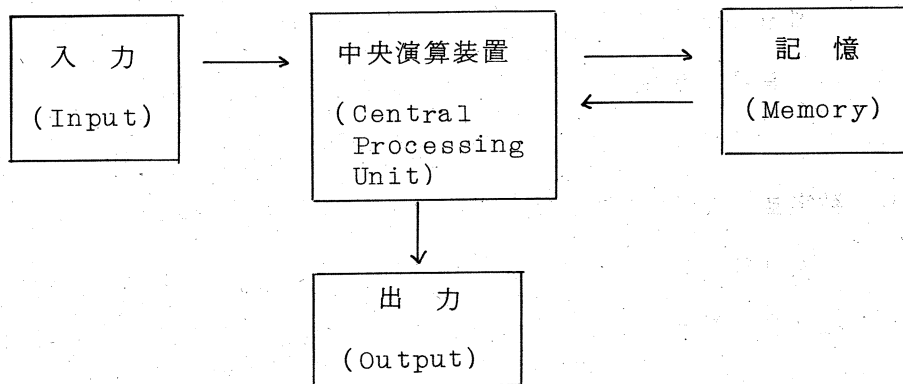
시스템을 도표로 表示하면 아래와 같이 될 것이다.



계산기를 使用하는 사람은 어떤 주어진 問題를 가지고 있다.

다시 말하면, 問題가 시스템 속으로 들어가는 것이다. 만일 問題가 複雜하면 계산기 이용자는 中間結果를 記錄해 두어야 한다. 이 中間結果는 메모지에 記錄하였다가 후에 必要한 계산시 使用한다. 最終的으로 問題가 完全히 계산되면 해답이 나오게 된다. 이것이 “시스템 (Output)”에서 나온 결과 (Output)이다.

컴퓨터에 의한 處理도 이와 아주 흡사하다. 컴퓨터를 몇개의 部分으로 나누어 圖表化하면 아래와 같다.



資料는 위에서 說明할 入力装置를 통하여 일정한 入力形態로 시스템 속으로 들어 오게 된다. 계산은 中央演算装置에서 遂行되며 中間結果는 記憶装置속에 一定한 形態로 저장된다. 나중에 최종 결과는 出力装置를 통하여 시스템 밖으로 나오게 된다.

따라서 컴퓨터處理는 우리들이 계산기를 使用하는 것과 類似한 方法으로 이루어짐을 알 수 있다. 그러나 양자 사이에는 아주 중요한 차이가 있으며 몇가지 예를 들면 아래와 같다.

(a) 컴퓨터의 處理速度는 계산기 보다 아주 빠르다. 계산기를

완벽하게 사용할 수 있는 사람이면 1초에 간단한 덧셈 2~3개를 할 수 있을 정도이나 컴퓨터는 1초에 수 백만개의 덧셈을 할 수 있다.

(b) 컴퓨터는 엄청난 량의 정보를 축적할 수 있고 어떠한 항목이라도 즉시 검색이 가능하다. 예를 들면 최신의 중형 컴퓨터는 자기 디스크 또는 자기테이프 같은 외부 기억장치속에 수백만 항목을 축적할 수 있다.

(c) 계산기에서는 숫자로 된 정보의 처리만 가능하나 컴퓨터는 숫자는 물론 문자로 된 정보의 처리도 가능하다. 중앙演算装置内에서는 비록 숫자의 처리만이 가능할지라도 문자를自動的に 숫자로 번역할 수 있는 기능을 갖추고 있다.

(d) 현대의 컴퓨터시스템은 複數의 利用者가 同時に 다른作業을 處理할 수가 있다. 계산기로 동시에 여러개의 問題를 處理한다면 利用者는 즉시 混同을 일으킬 것이다. 그러나 컴퓨터는 초고속처리가 가능하므로 많은 問題를 同時に 處理할 수 있다.

#### 8.4.3 (c) 컴퓨터 프로그램

계산기 이용자가 풀려고 하는 問題를 가지고 있다면 問題 自体가 곧 프로그램이다. 즉 問題가 무엇을 해야 하는가를 지시해 주는 것이다. 예를들면 問題  $(3.2 + 4.5) \times 3.6$  은 이용자가 계산기 속에 3.2를 入力시켜 4.5를 더하고 그 結果에 3.6을 곱하라는 지시를 하고 있는 셈이다.

같은 方法으로 컴퓨터는 프로그램의 지시하에 作業을 處理한다. 프로그램이란 資料와 마찬가지로 컴퓨터内部에 들어가 있는

命令의 目錄이다. 현대의 컴퓨터에서의 프로그램은 특수한 “言語(Language)”로 작성된다. 이러한 言語는 컴퓨터内部에서 단순한 命令文으로 變換된다. 단어나 문자도 入力될 수 있으므로 命令文은 흔히 英語와 흡사한 形態를 취하기도 한다. 이러한 것이 컴퓨터의 사용과 조작성이 더욱 용이한 점이다.

컴퓨터용어에 있어서 실제로 작업을 수행하는 機械를 우리는 “Hardware”라고 부른다. 機械에 주어지는 命令을 Software라고 한다. Software는 기계로 하여금 어떻게 處理하고 프로그램을 어떻게 단순한 명령으로 번역하는가를 명령하는 것을 의미하며 또한 入力裝置, 中央演算裝置, 記憶裝置와 出力裝置間에 情報를 어떻게 變換시킬것인가를 명령하는 것도 包含한다. 컴퓨터에서는 자료검증, 결과표작성과 같은 특별한 작업처리를 可能케 하기 위하여 때때로 프로그램 팩키지(Package)를 使用하기도 하는데, 이러한 특별한 프로그램들도 역시 Software속에 包含된다.

컴퓨터가 아무리 빠르고 正確하다 하더라도 틀린 명령을 내리게 되면 틀린 結果를 만들어 낸다는 점을 명심하는 것이 중요하다. 이는 마치 計算機의 키를 잘 못 누르는 것과 마찬가지로 이다.

#### 8.4.3 (d) 資料의 入力과 出力

資料는 컴퓨터속으로 入力될 때 여러가지 形態를 취할 수 있다. 가장 一般的인 形態를 보면

- (a) 80 칼럼의 카드에 의한 천공 입력
- (b) 천공종이 테이프
- (c) 자기테이프
- (d) 자기디스크
- (e) 단말기(Terminal)를 통한 입력

(a)에서 (d)까지의 방법에 의해서는 資料가 펀치 또는 특수 장치를 利用해서 入力된다. 이들은 타이프라이터와 흡사한 자리판을 가지고 있어 資料의 검증, 合計의 검증등과 같은 기능을 自動적으로 수행할 수도 있다. (e)의 方法에서 資料는 주전산기(Main computer)에 케이블 또는 전화선으로 연결된 단말기(Terminal)에 의해서 직접 入力된다. 이러한 터미널은 텔레비전과 같은 화면으로 된 경우도 있으며 入力된 資料를 비춰보거나 컴퓨터가 보내주는 메시지를 보여주기도 한다.

資料의 出力도 비슷한 方法으로 수행된다. 資料는 直接 단말기나 텔리타이프라이터에 의해서 스크린이나 종이에 타이프되어 출력된다. 천공카드, 종이테이프, 자기테이프 또는 자기디스크를 통하여 출력되기도 한다. 이들은 펀칭기계로 입력되어 必要한 情報를 화면을 통하여 또는 印刷裝置를 통하여 印刷해준다. 또한 資料는 고속프린터 또는 그래프나 도표를 제작할 수 있는 기계(Plotter)를 통하여 출력할 수도 있다.

컴퓨터의 장점의 하나는 출력형식이 간행물 발간에 適合한 形態로 設計되어 있다. 그래프와 다이어그램은 기초자료로 부터도 作成될 수 있다.

資料入力の 또 다른 하나의 方法으로서 Optical Mark Sensing을 들 수 있다. 이 方法은 약간 특이하게 사용된다. 그러나 가장 一般的인 方法은 흑연연필이나 유사한 수단을 이용하여 특별히 고안된 종이에 표시를 하는 것인데 이는 천공이나 다른 入力作業을 거치지 않고 직접 읽혀서 컴퓨터속으로 입력된다. 따라서 조사표가 이러한 특수종이로 만들어진다면 資料蒐集을 위한 어떤 중간과정도 거칠 必要없이 이것이 직접 입력수단으로 使用될 수 있다. 이 方法은 資料를 컴퓨터속에 入力시키는 다른 어떤 方法에 비해서도 경비가 적게 들고 신속하므로 아주 큰 장점을 갖고 있다. 그러나 여기에도 다음과 같은 몇 가지 단점이 있다.

(a) 종이가 기후에 비교적 민감하며 기술개발에 의해서 많은 개선이 이루어지긴 했으나 열대지방에는 역시 적합치 못한것 같다.

(b) 표시가 아주 正確하지 않거나 조사표에 우연히 不必要한 表示가 되거나 조사표 사용중에 닳거나 損傷되는 경우등에는 직접 입력이 불가능해 진다.

(c) 이 方法은 調査票가 完全하게 미리 코드화되어 있어야 한다. 따라서 앞에서 미리 言及된 미리 코드를 조사표상에 부여하는데 따른 단점의 적용을 받게 된다.

(d) 조사표는 자료를 읽어내는 기계의 처리능력에 맞추어 設計 되어야 한다. 이것은 모든 調査項目을 包含시킬 수 없다는 의미가 되며 또한 個別項目에 대한 코드를 우리가 원하는 만큼

많이 넣을 수 없다는 것을 뜻하는 것이다. 현재로서는 Optical Mark Sensing의 장점에 대해서 많은 論議가 대두되고 있는 실정이다. 이 方法은 여러나라가 인구센서스에 시도해 왔는데 성공정도가 나라에 따라 다르며 Bangladesh 같은 나라는 다음 센서스부터 이 方法을 채택하기로 計劃하고 있다.

#### 8.4.3(e) 컴퓨터에 의한 資料處理

여기 마지막 節에서는 컴퓨터에 의한 資料處理의 이모저모를 簡略히 살펴보고자 한다. 이것은 굉장히 複雜한 內容이며, 컴퓨터시스템을 어떻게 利用할 것인가를 分析하는 사람, 또한 컴퓨터에 명령하여 어느 작업을 수행할 것인가를 지시하는 사람들도 있다. 이와 같은 사람들을 각각 업무분석자(Systems Analysts), 프로그래머라고 부른다. 여기서 설명할 內容은 컴퓨터가 利用되는 方法을 몇가지 소개하는 것이며 자세한 說明은 省略코자한다.

컴퓨터는 아주 많은 量의 資料를 處理할 수 있고 산술연산을 신속하게 할 수 있는 能力이 있기 때문에 이는 資料處理에 있어서 가장 이상적인 方法이라 할 수 있다. 實際적으로 컴퓨터가 어떻게 利用되는가를 살펴보기 위하여 人口센서스에서 얻어진 情報를 分析하는 경우를 생각해 보기로 하자.

거의 모든 나라에서 人口센서스는 가장 많은 資料를 蒐集하는 調査로 되어있다. 人口센서스에서는 一般적으로 그나라의 모든 사람에 대해서 性, 연령, 婚姻狀態, 教育程度, 子女數, 出生地, 就業狀態, 宗教 및 其他 蒐集possible한 情報를 蒐集되고 있다. 印度와



같이 아주 큰 나라에 있어서는 처리해야 할 자료의量は 20個項目×6億人口이다. 다시 말하면 처리해야할項目數가 120億以上이 된다는 의미이다. 아시아에는項目數가 10億個以上되는國家가 상당수 있다. 이론상으로는 사람의 손으로 이러한業務處理가 可能하지만 處理量이 워낙 방대해서 시도하기 조차 어렵다. 지금부터 컴퓨터가 어떻게 資料處理에 利用되는가를 살펴보기로 한다.

우선 資料處理 方式과 必要한 結果表가 計劃段階에서 決定되어야 한다는 점을 인식하는 것이 重要하다. 이렇게 하여 調査票가 資料를 컴퓨터속에 能率的으로 入力시킬 수 있도록 設計될 수 있다. 조사표 양식은 資料入力を 可能한 限 단순화시킬 수 있도록 만들어져야 한다. 우리가 지금까지 보아온 바와 같이 現實的인 資料入力方法에는 여러가지가 있다. 즉, 펀치카드, 종이테이프, 자기테이프, 자기디스크 또는 단말기를 利用한 직접입력이 그것이다. 이 모든 方法은 一般 타이프라이터와 흡사한 자리판을 使用한다. 즉, 이것은 문자와 數字를 동시에 갖고 있는데 문자, 數字 또는 문자와 數字를 混合된 情報를 入力시킬 수가 있다. 예컨대 어떤 사람의 性이 M(남자) 또는 F(여자)로 記錄된다면 性에 대해서 미리 코드를 부여할 必要는 없다. 직접 “M” 또는 “F”로 入力시키면 되는 것이다. 이러한 情報를 處理하여 分類(Sorting)할 때, 컴퓨터는 프로그램에 따라 “M”과 “F”를 數字라 하는데 이때 自動적으로 이것이 행해지며 사람이 參與할 必要가 없다.

컴퓨터는 資料確認이나 修正도 할 수 있으며 이는 크게 두가지 方法으로 수행된다. 우선 資料入力에 使用되는 機械는 그 自体가 하나의 컴퓨터와 같이 작용한다. 이들 기계는 入力資料를 確認하여 資料가 받아 들일 수 있는 形態인가 아닌가를 판단한다. 예컨대 모든 사람에 대한 세번째 項目이 性으로서 이것이 入力되면 機械는 이것이 “ M ”이나 “ F ”냐를 프로그램의 지시에 따라 판단한다. 만약 M이나 F가 아닌 다른 數字나 표현으로 판단하면 오류가 발생했다는 메시지를 보내주게 된다.

두번째 검증형태는 컴퓨터 自体内에서 수행된다. 즉, 우리가 前章에서 살펴본 종류의 검증을 하게되는 것이다. 個人에 對한 情報는 다른 사람의 情報와 比較할 때 一致性이 있다고 確信하는 식으로 검증된다. 이때 오류가 發生되면 컴퓨터는 오퍼레이터에게 正確한 情報를 알려 달라는 內容의 메시지를 印刷하게 된다. 컴퓨터가 오류를 正정할 수 있는 길은 일련의 명령문에 따라서 可能하다.

모든 오류의 수정이 끝나고 完全한 情報가 作成되면 컴퓨터는 資料를 적절한 方法으로 저장한다. 실제 조사표를 보관하자면 센서스같은 경우 아주 큰 건물 하나를 요한다. 하나의 컴퓨터 시스템内에서 情報의 種類에 따라 몇개의 자기테이프, 자기디스크 속에 저장될 수 있다.

이와 같이 저장된 情報는 必要한 모든 結果표작성에 利用된다. 컴퓨터는 매우 融通性이 높아서 많은量的 상이한 結果表를 신속하게 작성할 수 있다. 예컨대 사람손으로 處理하는 데는 하나

의 結果表作成에도 수개월이 所要될 것이다. 컴퓨터作業으로는 몇분 또는 몇시간의 분량에 불과하다.

結果表가 作成되고 나면 印刷過程으로 넘어간다. 結果表 樣式은 프로그램으로 조정되는데, 이는 그들 결과표가 직접 作成될 수 있다는 것을 의미한다. 資料를 다시 타이프할 必要가 없는 것이다. 어떤 國家에서는 심지어 컴퓨터와 연결된 印刷機械가 있어서 책판권이 직접 만들어 질 수도 있다. 타이프에 의한 원고를 별도로 작성할 必要가 없으므로 印刷에 수반되는 오류가 훨씬 줄어들 수 밖에 없다.

일단 주요 결과가 作成되면 基本資料와 集計結果는 자기테이프에 保存되어 장차 연구, 분석업무에 수시로 活用될 수 있다. 모든 資料處理가 사람손으로 完成된 경우에는 處理結果의 利用者들에게 資料를 複写하여 제공하는 것은 상당히 힘든 일이다. 자기테이프를 利用하면 複写가 아주 단순하며 이를 바로 우편으로 보낼 수가 있다. 컴퓨터는 전화선을 이용하여 컴퓨터 상호간에 資料를 교환할 수 있고 또한 단말기를 이용하여 신속·용이한 전달이 可能하다.

물론 資料의 交換時에는 資料의 보안문제를 고려해야 한다. 만일 컴퓨터테이프가 一般 利用者들에게 공개되는 경우에는 개인의 이름이나 住所가 식별되어서는 안되며 어느 項目이 어느 個人의 것이라는 것도 알려지지 않도록 해야한다. 이러한 테이프는 개인단위의 資料를 包含하고 있다는 것을 알려주기 위해 "Unit record tapes" 라고 표시된다. 그러나 國家에 따라서

는 매우 嚴格한 규정때문에 이러한 정도의 비밀유지로는 資料를 공개할 수 없다.

앞에서 言及한 것과 같이 컴퓨터로 수행되는 이러한 모든 資料處理過程은 명령(또는 프로그램)에 의해서 조정된다. 프로그램을 작성하는 것은 매우 複雜한 作業으로서 이에 종사하는 사람은 특별한 지식을 요한다. 그러나 근래에는 여러가지 問題의 해결을 예상하여 미리 만들어진 소위 컴퓨터팩키지(Computer package)가 있는데 이것은 포괄적으로 작성되어 어떠한 要求도 만족시킬 수 있도록 되어 있다. 人口센서스 分析用 Package가 開發되어 이들 중 一部는 이미 몇몇 아시아 國家들이 利用하여 성공을 거둔 바 있다.

#### 8.4.3(f) 結 論

지금까지 컴퓨터는 주요한 기능의 하나로서 統計科學分野에 크게 기여한 아주 有用한 機械라는 점을 알았다. 아무리 그것이 有用하다 하더라도 또한 고속처리가 가능하고 방대한 資料處理를 할 수 있다 하더라도 그것은 어디까지나 기계에 불과하다. 컴퓨터를 사용한다고 해서 調査나 센서스가 바람직하게 수행되거나 結果가 완벽해 지는 것은 아니다. 統計 調査 企劃에서 實查要員에 이르기까지 모든 統計業務 종사자들은 必要한 資料의 수집을 위해서 設計하고 計劃하고 이것이 적절한 양식인가 만족스러운 正確性을 유지하는가 그리하여 컴퓨터로 處理되었을 때 信賴性이 과연 있는가를 確信할 수 있도록 해야한다.

아마도 우리는 컴퓨터에 대해서 아주 중요한 암시를 주는 표현을 음미하면서 컴퓨터에 대한 설명을 마쳐야 될 것 같다.

Acronyms (머릿글자를 따라 만든 용어)를 사용하기 좋아하는 컴퓨터 전문가들이 “쓰레기를 넣으면 쓰레기가 나온다.”는 뜻의 GIGO 라는 용어를 만들어 냈다. 이 말의 뜻은 우리가 컴퓨터에 형편없는 情報를 入力시키면 역시 형편없는 結果를 받아 볼 수 밖에 없다는 것이다. 컴퓨터는 우리에게 많은 것을 도와 줄 수가 있다. 그러나 신빙성이 없는 資料를 가지고는 좋은 統計를 기대할 수가 없다.

## 第 9 章 統計資料의 提供과 公表

### 9.1 概 要

統計業務 종사자들에 있어서 가장 중요한 일의 하나는 그들이 蒐集한 資料를 必要로 하는 사람들에게 提供하는 것이다. 이들 이용자에는 政府機關, 法定機關, 企業體, 研究者 및 일반이용자가 포함된다. 調査設計가 아무리 잘 되고 分析이나 資料處理方式이 아무리 복잡하더라도 調査結果가 이용되지 못한다면 모든 것이 낭비에 불과한 것이 되고 만다. 調査結果가 公表될 때에는 理解가 용이하고 活用될 수 있는 형태를 갖추어야만 한다.

여기서는 統計情報가 提供되고 公表되는 세가지 주요 형태를 살펴보고자 한다. 즉, 통계표, 그래프와 도표, 보고서 및 기타방법이 그것이다. 공표의 세가지 방법을 각기 Tabular, Graphical 및 Textual 로 명명하기로 하자.

### 9.2 Tabular 방식의 資料提供

統計表는 가장 일반적으로 사용되는 統計提供方法이다. 이것은 가장 편리하고 정연하고 効果的인 결과제공 方法으로서 통계조사결과 제공에 있어서 가장 비용이 적게 든다.

統計表란 행과 열에 숫자를 배열하여 일종의 분류체계에 의한 設計로 통계결과를 체계적으로 제공한 것을 말한다. 統計表는 資料의 배치를 용이하게 하고 통계의 비교와 해석을 편리하게 한다.

9.2.1 통계표의 構成要素

統計表의 設計와 구성요소의 명칭에 관해서는 지금까지 여러가지 협의가 이루어져 왔다. 현지에서 統計業務에 종사하는 사람들은 統計表에 익숙해야 한다. 이러한 統計表를 자세히 관찰하여 보면 統計担当者들과 이용자 누구에게나 도움이 될 수 있도록 통계표구조의 통일성을 읽을 수 있다. 통계표의 구성요소와 그들의 명칭은 다음 예와 같다.

番 號

표 16. Utopia 國의 地域別 國立學校 教育現況

題 目

1979.

Headnote

{ 學校, 教師, 學生數는 1979年 8月 1日 現在  
國家의 教育費支給豫算은 1979年中 數值임 }

Captions

Spanner heading

Sub-heading

Column

Stub-heading

| 地域    | 學校    | 教師 <sup>1/</sup> / 學生 |         | 國家의 教育費支給 <sup>2/</sup> |        |        |
|-------|-------|-----------------------|---------|-------------------------|--------|--------|
|       |       | 國民學校                  | 中學校     | 計                       |        |        |
| (1)   | (2)   | (3)                   | (4)     | (5)                     | (6)    | (7)    |
|       | 個數    | 名                     | 名       | 千弗                      | 千弗     | 千弗     |
| 1. 中道 | 2,963 | 35,216                | 481,239 | 26,211                  | 7,944  | 34,155 |
| 2. 北道 | 284   | 8,941                 | 129,366 | 6,318                   | 1,690  | 8,008  |
| 3. 東道 | 405   | 10,206                | 166,221 | 8,067                   | 1,523  | 9,590  |
| 4. 南道 | 196   | 5,785                 | 87,910  | 4,336                   | 772    | 5,108  |
| 5. 西道 | 178   | 4,493                 | 83,865  | 4,058                   | 482    | 4,540  |
| 全 國   | 4,026 | 64,641                | 947,621 | 48,990                  | 12,411 | 61,401 |

Boxhead

Row

Body or Field

Column numbers

Specifica - tion of units

Stub

Footnote

1/ 補助教師 및 時間講士包含

Source note

2/ 1979年 12月 31日까지의 年間數值임  
資料: Utopia 文教部 1979年版 通計年報

다음은 統計表에서 중요성이 비교적 큰 통계표 구성요소에 대한 해설이다.

(a) 표번호 ( Table Number ) : 이것은 간행물 내용중에서 統計表를 식별해 주며 뒤에 통계표제목이 따른다. 이것은 統計表中 가장 용이한 참고사항이다.

(b) 제목 ( Title ) : 題目은 統計表의 Main Body 위에 위치한다. 題目은 간략하고 壓縮된 표현을 써서 統計表에 있는 資料의 種類, 資料의 分類, 資料의 응용분야 資料의 작성시점등에 관한 情報를 제공한다. 題目이 너무 긴 경우에는 題目을 짧게 사용하고 나머지 情報는 Headnote 에 넣는 것이 일반적인 경향이다.

(아래(c)참조), 위의 예에서 보는 바와 같이 題目을 설정함에 있어 원칙은 대체로 統計表 內容의 검토로 명백해지는 부분은 언급하지 않는다는 것이다. 따라서 “ 1979 Utopia의 국립학교 教育 ”이라는 題目으로 충분한 것이다. 그러나 간행물에 들어가는 統計表內容은 보다 상세한 情報가 필요할지도 모른다.

(c) Headnote : 표에 따라서는 Headnote가 필요한 경우가 있다. 이것은 題目보다 적은 활자로 쓰는 것이 보통이며 統計表 全體를 완전하게 이해하는데 필요한 解設資料를 提供한다.

(d) Captions : Caption은 각 칼럼의 제일윗쪽에 위치하며 각 칼럼이 무엇을 나타내는가를 설명해 준다. 일반적으로 하나의 Caption이나 Heading은 2개이상의 칼럼과 관련되는 경우가 많은데 이것은 Spanner heading을 의미한다. 따라서 각 칼럼에는 Sub-heading이 있다고 할 수 있다.



(e) Boxhead : 이것은 Stub-heading 을 제외한 모든 Caption 을 포함한다. 즉, 숫자를 포함하고 있는 모든 칼럼이 포함된다.

(f) Stub : 이것은 가장 왼쪽에 있는 칼럼과 Caption (즉 Stub heading 을 말함) 을 합친 부분으로서 統計表의 모든 행 ( Row ) 의 명칭을 제공하고 있다.

(g) Body : 統計表의 Body 는 행과 열의 Heading 에 연결된 하나 하나의 개별 숫자로 構成되어 있다. 이것은 간혹 統計表의 Field 라고 표현되기도 한다.

(h) Footnote(S) : 이것은 개별 숫자, 칼럼 또는 행 ( Row ) 에 관한 설명이다. Footnote 는 표의 제일 아래에 위치하며 작은 활자로 쓰는 것이 보통이다. 이것은 대개 번호가 부여되어 있는데 같은 번호가 해당 숫자 또는 Caption 에 붙는다. General note 가 사용되기도 한다.

(i) Source note : 이는 資料를 提供하는 개인, 기관 또는 기구를 말하며 資料出處로 사용된 간행물 명칭을 말하기도 한다.

### 9.2.2 統計表의 形態

가장 간단한 統計表 形態는 단일 칼럼 ( 또는 행 ) 의  
> 숫자로 構成된 것이다. 예를들면,

(57)

道別 유토피아 인구

1980.6.30 現在

| 道   | 人 口        |
|-----|------------|
| 中 道 | 4,768,600  |
| 北 道 | 2,911,400  |
| 西 道 | 3,117,700  |
| 南 道 | 845,300    |
| 東 道 | 1,536,600  |
| 全 国 | 13,179,600 |

이것은 One-way Table ( 또는 One-way Classification ) 이라 칭하는데 人口資料가 단지 거주지역에 의해서만 分類되었기 때문이다.

이러한 간단한 統計表를 단순히 연장할 수가 있는데 이것은 역시 地域別로 人口가 아닌 다른 資料를 하나이상의 칼럼으로 추가하는 것을 의미한다. 예컨대 各 地域別 住宅數를 포함하는 하나의 칼럼을 위 표에 추가하였다고 하자 이경우 두개의 칼럼에 資料가 있게 되는데 이는 역시 One-way Classification 일 뿐이다. 各 변수들 ( 人口와 住宅 ) 은 단지 하나의 특성 ( Characteristic ) 에 의해서 分類되었기 때문이다. Character

그러나, Two-way Classification 으로 資料를 제공하는 것이 可能하고 또 바람직한 경우가 흔히 있다. 위의 예에서 나타난 人口資料를 이를 다시 연령계급별로 分類하여 統計表를 만들면 다음과 같이 될것이다.

Utopia 의 地域別 年齡別 人口

Variables

1980.6.30 現在

*two way*

| 地 域 | 연 령 계 급   |           |           |           |           |           |           | 計          |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
|     | 0 ~ 9     | 10 ~ 19   | 20 ~ 29   | 30 ~ 39   | 40 ~ 49   | 50 ~ 59   | 60 以上     |            |
| 中 道 | 1,107,900 | 923,500   | 714,600   | 622,800   | 557,000   | 481,100   | 361,700   | 4,768,600  |
| 北 道 | 667,300   | 560,700   | 439,200   | 373,800   | 334,700   | 286,000   | 239,700   | 2,911,400  |
| 西 道 | 702,400   | 571,100   | 478,500   | 422,100   | 359,900   | 298,700   | 285,000   | 3,117,700  |
| 南 道 | 196,300   | 161,400   | 128,200   | 105,600   | 97,500    | 89,200    | 67,100    | 845,300    |
| 東 道 | 342,700   | 287,000   | 228,600   | 201,400   | 186,900   | 162,700   | 127,300   | 1,536,600  |
| 全 国 | 3,016,600 | 2,503,700 | 1,989,100 | 1,725,700 | 1,546,000 | 1,317,700 | 1,080,800 | 13,179,600 |

이와같은 Two-way Classification은 흔히 Cross-Classification으로 불리며 地域別로 分類된 資料가 다시 연령계급별 표에 의해서 세분된 것을 알수 있다.

이것은 다시 아래 統計表와 같이 행과 열을 바꾸어 이용할 수도 있다. 즉

Utopia 國의 地域別, 性別人口

1980. 6. 30 現在

| 年 齡<br>階 級 | 地 域       |           |           |         |           | 全 國        |
|------------|-----------|-----------|-----------|---------|-----------|------------|
|            | 中 道       | 北 道       | 西 道       | 南 道     | 東 道       |            |
| 0 ~ 9      | 1,107,900 | 667,300   | 702,400   | 196,300 | 342,700   | 3,016,600  |
| 10 ~ 19    | 923,500   | 560,700   | 571,100   | 161,400 | 287,000   | 2,503,700  |
| 20 ~ 29    | 714,600   | 439,200   | 478,800   | 128,200 | 228,600   | 1,989,100  |
| 30 ~ 39    | 622,800   | 373,800   | 422,100   | 105,600 | 201,400   | 1,725,700  |
| 40 ~ 49    | 557,000   | 344,700   | 389,900   | 97,500  | 186,900   | 1,546,000  |
| 50 ~ 59    | 481,100   | 286,000   | 298,700   | 89,200  | 162,700   | 1,317,700  |
| 60 以上      | 361,700   | 239,700   | 285,000   | 67,100  | 127,300   | 1,080,800  |
| 計          | 4,768,600 | 2,911,400 | 3,117,700 | 845,300 | 1,536,600 | 13,179,600 |

이와 같이 두개의 선택가능한 統計表가 작성되는 統計資料는 서로 별개라 볼 수 있는데 위에서 본 경우는 어떤 統計表 形式을 취하든 별다른 問題가 없다. 그러나 때에 따라서는 어느 分類를 Box head 에 두고 Stub 에 두는가에 따라 統計表의 모양과 이용자들의 資料理解의 용이도가 크게 달라지기도 한다.

114 統計表의 모양을 선택하는데 있어서 결정요인은 여러가지가 있다. 그러나 이와 같은 요소는 統計表設計時에 신중히 고려해야 한다. ① 統計表는 개방적이고 산만해서는 안된다는 점이 가장 중요한데 이러한 면을 갖추어야만 이용자가 資料를 쉽게 이해할 수 있기 때문이다. ② 특별히 고려에 넣어야 할 사항으로서는 너무 많은 칼럼을 집어 넣지 않는 것이 바람직하다는 것이다. 이는 統計表가 매우 산만해 보이기 쉽고 각 칼럼의 마지막 자리의 숫자와 다음 숫자의 첫번째 자리에 있는 수가 거의 맞붙게 되기 때문이다. 앞의 예에서 보면 연령계급을 Box head에 둔 표는 8칼럼으로 되어 있고 그 반대의 표는 6칼럼으로 構成되었음을 알 수 있다. 비록 앞의 표가 선택될 수 있는 소지가 있긴 하나 뒤의 표가 모양이 보다 양호하고 개방적이다. 그러나 만약 연령계급이 2개가 더 추가된다면 처음 표의 칼럼 수는 8개에서 10개로 늘어나게 되고 이는 선택하기 곤란한 혼잡한 형태가 되어 표의 행과 열을 완전히 바꾸지 않을 수 없을 것이다.

③ 8개의 경우에 행의 수가 칼럼수보다 많은 것이 보통인데 이러한 사실은 統計表의 양식을 결정할때 명심해야 할 사항이다.

또 다른 요소는 Boxhead보다 Stub에 필요한 단어를 더 많이 넣을 수 있는 여유가 있다는 것이다. 따라서 긴 항목을 가진 分類(예컨대 職業, 産業 또는 貿易商品分類)는 統計表의 Stub에 넣는 것이 정상적이다.

지금까지는 One-way와 Two-way 分類만 보아 왔으나 보다 복잡한 양식이 가능한데 이들은 Multiple Classification Tab-

le 이라 칭한다. 이와 같은 構造를 가진 통계표는 다양한 方法으로 만들 수가 있다. 예를 들면, 세번째 特性 (Characteristic) 별로 Sub-Classification이 만들어질 수 있는데 이는 몇개의 칼럼을 묶어서 Spanner Heading을 만드는 것과 같다. 이러한 構造의 표는 보다 상세한 情報를 제공할 수 있으나 필연적으로 복잡해지기 마련이다. 또한 이용자들은 자료이용시 혼란에 빠지는 것을 원치 않는다는 점을 명심해야 한다. Multiple Classification을 이용하여 비교적 단순한 統計表를 작성할 수가 있다. 지금 앞에서 인용한 人口資料를 性別 (세번째 특성)로 分類하여 가장 一般的인 Three-way table의 하나를 소개하면 다음과 같다.

Utopia 國의 性別, 年齡別, 地域別人口

1980.6.30 現在

| 年齡階級    | 地 域 |     |     |     |     | 全 國 |
|---------|-----|-----|-----|-----|-----|-----|
|         | 中 道 | 北 道 | 西 道 | 南 道 | 東 道 |     |
|         | 男 子 |     |     |     |     |     |
| 0 ~ 9   |     |     |     |     |     |     |
| 10 ~ 19 |     |     |     |     |     |     |
| ·       |     |     |     |     |     |     |
|         | 女 子 |     |     |     |     |     |
| 0 ~ 9   |     |     |     |     |     |     |
| 10 ~ 19 |     |     |     |     |     |     |
| ·       |     |     |     |     |     |     |
|         | 總 計 |     |     |     |     |     |
| 0 ~ 9   |     |     |     |     |     |     |
| 10 ~ 19 |     |     |     |     |     |     |
| ·       |     |     |     |     |     |     |

### 9.2.3 統計表作成要領

統計表가 별도의 설명문을 수반하든 아니든 統計表 자체는 가능한 한 완전해야 한다. Headnote나 Footnote는 統計表内の 모든 사항을 명백히 할 수 있는 충분한 情報를 제공해야만 한다.

統計表의 모든 부분은,

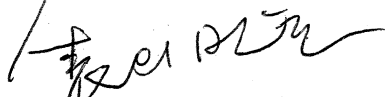
- 중요한 항목을 강조하고
- 비교를 용이하게 그리고 관계를 분명하게 하고
- 중요한 비교사항을 강조하고
- 크기를 최소화하여
- 산뜻한 모양을 갖추 수 있도록 배치시켜야 한다.

좋은 統計表를 作成하기 위해서는 다음사항을 명심해야 한다.

- 표의 각 부분에 대한 별도의 고려
- Stub와 Caption項目的 선택과 적절한 배치
- 칼럼과 행(行)에 번호부여(쉽게 참조가능하도록)
- 여백, 줄꼭선(Indentation), 고딕체, 쌍선과 중요한 사항을 눈에 잘 띄게 할 수 있는 기타방법의 동원
- 小計와 合計의 적절한 배치
- 資料의 출처와 水準에 대한 언급
- 전체적인 간결성과 균형의 유지

統計表를 작성할 때 다음 원칙을 염두에 두어야 한다.

- (a) 간행물 등에 두개 이상의 통계표가 수록될 때 번호부여(로마자 아닌 아라비아 숫자를 사용할 것)



(b) 題目은 간결하고 표현이 명확할것. 그러나 統計表의 內容을 적절히 서술할것, 필요시 Headnote를 사용할 것. 적절한 대 상 期間을 명시할 것.

(c) 표의 적당한 곳에 단위를 꼭 표시할 것. 예컨대, 모든 資料의 단위가 공통이면 이를 Headnote에 포함시킬 것. 칼럼마다 단위가 다르면 Caption이나 Box-head 바로 아래에 있는 Column의 제일 위에 단위를 표시하고 行마다 단위가 다르면 Stub의 바로 다음 첫번째 칼럼을 單位表示 장소로 이용할 것.

(d) 수치가 사사오입된 경우에는 사사오입된 자리수를 표시할 것 1,000 달러, 백만달러, 1,000 개등과 같이 표시하는 것이 좋다.

(e) 合計, 小計, 平均, 백분비등은 그들이 유용한 情報가 된다고 판단되면 반드시 作成할 것.

(f) 숫자가 들어가는 공간은 어떠한 표시라도 있을것 절대로 빈 칸으로 남기지 말고 다음과 같은 기호를 사용할 것,

= 해당없음

...해당 숫자없음

-- 단위 미만

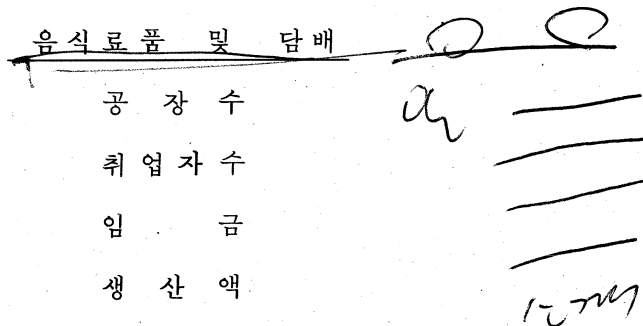
..미 상

정해진 약속은 없으나 유사한 형태의 표기가 모든 統計表에 적용할 것. 중앙통계국에서 정한 표준통계표 양식을 준수할 것. 使用된 기호는 일러두기에서 반드시 설명해 줄것.

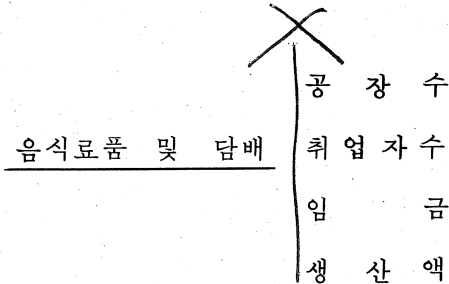
(g) Stub에 있는 어떤 항목이 아주 긴 경우에는 2 Line 이상을 使用해서 表示할 것. Stub상의 項目은 그룹으로 묶을 수



있으나 아래 예와 같이 중분류와 소분류간에 縱으로 선을 가르  
지 말것.



위와 같이 하는 것이 아래와 같이 분리하는 것보다 더 좋은 構  
成이다 (본다 짜임새 있고 Heading과 項目間的 관계가 분명 하  
다)



(h) 개별 수치나 칼럼 또는 行에 관한 설명은 적절한 番号를  
부여하여 Footnote에 배치할것 (예컨대 (1), (2), (3)과 같이 하고  
순서는 ~~왼쪽에서 오른쪽으로, 위에서 아래로 할것~~)

(i) 資料出處는 Footnote아래, 경우에 따라서는 Headnote 아  
래에 表示할 것, 여기에는 資料가 수록된 간행물 명이나 資料作  
成 機關名을 기록할 것 간행물인 경우 간행물 호수, 페이지, 출  
판사 또는 발간 일자도 기재할 것.

(j) 표의 양쪽 끝에 수직선을 사용하지 말 것. 횡선은 표의  
중간에 사용하지 말것 (합이나 区分이 필요한 때에는 예외) 종선

은 Boxhead 의 아래와 위 그리고 Body 또는 Field 의 아래에 使用할것.

#### 9.2.4 Stub 와 Boxhead 상의 項目 배열

Stub 나 Boxhead 에 項目을 배열할 때에는 선택적으로 할 수 있다. 예컨대 地域은 알파벳順, 地理的인 순서, 크기와 중요도順 또는 이러한 方法을 조합하여 배열할 수 있다.

① 分類는 하나의 고정된 배열순서를 가지는 것이 바람직한데 특히, 같은 分類가 여러 통계표에 나타날 때는 특히 그렇다. 가령 어떤 통계표는 알파벳 순서로 分類하고 다른 표에서는 地理的인 순서로 分類한다면 統計表間의 비교가 어렵고 통계이용자들에게 혼동을 일으킬 수도 있다.

② 統計作成機關에서 공통적으로 使用하는 分類의 표준순서를 정해두면 많은 장점이 있다. 統計作成機關은 물론이고 政府機關 또는 지방관서에서는 통일성을 기하기 위해서 같은 分類順序를 使用해야 할 것이다. 통상적으로 使用되는 分類中에는 국제기구에서 使用하는 분류체계를 따르는 것도 있다. 예컨대 國際標準産業分類 (ISIC), 國際標準貿易商品分類 (SITC) 등이 그것인데 이들 分類는 항상 같은 순서로 되어있는 표준코드를 가지고 있다. 따라서 이들 分類를 使用하는 國家間에는 상호 비교가 可能하다.

(SIC) 0 이하 행정구역  
1. 1차 산업  
2. 2차 산업  
3. 3차 산업  
4. 기타 서비스  
-298-  
SITC  
Standard Korean  
Trade Classification

### 9.3 圖表作成

#### 9.3.1 目的

統計調査結果는 항상 일련의 統計表로 提供되나 때로는 統計結果의 주요한 特性을 설명하고 弘報하기 위하여 다양한 형태의 圖表로 作成되기도 한다. 여기서는 이와 같은 圖表作成의 원칙과 目的 및 圖表作成方法의 內容을 다루고자 한다.

圖表作成의 目的은 중요한 統計結果를 용이하게 提供하기 위하여 그림형태로 만드는 데 있다. 이와 같은 方法은 중요한 情報을 아주 쉽게 전달되도록 즉각적인 시각효과를 줄 수 있고 또한 쉽게 記憶될 수 있는 형태로 만들어 진다. 이들 方法은 래프차트 또는 모림과 같은 형태를 취한다. 이들 세가지 형태를 총칭하여 “圖表 (Diagram)”라 부른다.

圖表는 유용하기는 하지만 반면에 단점이 없는 것은 아니다. 모든 사람들의 처음에 받는 인상이 오래 持續되며 圖表構成要素의 크기 또는 곡선의 완급에 의해서 結果의 特性을 이해하게 된다. 그러나 수반되어야 할 資料의 質이나 설명이 생략되어 있다. 이러한 表現方法은 全體的인 윤곽을 把握하기 위하여 效果的인 최적의 方法이다.

#### 9.3.2 基本的인 圖表

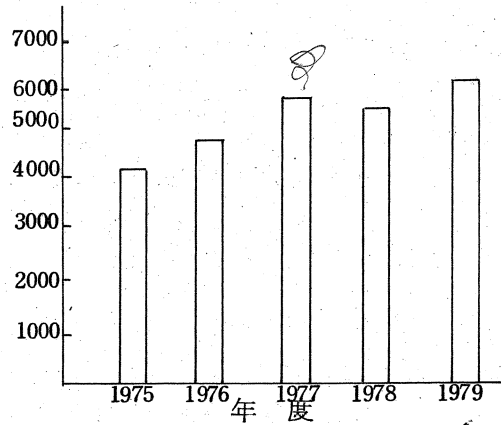
우선 어떠한 형태의 圖表作成에 있어서도 基本的인 몇 가지 問題가 고려되어야 할 것이다. 圖表는 統計表의 圖示로 간주될 수 있기 때문에 統計表作成時의 원칙이 많이 적용될 수

있다. 統計表와 마찬가지로 圖表은 가능한 한 그 자체로 설명이 가능해야 한다. 즉 圖表가 나타내는 統計結果는 도표라는 제한된 범위내에서도 설명될 수 있어야 한다. 여기서 “제한된 범위내에서”라는 문구는 도표상으로는 실제로 調査된 資料의 근사치에 불과하기 때문이다. 圖表은 統計表와 같은 정확한 情報를 提供하지는 못한다. 따라서 도표와 統計表는 같이 수록되는 것이 보통이며 도표는 資料의 즉각적인 인상을 傳達하고 統計表는 정확한 情報를 제공한다.

지금 간단한 米穀生産量 統計表를 가지고 아래 예와 같이 실제 자료를 왼쪽에 같이 表示하여 圖表作成을 설명하고자 한다.

<表 1> Utopia 國의 쌀生産量 <圖表 1> Utopia 國의 쌀生産  
1975 ~ 1979 1975 ~ 1979

| 年 度  | 生 産 量<br>(1,000吨) |
|------|-------------------|
| 1975 | 4,270             |
| 1976 | 4,860             |
| 1977 | 5,910             |
| 1978 | 5,660             |
| 1979 | 6,200             |



앞에서 統計表의 構成要素를 설명한 것과 같이 아래와 같은 圖表의 構成要素를 정리하여 보면,

- (a) 図表番号 : 図表가 여러개인 경우 使用된다.
- (b) 題目 : 統計表와 마찬가지로 図表에 表示되는 資料를 설명하는 데 使用된다.
- (c) Headnote : 題目을 한정하거나 보충 설명하는데 필요하다.
- (d) 資料出處 : 統計表와 꼭 같은 것으로 資料의 源을 表示하는데 이용된다.
- (e) 보기 또는 범례 : 図表上에 나타나는 변수가 둘이상 (예컨대 통화와 물가)일 때 도표의 어느 부분이 어떤 변수인지를 表示할때 使用함.
- (f) 수평축 (Ox) : 하나의 변수를 表示하기 위하여 사용되는데 보통 같은 크기로 분할된다. 보통 수평축 아래 주기가 붙는다.
- (g) 수직축 (Oy) : 두번째 변수를 表示하는 데 使用된다. 변수에 관한 설명이 수직축 옆에 올 수 있다.
- (h) 배열 : 변수의 一般的인 배치는 각 축의 왼쪽에서 오른쪽으로 아래에서 위로 해야한다. 단위 또는 크기가 작은 것부터 각 축의 왼쪽끝 또는 제일 아랫쪽에 배열해야 한다.
- (i) 눈금 : 눈금은 변수와 종이의 크기에 따라 달리할 수 있다. 눈금은 편리한 方法으로 조정하되 表示되는 점들이 쉽게 식별될 수 있도록 해야한다. 때로 로그 (Logarithms) 같은 특수한 눈금도 있으나 대개는 등분 (等分)이 보통이다.
- (j) 비례 : 도표의 양축은 같은 크기가 되어야 한다. 어느 한쪽이 다른 한쪽에 비해 크게 차이 나면 資料가 歪曲되거나 경향이 달라 보이게 되는 경향이 있다.

⑤ 도표에서는 대개 두개의 변수간의 관계 즉 하나의 변수가 다른 변수에 의해서 어떤 영향을 받는가를 나타내는데 사용된다.

예컨대 어느 國家의 米穀生産이 일정한 期間 즉 5年間 어떻게 변동되는가를 도표로 표시하는 것을 들수 있다. 이 경우 時間을 독립변수라 부르며 米穀生産을 종속변수라 부른다. 대개는

독립변수의 주어진 일정한 변동에 따라 종속변수가 어떻게 달라지는가에 관심을 가지는 경우가 많다. 통상적으로 독립변수는 수평축에 表示되며 종속변수는 수직축에 놓이는 것이 원칙이다.

따라서 그래프에서 일정기간 (Time Series라 부른다)의 변동을 나타내기 위해서 시점은 年間, 月間 또는 어떤 時間單位를 선택하든 항상 수평축에 오며 수직축에 오는 경우는 결코 없다.

### 9.3.3 圖表의 주요한 형태

도표에는 참으로 많은 형태가 있다. 여기서는 공통적으로 사용되는 4가지 基本形態를 설명하고 한 두개의 복잡한 도표를 소개하고자 한다.

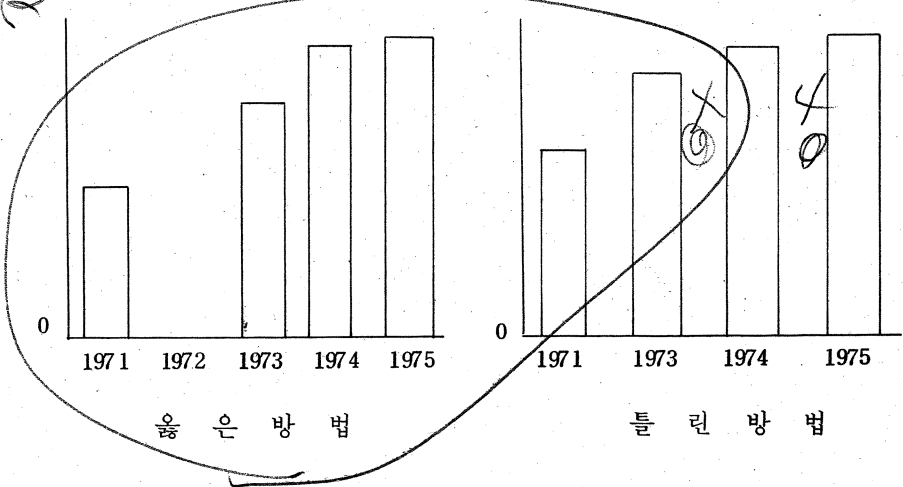
#### (a) Bar Chart

Bar Chart를 作成하는 目的은 資料의 절대크기를 나타내기 위함이다. Bar Chart는 일련의 봉 (Bar)으로 구성되는데 이것의 길이가 수의 크기에 비례한다. 앞의 예가 Bar Chart의 일종이었다. 종속변수인 천단위 年間 米穀生産량이 각 Bar의 높이로 표시되어 있다.

Bar Chart나 다른 어떤 種類의 圖表를 作成할 때에도 특수

한 제도기를 使用하지 않고 그릴 수 있는 가장 적은 單位로서 통상 밀리미터를 使用해야 한다.

9 각 Bar 간의 거리가 Bar의 넓이의 반정도가 될 때 가장 보기가 좋을 것 같다 (비록 공간의 이용에 어려움이 없는 경우에도 그러하다) 어느 경우에도 Bar 간의 거리와 각 Bar의 넓이 각 각각 일정해야 한다. Bar 자체는 눈에 띄게 하기 위하여 음영을 넣거나 평행되는 교차선을 넣을 수가 있다. 색깔은 특히 복합 Bar Chart나 각 Bar를 分割하여 作成되는 Bar Chart에서 效果가 크다 (여기에 대해서는 뒤에 따로 설명될 것이다) 만일 Bar Chart가 균등한 간격의 시계열이나 다른 변수를 포함할 때는 눈금을 빠뜨리지 말고 표시해야 한다. 예를들면 어떤 계열에서 어느 한 연도의 資料가 없을 때에는 그 年度에는 공백으로 남겨두어야 한다. 아래 예에서 보면 오른쪽 도표는 1973年과 1975年(2개년도) 사이의 간격이 1975年과 1976年(1個年度) 사이의 간격과 같다. 이것은 歪曲될 가능성이 크다 따라서 이러한 형태는 피하는 것이 좋다.



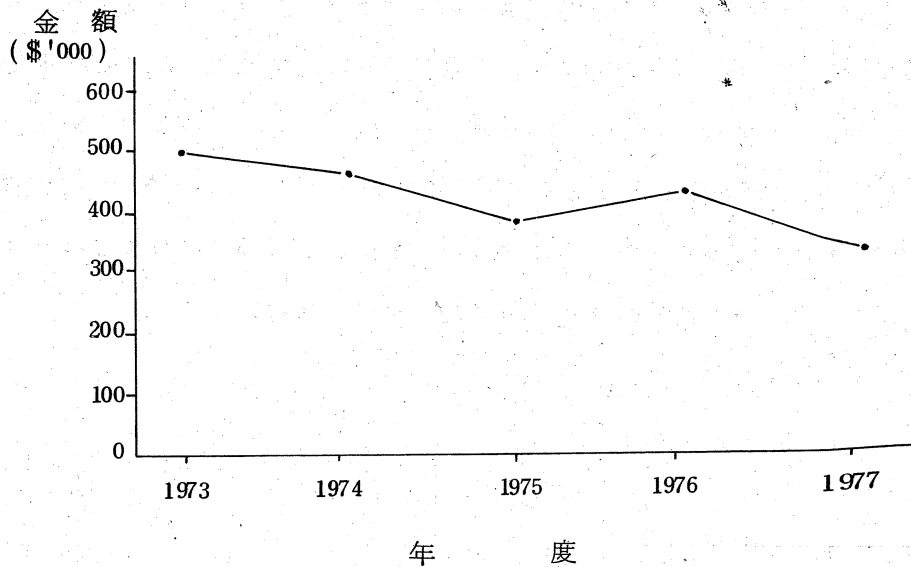
\* Bar Chart에서 수직축의 눈금은 항상 "0"부터 시작되어야 한다. 다음에 설명할 線圖表에서는 수직축의 일부를 생략할 수 있으나 Bar Chart에서는 생략될 수 없다.

Bar Chart는 수직이나 수평 어느쪽으로도 그릴 수가 있는데 수직형태가 보다 一般的이다. 수평 Bar Chart는 앞으로 복잡한 도표의 설명시 예시될 것이다.

(b) 線圖表

다음은 線圖表의 예이다.

<圖表> Utopia 國의 物產 輸出額 1973-1977



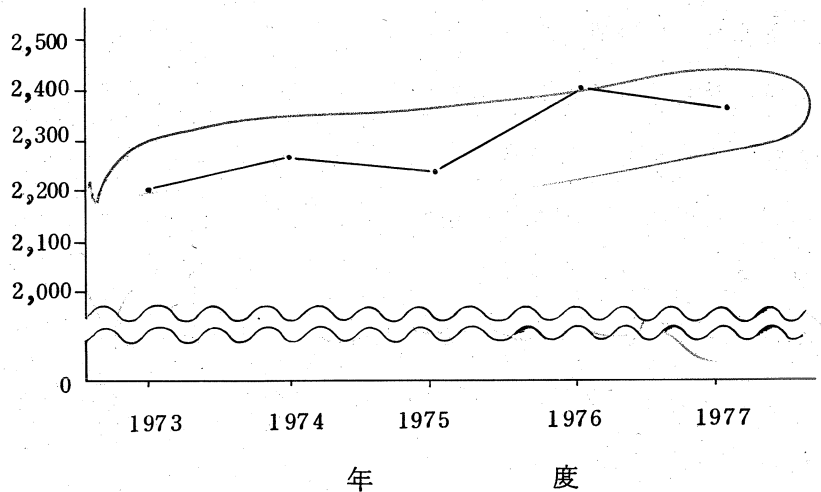
Bar Chart가 절대치의 크기를 나타내는 반면, 그래프는 경향의 변동이나 特徵을 표시하는 데 目的이 있다. 수치는 두변수의 값에 따라 그래프상에 점으로 나타난다. 이러한 점들을 보



고 변동을 파악할 수 있다. 예를들면, 위의 예에서 1975년에  
 比하여 1976年の 輸出이 增加하긴 했으나 1973년부터 1977  
 년까지는 輸出이 減少되고 있음을 쉽게 알 수 있다.

그래프는 추세나 경향을 나타내기 때문에 만약에 그래프를 편  
 리하게 볼 수 있거나 공간을 절약할 수 있는 경우에는 수평축  
 의 일부를 생략해도 좋다. 그러나 이 경우 그래프에서 수직눈  
 금을 잘라낼 때는 아래 예와 같이 두개의 선을 사용해야 한다.

<圖表> 南道の 유아사망 1973~1977



지금 위의 예에서 눈금의 일부를 잘라내지 않았다면 필연적으  
 로 수직으로 긴 그래프가 되어 곡선과 수평선사이에 긴 공간이  
 생겼을 것이다. 상기 그래프를 자세히 보면 모든 점이 2,200  
 에서 2,400 사이에 집중되어 있기 때문에 이 범위내의 경향파  
 악은 아랫부분을 잘라내고 보는 것이 더 쉽다는 것을 알 수

있다.

이러한 형태의 그래프는 곡선도표라고 칭한다. 그러나 자료점 간의 값을 알 수 없기 때문에 점을 직선으로 서로 연결하여야 한다. 이런 형태의 그래프는 점을 모두 합쳐서 볼 때만이 의미가 있다. 이것은 시간의 흐름에 따른 變動을 나타내기 위해 作成되는 것이 보통이다.

때로는 변수를 X와 Y로 표시하여 X의 變動에 따르는 Y의 變動을 合理的으로 파악할 수 있다. X의 값을 임의로 주어 Y의 값을 예측할 수도 있다. 이러한 추정은 X와 Y간의 관계를 곡선으로 나타낼 수 있다는 가정하에 可能하게 된다. 이것은 X의 값에 対応하는 Y의 값을 그래프에서 읽음으로써 할 수 있다. 우리는 이것을 "Intrapolation"이라 부른다.

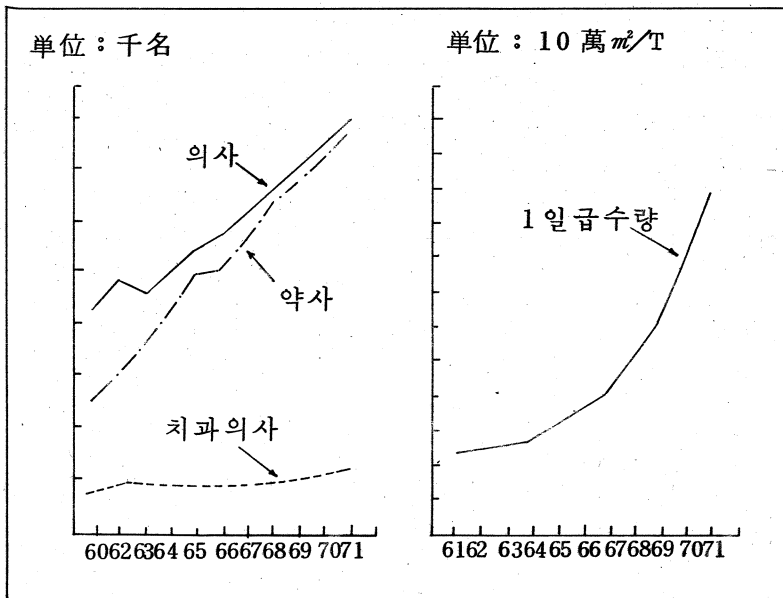
경우에 따라서 時間의 變動에 따른 종속변수의 분명한 추세를 決定할 수도 있다. 이러한 추세는 장래에 우리가 기대하는 값을 추정하는 데 이용된다. 그러한 추세가 지속이 될것을 가정하여 종속변수의 값을 추정하는 데 使用한다. 이것을 Extrapolation이라 한다.

Intrapolation과 Extrapolation方法은 決定되는 종속변수의 값을 추정하는 데 유용한데, 보다 精確한 값을 계산해 내기 위하여 수학적인 기법을 使用하는 것이 더 바람직하다. 잘 알려진 수학적인 기법으로는 "Best fitting"으로서 最精確曲線을 決定하여 必要한 추정을 할 수 있다. 이것을 회귀분석 (Regression Analysis)이라고 하며 직선적합인 경우에 직선회귀

*minimize* (X-X)

( Linear regression )라 부른다. 회귀에 대해서는 10장 7절에서 설명될 것이다.

② 時間의 進行에 따르는 여러개 변수의 變動을 보기위해서 같은 그래프에 이들을 모두 그려 넣는 경우도 있다. 보통은 3개 내지 4개의 변수가 최적이며 그 이상은 매우 혼동을 가져 오거나 해석이 어렵다. 아래 그래프에서 보는 바와 같이 여기에서는 의사수와 약사수는 치과의사의 수에 비하여 훨씬 큰 比率로 增加하고 있음을 분명히 나타내 주고 있다.



단순 및 복합 線圖表의 例

( 資料 : 經濟企劃院 (1972) : 圖表로 본 主要統計 P.54 )

(c) 파이 (Pie) 圖表

統計情報를 圖表에 나타내는 또하나의 方法은 變數의 값을 어떤 特別한 모양에 表示하는 것이다. 모양에는 사각형, 원형, 삼각형 또 그 밖의 다른 形態가 있다. 우리는 이것을 總稱하여 “Area Diagram”이라고 한다. 가령 한변이 1 cm 인 사각형이 1,000 평방미터의 곡물생산을 나타내는 것을 생각할 수 있는데 2 cm의 사각형에서는 4,000 평방미터를 나타내게 된다.

(面積이 4 배로 되기 때문이다)

크기가 서로 다른 特性을 比較하는데 있어서는 Area Diagram보다는 Bar Diagram이 그리기도 쉽고 이해하기도 쉽다.

그러나 Area Diagram의 특수한 形態로서 아주 廣範하게 使用되는 것이 있는데 이것이 Pie 圖表로서 여기에서는 이것에 대해서만 言及코자 한다.

이것의 名稱은 마치 둥근 파이와 같이 생겼기 때문에 붙여진 것이다.

Pie 圖表에서는 가령 全體 人口를 전체 원으로 表示하고 地域別 人口를 원의 部分으로 나타내는 형식을 취한다. Pie 圖表의 作成方法을 說明하기 위하여 다음 資料를 使用하여 한 나라의 人口分佈를 Pie Chart화 하겠다.

| 地 域 | 人 口 ('000) | 地 域 | 人 口 ('000) |
|-----|------------|-----|------------|
| A   | 7,709      | D   | 7,814      |
| B   | 6,521      | E   | 2,763      |
| C   | 5,991      | F   | 450        |
|     |            | 合 計 | 31,248     |

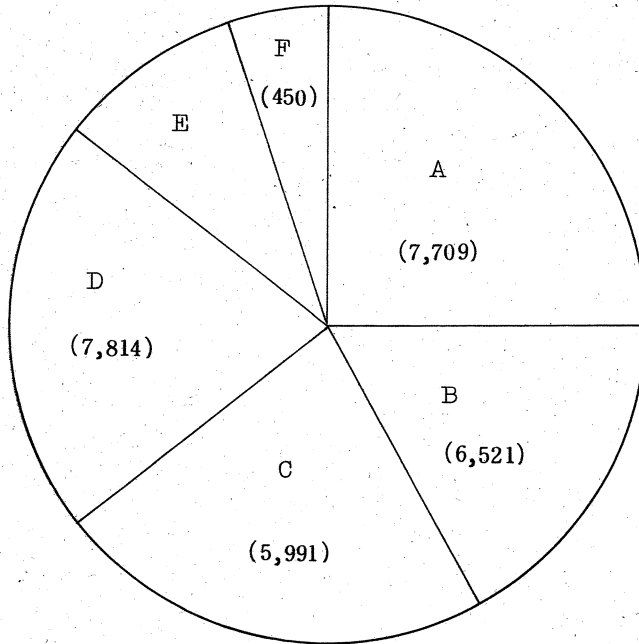
지금 원 全体를 総人口로 나타낸다고 하면, 원은,  $360^\circ$  이므로 아래와 같이 各 地域人口의 度数를 計算할 수 있다.

| 地 域 | 人 口<br>( ' 000 ) | 度 数                            |
|-----|------------------|--------------------------------|
| A   | 7,709            | $7,709 \times 360/31,248 = 89$ |
| B   | 6,521            | $6,521 \times 360/31,248 = 75$ |
| C   | 5,991            | $5,991 \times 360/31,248 = 69$ |
| D   | 7,814            | $7,814 \times 360/31,248 = 90$ |
| E   | 2,763            | $2,763 \times 360/31,248 = 32$ |
| F   | 450              | $450 \times 360/31,248 = 5$    |
| 計   | 31,248           | 360                            |

여기서 度数는 사사오입한다. 적당한 크기로 원을 그리고 원의 제일 위에서 부터 시작하여 시계方向으로 各 地域別로 計算된 度数를 配分함으로써 원을 분할하여 表示한다. 아래 Pie 圖表에서는 地域의 순서에 따라 圖表가 作成되었으나 實際로는 크기순으로 재배열한 후에 그리는 것이 훨씬 보편적인 方法이다. 이와같이 하면 圖表가 보다 아담해 보인다. 그러나 여기에 대해서는 정해진 原則이 있는 것은 아니다.

地域別 人口는 総人口에 대한 백분비로 表示할 수가 있으며 Pie 圖表도 역시 백분비로 그릴 수 있다. Pie 圖表는 統計資料가 比例的으로 細分되는 것을 나타내는데 가장 有用한 方法으로서 원 전체를 100%로 간주하여 各 部分은 百分比로 計

地域別 人口 ('000)



算하여 크기에 따라 表示하게 된다. 위 圖表에서 A地域의 比率은  $7,709 \times 100 / 31,248 = 24.7\%$ 이다.

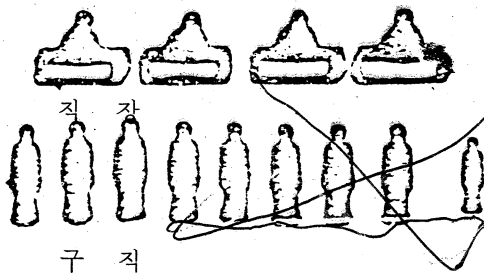
一般的으로 원을 너무 많은 部分으로 나누면 너무 細分되어 理解하기가 어렵게 된다. 一般的인 原則으로는 6~8개가 最大한 細分할 수 있는 한계이다. 또한 두개이상의 Pie 圖表를 同時에 그릴 수도 있는데 (예컨대 시점별로 작성) 이때는 모든 Pie 圖表의 變수가 같아야 한다. 그러면 이경우에 각 시점별로 원의 크기는 총계의 크기에 比例할 것이며 각 部分의 크기는 가령 地域別 人口의 크기에 比例할 것이다.

(d) 그림에 의한 圖示

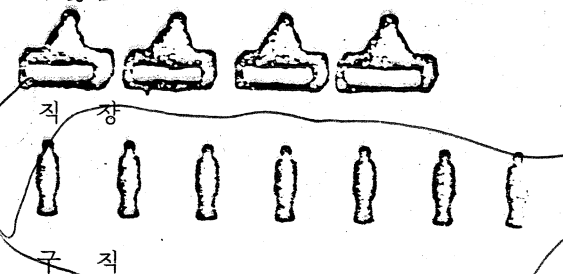
상징적인 表示와 그림은 圖表에 자주 利用된다. 이들 圖表의 目的은 딱딱한 선으로 그려진 圖表를 기피하는 사람들의 관심을 끌어 들이는데 있다. 이들 圖表가 적절히 設計된다면 統計, 經濟 및 社會的인 測定 結果의 提供을 補完 또는 改善하는 效果를 거둘 수 있다. 우리가 알고있는 바와 마찬가지로 시각적인 전달수단은 오래 記憶에 남을 뿐만 아니라 言語에 의한 전달보다는 效果가 있다. 이것에 의해서는 관심도가 약한 내용도 效果的으로 전달될 수 있다. 그림에 의한 圖表는 教育用, 宣傳用 및 여러가지 形態의 발표용으로 좋은 수단이 되고 있다. 따라서 보다 技術的인 性格을 갖고 있는 統計分野에 잘 適用되지 않는다는 것은 分명한 것 같다.

統計情報를 提供할 때 그림에 의한 方法이 몇가지 있다. 그러나 가장 보편적이고 有用한 方法은 提供하고자 하는 變數의 절대 수치를 작은 그림을 使用하는 것이다. 아래 그림에 의한 圖表는 職場數와 求職數를 比較하는 고용통계를 表示하는데 있어서 옳은 方法과 틀린 方法을 보여주고 있다. 이러한 形態의 圖表作成方法은 "Pictogram"이라 일컬어 진다.

A (틀림)



B (옳음)



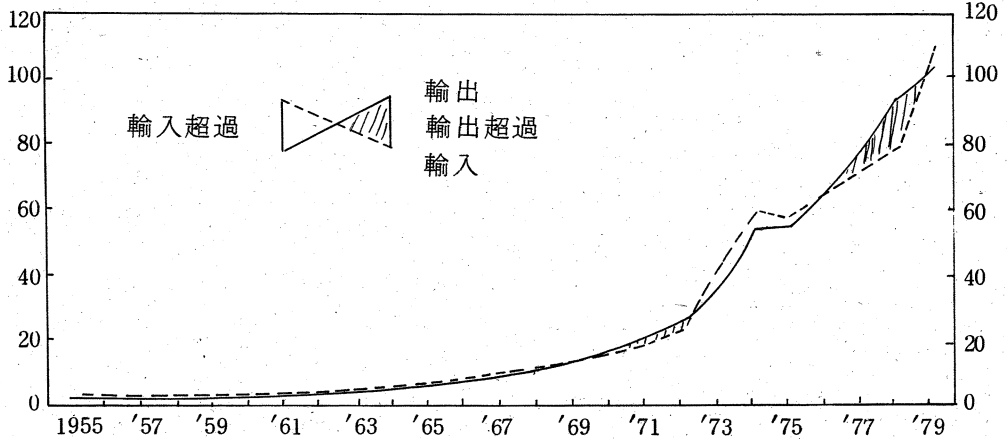
여기서 그림은 一定한 간격으로 또한 같은 크기로 배열되어야 한다. 이것은 서로 다른 두개이상의 그림群을 비교할 때 매우 중요한 사항이 된다. 따라서 위의 그림에서 왼쪽의 形態가 나쁜 이유가 여기에 있다. (위의 그림에서 보면 求職數가 職場數보다 많음에도 불구하고 分布된 全体길이는 같게 되어 있다)

(e) 보다 複雜한 圖表

우리는 이미 앞에서 基本的인 形態의 圖表에 어떤 것이 있는가를 알았다. 여기서는 이러한 基本形態가 아주 複雜한 情報의 表現을 위해 어떻게 상호 결합되는가를 알게 될 것이다. 이것은 하나의 (圖表에 많은 數의 變數가 나타난다는 意味가 될 것이다.

貿易推移

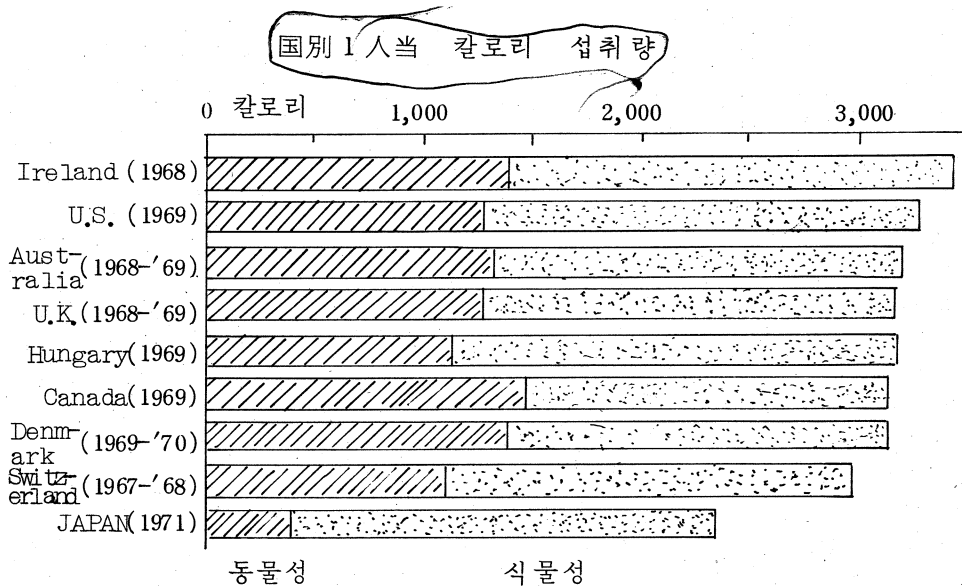
單位：10 億 弗



輸出入 均衡狀態를 表示하는 圖表의 例

(資料：日本統計局의 1980年版 統計年報)





복합 Bar chart의 예

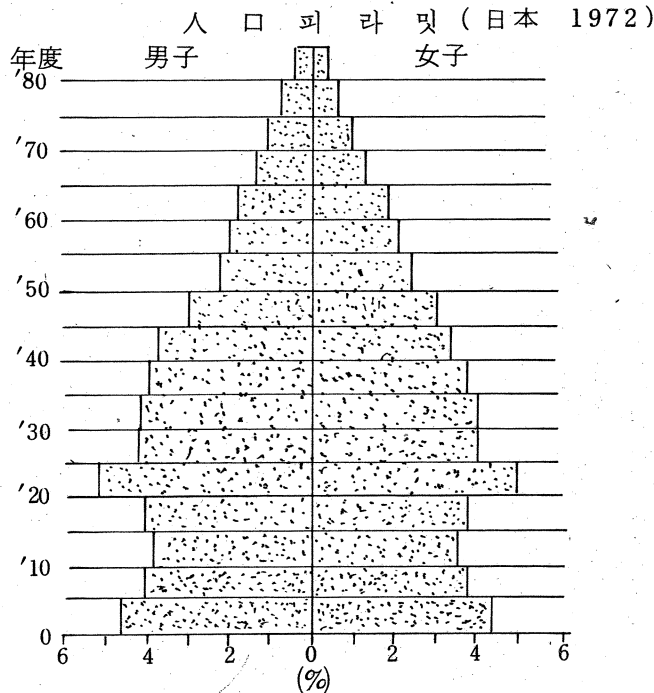
(資料: 日本統計局 1973年版 統計핸드북)

Bar Chart의 특수한 形態로서 性別·年齡別 人口構造를 나타내는 圖表가 있다. 이러한 圖表는 “Profile Chart” 또는 보다 一般的으로는 “Population Pyramid”라 불린다. 여기서는 平行으로 배열되는 Bar Chart가 使用되며 수직 Bar는 결코 使用되지 않는다. 高年齡層의 死亡率이 높기 때문에 이 圖表는 特徵的으로 피라밋형을 취하게 된다.

이러한 人口피라밋은 일정 시점에 있어서의 性別·年齡別 人口構造를 反映한다. 시점별로 여러개의 피라밋을 作成하면 시점간의 性別·年齡別 構造의 變動을 나타낼 수 있다. 피라밋의 모양에서 들어가고 나온部分은 戰爭, 經濟不況(이 경우 一般적으로 出生率이 떨어짐), 戰後베이비붐(Baby Boom), 家族計劃, 變化하는 經濟狀況 및 社会的인 특성과 같은 요인들이 人口成長에

미치는 影響을 반영하고 있다. 왼쪽과 오른쪽에 있는 Bar의 상대적인 크기는 성별 균형 또는 不均衡 程度를 表示하고 있다.

다음 圖表에서는 人口가 매년 增加하고 있으나 피라미트의 아래쪽 Bar의 길이가 긴 것과 같이 老人層 보다는 젊은層의 사람이 훨씬 많다는 것을 알 수 있다.



두개의 Bar chart를 하나로 統合시킨 圖表의 例  
 (資料：日本統計局의 1973年版 統計핸드북)

마지막으로 言及해야 할 圖表形態로는 準로그 그래프가 있다. 여기서의 수직축의 눈금이 등분이 아니고 로그로 되어있다. 이것은 不必要하게 複雜性을 띄고 있는 것 같이 볼 수가 있으나 실제로 이런 形態의 그래프는 어떤 特定狀況에 있어서의 趨勢

를 아주 現實的으로 나타낸다. 여기서는 이것을 더이상 言及하지 않겠으나 11章에서 指數를 說明할 때 로그그래프의 作成方法과 一般그래프보다 바람직한 경우등에 대해서 論議하고자 한다.

#### 9.3.4 가장 적절한 形態의 圖表選定

어떤 統計表를 나타낼 圖表가 作成되어야 한다는 決定이 이루어지면 가장 적절한 形態의 圖表가 무엇인가에 對한 要求가 發生될 것이다. 目的에 따라서는 圖表의 形態에 따라 適合한 것이 있고 그렇지 못한 경우가 있다. ~~예컨대 Pie 圖表는 總人口 및 地域別 人口에 對한 2~3개 시점에 對한 變動을 나타낼 수 있다. 그러나 많은 시계열에 對해서는 적합치 않다. 總人口 또는 人口 構成要素들의 變動이 極히 적을 때는 어떠한 形態도 適合치 않다. 이것은 시각적인 印象이 전달되지 못하기 때문이다. 線圖表는 시계열자료의 提供方法으로서 또는 적은 變動을 比較하는데 아주 適合하다.~~

어떤 特別한 統計表에 適合한 圖表의 形態는 可能할 수 있다. 그러나 嚴格하게 규정된 것은 없다. ~~특별한 目的에 가장 適合한 形態의 圖表를 選定하기 위하여는 아래와 같은 面을 自問해 보아야 할 것이다.~~

- (a) 선정된 圖表가 資料를 포괄적으로 나타내고 있는가?
- (b) 圖表가 大體的으로 結果를 正確하게 提供하고 있는가?
- (c) 圖表의 크기가 모든 利用者들을 만족시키는가?
- (d) 印刷에 의해서 제공되는 경우 시각적으로 均衡을 이루고

있는가?

(e) 수직과 수평 축의 크기가 均衡을 이루는가? 또한 글자배열이 조화를 이루는가?

(f) 圖表가 収録된 책자의 재작성이 可能한가? 圖表의 印刷가 容易한가?

(g) 당초 의도한 內容이 시각적인 수단을 통하여 效果的으로 反映되는가?

#### 9.4 報告書 解説의 作成

##### 9.4.1 概要

統計月報, 年鑑 등과 같이 定期的으로 發刊되는 綜合統計 刊行物과는 別개로 大개의 統計調査結果는 調査가 完了되면 報告書로 發刊되는 것이 보통이다. 여기에는 定期報告書(예컨대 年間報告書) 또는 一回限 報告書가 있다.

統計情報의 作成이 完了됨과 同時에 加급적 迅速하게 公표하기 위하여 本報告書가 作成되기 前에 미리 新聞이나 라디오를 통해서 發表하는 경우가 있다. 이것은 후에 全體內容을 상세하게 収録한 本報告書의 發刊에 앞서 “暫定報告書”를 통하여 이루어진다. 新聞에도 統計表가 실리긴 하지만 新聞에 通常的으로 나타나는 提供方式은 해설원고형태(Textual)이다. 이것은 해설원고에 의한 資料만 使用하는 라디오放送에서는 더욱 그러하다. 따라서 뉴스公表方式에 있어서는 1~2개 정도의 統計表를 附錄

으로 한 해설원고를 利用하는 경우가 大部分이다. 統計結果는 해설원고에 가령 다음과 같은 內容으로 정리되어 公표된다. 즉 “政府統計에 의하면 5月달의 出生은 4月の 198에 비하여 8.1%가 增加한 214로 밝혀졌다.”

#### 9.4.2 報告書 作成

結果報告書는 經常 統計系列의 作成보다는 特別調查에 대해서 作成되는 것이 보통이다. 이렇게 하는 이유중의 하나는 經商 통계계열은 주기적으로 發表되기 때문에 結果報告書에 매 發表時마다 解説을 収録한다면 業務量이 엄청나게 커지가 때문이다. 다음으로는 대개의 經商 통계계열은 포괄적인 目的을 갖고 있어서 結果에 대한 各양各색의 利用者를 모두 만족시킬 수 있는 報告書의 作成이 아주 힘들기 때문이다. 나아가서는, 報告書의 作成에는 많은 時間이 所要되므로 報告書가 完了될 때까지 規則적으로 作成되는 통계계열의 公표를 늦추는 것은 바람직하지 못하기 때문이다.

報告書에는 解説과 結論에 따른 分析結果의 収録외에도 調查의 行政的인 또는 技術的인 측면에 대한 서술을 包含해야 한다. 여  
각서는 다음과 같은 內容이 包含되어야 할 것이다.

- (a) 調查의 배경과 目的을 서술하는 調查概要
- (b) 資料의 範圍, 對象, 概念, 定義, 分類 및 項目이 說明되는 調查設計, 만일 標本調査인 경우는 標本設計도 여기에 包含되어야 한다.

- (c) 資料蒐集方法
- (d) 資料의 正確性 및 限界
- (e) 統計表, 函表, 結果要約과 結果에 대한 技術的인 註記等을 包含하는 結果分析
- (f) 結 論
- (g) 調査票明細 및 調査員에 대한 설명과 같은 其他 情報를 包含하는 附錄

統計報告書는 論理的이고 体系的인 形態로 쉽게 參照가 되도록 하는 것이 바람직하다. 예컨대 本 教育教材는 章, 節 等과 같이 作成되어 있음을 알 수 있다. 지금 간단히 보면 지금 說明되고 있는 부분은 第9章 4節의 2에 속해있고, 9.4.2라는 번호가 부여되어 있다. 第8章에 言及된 바 있는 Hierarc-hical code system과 같은 原則을 適用하고 있는 이러한 種類의 作成方法을 統計報告書에서는 可能한 限 適用해야 한다.

처음에는 關係專門家들의 意見을 들어보기 위한 報告書의 초안을 作成하는 것이 一般的인 경향이다. 이러한 意見들이 綜合되면 最終報告書가 作成, 發表된다.

#### 9.5 統計刊行物の 作成

統計刊行物은 統計調査의 最終結果로서 여기에는 統計作成의 3가지 方法이 混合되어 있다. 報告書가 發刊될 때는 信賴度가 明示되어야 한다. 이렇게 하여 政府機關의 報告書는 公式統計가 되는 것이다. 따라서 信賴性있는 統計情報의 迅速, 적

개발안은 모든 統計組織의 窮極的인 目標이다.

統計刊行物は 다음과 같이 4 가지로 大別할 수가 있다.

(a) 季報, 年報와 같이 定期的(月刊, 分期, 年刊)으로 發行되는 一般刊行物

(b) 貿易統計, 物價指數, 人口動態統計, 고용통계와 같이 주기적으로 發行되는 調査別 다일보고서

(c) 5年 또는 10年 간격으로 作成되는 인구센서스報告書보다 細分된 情報를 여러 권에 나누어 収録한 最終報告書에 앞서 速報로 發刊되는 暫定報告書

(d) 必要時 隨時로 發刊되는 特別報告書

하나의 간행물을 作成하기 위해서는 고려해야 할 要素가 많이 있는데 다음과 같은 점이 특히 重要시 되고 있다.

(a) 可能하다면 해설과 統計表 및 圖表를 經驗 是 統計專門家로 하여금 檢討할 수 있도록 조정

(b) 統計表, 圖表, 計算된 結果, 해설과 技術的인 註記 및 結論은 가장 유용한 形態로 作成되어야 한다는 점을 명심

(c) 表記의 形態, 제목선정 및 번호부여방법, 그래프, 統計表와 圖表가 適合한지의 여부검토

(d) 오류검사, 정오표작성, 배부와 판매에 대한 準備, 最終 檢査 단계에서 모든 統計表가 Cross Check되어야 함을 명심, 이는 초안작성 및 타자時 발견하지 못한 오류를 발견하기 위함이다.

(e) 資料의 基準年度, 단위, 정의 및 한계나 일러두기 또는 報告書의 概覽부분에서 자세히 설명되어야 함을 認識 統計作成機

關에서는 흔히 매회 간행물 발간시마다 概念이나 定義를 反復하여 수록하지 않고 있다. 이러한 경우에는 이들이 수록된 간행물호수(회수)를 명시해 주어야 한다.

(f) 천단위 또는 백만단위까지의 사사오입이 적절하다고 판단되는 경우 사사오입하여 収録할 것

(g) 統計調査의 調査基準일 또는 対象期間을 表示할 것. 예컨대 人口資料는 1月1日 또는 7月1日 基準과 같이 表示될 것이며 재정자료는 歷年이 아닌 “會計年度”로 나타날 것이다.

(h) 편집방향이 바뀐 다음에 작성되는 資料는 그 이전 資料와 嚴格하게 比較할 수가 없으므로 적절한 주기를 달아서 利用上 不便이 없도록 해야 한다. 경우에 따라서는 資料가 없는 계열을 표시하기 위하여 관련부호를 使用하기도 한다.

(i) 暫定數值가 包含된 경우 이들은 후에 조정된 數字로 바뀌게 되므로 적절한 주기가 必要하다. 유사한 경우로서 前號의 暫定數值가 後號에서 다른 數值로 대체된 때에도 적절한 주기가 必要하다.

(j) 해설에도 통계표나 도표가 包含된 다는 점을 명심할 것. 해설내용의 어떠한 結論에도 統計表가 必要하다.

(k) 数学에서 使用되는 기호나 公式은 부록에 包含시키는 것이 바람직하다.

#### 9.6 統計資料의 秘密保護

資料가 蒐集되고 統計가 作成되어 발간될 때 가령 어떤

統計資料 사용



人口의 全般的인 構造를 數字的으로 나타내하고자 한다. 그러기 위하여는 개개 조사단위별로 모든 事項 또는 標本으로 選定된 事項을 調査하게 된다. 이것은 大部分의 경우 個人的인 또는 事業上의 情報를 蒐集하는 것을 意味하게 된다. 統計業務에 종사하는 사람들이 이러한 個人 또는 事業에 관련된 세밀한 內容을 包含하거나 公표된 資料로부터 이러한 細部事項의 유도가 가능한 情報를 결코 발간(또는 公표)하지 않겠다는 것을 단호하게 確信한다는 것은 매우 重要한 責任이다.

情報의 蒐集은 때로 법령의 效力에 의해서 蒐集된다. ( 통계법 規定에 의한 직접수집 또는 統計가 부산물로 作成되는 行政遂行을 위한 법령으로부터의 間접수집 ) 이러한 법령에는 情報에 대한 비밀보호와 위반시 罰則이 規定되어 있다. 그러나 統計情報은 거의 大部分이 自發적으로 수집되는데, 법령에 의한 강제규정이 없더라도 応答者의 협조에 의해서 수집된다. 이렇게 하여 資料가 蒐集될 때는 ~~응答者가 提供한 資料는 누설될 수 있는 어떠한 경우에도 사용되지 않는다는 점을 약속하는 것이 通例이다.~~

이러한 비밀을 어떻게 해서라도 지킬 수 있도록 주의하는 것이 重要하다. ~~아무도 특정인 또는 특정기업으로 부터 蒐集한 內容과 이와 유사한 內容이 수록된 책자는 항상 비밀이 보장되어야 한다는 점에 대해서 비판을 해서는 안된다.~~ 이러한 資料를 다른 政府機關이나 여타 機關에 提供해서는 안된다는 것도 사실이다.

統計表作成에 있어서 단지 몇사람이 応答한 情報은 公開할 수 없다는 規定이 있어야 할 것이다. 예컨대 특정 제품을 生産하는 두개의 企業이 있다면 各 企業의 立場에서 보면 總計에서 자기기업分을 빼면 상대방의 情報을 거의 正確하게 추정할 수 있다. 또한 一般에게 公開되기 전에 미리 個人的인 資料要求에 따라 提供할 수 있는지의 여부에 대해서도 規定이 있어야 할 것이다. 이러한 規定위반시에는 政治的인 制裁, 또는 不리한 營業上의 措置등을 가할 수 있을 것이다.

資料의 秘密保護와 發表에 대해서 대부분의 國家에서는 상세한 規定을 두고 있다. 여러분들은 우리나라의 規定은 어떠한가를 알아야 하며 세밀히 관찰해 보아야 한다.

#### 9.7 其他의 統計結果 提供方法

統計刊行物은 센서스, 統計調査, 其他調査의 結果를 一般에게 公표하는 가장 보편적인 方法이다. 이 方法은 比較的 經費가 적게 들며 利用이 쉽다. 그러나 印刷할 수 있는 資料量이 制約되어 있고 限定된 紙面의 크기때문에 간행물은 모든 資料의 収録이 어렵다. 예를 들면 人口센서스에는 수많은 Cross tabulation이 可能하고 이들중 一部는 量이 엄청나다. 예컨대, 職業(1,200) × 年齡(12個의 5歲年齡 階級群) × 性別(男子, 女子, 合計) = 43,200개의 資料(Information cell)가 생긴다. 지금 페이지당 600개의 Cell을 収録한다고 가정할 때 이 統計表는 70페이지의 分量이 된다. 더구나 이 資料가 300개의

地域別로 作成된다고 하면 20,000 페이지 所要될 것이다.

刊行物에 収録되는 資料의 量은 管理가 可能하고 가장 관심의 대상이 되고 利用頻도가 높은 정도로 한정할 수 밖에 없다.

한편, 아주 세분된 資料는 小地域開發計劃과 같은 특별한 目的을 위해서는 重要的 資料가 될 수 있다. 따라서 이와 같은 資料는 어떤 다른 形態로 이용이 可能하거나 컴퓨터處理가 可能해야 한다. 아주 특수한 統計表의 利用도 또한 必要時 作成이 可能해야 한다.

여기에 대해서는 다음과 같은 解決方案을 고려할 수 있다.

(a) 統計作成機關에서 作成되는 아주 세분된 資料는 컴퓨터 Print out 形態로 保管된다. 이용자는 그들이 必要로 하는 特殊한 資料를 求하기 위하여 統計作成機關을 訪問할 수 있다. 이것은 시간이 걸리고 귀찮은 일이다. 그러나 여건에 따라서는 이것이 유일한 方法이기도 하다. 政府部處와 같이 重要的 利用者들에게는 Print-out의 사본을 제공할 수도 있다.

(b) Microfilm과 Microfiche와 같은 形態인 Microphotography가 広範하게 使用되는데 여기에는 細分된 資料, 調査表 또는 다른형태, 심지어 統計刊行物의 內容까지도 수록·保存이 可能하다. 統計資料는 이와같은 形態로 提供하는 것이 아주 편리하다. 그러나 여기에 記錄된 資料를 실제로 이용할 때는 특수한 形態의 擴大裝置 또는 읽어내는 裝置가 必要하다.

(c) 또한 자기테이프나 Diskette로도 資料의 提供이 可能하다. 이런 形態로 資料가 提供되면 利用者は 자신의 컴퓨터에 쉽게

입력하여 必要時에는 分析 또는 다른 目的으로 出力해 낼 수 있다.

(d) 경우에 따라 이용자는 자기 사무실에 설치된 Terminal (이는 每 컴퓨터에 연결되어 있음)을 통하여 저장된 資料를 직접 利用할 수도 있다.

(e) 特殊目的을 위한 分析: 統計調査의 集計用 프로그램은 利用者가 必要로 할 것으로 예상되는 모든 統計表(또는 적어도 重要的 統計表)를 包含하고 있다.

이러한 統計表의 作成에는 많은 時間이 所要되는데 센서스 같은 큰 프로그램의 작성은 몇년이 걸릴 수도 있다. 이러한 与件下에서는 統計表의 作成이 프로그램化되어 있지 않다면 特殊目的을 위한 統計表를 기대하기가 매우 어렵다. 따라서 集計프로그램의 작성이전에 가급적 여러 이용자와 협의하여 必要的 統計表를 선정하는 것이 매우 중요하다.

이와 같이 協의를 거치더라도 특수목적의 統計表가 必要하다는 認識은 그 후에 발생하는 경우가 있다. 이러한 특수한 수요에 대비해서 자기데이터에 基本統計資料(秘密保護를 위한 措置가 이미 끝난 資料)를 準備해 두는 것을 생각할 수 있다. 特殊目的을 위한 分析을 원하는 사람들은 이 資料를 利用하여 目的을 달성할 수 있다.

## 第 10 章 統 計 分 析

### 10.1 概 要

앞장에서 統計 發刊과 公表의 一般的인 事項에 關係 論議했다. 여기서는 보다 専門적인 견지에서 살펴보고 資料가 어떻게 分析을 위해 적절한 形態로 表示될 수 있는가 그리고 가장 一般的인 分析 技法을 몇가지 알아보자.

本章에서 過程의 첫段階로 数学的 表記法을 알아보자. 그러나 상당히 單純한 概念만 알아보고, 複雜한 数学을 사용하지 않고 소개해 보겠다.

아래 6가지 表題로 論議할 主題를 分類한다.

- 度数分布
- 平均과 여러 位置測度
- 散布度
- 標本誤差
- 標本데이터로 부터 母數推定
- 推定과 豫測을 위한 統計資料의 利用

### 10.2 度数分布

#### 10.2.1 基本概念과 原理

統計調査의 結果로써, 單位集合의 하나 혹은 많은 特性에 關한 一連의 觀測值를 갖는다. 이런 單位集合은 全体 母

集團을 意味할 수도 있고 標本일 수도 있다. 이 資料 (Data) 를 어떤 生生한 情報의 損失없이, 쉽게 理解할 수 있도록 하기 위하여 要約하는 方法이 必要하다. 例를들면, 시골 農家調查에서 우리는 헥타르 (Hectare)로서 各家口가 所有하는 總面積을 나타내는 一連의 測定値를 얻을수 있다. 여기서 거의 수천의 相異한 數字가 나타나는 리스트를 이해하는데 어려움이 있다. 우리는 리스트를 이해할 수 있는 方法으로 나타내 보고자 한다. 度数分布는 이때 가장 一般的으로 使用된 方法이다. 아래表 (Table) 는, 한지방의 全体農場面積別 코프라 (Copra) 農場의 數를 要約한 度数分布를 나타낸다.

| 농장면적 (헥타르) | 농장수   |
|------------|-------|
| 0 ~ 49     | 1,467 |
| 50 ~ 99    | 512   |
| 100 ~ 199  | 656   |
| 200 ~ 299  | 529   |
| 300 ~ 399  | 374   |
| 400 ~ 599  | 321   |
| 600 ~ 799  | 118   |
| 800 ~ 999  | 59    |
| 1,000 이상   | 35    |
| 총 계        | 4,071 |

이 度数分布는 各農場面積을 나타낸 4,071이라는 要約된 數值

를 갖는다. 確實히 上記表는 이러한 數들이 단지 크기순 으로  
만 分類되어 있다 하더라도 4,071 개의 리스트가 나타내는것 보  
다 훨씬 더 쉽게 이해할 수 있다.

전형적인 度数分布는, 우리가 觀察한 特性値의 값의 範圍를 相  
異한 階級으로 나눈다음 각 階級內的 觀測値의 數나 度数를 헤  
아린다. 우리는 變數와 計數로서 度数分布를 만들수 있다. 그  
러나 이 表題에서 거의 대부분이 變數에만 관한 것이다. 나중  
에 變數의 두가지 基本形態를 區別할 것이다.

度数分布는 어떻게 變數의 값이 分布되어 있는가를 把握하는데  
매우 有用하다. 즉 얼마나 높은값과, 낮은값이 觀察되었는가, 가  
장 빈번히 나타나는 값은 어떤것인가, 그리고 값이 範圍간에 對  
稱的으로 分布되었는지 혹은 大部分이 한쪽끝에 分布되어 있는지  
를 일목요연 하게 나타낸다.

주어진 基本的인 데이터의 리스트로 度数分布를 만드는데 4段  
階로 구성된다.

- (a) 分類하기 위하여 特性値를 定義한다.
- (b) 데이터를 그룹化하여 階級을 明示한다.
- (c) 이런 階級으로 데이터를 分類한다.
- (d) 各階級內에서의 觀測回數를 把握

이 마지막 두단계는 아주 간단하지만 우리가 必要한 階級の 數  
와 階級에 대한 값의 범위를 결정하는데는 많은 어려움이 있다.  
農場面積을 나타내는 앞 테이블에서 9개의 階級을 갖고 級區間  
( 혹은 範圍 )은 첫번째 두 階級에서 50헥타르 이고 다음 세

階級은 100 헥타르 이며 마지막 階級은 限界가 없다는 것을 알 수 있다. (다시말하면 이것은 끝이 열려있는 것이다. 1,000 헥타르나 그 이상의 觀測値는 여기에 포함되어 있을 것이다)

原則에도 불구하고 우리는 分布에 적합한 階級の 어떤 形態를 決定할 수 있다. 그리고 各 階級の 定義는 分布의 目的에 달려 있을것이다. 다음의 有用한 몇가지 原則이 있다.

(a) 階級은 가장 낮은값 부터 가장 높은값까지 모든 觀測値를 충분히 포함하는 範圍를 選擇해야 한다.

(b) 階級間에 間隔이나 重複이 없어야 한다. 各 階級은 分離되고 個別的이어야 한다. 各 階級の 範圍는 各 觀測値가 단정한 階級에만 내포될 수 있도록 定義하는것이 특히 重要하다.

만약, 코프라 農場에 關한 前例에서 0 ~ 50, 50 ~ 100, 100 ~ 200, 등으로 階級이 부주의하게 記述되어 있다면 우리는 正確히 100 헥타르의 農場이 두階級에 다 포함될 수 있기 때문에 어떻게 分類할지 알지 못할 것이다. 階級은 모호함이 없이 確實히 해야한다. 그리고 두階級の 境界에 있는 觀測値들은 두階級中 어느 하나에 포함되어야 한다.

(c) 너무 많은 階級이 있어서는 안된다. (原資料에 關한 度数 分布의 利點을 상실할 것이다) 또한 너무 적어서도 안된다 (매우많은 情報를 상실하기 때문이다) 即, 一般的으로 階級이 5 이상이거나 16 以下로 분할되는것이 適當하다는 結論을 意味한다. 그러나 이와같은 것은 確固한 制限이 아니고 階級數의 構成은 資料의 性質에 달려있다.



(d) 길이가 같은 階級間隔은 실제 利用에 有理하다. 均等한 階級區間은 分布를 理解하고 적절한 圖表를 作成 하는데 尙당히 편리하다. 만약 同一하지 않는 階級을 사용하면 한階級の 度数와 다른 階級の 度数를 比較하는데 많은 어려움이 있다. 하지만 때로는 資料의 다양성과 資料 利用의 信賴性에 대한 必要때문에 同一하지 않는 階級을 使用해야할 때가 있다.

(e) 階級の 마지막값을 정할수 없는 分布의 마지막 階級인 端開區間の 使用은 可能한 限 피해야 한다.

이미 觀測했듯이 (d)항에서 言及한것처럼 農場面積의 度数分布는 同一한 階級區間을 갖지 않는다. 여기서 同一한 間隔, 즉 200 헥타르를 갖는 階級으로 데이터를 나타낼 수 있다. 그러면 그 度数分布는 다음과 같다.

| 농장면적 (헥타르) | 농장수   |
|------------|-------|
| 0 ~ 199    | 2,635 |
| 200 ~ 399  | 903   |
| 400 ~ 599  | 321   |
| 600 ~ 799  | 118   |
| 800 ~ 999  | 59    |
| 1,000 이상   | 35    |
| 총 계        | 4,071 |

이 度数分布는 첫번째 階級이 모든 階級の 約  $2/3$  (4,071 중 2,635)를 포함하고 있기 때문에 처음 度数分布表 만큼 더

有用한 情報을 나타내지는 않는다. 처음에 보았던 것처럼 이 階級을 더 작은 階級으로 분할하는것은 이렇게 큰 度数에서 農場의 크기에 關한 追加的인 情報을 提供하기 위해 바람직한 일이다. 一般的으로 觀測值들이 均等하게 分布되어 있을수록 훨씬 더 쉽게 同一한 階級區間을 갖는 度数分布를 만들수 있을것이다.

農場 度数分布의 最終階級은 上記(e) 原則의 권고에서 불구하고 端開(Open-ended)인 것에 주의해야 할것이다. 그러나 端開를 使用치 않는다는 것은 사실상 不可能하다. 例를들어 15,000 헥타르의 한두개 매우 큰 農場을 가지고 있다면 우리가 使用했던 端開階級을 피하기 위하여 마지막 階級을 상이한 區間을 갖는 많은 階級으로 분할 해야만 할 것이다. 이러한 觀測值는 이미 이 階級內 모든 觀測值의 1%보다 적기 때문에 더이상 분할하는 것은 바람직 하지않다.

度数分布가 어떻게 이루어 지는가를 例를 들기 위하여 상이한 系列의 資料를 利用하고 이章을 공부할때 수시로 이 例를 言及할 것이다.

63 마리 魚族 무게 (kg)

|     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 4.6 | 3.9 | 2.8 | 6.6 | 4.2 | 3.7 | 3.7 | 5.9 |
| 3.2 | 2.2 | 3.2 | 4.1 | 3.1 | 3.0 | 4.8 | 4.1 |
| 2.1 | 4.2 | 5.0 | 4.6 | 5.4 | 2.4 | 6.3 | 2.9 |
| 5.3 | 4.0 | 4.7 | 3.6 | 3.3 | 6.9 | 4.5 | 2.5 |
| 5.4 | 5.7 | 3.8 | 4.1 | 5.6 | 6.2 | 3.0 | 3.3 |
| 5.0 | 5.4 | 3.4 | 4.4 | 4.0 | 3.6 | 5.0 | 4.1 |
| 4.8 | 7.2 | 6.4 | 3.0 | 3.5 | 5.8 | 7.7 | 3.9 |
| 2.6 | 7.9 | 3.3 | 5.5 | 4.3 | 3.9 | 6.3 |     |

상이한 階級을 決定하기 위하여 무엇보다도 먼저 가장 낮은값 부터 가장 높은값 까지의 範圍를 알아야 한다. 이 경우에 가장 낮은값은 2.1 kg이고 가장 높은값은 7.9 kg이므로 그 範圍는 5.8 kg이다. 範圍를 많은 階級으로 区分하고 觀測值들이 매우 均等하게 分布되어 있기 때문에 不均等한 區間을 갖는 階級을 扨할 必要가 없다. 적당한 分布가 다음表와 같이 주어진다.

魚族 무게 資料의 度数分布

| 무 계 (kg)  | 도 수 체 크     | 도 수 |
|-----------|-------------|-----|
| 2.0 ~ 2.9 | ###         | 7   |
| 3.0 ~ 3.9 | ### ### ### | 19  |
| 4.0 ~ 4.9 | ### ### ### | 16  |
| 5.0 ~ 5.9 | ### ###     | 12  |
| 6.0 ~ 6.9 | ###         | 6   |
| 7.0 ~ 7.9 |             | 3   |
| 계         |             | 63  |

### 10.2.2 연속 데이터와 離散 데이터

度数分布 학습에 있어 이 대목에서 度数分布의 作成과 分布의 圖表作成은 變數의 形態에 따라 매우 상이하기 때문에 變數의 상이한 두 形態를 考察해볼 必要가 있다.

一般的으로 變數는 두 形態로 區別되는데 變數는 첫째, 定해진 範圍內에서 어떤값을 취할수도 있으며, 둘째, 一定한

값만을 취할수도 있다. 첫번째 形態를 연속형 變數라 하고 또 연속형 데이터라고도 하며 두번째 形態를 離散型 變數라 한다.

연속變數의 例

- (a) 각개인의 신장
- (b) 農家耕地面積
- (c) 魚族의 무게
- (d) 피용자의 年間所得
- (e) 日中 最高氣溫

離散變數의 例

- (a) 家口當 家族數
- (b) 住宅當 房數
- (c) 가축에서 소의數
- (d) 會社當 피용자數
- (e) 두개 주사위를 던졌을때 얻어진 점수

大部分의 경우 離散變數는 필수적이 아니라 할지라도 정수값을 취한다.

실제 연속 데이터는 變數를 測定하고 기록함에 있어 制限이 있기 때문에 절대 연속이 아니다. 예를들면 우리는 정밀한 기구를 가지고 있으면 사람의 신장을 mm (밀리미터)까지 測定할 수 있을것이다. 그러나 실제로는 cm (센티미터)까지만 올바르게 測定하고 있다. 年間所得은 理論上 거의 Cent (센트)까지 把握할 수 있으나 記錄할때는 大部分 Dollar (달러)까지만 記錄한다.

그래서 연속적인 데이터로 부터 度数分布를 作成할때 여러 階級을 分類하는데 상당한 주의를 기울여야 하고 이미 시도한 반올림에 대해서도 고려를 해야 한다. 前例와 같이 무게는 측정 단위(kg)의 상하 10번째 값을 취하는데 실제로 그 무게는 어떤 값을 갖게 된다. 그래서 階級 2.0~2.9 kg에는 고기의 무게가 반올림후 2.0 kg 이상 2.9 kg 미만인 모든 魚族이 포함된다. 즉 실제 階級 區間은 1.95 kg 부터 2.59 kg까지다. 고기무게가 1.96 kg 인것은 2.0 으로 반올림 될것이고 이 集團에 포함된다.

離散的 데이터로 부터 度数分布를 作成하려면 여러가지 問題가 나타난다. 한 階級內에 한값의 變數를 갖는한 별로 問題가 되지않는다. 예를들면 住宅調査에서 1, 2, 3, 4 개 ... 등등의 房數別 住宅의 數를 要約해서 데이터를 나타낼 수 있다. 그러나 가끔 값의 範圍는 너무커서 各 階級間의 값을 舍해야 한다. 예를들면, 牛數別 牛集團數의 分布는 아래와 같다.

集團의 規模別 牛集團數

| 集團規模 (牛數) | 集團의 數 |
|-----------|-------|
| 1 ~ 10    | 366   |
| 11 ~ 20   | 118   |
| 21 ~ 50   | 55    |
| 51 ~ 100  | 11    |
| 101 ~ 500 | 19    |
| 500 以上    | 5     |
| 計         | 574   |

上記表는 魚族무계分布에 관한 表와 같이 보이지만 資料가 離散的인 것에 주의해야 한다. 예를들면 12.5 마리 소의 集團크기에 관해 說明할 수가 없다.

### 10.2.3 用語의 定義

度数分布에 관해 論議할때 어떤 用語를 定義해야 할 必要가 있다. 이러한 用語는 이미 使用한 것이지만 이 節에서 보다 더 깊이 考察해 보자.

#### (a) 階級度数

한分布에서 階級度수는 特定階級内に 있는 觀測值의 數이다. 表形式으로 度数分布를 나타낼때 階級은 항상 왼쪽난에, 階級度수는 오른쪽난에 記錄한다.

#### (b) 階級限界

주어진 階級内の 가장 작은값과 가장 큰값을 階級限界라 한다. 魚族무계 테이블(表)에서 階級限界는 2.0, 2.9, 3.0, 3.9 등이다. 下限階級限界(2.0, 3.0, 4.0 등)와 上限階級限界(2.9, 3.9, 4.9 등)는 區別되고 있다.

#### (c) 階級值

階級值는 階級の 中央点이고 上限과 下限 階級限界의 算術平均으로 얻어진다. 例에서 階級值는 2.45, 3.45, 4.45 등이며 이것은 中央值, 中央点, 中央값등으로 불리운다.

(d) 階級區間(範圍)

階級區間은 어떤階級の 길이이며 값의 範圍를 나타낸다. 한階級の 階級區間은 그階級の 下限階級限界와 다음階級の 下限階級限界사이의 差이다. 만일 모든 間隔이 同一하다면 연속적인 階級值間의 差異역시 同一하다. 階級の 上限과 下限限界 사이의 差는 반드시 階級區間이 아님에 유의하라. (魚族表에서 階級間隔은 0.9 kg으로 同一하다)

10.2.4 累積度數分布

度數分布는 여러 階級の 觀測數에 關한 情報를 提供한다. 表에서 얼마나 많은 작은 規模와 큰規模의 觀測值가 있는지 첫눈에 알 수 있다. 그러나 가끔 약간씩 다른 情報가 必要的인 경우가 있다. 이경우에 중요한것은 얼마나 많은 觀測值가 어떤 特定값보다 크거나 작은지를 찾는 것이다. 예를들어 農家所有 面積分布를 把握할때 5헥타르보다 작은 가구가 얼마나 되고 또 100헥타르 보다 큰 가구가 얼마나 되는지를 알아보는 것이다. 所得分布를 고려할때 年間所得이 500 \$ 以下인 家口가 얼마나 되는지 찾기를 원할것이다. 이 形態의 情報는 累積度數分布에서 쉽게 얻을수 있다. 앞에서 주어진 코코넛 農場面積의 데이터에서 累積度數分布를 다음 方法으로 만들수 있다.

| 農場面積 (헥타르) | 農場 數  | 累 積 度 數 |       |
|------------|-------|---------|-------|
|            |       | 以 下     | 以 上   |
| 0 ~ 49     | 1,467 | 1,467   | 4,071 |
| 50 ~ 99    | 512   | 1,979   | 2,604 |
| 100 ~ 199  | 656   | 2,635   | 2,092 |
| 200 ~ 299  | 529   | 3,164   | 1,436 |
| 300 ~ 399  | 374   | 3,538   | 907   |
| 400 ~ 599  | 321   | 3,859   | 533   |
| 600 ~ 799  | 118   | 3,977   | 212   |
| 800 ~ 999  | 59    | 4,036   | 94    |
| 1,000 이 상  | 35    | 4,071   | 35    |
| 計          | 4,071 |         |       |

累積度數分布은 各 階級度數의 累積計算으로 얻어진다. 上記表의 例처럼 두가지 方法중 하나로 얻어진다. 以下 累積度數分布은 첫번째 階級으로부터 計算을 시작하고 마지막 階級까지 各 階級の 累積度數를 더해 나간다. 以上累積度數分布은 마지막 階級부터 計算을 시작하여 위로 進行한다.

그 名稱이 암시하듯이 두 累積分布은 다음 문제에 대한 答으로 利用된다. 즉, 어떤값보다 큰 觀測値는 얼마나 되는가? 혹은 어떤값보다 작은 觀測値는 얼마나 되는가? 테이블에서 300 헥타르 以上되는 面積을 갖는 農場은 907個이고 100 헥타르 보다 작은 農場은 1,979個라는 것을 쉽게 알 수 있다.

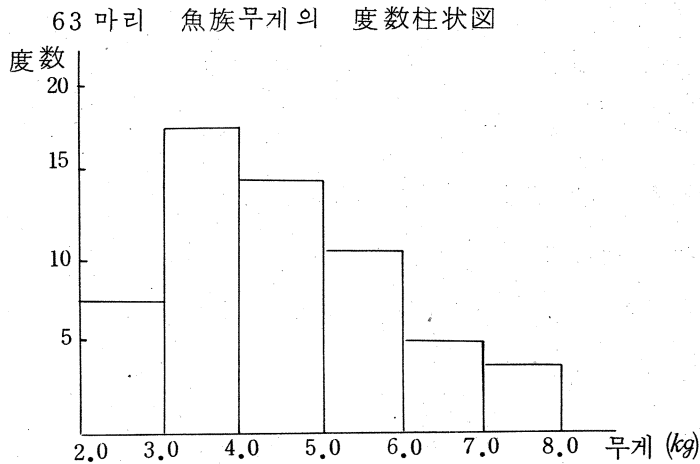


### 10.2.5 度数分布図表

度数分布는 特別한 特性의 觀測수가 값의 範圍에 따라 分布되어 있거나 퍼져있는 形式으로 情報를 준다. 이러한 分布表를 만드는외에 그림表로 나타내는 方法은 다른 形態의 資料를 쉽게 볼수 있기 때문에 역시 重要하다.

#### (a) 度数柱状図表 (Frequency Histogram)

度数分布를 나타내는 가장 一般的인 方法은 度数柱状圖이다. 魚族무게에 關한 앞 데이터의 柱状圖는 아래와 같다.



度数柱状圖를 그리기 위하여 다음 一般原則을 따라야 한다.

- (i) 수평축(OX)을 따라 크기를 그린다.
- (ii) 수직축(OY)을 따라 度数를 그린다.
- (iii) 모든 柱状圖의 一般的인 關係는 다음에 따라야 한다.

表題, 脚註와 양측의 자세한 說明을 적절한곳에 나타내야 한다.

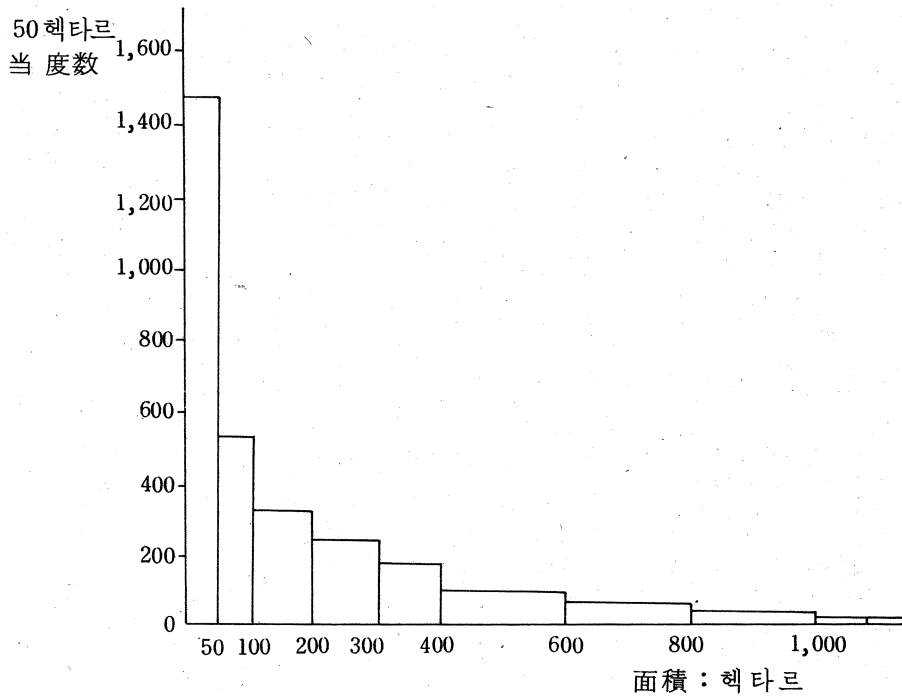
(iv) 데이터가 연속적이면 度数는 全体区間을 포함하는 四角形으로 나타내야한다. 離散型柱狀圖는 이節의 다음에 나타나는것 처럼 약간 다르다.

(v) 연속分布를 나타낼때 度数는 이에 비례하는 높이가 아니라 直四角形의 面積이다. 이것은 度数가 높이에 比例하는 以前의 圖表에서와 같이 級間이 同一한 경우에만 適用된다.

同一하지 않는 区間을 갖는 分布를 어떻게 處理해야 하는가 例를 들기위해 코코넛 農場의 例를보면 이경우에 農場面積에 대한 상이한 여러 階級이 있다. 各四角形의 面積이 階級区間에 비례가 되도록, 그러나 그 面積이 觀測度数에 비례가 되도록 나타내야 한다. 이러한 것을 해결하기 위한 最善의 方法은 이 테이블에서 가장작은 階級区間에 대하여(즉, 50 헥타르) 同一하게 各 階級度数로 計算하는 것이다. 그래서 구간 100 ~ 199 헥타르에 대한 656개 觀測數는 각 50헥타르 区間마다  $656 \times 50 / 100 = 328$ 개의 觀測數와 同一하다고 말할수 있다. 이런方法으로 모든수는 共通的인 바탕위에서 다루어지며 柱狀圖의 높이는 이에 비례할 것이다. 計算結果는 다음과 같다.

| 階   | 級     | 度 數   | 50 헥타르當 度数                  |
|-----|-------|-------|-----------------------------|
| 1   | ~ 49  | 1,467 | 1,467                       |
| 50  | ~ 99  | 512   | 512                         |
| 100 | ~ 199 | 656   | $656 \times 50 / 100 = 328$ |
| 200 | ~ 299 | 529   | $529 \times 50 / 100 = 264$ |

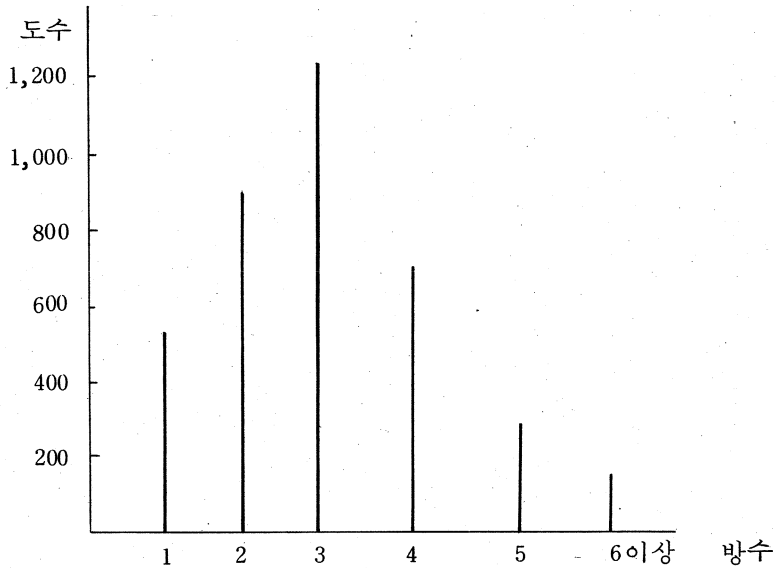
| 階 級       | 度 數 | 50 헥타르當 度數                |
|-----------|-----|---------------------------|
| 300 ~ 399 | 374 | $374 \times 50/100 = 187$ |
| 400 ~ 599 | 321 | $321 \times 50/200 = 80$  |
| 600 ~ 799 | 118 | $118 \times 50/200 = 30$  |
| 800 ~ 999 | 59  | $59 \times 50/200 = 15$   |
| 1,000 以上  | 35  | -                         |



이分布가 同一하지 않는 間隔을 갖는다는 사실과는 別個로 興味 있는 하나의 重要な 事項이 있다. 1,000 以上 階級の 경우인 端開階級을 어떻게 처리할것인가를 決定하는 問題이다. 즉 이 階級の 넓이를 알수 없기 때문에 全体度數에서 이部分을 나타내기위한 사각형의 높이를 計算할 수 없다. 요컨대 이것을 그

대로 놔두는것은 좋지않을 것이다. 그래서 어떻게 처리해야 할 것인가를 결정해야 한다. 根本적으로 이問題는 두가지方法中 하나로 해결할 수 있다. 첫째 分布의 上限을 가정하고 그에 맞추어 사각형을 그릴수 있다. 둘째方案은 약간의 높이를 갖는 사각형을 그리고 上記 柱狀圖처럼 끝을 열어둔다. 이런方法은 端開區間임을 나타내는 것이다. 두번째 方法은 端開區間內的 度数가 적을때만 사용하고 이러한 條件下에서는 아마 보다 바람직한 圖示法일 것이다. 離散데이터를 처리할때 觀測值의 크기에 비하여 測定單位가 적다면 이산데이터를 때때로 근사적으로 연속이라고 간주할 수 있다. 예를들면, 각마을의 住民數別 마을수의 分布를 알수 있다. 간단히 말하면 이分布는 사람을 小數로 나타낼수 없기 때문에 離散分布이다. 그러나 실제로 測定單位가 예를들어 데이터의 範圍가 1,000명 이라는 것과 비교하면 상당히 적은 1명이기 때문에 연속적인 데이터로 처리할 수 있을것이다. 測定單位가 값의 範圍와 비교해서 큰(예를들면 가구에 있어서 방수와같은) 離散分布를 할때 연속형 柱狀圖로 나타낼 수 없다. 그대신에 앞장에서 공부한 단순막대그림표 形態로 度数를 나타낸다. 때때로 아래 그림의 例처럼 막대 대신에 선을 사용한다.

도표 : 방수별 가구수, 솔로몬 군도 - Honiara, 1976

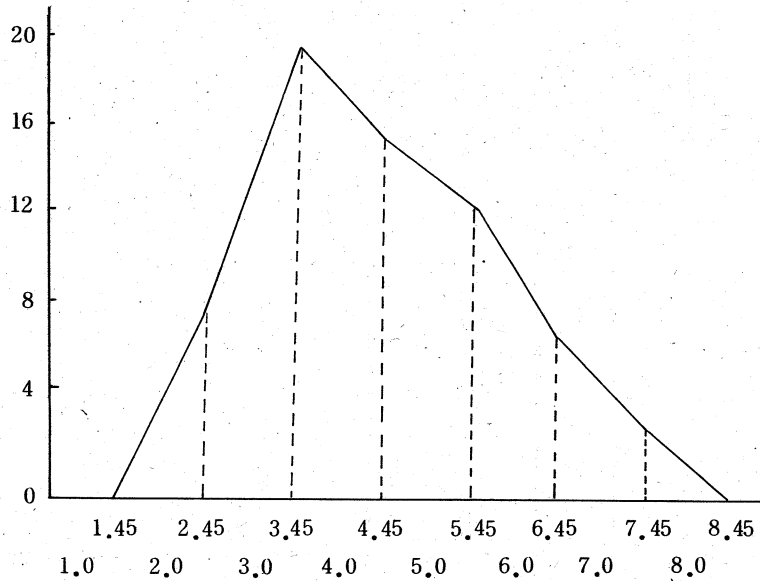


6 이상 階級이 갖는 問題를 處理한 方法에 주목하라.

(b) 度数多角形 (Frequency Polygon)

연속형 데이터나 거의 연속이라고 할 수 있는 離散型 데이터에 적합한 다이어그램 (Diagram) 代身の 形態는 度数多角形이다. 이 形態에서 各 階級度수는 階級值에 表示되고 연속적인 점을 連結하여 直線으로 만든다. 魚族데이터에 대한 度数多角形의 例는 아래와 같다.

魚族무게 데이터에 대한 度数多角形

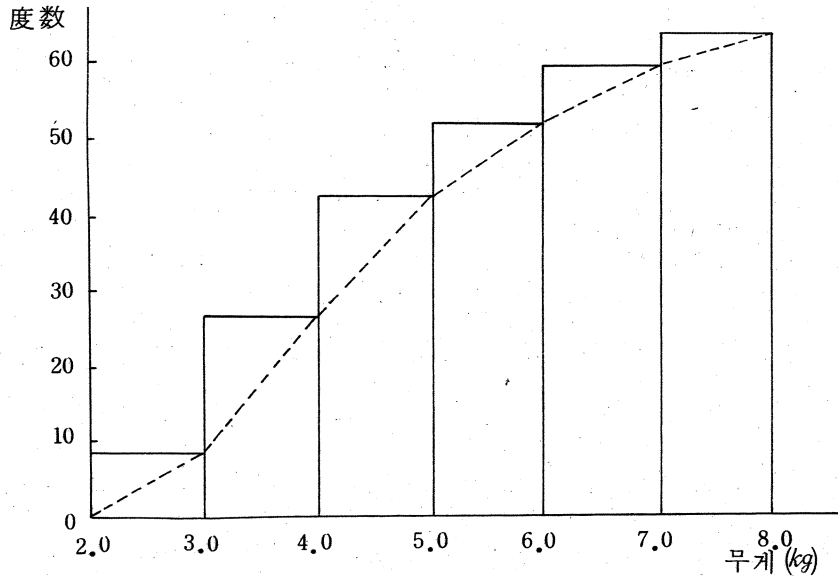


多角形의 시작과 끝은(階級の 上限과 下限의 中央点에서) OX 線上에 맞닿아야 한다. 多角形内の 面積은 앞柱狀圖의 四角形 面積과 같다. 柱狀圖에서 이面積은 全体度数를 나타낸다.

(c) Ogive (累積度数分布圖)

어떤 目的을 위해서는 꼭 累積度数分布를 만드는 것이 훨씬 더 좋다. 따라서 이를 그래프로 나타내는것이 有用하다는것을 알 수 있다. 만약 魚族데이터에 對한 累積度数分布를 만들고 柱狀圖를 그리면 다음 다이어그램(Diagram)이 보이는것 처럼 일련의 多角形으로 나타날 것이다.

魚族 무게에 관한 累積度数柱状圖



이 다이어그램 (Diagram)에서 점선은 累積度数多角形을 나타내고 이것을 Ogive (발음은 Oh-jive)라 부른다. 이것은 각 四角形의 오른쪽 꼭대기 모서리점을 이은 것이다. 이 圖表는 (以下) 累積度数分布를 나타내고 있다. 같은 方法으로 (以上) 累積度数分布에 대한 Ogive를 그릴수 있다.

앞 圖表에서 나타낸것처럼 OY축에 실제적인 度数를 나타내는 대신에 度数를 퍼센트로 변환시켜 比較度数로 나타낼수가 있다. Ogive의 形態는 實際적으로 같지만 OY축은 고기수 대신에 0부터 100까지의 퍼센트로 표시되어 진다. Ogive는 특히 퍼센트를 기반으로 해서 나타낼때 여러가지 目的에 有用하다. 地方事務所에서의 가장 實際적인 利用의 하나(이 課程에서 일찌기 言及한)는 統計수집에서 回答을 받는 過程을 나타내는 것이다. 이

Ogive에서 OY축에 퍼센트를 나타내고 조사시작 이후의 시간( 거의 週別로 )은 OX축에 그린다. Ogive는 調査期間에 접수된 回收率을 累進的으로 나타낸다. 이전 調査에 關한 資料를 미리 그래프 위에 나타낸다면 이전의 回收率보다 더 좋은지 나쁜지를 아는것은 어려운일이 아니다. 다음節에서 Ogive는 어떤 測定值 특히 중앙치(Median)를 매우 쉽게 계산할 수 있다는 것을 알게 될것이다.

#### 10.2.6 母集團의 分布

때때로 資料를 度数分布로 만들때 대상이 되는 全体 母集團을 使用한다. 그러나 우리가 가지고 있는 資料는 母集團의 한부분이고 표본으로 부터 母集團을 推定하고자 할 것이다. 예를들면 앞에서 다루었던 63마리의 어류의 度数分布에 관심을 가지고 있을뿐만 아니라, 유사한 魚族의 全体 母集團에 關해 어떤 情報를 얻기를 원한다. 이節에서 크기에 限定이 없는 母集團의 度数分布 概念을 어떻게 제공할 수 있는가를 매우 간단한 方法으로 다룰것이다. 母集團의 모든項을 모두 셀 수 없는 集團을 無限母集團이라 한다. 대부분의 경우에 無限이라고 생각할 수 있는 많은 사람으로 구성된 有限母集團의 경우에도 같은 概念을 適用할 수 있다.

앞에 나타난 魚族資料에 關한 度数多角形을 생각해 보자. 6개 階級대신에 두배의 魚族數를 갖는 12개 階級을 갖는다고 생각해보자. 이 데이터로 度数多角形을 그리면 앞 다이어그램과 大

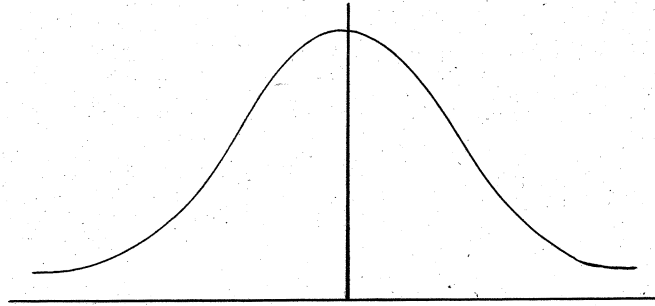


略 같은 모양이 될것이나, 더 많은 線으로 인하여 완만한 型이 될것이다. 이런과정을 다시한번 反復하면 많은 階級과 많은 觀測數를 갖게되고 이 度数多角形은 한층 더 완만한 型이 될것이다. 결국 점점 더 많은 階級을 취하고 더욱더 觀測數를 增加시키면 매우 완만한 曲線이 된다. 이 曲線은 魚族의 母集團分布를 나타낸다. 母集團分布 概念은 統計學에서 매우 重要하다. 이것은 高級統計理論을 다루는 기초를 이룬다. 이 理論을 상세히 다룰 시간이 없으며 약간의 概念만을 소개하고자 한다.

많은 상이한 출처에서 資料를 蒐集하고 度数分布表를 만들고 히스토그램과 度数多角形을 그릴때 많은경우에 히스토그램의 形態는 매우 規則的인것을 알게된다. 지금까지는 히스토그램의 가장 단순한 形態만 考察했다. 히스토그램은 統計人이 이런 단순한 形態를 설명하거나 나타내는 간단한 수학적 함수를 利用토록 하고있다. 연속형 데이터의 경우에 數學的 函數는 완만해지고 母集團의 分布를 나타낼 것이다. 만약 많은 상이한 母集團이 대개 같은 型으로 分布되어 있다면 母集團에 대한 重要한 質問의 答으로 數學的 函數式을 使用할 수 있다. 例를들면, 一般的으로 한 海역의 魚類가 다른 海역의 魚類보다 더 큰가? 한집단의 사람들이 타집단보다 더 많은 所得을 갖는가? 새로운 바나나 品種의 수확량 範圍는 얼마인가 등이다.

母集團分布에서 가장 重要한 形態中 하나는 正規分布이다.

“정규”(Normal)라는 單語는 실제로 자주 나타나는형 이라는 것을 의미한다. 전형적인 正規分布는 다음에 나타나 있다.



평균

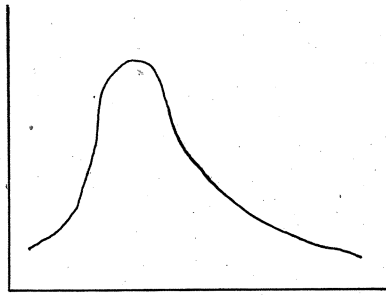
正規分布은 左右对称이고 양측은 같은형을 이룬다. 이러한 형을 갖는 分布를 例를 들면

- (a) 신장별 成人數
- (b) 한 作物의 農場別 生産量
- (c) 젖소의 우유 産出量
- (d) 年中 강우일수

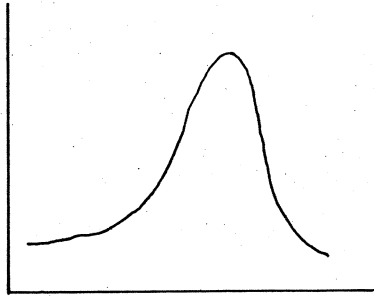
이러한 例에서 觀測值의 一部는 매우 높거나 매우 낮지만 大部分의 觀測值는 “平均” 값을 갖는다.

正規分布은 많은 相異한 方法으로 統計理論에서 매우 重要하다. 統計學을 보다 상세히 공부하려는 사람은 正規分布에 關係 보다 더 많이 공부해야할 것이다.

母集團分布에는 正規分布이외의 많은형이 있지만 여기서는 歪分布 (Skewed distribution)만 言及하자. 歪分布은 正規分布가 한쪽으로 치우친 형태이기 때문에 对称이 아니고, 그러보면 다음과 같다.



오른쪽으로 찌그러짐 (우경)



왼쪽으로 찌그러짐 (좌경)

右傾分布의 例는 母集團의 所得分布이고 많은나라에서 상당히 낮은水準의 所得이 集中되어 있다. 그리고 그래프의 긴 꼬리는 우측으로 뻗어있고 極小數의 사람들이 매우높은 所得을 갖는것을 나타내고 있다.

### 10.3 平均과 여러 位置測度

#### 10.3.1 平均의 形態

“平均”이라는 單語는 영어에서 일상적으로 자주 사용된다. 우리는 가끔 보통사람, 平均 이상의 달성, 平均以下の 기온 등의 말을 한다. “平均”(Average)은 “전형적인”, “일상적인”, “정상적인”의 意味로 사용된다. 統計學에서도 역시 “平均”(Average)이란 單語를 使用하지만 그 意味는 같지 않다. 大部分의 統計職員은 數集團의 平均을 計算하는 法을 알고 있다. 그數를 모두 합하고 몇개의 數가 存在하는가에 따라서 그합을 나눈다. 그러나, 統計學에서는 여러가지 形態의 平均을 使

用한다. 위에서 言及한것은 算術平均으로 알려져 있으며 여기에서는 이러한 平均들에 關係 考察해 보고 어떻게 計算이 이루어지고 어떤 狀況下에서 利用되는지를 알아보자.

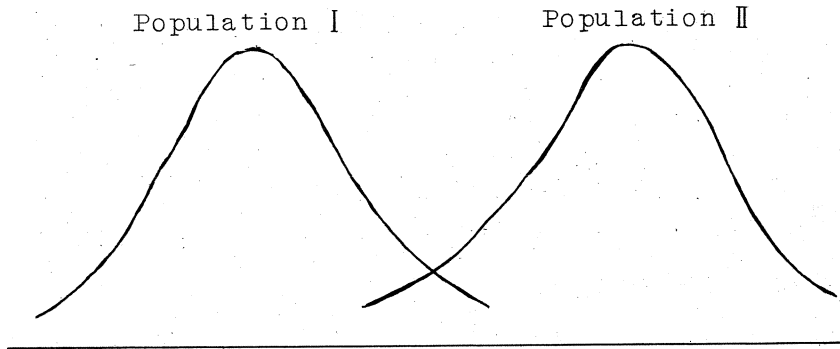
統計學에서는 상이한 母集團을 서로 比較하는 경우가 많다. 事實上 統計理論의 大部分은 이問題에 關係가 있는 것이다. 예를들면 地域別 家口所得, 國民學生의 階層別 虫菌 發生率, 年中 어획량 등을 比較하고자 할것이다.

만약 매우 相異한 分布形態를 갖는 두母集團을 比較하고자 한다면 이런 比較는 매우 어렵다. 그러나 大部分의 경우에 두分布의 型은 매우 유사하다는 것을 알수있고 이런 경우의 比較는 아주 쉬워진다. 우리가 必要로 하는것은 母集團을 說明하는 한 두가지 方法을 찾거나 어떤 特性值別로 分布를 要約하는 것이다.

실제로 母集團의 두개의 가장 重要한 特性值는 分布의 位置測度와 散布度이다. 이節에서 位置測度を 考察하고 散布度는 다음節에서 다루겠다.

分布의 位置測度에 의해 어떻게든지 分布의 모든 相異한값의 크기를 要約한 수치를 알고자 할것이다. 어떤學者들은 중심경향의 측도(Measures of central tendency)나 중앙위치 측도(Measures of central location)라고 한다. 이러한 것은 같은 概念을 다른 方法으로 表示한 것이다. 우리는 位置測度(Measure of location)라는 用語를 使用하도록 하자. 예를들어 만약 分布의 다이아그램(Diagram)을 그린다면 位置測度는

母集團값이 OX 축위의 어디에 集中되어 있는가를 나타낸다. 아래 圖表에서 一般的으로 母集團 II의 값이 母集團 I의 값보다 크다는 것을 알수있다.



位置測度로써 다음의 平均形態를 알아 보자.

- 산술평균 (The arithmetic mean)
- 中位数 (The median)
- 最頻數 (The mode)
- 幾何平均 (The geometric mean)

四分位數 (Quartiles) 라는 또다른 位置測度를 言及할 것이다. 끝으로 이節에서 標本과 母集團간의 差異를 다루고 位置測度에 關係 言及할때 이런 差異를 處理하는 方法을 공부할 必要가 있다.

一般的으로, 統計人들은 標本과 母集團을 區別하는 特別한 表示法을 使用한다. 이 表示法에 의해 특수문자는 항상 같은것을 나타낸다. 이런 문자의 使用은 간단히 전달하는 間接表記法의 일종이다.

만약, 母集團에서 標本을 抽出할때 標本觀測值의 數를  $n$  으로 나타내면 얻어진 特定變數의 값은  $X_1, X_2, X_3, \dots, X_n$  (…의 뜻은 등등)으로 나타낸다. 따라서  $X_3$ 는 標本에서 3번째 값을 意味한다. 標本을 抽出한 母集團은 두 形態중 하나일 것이다. 첫째로 母集團의 모든 원소를 셀수있는 한정된 크기를 가질것이다. 이경우 有限母集團이라 하고 母集團의 크기는  $N$ 으로 나타낸다. 이러한 例는 国内居住國民數, 都市内 經營企業數, 特定地域内の 모든 農場數 등이다. 두번째 形態의 母集團은 그 크기에 制限이 없고 個個의 數를 셀수 없다. 이러한 母集團을 無限母集團이라 하며 例를들면 어떤나라의 總 쌀 生産量, 어떤海域의 總 魚獲量 등이다. 만약 母集團이 無限이면 母集團크기를 나타내는 부호가 없다.

一般的으로 標本값보다 位置測度와 散布度의 母集團 값을 더 重要視 한다. 母集團값은 그리스 문자로 表示하며 母集團의 算術平均은  $\mu$  (그리스문자 “mu”)로 나타낸다.

이런 母集團값을 母集團의 “母數”(Parameters)라 하며 標本으로 부터 산출된 값과는 區別된다. 母集團에 관한 情報는 거의 없지만 標本에 의해 一連의 값을 갖는다. 여기서 標本 統計量을 計算하여서 모수를 推定하는 것이다. 平均의 여러 形態를 言及할때 이점에 대해 複習할 것이다.

### 10.3.2 算術平均(The arithmetic mean)

特定變數가  $n$ 個의 標本值를 갖고 이것은  $X_1, X_2, \dots$

$X_n$ 으로 나타내면 이 標本의 算術平均은  $\bar{X}$  (발음 X bar)로 나타내고

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

로 정의한다. 그리스문자  $\Sigma$  (Capital sigma)는 合計를 나타내고  $\sum_{i=1}^n X_i$ 는  $X_1 + X_2 + \dots + X_n$ 을 나타낸다. 다시 말

하면 이公式은 算術平均에 대한 簡便表記法이다. “ $n$ 個數의 平均은 모든값의 和를  $n$ 으로 나눈것이다.” 算術平均의 上記定義 때문에 모든값은 같은 加重值  $1/n$ 을 가지며 이平均은 單純算術平均이라 한다. 度數分布의 경우에 算術平均은 加重平均으로서 얻을수 있으며 다음에 언급할 것이다.

첨자  $i$ 의 使用은 고려한 觀測值를 나타내기 위한 약속이다.

그래서  $\sum_{i=1}^n$ 은 첫번째부터  $n$ 번째까지 모든觀測值의 合計를 意味하며 마찬가지로  $\sum_{i=2}^4$ 는 2번째, 3번째, 4번째 觀測值의 合計를 意味한다.

이章의 다음부터는  $\sum_{i=1}^n$ 은 항상  $n$ 번째, 觀測值까지의 合計를 나타낸다. 때때로 간소화하기 위해 이것을  $\Sigma$  부호만으로 나타내며 이렇게 사용할때는  $\sum_{i=1}^n$ 을 간단히 표시한

것이라고 이해해야 한다.

一般的으로 일련의수에 대해 算術平均을 計算하고 그 結果를 이용할때 실제로 나타날수 있는 어떤수를 얻는다는 보장은 없다. 예를들면 가구당 가구원수의 算術平均을 計算하면 4.5명이 될 것이다.

사람에 있어서 0.5명이라는 것은 存在하지 않기 때문에 이런

종류의 결과가 混同을 야기시키며 중요한것은 算術平均의 인위적인 概念이라는 점이다. 算術平均은 計算이 간편하며, 다른 특성치를 알고자 할때 수학적으로 편리하기 때문에 이용된다. 위의 例示에서조차도 가구의 規模는 4명 내지 5명이라고 말할 수 있는 편리한 方法이다.

定義에 따라 연속데이터는 가능한 範圍內에 어떤값을 실제로 가질수 있기 때문에 이러한 問題는 離散데이터의 경우에만 발생할 것이다.

그래서 어떤 그룹의 “남자의 신장의 算術平均은 1.75 m이다.”라는 말이 의미하는것이 무엇인지 이해하는것은 어렵지 않다. 離散資料의 算術平均에 관해 언급할때 주의를 해야한다. 算術平均은 離散資料로 母集團의 位置測度인 平均의 한形態인것을 알아야 한다. 그러나 실제로 거의 비슷한 數値가 나타나거나 혹은 모두 똑같은 값이 나타날지라도 그것은 平均이 아니다.

#### 度數分布에서의 算術平均의 計算

무엇보다도 먼저 離散度數分布의 단순한 경우를 살펴보자. 예를들면 住宅調查結果로서 나타난 가구당 방수를 나타내는 資料가 다음과 같이 주어졌다.

| 방 수   | 가 구 수 |
|-------|-------|
| 1     | 476   |
| 2     | 813   |
| 3     | 1,189 |
| 4     | 507   |
| 5     | 191   |
| 6 이 상 | 25    |
| 총 계   | 3,201 |



가구당 방수의 平均을 計算하기 위해 방의數를 合計해서 階級數로 나눈것은 적합하지 않다. 5개의 방을 갖는 家口보다 3개의 방을 갖는 가구가 더 많다는 狀況을 計算에 考慮해야 하며 마지막 階級이 6개이상 이라는 것을 처리해야 한다. 마지막 階級の 度數는 적기때문에 이를 處理하는 方法으로 인하여 매우 큰오차는 나타나지 않을것이다. 이例에서 마지막階級の 모든 가구가 平均 7개의 방을 갖는다고 가정할 때 그러면 이 計算은

$$\begin{aligned} \text{算術平均} &= \frac{(1 \times 476) + (2 \times 813) + (3 \times 1,189) + (4 \times 507) + (5 \times 191)}{3,201} \\ &\quad + \frac{(7 \times 25)}{3,201} = \frac{8,827}{3,201} = 2.76 \end{aligned}$$

바로 數集合의 算術平均에 대한 수식으로써 度數分布에도 유사한 公式을 유도할 수 있다. 이경우에 階級の 數는  $k$ 로 나타내고 各階級の 값은  $X$ 로 나타내며 各階級の 度數는  $f$ 로 나타낸다. 平均  $\bar{X}$ 에 대한 公式은

$$\bar{X} = \frac{\sum_{i=1}^k f_i X_i}{\sum_{i=1}^k f_i}$$

이것을 말로 표현하면 度數分布의 平均은 各階級の 變數값에 그 階級の 度數를 곱하여 合計한것을 全體度數로 나눈것으로 나타난다.

연속分布를 다룰때 앞절에서의 魚族資料를 이용한 다음예처럼  $X_i$ 로써 階級值를 사용한다.

| 階 級 (kg)  | 階級値 (Xi) | 度 数 (fi) | 度数×階級値 (fi Xi) |
|-----------|----------|----------|----------------|
| 2.0 ~ 2.9 | 2.45     | 7        | 17.15          |
| 3.0 ~ 3.9 | 3.45     | 19       | 65.55          |
| 4.0 ~ 4.9 | 4.45     | 16       | 71.20          |
| 5.0 ~ 5.9 | 5.45     | 12       | 65.40          |
| 6.0 ~ 6.9 | 6.45     | 6        | 38.70          |
| 7.0 ~ 7.9 | 7.45     | 3        | 22.35          |
| 計         |          | 63       | 280.35         |

$$\bar{X} = \Sigma fi Xi / \Sigma fi = 280.35 / 63 = 4.45$$

이 경우에 算術平均은 階級値나 階級區間 中間點의 加重平均으로 써 얻어지며 加重値는 度数  $fi$  혹은 相對度数  $fi / \Sigma fi$  이고 階級區間의 모든 單位는 그 값의 中央點으로 일치시킨다고 가정하기 때문이다.

平均으로 算術平均을 使用하는 主된 이유중 하나는 數學적으로 간편하다는 것이고 算術平均은 平均의 다른 形態보다 母集團을 比較할 때 母集團 平均間의 實際的인 差異가 있는지 없는지를 決定하기가 매우 쉽기 때문이다. 一般적으로 母集團의 算術平均은  $\mu$ 로 나타내고 標本值의 算術平均은  $\bar{X}$ 로 나타내며 標本平均은 母平均  $\mu$ 의 推定值이다.

### 10.3.3 중위수 (The median)

중위수 (Median)은 數集合에 관한 平均의 한 形態이

며 位置測度이며 매우 간단한 概念이다. 中위수는 數集을 同一한 兩部分으로 나누는 바로 그런값이다. 즉 數들을 크기順으로 배열했을때 그 中間값이다. 中間값이 없을때 즉 集의 數가 홀수일때 數集의 中위수를 決定하고자 하면 문제가 있으며 이때는 두 中間값을 算術平均함으로써 中위수를 얻는다. 中위수는 中央數의 값이나 系列數에 달려있고 極端的으로 높거나 낮은값에 영향을 받지 않는다. 이것은 中위수의 매우 重要한 性質이며 아주 비정상적인 分布를 갖는 數集의 中위수와 平均(Mean)을 계산해서 예를 들어보자. 어떤地域에서 그중하나는 큰 企業農場인 반면 나머지 農場은 農民이 所有하고 있는 조그만 農場인 15개 코코넛 農場의 각 面積을 測定했다고 하자. 헥타르로 나타낸 面積은 다음과 같다.

|     |     |     |     |       |
|-----|-----|-----|-----|-------|
| 1.3 | 1.3 | 1.5 | 1.7 | 2.0   |
| 2.1 | 2.3 | 2.7 | 2.7 | 2.8   |
| 3.0 | 3.7 | 5.5 | 8.5 | 120.3 |

中위수(15개 觀測值의 8번째인)는 즉시 2.7헥타르라고 알 수있을 것이다.

算術平均은  $1,614/15 = 10.8$ 헥타르로 쉽게 計算할 수 있다. 위와같이 두平均사이에 매우 큰 差異가 있고 企業農場이 算術平均에 매우 크게 影響을 미치지만 中위수에서는 그렇지 않기때문에 이러한 차이의 중요한 요인이 된다. 이러한 극단적인 농장을 제외한다면 中위수는 14개 觀測值의 7번째와 8번째값의 平均이 될 것이다. 즉  $\frac{2.3 + 2.7}{2} = 2.5$  반면에 算術平均은  $41.1/14 =$

2.9 톤 이하이다.

중위수는 매우 크거나 매우 작은 觀測值 값에 影響을 받지 않지만 算術平均은 그렇지 않다. 測定值의 양 극단에서 觀測值의 精度에 의심이 있다면 중위수(Median)는 算術平均보다 더 나은 平均(Average)이 된다.

度数分布로 부터 중위수의 計算은 두가지 方法을 使用할 수 있는데 첫째 테이블(表)에서 직접 計算하는 方法이고 둘째 Ogive의 正確한 函表를 使用하는 것이다.

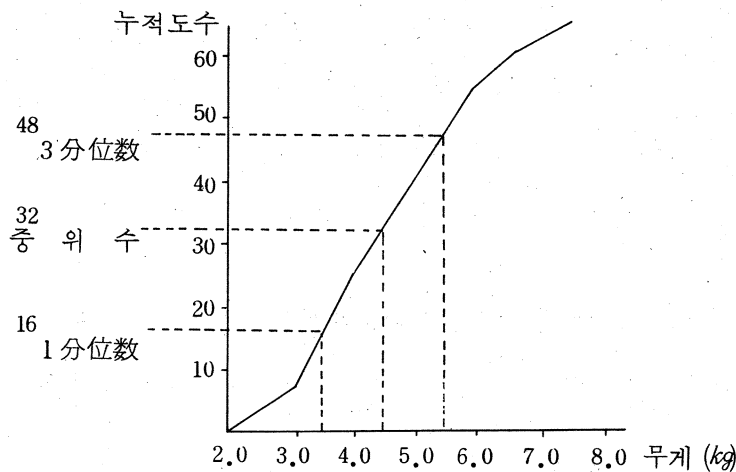
첫번째 方法을 例로들기 위하여 魚族資料를 利用하여 다음과 같이 쓸 수 있다.

| 階 級 ( kg ) | 度 數 | 累積度數 |
|------------|-----|------|
| 2.0 ~ 2.9  | 7   | 7    |
| 3.0 ~ 3.9  | 19  | 26   |
| 4.0 ~ 4.9  | 16  | 42   |
| 5.0 ~ 5.9  | 12  | 54   |
| 6.0 ~ 6.9  | 6   | 60   |
| 7.0 ~ 7.9  | 3   | 63   |
| 計          | 63  |      |

63個 觀測值가 있으므로 메디안觀測值는 中央의 한값 즉 32 번째 값 ( $(n+1)/2$ )이다. 累積度數 行으로부터 32 번째 觀測值는 4.0 ~ 4.9 kg 階級內에 있다는것을 알수있다.

4.0으로 시작되는 구간 이전에 26개 觀測值가 있고 이區間

内に 16개 觀測値가 있다. 그래서 中位數 加重値는  $4.0 + (32-16)/16 \times 1.0$  으로 계산한다. 따라서 中位數는 4.4 kg이다. 離散分市의 計算은 훨씬 간단하다. 累積度數分布는 計算되었고 메디안 觀測數는 ( $n+1/2$  번째) 결정되었다. 中位數 값은 中位數 觀測數가 있는 階級값이기 때문에 간단하다. 예를들면 가구당 방수의 앞例에서 全体 3,201 觀測數의 1,601 번째 階級은 3이기 때문에 中位數는 3이다. 中位數는 決定하는 다른 方法은 度數分布의 Ogive로부터 직접 決定된다. 단지 이경우에 Median의 精度는 그래프線의 精度에 의해 決定된다. 中位數 觀測數에 對應하는 점에서 OY축으로 부터 수평선을 긋고 이線이 Ogive를 자르는 점으로부터 수직선을 그린다. 이 수직선이 OX축과 만나는점이 Median값이다. 다음 圖表는 앞에서 그려진 魚族資料의 Ogive이다. 이 Ogive가 正確하게 나타났다면 Median값은 이미 計算한 4.4 kg (다시 말해서 32 번째 값) 인것을 알 수 있다. 이 Ogive는 上, 下 四分位數가 얼마인가를 그래프에서 直接 알 수 있다.



#### 四分位数(Quartiles)

중위수는 總度数를 同一한 두部分으로 나누는 觀測值의 값이다. 같은方法으로 度数를 여러部分으로 나누는 어떤값을 決定할 수 있는데 그중 가장 重要的 것이 四分位数이다. 四分位数은 그 名稱에서 암시하듯 全体度数를 네개의 同一한 部分으로 나누는것이다. 첫번째 第1四分位数 혹은 下四分位数은 이값보다 작은 觀測度数가  $1/4$  이고 더큰 觀測度수는  $3/4$  를 갖는다. 第2四分位数은 중위수와 같다. 第3四分位数은 이값보다 작은 觀測度数가  $3/4$  이고  $1/4$  은 더크다. 觀測值로 나타내면 下四分位数은  $n+1/4$  번째값이고 中四分位数(메디안과 일치)는  $n+1/2$  번째값 第3四分位数은  $3(n+1)/4$  번째값이다.

四分位数은 平均은 아니지만 觀測值의 位置測度라 할 수 있으며 다음 論題에서 分散度나 觀測值의 變異性을 決定하기 위해 四分位数을 使用할 것이다.

全体度数를 4 등분하는 것과는 별개로, 여러부분으로도 나눌 수 있는데 이러한 것의 名稱은 分割한 數와 관련되어있다. 아마 가장 一般的으로 使用되는 것은 10分位数(Deciles) (全体度数를 10개 그룹으로 나누는)와 五分位数(5개 그룹으로 나누는)이다. 그러나, 이러한 分割은 널리 使用되지 않으므로 여기서 다루지 않겠다.

#### 10.3.4 最頻數(The mode)

만일 어떤 母集團分布가 一定時点에서 分布函数의 最高点을 가지면 이점에서 最頻數(Mode)라 한다. 算術平均이나

中位数처럼 最頻數는 平均의 한 形態이며 또한 分布의 位置測度이다.

標本觀測值를 다룰때 最頻數의 概念은 度數分布와 관련되어 가장 有用하게 쓰이며 離散分布에서 最頻數는 가장 많이 나타난 값이다. 例를들면, 가구당 방수의 實例에서, 最頻數는 가장 많이 나타난 값이기 때문에 가구당 방수는 3개이다. 이것은 非統計人에게는 무의미한 2.76개의 방이라는 算術平均보다는 情報를 매우 有用하게 要約한 것이다.

연속分布에서 最頻數를 決定하는것은 매우 複雜하다. 그래서 단지 最頻數의 Group (그룹)이나 階級만을 다를것이다. 最頻數는 同一한 階級間隔을 갖는 分布에서 가장 높은 度數를 갖는 階級이다.

그러나, 특히 그分布가 母集團으로 부터의 標本인 경우 연속度數分布에 대한 最頻數 그룹을 決定할때에 자주 問題가 發生한다. 最頻數그룹은 그階級들이 정의되는 方法에 따라, 그리고 階級間的 均等한 分布를 갖는 데이터에 크게 달려있으며 階級間的 定義가 變하면 最頻數그룹이 變할수도 있다. 이러한 理由때문에 最頻數는 限定된 값이고 주의깊게 使用해야 한다.

### 10.3.5 기하평균(The geometric mean)

$n$ 개 觀測值  $X_1, X_2, X_3, \dots, X_n$ 의 기하平均은 다음과 같이 定義된다.

$$\text{기하평균} = (X_1 \cdot X_2 \cdot \dots \cdot X_n)^{1/n}$$

간단히 表記하면

$$\text{기하평균} = \left( \prod_{i=1}^n x_i \right)^{1/n}$$

부호  $\Pi$  (그리스문자 pi) 는  $\Sigma$  가 합을 뜻하는 것처럼 수의 곱을 의미한다. 例로서 3, 7, 9의 기하평균은  $(3 \times 7 \times 9)^{1/3} = (189)^{1/3} = 5.74$

기하평균은 計算하기가 매우 어렵고 특히 많은數 일때는 더욱 어렵다. 어떤 計算機는 이計算에 도움을 주는 特殊한 機能을 가지고 있지만 그렇지 못하면 가장 쉬운 方法은 对數表를 使用하는 것이다.

만일 어떤 값이 음수이면 幾何平均은 計算할 수 없고, 어떤 값이 0이면 幾何平均은 0이 되고, 모든數의 平均으로서의 幾何平均의 概念은 意味가 없게 된다. 이러한 理由때문에 幾何平均의 使用은 相當히 制限을 받고 있으나 指數와 成長率을 다룰때에는 매우 有用하다.

例를들어 어떤나라의 小売物價指數가 다음과 같이 해마다 增加한다고 하자. 즉 1973年에서 1974年까지 16.7%, 1974 ~ 1975는 10.8%, 1975 ~ 1976은 2.1%, 1976 ~ 1977은 4.5%이다. 1973년부터 1977년까지 平均年間增加率은 이數의 幾何平均으로 얻어지며 算術平均에서 얻어지는 것은 아니다. 첫해의 16.7%의 增加는 100에서 116.7의 增加와 같고 이것은 比例적으로 1에서 1.167의 增加와 같다. 환언하면 첫해다음해의 水準을 求하기 위해 指數에 1.167을 곱한다. 마찬가지로



10.8%의 증가에 대하여 1.108을 곱한다.

이와같은 방식으로 각퍼센트를 變換하면 幾何平均은 다음과 같이 計算할 수 있다.

$$(1.167 \times 1.108 \times 1.021 \times 1.045)^{1/4} = (1.380)^{1/4} = 1.0838$$

고로 年間平均成長率은 8.38%이다.

이것을 8.4%로 반올림하기 때문에 원래의 데이터보다 더 意味있는 수치를 나타내는 것은 아니다. 算術平均이 8.5%라는 것을 注目하라. 이것은 幾何平均이 算術平均보다 작거나 같다는 것을 나타낸다.

#### 10.3.6 平均의 여러形態와 그 使用에 關한 要約

이 論題를 完成하기 위하여 이제까지 論議하고 要約한 平均의 여러 形態에 關한 長點과 短點 그리고 平均이 使用되는 狀況을 살펴보아야 할 것이다.

##### (a) 算術平均

算術平均의 長點은

- (i) 모든사람이 利用할 수 있다.
- (ii) 모든경우에 쉽게 計算되고 測定할 수 있다.
- (iii) 系列의 모든값을 고려한다.
- (iv) 對數的으로 處理하기가 용이한것과 같이 數學的으로 重要하다.

短點은

- (i) 極端的으로 높은값과 낮은값에 影響을 받는다.

(ii) 실제로 나타나지 않는 平均값이 算出될 수 있다.

算術平均은 거의 모든 狀況에서 使用할 수 있는데 특히 相異한 母集團을 비교할 때, 標本값에서 母數를 推定할 때 有用하다.

(b) 中位数(Median)

中位数의 長點은

(i) 계산이 쉽다.

(ii) 極端的인 값에 影響을 받지 않는다.

(iii) 모든 狀況에서는 測定할 수 있다.

短點은

(i) 標本の 모든 값을 使用하지 않는다.

(ii) 對數的으로 다루기가 어렵다.

中位数는 또한 많은 경우에 使用할 수 있는데 歪分布(Skewed distribution)에서 觀測值를 다룰 때 가장 많이 使用된다. 왜냐하면 이 경우에는 算術平均보다 더 有用한 것이기 때문이다.

(c) 最頻數(Mode)

最頻數의 長點은

(i) 어떤 資料에 대해서도 計算이 용이하다.

(ii) 많은 狀況에서 명백한 의미가 있다.

(iii) 極端的인 값에 影響을 받지 않는다.

短點은

(i) 연속度數分布에서는 計算이 어렵다.

(ii) 때때로 存在하지 않을 수도 있다.

(iii) 度数分布에서 定한 階級間隔의 크기에 따라 달라진다.

(iv) 数学的으로 處理하기가 적합치 않다.

(v) 系列의 모든資料를 使用치 않는다.

最頻數는 度数分布의 경우 가장 一般的인 階級에서의 情報나 가장 큰 度数가 나타나는 階級이 必要할때 有用하며 연속적인 資料를 다룰때보다 離散資料를 다룰때 훨씬 더 有用한 概念이다.

(d) 幾何平均

幾何平均의 長點은

(i) 系列의 모든資料를 利用한다.

(ii) 算術平均보다는 極端的으로 높은값에 影響을 덜 받는다.

(iii) 对数的으로 處理할 수 있다.

短點은

(i) 計算이 어렵다.

(ii) 음수일때는 存在하지 않으며 0일때는 쓸모가 없다.

(iii) 매우낮은값( 거의 0에 가까운)에 影響을 많이 받는다.

(iv) 一般的으로 이해가 容易하지 않다.

幾何平均은 成長率이나 指數, 퍼센트變化등의 平均을 計算하는데 使用된다.

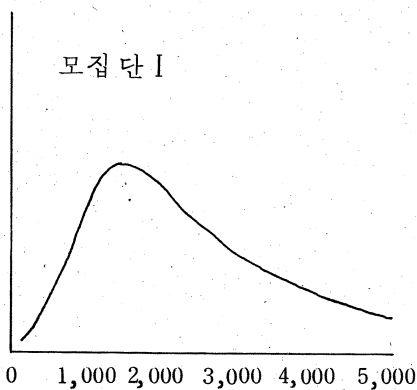
## 10.4 散布度

### 10.4.1 基本原理

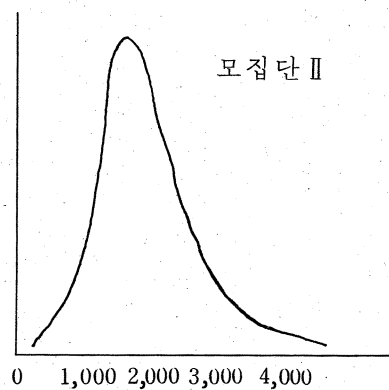
앞 두節에서 統計的 데이터를 要約하는 方法과 아주 쉽게 理解할 수 있는 올바른 方法으로 나타내는 法을 論議하였다. 度數分布는 圖表形式으로나 그래프 形式으로 情報를 要約하는 方法이며 반면에 位置測度(算術平均과 같은)는 이 情報를 하나의 數字로 要約한 것이다.

그러나, 要約할때 強調할 必要가 있는 必須的인 特徵을 大量의 데이터에서 찾아내지만 이렇게 하다보면 항상 얼마간의 情報를 상실한다는것을 인식해야한다. 우리는 要約할때 狀況을 적절히 이해하는데 必要한 情報에서 너무 벗어나거나 간단히 要約해서는 안된다는 점에 留意해야한다. 이節에서 位置測度가 母集團을 적절하게 說明하는데는 상당히 不充分하다는 것을 알게될것이다. 특히 相異한 母集團의 特性值를 비교하고자 할때 位置測度이외에 다른 어떤 測度가 必要하다.

이것을 例를들어 살펴보자. 다음 두그래프는 相異한 두地域의 住民年間所得分布를 나타내고 있다.



소득 (\$)



소득 (\$)

母集團 I 과 母集團 II 는 둘다 平均所得이 年間 \$ 1,800 라는 것을 알 수 있으나 이 두 分布가 같다고 할수없다. 母集團 I 은 母集團 II 보다 널리 흩어져 있다. 母集團 I 이 훨씬 더 큰 變異性和 分散을 갖고있다. 만약 “各母集團의 1人當年間平均所得이 \$ 1,800 로 같다” 라고 要約하면 아주 부적절한 說明이 될것이다. 대신에 母集團 I 의 1人當年間平均所得은 \$ 1,800 이고 個個의 所得은 平均의 오른쪽에 매우 넓게 分布되어 있다. 母集團 II 의 平均은 \$ 1,800 이지만 平均에 매우 가깝게 密集되어 있다. 觀察한 것을 훨씬 더 좋게 기술할 수 있다. 그러나 이것은 매우 애매모호하고 “매우 넓게 퍼져있다” 는 말은 正確하지 않으면 必要한 것은 이러한 分散을 數量化하는 方法이며 이것은 散布度를 意味한다.

이 論題에서 4 가지 相異한 測度를 考察할 것이지만 根本적으로 두가지로 大別할 수 있다.

(a) 母集團의 代表值간의 거리測度

(b) 어떤 特定 中央值로부터 母集團의 모든數에 대한 偏差測度

散布度の 첫번째 形態의 例로써 範圍와 4分位範圍를 살펴보고 두번째 形態의 例로서 平均偏差와 標準偏差(이것의 제곱은 分散이다)를 살펴보자.

#### 10.4.2 範圍

數集合의 範圍는 數集合에서 가장 작은값과 가장큰값간의 差異나 거리로 나타난다. 그러므로 計算이 容易하고 어떤

數集合에서도 찾을수 있다.

度數分布로부터 範圍는 보통 計算하지 않지만 마지막 階級上限과 첫번째階級下限사이의 差異로서 近似값을 얻을수 있다. 이階級の 어느한쪽이나 양쪽이 端開이면 範圍는 찾을수 없다.

범위는 求하기는 쉽지만 가장높은값과 가장낮은값 사이의 모든 資料를 무시한다는 뚜렷한 短点이 있다. 다음 3개의 數集合을 살펴보자.

(i) 5, 17, 17, 17, 17, 17, 17, 17, 17, 17, 17

(ii) 5, 5, 5, 5, 5, 17, 17, 17, 17, 17

(iii) 5, 6, 7, 8, 10, 11, 14, 14, 15, 16, 17

세개 모든 같은범위 ( $17-5 = 12$ )를 갖지만 그 數의 分散은 결코 같지 않다.

範圍는 數集合에 포함된 가장높은값과 가장낮은값에 影響을 받는다. 고로 범위는 信賴性없는 散布度이다.

#### 10.4.3 四分位 範圍 (Inter-quartile range)

매우 單純한 計算方法이지만 範圍의 어떤 難点을 克服하고자 할때 四分位範圍가 散布度로써 利用되고 있다. 앞節에서 살펴보았듯이 四分位數는 觀測值를 4개의 똑같은 그룹으로 나눈값이다. 第2四分位數는 中위수와 같고 四分位範圍는 제3四分位數와 제1四分位數의 差異이다.

그러므로 이 차이는 觀測值의 半을 포함하지만 제3四分位數와 제1四分位數를 제외함으로써 매우 높은값이나 매우 낮은값에 영

향을 받지 않는다. 이와같은 差異는 範圍보다 더 安定的이다.

이節에서 주어진 3개의 數集을 살펴보면 각각의 四分位範圍를 計算할 수 있다. 그集에  $n$ 개의 觀測值가 있다면 제1 四分位數  $Q_L$ 은

$$Q_L = n+1/4 \text{ 번째 觀測值나 값}$$

제3 四分位數  $Q_u$ 는

$$Q_u = 3(n+1)/4 \text{ 번째 觀測值나 값}$$

세개 數集에서  $n=11$ 이므로 제1 四分位數는 3번째 觀測值이고 제3 四分位數는 9번째 觀測值이다.

각 그룹의 4分位數範圍는

$$i) 17 - 17 = 0$$

$$ii) 17 - 5 = 12$$

$$iii) 15 - 7 = 8$$

그래서 각 그룹의 散布度는 다르다는것을 알 수 있다.

四分位範圍는 度數分布에서 아주쉽게 計算할 수 있다. 四分位數는 Ogive에서 직접 구하거나 또는 中位數를 구하기 위하여 사용한 方法과 비슷하게 계산된다.

이測度는 四分位數의 어느한쪽이 端開階級(Open-ended Classes)內에 포함되지 않는다면 端開階級을 갖는 度數分布에 대해서도 計算할 수 있다는 장점을 가지고 있다. 예를들면 이章의 처음에 나타난 코코넛 農場의 例에서 四分位範圍를 計算할 수 있다. 이경우에 다른 세개의 散布度를 正確히 計算할 수 없다는 것은 注目해야할 일이다.

왜냐하면 마지막 端開階級에서의 값의 分布에 의하여 影響을 받기 때문이다. 반면에 四分位範圍는 그렇지 않다. 사실상 마지막 階級에서 近似 階級值나 中央值를 취함으로써 平均偏差나 標準偏差를 推定할 수 있다. 그러나 例外的으로 큰 農場이 있는 경우에 그 推定值는 妥當치 않다.

四分位範圍는 母集團이나 標本에서 모든 數值를 고려하지 않는 短點을 가지고 있고 標本에서 母集團값을 推定하려 할때 數學的으로 處理하기 힘들다. 그러나 利用이 簡便하고 迅速하게 구할 수 있는 것은 散布度이다.

#### 10.4.4 平均偏差(The mean deviation)

여기서 간단히 공부할 平均偏差나 標準偏差는 標本에서 母集團의 모든數值나 觀測值를 고려하는 散布度이다. 이 두測度는 어떤中央값에서 各 集合의 數의 偏差에 근거를 두고 있다. 사실상 理論上으로는 偏差를 測定하기 위하여 어떤 中央值를 選擇할 수 있으나 實제는 算術平均을 利用한다. 어떤 數集合에도 算術平均은 있기 마련이며 數學的으로 處理하기 때문이다.

아래表에 주어진 數值들을 살펴보자. 數의 合計는 72,091 이고 10年의 期間이기 때문에 平均은 7,209.1이다. 다음統計表는 各 數值의 偏差를 나타내고 있다.



| 年 度     | 쌀 수 입 량<br>( ton ) | 平均偏差<br>( $X_i - \bar{X}$ ) | 平均偏差<br>( 부 호 무 시 ) |
|---------|--------------------|-----------------------------|---------------------|
| 1 9 6 7 | 10,841             | + 3,631.9                   | 3,631.9             |
| 1 9 6 8 | 5,090              | - 2,119.1                   | 2,119.1             |
| 1 9 6 9 | 7,825              | + 615.9                     | 615.9               |
| 1 9 7 0 | 5,735              | - 1,471.1                   | 1,471.1             |
| 1 9 7 1 | 8,421              | + 1,211.9                   | 1,211.9             |
| 1 9 7 2 | 5,066              | - 2,143.1                   | 2,143.1             |
| 1 9 7 3 | 6,355              | - 854.1                     | 854.1               |
| 1 9 7 4 | 11,657             | + 4,447.9                   | 4,447.9             |
| 1 9 7 5 | 5,575              | - 1,634.1                   | 1,634.1             |
| 1 9 7 6 | 5,523              | - 1,686.1                   | 1,686.1             |
| 計       | 72,091             | 0                           | 19,815.2            |

平均으로부터 偏差의 총합은 0이다.

平均 (Mean) 의 定義에 따라 모든 數集合에서 平均偏差의 合은 항상 0이다

平均偏差의 어떤값은 양수이고 어떤값은 음수이다. 데이터가 넓게 分布되어 있다면 한층더 그러할것이나 ( 平均偏差의 合은 0 ) 부호를 무시한 수의 合은 상당히 클것이다. 이것은 散布度로써 부호를 무시한 平均으로부터의 差를 合한 것이다. 即 平均偏差는 偏差의 絶对值를 合한것으로 알려져있고 表의 마지막項에 나타나 있다.

그값은 19,815.2이다. 數集合에서 더많은 年數를 고려한다면 더높은값이 나타날 것이다. 그래서 그값을 數集合이나 相異한 母集團의 크기로 比較하기 위해 觀測數로 나눈다. 그래서, 소위 平均偏差는 이 데이터의 集合에서 1,982.52 ton이다.

標本觀測值  $X_1, X_2, \dots, X_n$ 으로부터 平均偏差를 計算하는 方法은 다음과같이 쓸수있다.

$$\text{平均偏差} = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{x}|$$

$\bar{x}$ 는 觀測值의 平均이고  $| |$ 은 平均으로부터 偏差의 絶對值를 나타낸다.

平均偏差는 모든값을 計算에 고려하기 때문에 좋은 散布度이다. 그러나 平均偏差는 根本的인 결점을 가지고 있다. 絶對值의 使用은 特히 標本理論에서, 数学的으로 處理하기가 매우 어렵다. 이러한 이유에서 자주 使用되지 않는다.

#### 10.4.5 標準偏差(The standard deviation)

平均偏差를 使用하는데 있어서 主要한 결점은 平均偏差가 絶對值를 合한것이며 이 절대치를 数学的으로 處理하기가 어려운데 있는것이다. 標準偏差는 이와 똑같은 原理에 기초를 두고 있으나 平均으로부터의 偏差를 제공하여 偏差부호를 제거한다.( - 數值의 제공이 +이다 )

平均偏差를 計算하는데 使用한 똑같은 데이터를 利用하여 標準偏差를 계산하여 보자.

| 수 출 량  | 평균 편 차    | 편 차 자 승  |
|--------|-----------|----------|
| 10,841 | + 3,631.9 | 13190697 |
| 5,090  | - 2,119.1 | 4490585  |
| 7,825  | + 615.9   | 379333   |
| 5,738  | - 1,471.1 | 2164135  |
| 8,421  | + 1,211.9 | 1468702  |
| 5,066  | - 2,143.1 | 4592878  |
| 6,355  | - 854.1   | 729437   |
| 11,657 | + 4,447.9 | 19783814 |
| 5,575  | - 1,634.1 | 2670283  |
| 5,523  | - 1,686.1 | 2842933  |
| 計      | 0         | 52312847 |

다시 한번 편차자승의 합계를 觀測數로 나누면 5231284.7 라는 답을 얻으며 偏差의 제곱을 더했기 때문에 원래 單位로 환산하기 위해 이 數字의 平方근을 求해야 한다. 이것을 標準偏差라고 하며 이 경우에

$$\text{標準偏差} = \sqrt{(52312847)/10} = 2,287.2$$

平均으로서  $\mu$  를 갖는 N의 값  $X_1, X_2, \dots, X_n$  으로 구성되는 母集團에 關한 標準偏差는

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2}$$

으로 정의한다.

母集團의 標準偏差는  $\sigma$  (그리스문자 Small sigma)로 表示한다. 標準偏差의 제곱을 分散(Variance)이라고 하며  $\sigma^2$ 으로 表示한다.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$$

母集團값을 推定하기 위하여 標本을 處理하고 標本分散 (혹은 標準偏差)을 計算하고자 한다면 公式을 약간 變형하면 이경우에  $S^2$ 은 標本分散,  $\bar{X}$ 는 標本平均 그리고  $n$ 은 標本크기 이다.

이때 公式은

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

이公式은  $N$ 代身에  $n-1$ 이라는 除數를 사용한것 이외에는 母分散의 公式과 같다. 이 이유는 여기서는 說明할 수 없는 複雜한 數學的理論에 근거를 두고있다. 記憶해야할 重要한 事項은 標本の 分散이나 標準偏差를 計算할때에  $n-1$ 로 나누는 것이다. 母集團의 分散이나 標準偏差를 計算할 때에는  $N$ 으로 나눈다.

이 狀態대로 많은 數集畧의 分散을 計算하는것은 매우 複雜한 절차이다. 무엇보다도 먼저 數集畧의 平均을 求해야 한다.

그다음에 平均으로부터 各畧의 觀測值의 偏差를 計算하고 이것을 제곱하여 더한다. 계산기로도 各값들을 두번씩 計算해야되고 또 시간이 걸린다.

그러나 分散公式을 再配列하여 훨씬 쉽게 計算할 수 있다. 標本の 分散에 대해서는

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n X_i^2 - \frac{1}{n} \left( \sum_{i=1}^n X_i \right)^2 \right]$$

母集團의 分散은

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 = \frac{1}{N} \left[ \sum_{i=1}^N X_i^2 - \frac{1}{N} \left( \sum_{i=1}^N X_i \right)^2 \right]$$

이 두번째 公式은 複雜하게 보이지만 실제 계산기를 사용하기가 훨씬 용이하다. 계산기의 記憶機能을 使用하여 2번째 계산은 어떤 중간 결과치를 기재하지 않고서도 計算할 수 있다.

두번째 公式으로의 變換은 度數分布의 分散을 計算하는 상당히 간단한 方法이다.

이章의 앞에 주어진 魚族資料를 例로 들어보자. 이것을 다음 表에서 관련된 公式으로 다시 作成한 것이다.

| 무 계<br>(kg) | 階級值<br>(X) | 度 數<br>(f) | f X    | f X <sup>2</sup> |
|-------------|------------|------------|--------|------------------|
| 2.0 ~ 2.9   | 2.45       | 7          | 17.15  | 41.0175          |
| 3.0 ~ 3.9   | 3.45       | 19         | 65.55  | 226.1475         |
| 4.0 ~ 4.9   | 4.45       | 16         | 71.20  | 316.8400         |
| 5.0 ~ 5.9   | 5.45       | 12         | 65.40  | 356.4300         |
| 6.0 ~ 6.9   | 6.45       | 6          | 38.70  | 249.6150         |
| 7.0 ~ 7.9   | 7.45       | 3          | 22.35  | 166.5075         |
| 계           |            | 63         | 280.35 | 1357.5575        |

$$\begin{aligned}
 \text{分散 } S^2 &= \frac{1}{\sum f - 1} \left( \sum f x^2 - \frac{1}{\sum f} (\sum f x)^2 \right) \\
 &= 1/62(1357.5575 - 1247.5574) \\
 &= 1/62 \cdot 110.0001 = 1.77
 \end{aligned}$$

데이터가 표본이기 때문에 분모로써  $n - 1$  을 쓰는 대신에  $(\sum f - 1)$  을 사용한다.

표본인 경우의 度数分布에서의 分散公式은

$$S^2 = \frac{1}{\sum_{i=1}^k f_i - 1} \left[ \sum_{i=1}^k f_i X_i^2 - \frac{1}{\sum_{i=1}^k f_i} \cdot \sum_{i=1}^k (f_i X_i)^2 \right]$$

이公式은 다소 複雜하게 보이나 숙달이 되면 이公式이 使用하기 쉽다는 것을 알게 될 것이다.

正規分布에 대한 標準偏差의 중요한 特徵을 여기서 言及할 것이다. 만약 母集團이 平均에 대하여 正規적으로 分布되어 있다면 모든값의 약 68%는 平均의  $1\sigma$  内に 存在하고 약 95.5%는  $2\sigma$  内に 存在할 것이다. 그래서 平均 88 標準偏差 17 을 갖는 正規分布에서 모든값의 約 68%가 범위 88-17 에서 88 + 17 ( 즉 71 과 105 사이 ) 사이에 存在한다. 그리고 95.5%는 범위 54 와 122 사이에 놓여있다. 이것은 正規分布를 그래프로 그려 證明할 수 있다. 값 71 과 105에서 X축에 그은 수직선은 曲線아래 總面積의 68%를 포함하고 있다. 이러한 特別한 性質은 標準誤差를 이해하는데 매우 有用하다는 것이 立証될 것이며 이러한것은 다음節에서 論議할 것이다. 실제로 標準偏差는 4개의 散布度중에서 가장 広範圍하게 使用된다.

標準偏差는 또한 標本誤差를 計算하는데 使用되고 있음을 알아 보자. 標準偏差는 널리 使用되지만 모든면에서 우수하다는 것을 뜻하는것은 아니다. 이것의 결점은 平均偏差에서보다 훨씬 極端的인 값에 상당히 影響을 받는다. 이러한 影響은 標準偏差의 計算에서와같이 제공했을때 매우크게되는 平均偏差(極端的인 값인 경우에 상당히 크게되는)때문에 나타난다.

#### 10.4.6 散布度에 關한 要約

##### (a) 範圍

- (i) 度数分布의 경우를 除外하고는 計算이 용이하고 一般的으로 이해하기 쉽다.
- (ii) 2개의 極端的인 觀測值에 근거를 두고있으며 매우 不安定하다.
- (iii) 数学적으로 處理하기가 어렵다.
- (iv) 分布의 全般的인 特性에 대한 情報를 提供하지는 않는다.
- (v) 散布의 程度把握에 簡便하고도 신속한 지침으로만 使用된다.

##### (b) 四分位範圍

- (i) 計算이 용이하고 이해하기 쉽고 散布度로써 範圍보다 우수하다.
- (ii) 極端的인 값에 影響을 받지않는다.

(iii) 端開階級을 갖는 大部分의 度数分布에서 計算될 수 있다.

(iv) 數學적으로 다루기 어렵다.

(v) 모든 觀測值의 情報를 使用하지 않는다.

(vi) 散布의 程度把握에 有用하고 簡便한 지침이 된다.

(c) 平均偏差

(i) 모든 觀測值를 이용함으로써 分布全體의 形態를 나타낸다.

(ii) 標準偏差보다는 極端的인 값에 影響을 덜 받는다.

(iii) 絶對值에 근거를 두고 있기 때문에 數學적으로 使用하기에 편리하지 않고 이런 理由에서 널리 使用되지 않는다.

(d) 標準偏差, 分散

(i) 모든 觀測值의 情報를 利用하는 散布度이다.

(ii) 약간 調整을 하면 計算은 상당히 간단하다.

(iii) 極端的인 값에 크게 影響을 받는다.

(iv) 多方面의 數學的理論이나 統計的理論에서 主된 役割을 하며 매우 廣範圍하게 使用된다.

## 10.5 標本誤差(Sampling error)

### 10.5.1 標準誤差(Standard error)의 計算

5章에서 標本誤差를 다루었으며 이는 全數調查代身에 標本을 扞함으로써 치루어야 할 代價라고 強調한바 있다.



또한 주어진 任意抽出法을 使用하면 이 標本誤差가 얼마나 큰가를 數學적으로 推定할 수 있다는 것에 注目했다. 이節에서 單純任意抽出法 (Simple random sample)에 대한 標本誤差를 計算하는 公式을 表記할 것이다. 이것은 매우 複雜한 內容이므로 本教材에서는 상세히 다루지는 않겠다.

상당히 큰 (無限한) 母集團에서 抽出한 標本에 標準誤差의 計算을 먼저 論議하고 상당히 넓은 海域의 魚族全體母集團을 調査하는데 例로든 魚族資料를 使用할 것이다.

63 마리 魚類무게의 資料에 對해 다시 言及하면 平均무게가 4.43 kg이라는 것을 쉽게 計算할 수 있다. 이것은 63개 觀測值의 標本으로부터 나온 그 海域에서 魚族平均무게 ( $\mu$ )를 推定한 것이다. 이 推定이 아주 正確할 것이라는걸 기대할 수는 없다. 다른고기 63 마리를 잡아서 第2의 標本을 만든다면 그 標本の 平均무게가 다소 다르다는것 - 이를테면 4.69 kg - 을 알게될 것이다. 63 마리의 여러 相異한 標本을 취하는 이런 方法을 계속할 수 있고, 각경우에  $\bar{X}$ 를 計算할 수 있으며  $\bar{X}$ 의 分布값을 찾아내기 위하여 각각의  $\mu$ 를 별도로 推定한다. 이런過程을 무수히 反復하면 (다행스럽게 실제로 실시하지 않는)  $\bar{X}$ 의 理論的인 分布를 얻을수 있고 이分布의 標準偏差를 얻을수 있다.

理論적으로 이 結果를 나타내 보면 即 크기  $n$ 의 任意標本이 매우크거나 無限母集團으로부터 취해진다면  $\bar{X}$ 의 理論分布는  $\frac{\sigma}{\sqrt{n}}$ 의 標準偏差를 갖는다.

이것을  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$  으로 쓰고  $\sigma_{\bar{x}}$ 를 平均의 標準誤差라고 한다. 平均의 標準誤差는  $\bar{x}$ 에 關한 理論上 標本分布의 變異度이기 때문에 統計學에서 매우 重要한 役割을 한다. 다시말하면 이것은 얼마나 많은 標本平均들이 標本間에 變化하는가를 예측할 수 있다는것을 나타내준다. 分母  $\sqrt{n}$ 에 의해 平均의 標準誤差는 標本크기에 따라 감소하거나 증가하는것을 알수있다. 그래서  $n$ 을 크게하면 할수록  $\mu$ 의 推定量으로써  $\bar{x}$ 는 보다더 信賴性이 있을 것이다.

물론 이러한 모든것은 理論的이고 실제로는 상당히 큰 標本을 취할 수 없다. 一般的으로 단지 하나의 標本만을 갖는다. 보다 重要하게  $\sigma$  (母集團의 標準偏差)의 값을 알지못하고, 그 海域에서의 모든고기를 잡아 무게를 달지 않는한  $\sigma$ 를 實際로 計算할 수 없다. 다행히  $S$  (標本の 標準偏差)를 計算할 수 있기 때문에  $\sigma$ 를 얻을수 있다. 標本이 任意的이고 不偏性을 갖고 標本크기가 상당히 크다는 것을 전제로 하면  $\sigma$ 에 아주 密接하게 近似한  $S$ 를 기대할 수 있다. 그러나  $S$ 는 小標本에서는 얻을 수 없고 大標本에서 平均의 標準誤差는  $S/\sqrt{n}$ 과 같다고 推定한다.

계산기를 사용하여 魚族데이터에서  $S$ 는  $1.37\text{ kg}$ 이라고 計算할 수 있다. 標本으로부터 고기 全体母集團의 平均무게는  $4.43\text{ kg}$ 이고 標準誤差는  $1.37/\sqrt{63} = 0.17\text{ kg}$ 이라고 推定할 수 있다. 이제까지 無限 혹은 매우 큰 母集團으로부터 抽出된 標本에 대해 알아보았다. 이제 마을이나 촌락내의 가구와 같은 적

은 母集團에서 標本抽出할때는 公式을 수정해야하며 그公式은 다음과 같다.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

다시 말하면 앞公式에다  $\sqrt{\frac{N-n}{N-1}}$  항을 곱한 것이다. 이것은

“有限母集團修正係數”라고 한다. 이項이 平均標準誤差의 推定値에 미치는 영향을 例를들어 說明하기 위하여 이項의 값을 (a) 40,000 개의 全體母集團에서 200 개 標本觀測値를 抽出한것과 (b) 400 個에서 20 個를 抽出한것을 比較해 보자.

첫번째 경우의 有限母集團修正係數는

$$\sqrt{\frac{40,000 - 200}{40,000 - 1}} = \sqrt{\frac{39,800}{39,999}} = 0.998 \text{ 과 같다.}$$

이값은 1 에 거의 가깝기 때문에 얻는 답에는 實質적으로 影響을 미치지 못하며 앞에서 言及한 매우 큰 母集團에서의 標本抽出일때 이것을 무시한 이유이다. 그러나 두번째 경우에 有限母集團修正係數는

$$\sqrt{\frac{400 - 200}{400 - 1}} = \sqrt{\frac{200}{399}} = 0.701$$

修正되지 않은값에 이러한 修正係數를 곱하게 되면 平均에 對한 標準誤差의 推定値를 약 30%가량 낮아질 것이다. 이것은 매우 意味있는 것이어서 確實히 무시할 수 없다. 첫번째 例에서 그海域의 물고기 母集團은 事實상 無限이고 有限母集團修正係數는 1 과 같게되고 公式에 하등의 影響도 미치지 못한다.

실제로는 이러한 修正式을 약간 變化시켜 이용한다. 卽

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{N-1+1-n}{N-1}} = \sqrt{1 - \frac{n-1}{N-1}} \text{로 쓸수있고 } n$$

과  $N$ 이 매우적지 않다면 이식은  $\sqrt{1 - \frac{n}{N}}$ 이며 또는 다른말로  $\sqrt{1 - \text{標本抽出率}}$ 과 거의 같다. 一般的으로  $\sigma$ 의 값을 알지못하기 때문에 公式에서  $\sigma$ 대신  $S$ 로 대치시키면 平均標準誤差의 推定値公式은 다음과 같이 쓸수있다.

$$\tilde{\sigma}_{\bar{x}} = \frac{S}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

이公式은 실제로 가장 유용한 공식이다.

위에서 공부한 모든것은 單純任意抽出과 관련되어 있다. 別개의 標本設計에 기초를 둔 推定量의 標本誤差公式은 一般的으로 다르고 各경우에 適當한 公式을 使用하는것이 重要하다.

#### 10.5.2 信賴區間 (Confidence intervals)

信賴區間이 무엇을 意味하는지를 알아보자. 어떤 海域의 모든물고기의 平均 무게 ( 標準誤差는  $0.17 \text{ kg}$ 임 )는  $4.43 \text{ kg}$ 이라고 推定할 수 있다. 그러나 이러한 서술은 그자체에 한계가 있다. 統計利用者들은 平均  $4.43 \text{ kg}$ 이 나타내는 것이 무엇인지를 完全하게 理解할 것이다. 하지만 標準偏差의 값을 어떻게 說明할 수 있을것인가?

다행히도  $\bar{x}$ 推定値에 관한 理論標本分布의 또다른 重要한 特性이 도움을 줄것이다.  $\bar{x}$ 의分布를 計算하면 正規曲線에 매우 가깝게 접근할 것이고 母集團自体가 正規的인 分布를 하지않더라도

이것은 사실이다. 正規分布에 關한 標準偏差의 特性은 前節에서 言及한것에 따른다. 平均의 양쪽  $1\sigma$  (標準偏差) 内에 모든 標本推定値의 약 68%가 놓여있고  $2\sigma$  内에 95% 이상이 놓여있다.

물론 항상 여러개가 아닌 단지 하나의 標本推定値를 갖는다. 그래서 上記 內容을 보다더 有用하게 하기위해 다른 形態로 說明할 必要性이 있다. 실제로 주어진  $\bar{x}$ 는 母平均의 不偏推定量이라 할수 있으며 標本平均의  $\pm 1\sigma$ 에 母平均을 내포할 確率은 약 68%이고 標本平均  $\pm 2\sigma$ 내에 存在할 確率은 약 95%라고 한다.

魚族데이터에 이것을 적용하면 그 海域의 魚族平均무게는 範圍 (4.43 - 0.17 kg) 과 (4.43 + 0.17 kg) 즉 4.26 과 4.60 kg 사이에서 68%의 確率을 갖고 範圍 4.09 와 4.77 사이에는 95%의 確率을 갖는다. 이런 範圍를 信賴區間이라고 이러한 確率( 즉 68%, 95% 등 )을 信賴水準이라한다.

統計人들은 대부분 95% 信賴水準을 利用한다. 이것은 올바른것이라고 確信할 수 있는 충분한 數值이기 때문이다. 魚族資料에서 범위 4.09에서 4.77 kg내의 어디엔가 魚族의 實際平均무게가 95% 信賴度( 다른말로, 상당히 確實한 )를 갖는다고 平均에 대한 結果를 해석한다. 그러나 사실상 確實한것은 아니고 眞平均이 이 信賴區間内에 있다고 단언해서는 안된다.

만약 95% 信賴水準이 어떤目的에 부적당하다고 決定되면 다른 信賴水準에 대해 이와같은 計算을 해야 할것이다. 예를들어 99% 信賴水準은 標本平均 양쪽 標準偏差의 2.6배되는 區間에

속한다. 그래서 99% 水準을 갖고 魚族의 眞平均무게는  $(4.43 - 2.6 \times 0.7)$  과  $(4.43 + 2.6 \times 0.7)$  kg 사이 즉 3.99와 4.87 kg 사이에 있다고 信賴할 수 있다. 어떤 必要한 信賴水準에 대해서도 이와 비슷한 計算을 할 수 있고 特殊表를 通하여 어떤水準에 적절한 信賴水準을 구할 수 있다.

### 10.5.3 結 論

이節을 끝맺기전에 하나의 부수적인 特徵을 찾아야한다. 이것은 매우 重要한 것이다. 上記 모든計算에서는 標本誤差만 다루었고 그標本은 確率的이고 不偏性이라는 假定下에서 이루어졌다. 이책의 앞머리에서 偏倚性和 非標本誤差의 影響을 상세히 다루었다는것을 記憶해야 할것이다. 또한 어떤調査에는 무시할 수 없는 커다란 偏倚性이 있다는것을 알고있다. 때때로 이러한 非標本誤差는 推定할 어떤方法이 있다면 標本誤差보다 더 크다는것을 증명할 수 있을것이다. 만약  $\bar{x}$ 가  $\mu$ 의 不偏推定量이라면 信賴區間만으로 結果를 나타낼수 있다는것을 잊어서는 안된다. 이것은 調査에서 非標本誤差가 있다는것을 알고있거나 짐작할 수 있음을 의미한다. 아무制限없이 平均과 標準誤差에 의하여 실제로 그 結果를 나타내는것은 危險스럽고 誤導되기 쉬운 것이다. 標準誤差를 公表하는것은 統計值의 利用者에게 그 結果에 대해 誤差의 限界를 正確히 나타냈다는 信임을 주고 다른 誤差가 없다는것을 말해주며, 그렇지 않을경우 다른誤差가 있다는것을 명백히 나타낸다. 중요한 국가의 政策決定은 正確한 統計

에 근거를 둔다는 信念下에서 이루어질 것이고 만약 公表된 統計가 標本誤差는 작지만 非標本誤差가 상당히 크다는 事實을 은닉한다면 이러한 決定은 不合理하며 重大한 危險을 내포하게 된다. 우리가 알고있거나 짐작되는 상당한 非標本誤差를 認定하는데 絶對적으로 솔직해야할 책임을 가지고 있으며 公表하는 統計의 결과에 表示해야 한다.

이節에 나타난 公式은 單純任意抽出에서 平均의 標準誤差計算을 위한것이라는데 注意해야한다. 여러 標本調査에서 이보다 더 複雜한 算式에 接하게 되는 경우가 許多하다. 예를들면 系統抽出法이나 層化抽出法같은 것이 結合된 경우이다. 이러한 複雜한 設計의 目的은 同一標本크기의 單純任意抽出에서 얻은것보다 標本誤差를 더 낮은 水準으로 減少시키기 위해서이다.

이러한 狀況에 適合한 公式은 (設計할때 그 複雜함을 고려할만큼) 매우 複雜하게되고 이러한것은 현재 本教材의 範圍에서 벗어난 것이다.

## 10.6 標本데이터로부터 母數推定

### 10.6.1 不偏推定量(Unbiased estimator)

5章에서 標本抽出의 原理와 主要標本調査設計의 形態를 알아보았다. 한편 이章의 앞節에서는 標本平均에 의해 얻어진 推定值의 精度를 求하는 方法을 論議했다. 그러나 推定值 그自体를 求하는 技法을 考察하지 않았다. 말하자면 全体母集團에 대한 推定值를 求하기 위하여 標本으로부터 얻어진 資料를

擴張하는 方法이다. 이것은 事實上 매우 어려운 問題이고 主要한 調査(특히 層化, 多段確率比例抽出을 포함하는 複雜한 標本設計를 갖는)에 대한 計算은 매우 複雜하게 될 수 있으므로 計算處理는 거의 大部分 컴퓨터에 의하여 수행되어야 할 것이다.

이節에서는 關係된 主要原則만을 제시할 것이다.

母數推定을 위해 一般的으로 使用될 推定量의 形態는

$$\hat{Y} = \sum_{i=1}^n W_i y_i$$

여기서  $\hat{Y}$  (발음 Y hat)는 母集團 Y에 대한 特性值의 推定值를 나타내며  $y_i$ 는 最終段階 標本抽出單位에서  $i$  번째 抽出된 이 推定值의 값이며  $W_i$ 는 그單位에 적절한 加重值이다. 加重值(Weight)는 승수(Multiplier), 擴張要因(Expansion factor), 擴大乘數(Raising factor)로 다양하게 일컬어지며 推定에 많은 어려움을 준다. 말하자면 加重值는 最終段階(Final-stage) 標本單位에 대한 값으로써  $y_i$ 를 말한다. 이것은 상이한 段(Stage)에서 상이한 標本抽出單位를 갖을수있는, 예를들면 1段에서는 마을 2段에서는 가구와 같이 多段抽出때문에 이조건이 必要하다. 그래서 標本抽出單位가 무엇인지를 주의깊게 정의해야 한다)

모든 單位들이 抽出에 동일한 기회를 갖는 가장 간단한 標本形態에 대한 標本資料로부터 母集團을 推定하는것은 아주 간단하다. 이러한 狀況下에서, 예를들면 單純任意抽出이나 單純系統抽出에서, 乘數(Multiplier)는 모든 抽出單位에서 동일하게  $\frac{N}{n}$ 이



나, K抽出간격으로 주어진다. 이러한 모든것은 母数推定値를 얻기 위하여 標本値에 이러한 乘數를 곱할 必要性이 있다.

실제로, 國家統計機關이 수행한 標本調査는 同一確率을 갖는 一段抽出法보다 더 複雜한 것이다. 이런 複雜한 標本設計를 소개함으로써 모든 抽出單位에 適用할 수 있는 하나의 乘數(Multiplier)만은 없으며 여러개(아마도 수천개의)의 乘數를 必要로 할것이다. 이것은 3가지의 觀点에서 간단하게 고려해 보자.

(a) 多段設計(Multi stage design) : 多段抽出의 경우에 全体母集團의 推定量은 다음段階 標本抽出의 기초가 되는 各 단계에서 연속적으로 총수를 推定함으로써 얻어진다. 예를 들면, 二段標本抽出에서 全体母集團 Y의 推定量은 다음과 같다.

$$\hat{Y} = \sum_i W_i \sum_j W_{ij} y_{ij}$$

여기서  $y_{ij}$ 는 첫단계,  $i$  번째 標本에서 2 단계  $j$  번째 標本の 값을 나타낸다. 그리고  $W_i$ 와  $W_{ij}$ 는 각각 첫번째 단계와 두번째 단계에서의 乘數이다.

이러한 公式은 標本の 各段階에서의 設計나 標本크기를 決定하기 위하여 各段階에서 適用되어야 하는 同一確率法 또는 確率比例法이냐에 따라 여러가지 方法으로 쓸 수 있다. 正確한 設計에 상관없이 各段階에서 乘數가 使用되어야 한다는 것을 알게 될것이다. 一般的으로 標本設計가 二段以上으로 구성될때 점차 더 複雜하게 된다.

(b) 層化抽出法(Stratified Sampling)

推定値는 各層에서 個別的으로 必要할 것이므로 점차적으로 全体母集團에 대한 推定値를 設計할 必要가 있을 것이다.

全体母集團 Y의 推定値는 層化單純任意抽出의 가장 간단한 狀況下에서 다음과 같다.

$$\hat{Y} = \sum_{S=1}^K \frac{N_s}{n_s} \sum_{i=1}^{n_s} y_{si}$$

여기서  $N_s$ 와  $n_s$ 는 各各 S번째 層에서의 抽出單位의 數와 標本크기이고  $y_{si}$ 는 그層에서 i번째 抽出된 單位의 값이다.

- (c) 標本크기에 대한 確率比列抽出(p.p.s) 5章에서 p.p.s 標本抽出은 抽出單位의 바람직한 分布程度를 얻는데 있어 事實상 중요한 장점을 갖고있는걸 보았다. 그러나 이方法의 各抽出單位에서 抽出確率은 變하기 때문에 乘數역시 各單位마다 달라지게 된다.

大規模調査에서 많은 特性値를 推定하고자 할때 推定過程에서 利用될 必要가 있는 數千個의 상이한 乘數가 있는 設計라면 推定段階에서의 計算은 매우 어렵고 많은 時間을 소모할지도 모른다. 集計段階에서 운영의 容易性때문에 모든 抽出單位에 單一共通乘數를 갖는 標本設計를 찾는것이 매우 바람직한 것이다. 그러한 標本設計는 標本觀測値의 加重值부여에 거의 努力이 들지않기 때문에 自動加重標本(Self-weighting)이라 한다.

이러한 境遇가 理想的이지만 둘이상의 共通乘數가 적용된다 하더라도 共通乘數의 數가 아주 많지 않은 限 製表過程에 소요되는 상당한 時間을 줄일 수 있을 것이다. 이런 경우를 部分自動

加重設計라 한다. 共通乘數의 數가 적은 이런 狀況에서는 運營  
상 設計에서 일어나는 制限때문에 單一乘數에서보다 더많은 어려  
움에 직면하게 될것이다.

実査段階나 製表段階중에서 自動加重設計를 하는것이 可能하다.  
前者의 경우에 多段標本調査에서 最終段階의 標本(Final Stage  
Unit)의 抽出은 自動加重設計를 할 수 있도록 해주고 있다.  
後者の 경우 適用될수있는 여러技法이 있다. (때때로 사사오입법이  
라한다) 可能한限 実査過程에서 自動加重을 求하는것이 좋지만  
이것을 非現實的으로 만드는 不均等한 標本크기로 作業부담을 유  
발시킨다. 製表段階에서 自動加重을 求함으로써 뒤따르는 問題는  
어떠한 技法이 利用되던간에 얼마간의 偏倚를 유발시키거나 分散  
을 크게 만드는 것이다. 이러한 影響을 최소한으로 줄이기위해  
많은 노력을 기울여야한다.

#### 10.6.2 比率推定(Ratio Estimation)

두 特性値의 全体母集團의 比率은 매우 重要하고 때  
로는 全体母集團 그자체보다 훨씬 더 重要하다. 예를들면 標本  
調査로부터 임금과 勤勞時間에 関한 情報를 얻을수 있고 總賃金  
이나 總勤勞時間數의 推定보다 時間當平均所得의 推定에 더많은  
関心이 있다. 상이한 두時点에서의 調査에서 어떤 한時点의 水準  
을 測定하는것보다 總賃金이 上昇 혹은 下落하였는가를 알아내는  
것이 더 重要하다. 마찬가지로 農作物標本調査에서 平均수확량,  
가계조사에서 平均家口所得과 支出, 勞動力調査에서 非雇傭勞動力

의 퍼센트등의 推定値를 얻는것이 더 重要할지도 모른다.

두特性値의 不偏推定量의 比率에 의해 母集團比率을 推定하는 방법으로 比率推定量을 使用한다.  $\hat{Y}$ 와  $\hat{X}$ 는 각각  $Y$ 와  $X$ 의 不偏推定量이고 母集團比率  $R = \frac{Y}{X}$ 의 推定量은 比率推定量  $\frac{\hat{Y}}{\hat{X}} = R$ 에 의해 求할 수 있다.

分母  $X$ 에 대한 실제 母集團값을 알고있는 경우 바람직한 母比率( $R$ )을 推定하기위해, 분자  $Y$ 를 推定하는 것이 必要하다. 즉 推定된  $\hat{Y}$ 값을  $X$ 로 나누어  $R$ 을 구한다. ( $\frac{\hat{Y}}{X} \rightarrow R$ ) 分母와 分子의 推定量이 상당히 比例的이라면 (즉, 만일 두特性値가 原点을 지나는 直線으로서 매우 線型的으로 관련돼있다면) 分母와 分子의 推定量의 比率에 근거한 推定量은 매우 效果的인 方法일 것이다.

이것이 意味하는것이 무엇인지 例를들어 살펴보자. 어떤地方의 헥타르당 年間 米穀生産量을 推定하길 원한다고 假定하고 全体벼과중面積은 地方行政記錄에 나타나 있다고 하자. 그러면 상이한 두가지 推定方法을 생각해볼 수 있다.

(i) 標本調査로 總米穀生産量을 推定하고 이것을 과중面積으로 나눈다.

$Y$ 는 生産量을  $X$ 는 面積을 나타내면 이런推定の 直接的인 方法은  $\frac{\hat{Y}}{X} \rightarrow R$ 로 表示할 수 있다.

(ii) 標本調査로 부터 生産량과 面積을 推定함으로써 쌀生産率을 推定한다.

$$\text{즉 } \frac{\hat{Y}}{\hat{X}} = \hat{R} \rightarrow R$$

이 두 번째 方法에서 이 比率과 실제 알려진 쌀과종면적(X)를 곱하면 全体生産量の 推定量이 된다. 式으로 나타내면

$$\hat{Y}' = \hat{R}X = \left(\frac{\hat{Y}}{\hat{X}}\right) \cdot X \rightarrow Y$$

이 推定量은 앞에서 言及한 狀況下에서의  $\hat{Y}$ 보다 더 效率的이다. 이 推定過程을 比率推定法이라고 이것은 널리 이용되고 있다.

이 方法은 또한 많은기간동안 어떤特性値의 率의 變化를 測定하는데 매우 有用하다. 例를들면 한해의 센서스에서 全体 家畜數를 計算했다고 假定하고 다음해의 標本調査로 家畜數를 推定하길 원한다고 하자. 標本으로 抽出된 各農家の 家畜數( $y_i$ 로 表示) 뿐만 아니라 前年度의 센서스때 그農家の 全体家畜數( $x_i$ 로 表示)도 구할수 있다.

그러면 家畜數의 變化率은

$$\frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i x_i} - 1$$

그리고 當該年度의 全体家畜數의 推定量은

$$\hat{Y}' = \left( \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i x_i} \right) X$$

여기서 X는 센서스해의 全体家畜數이고 Y'는 標本調査實施해

의 比率推定數이다.

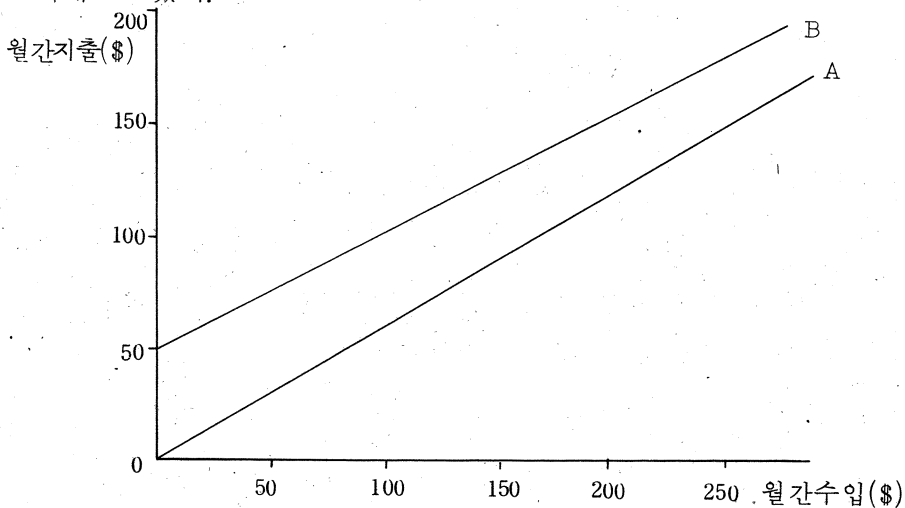
이것을 간단한 말로 나타내면 家畜調査는 다음과 같은 結論에 이른다. 지난해 센서스에서 가축수는 42 만이었다. 標本農場의 가축수는 그로부터 5% 增加되었다. 그래서 모든 農場에서 가축母集團은 5%로 增加되었다고 推定할 것이다. 그러므로 현재 家畜母集團을  $42,000 + 21,000 = 441,000$  이라고 推定한다. 이것은 比率推定에서의 推論의 限界線이다. 母數를 推定하기 위하여 標本으로부터 抽出한 어떤 比率을 計算한다. (이 경우에 센서스때의 가축수에 대한 현재 標本農場의 家畜數의 比率)

만약 推定量  $(\frac{Y}{X}) \cdot X$  를 고찰하면  $X$  는 標本에서 유도되지 않았다는 것은 명백하다. 그래서 標本誤差는 상수를 곱한 效果만을 갖는  $X$  와는 별개로 比率  $\frac{Y}{X}$  의 標本誤差에만 달려있다. 이것에 처한 實際的인 公式은 複雜하고 이책의 범위를 벗어난다. 그래서 어떤조건(다시 말하면 原点을 지나는 直線으로 나타난 두特性值간에 밀접한 線型關係를 가질 때) 하에서 標本誤差를 간단히 觀察할 것이다. 比率推定方法은 根本的으로 적은 標本誤差를 發生시킨다. 그러나 이것은 偏倚法이지만 적은 標本誤差를 얻을 수 있다. 다행히 偏倚는 大體的으로 큰 標本일 때 무시될 수 있다. 大部分의 實際的인 適用에 있어서 標本誤差를 減少시켜 얻은 利益에 比較해서 편의는 너무 적기 때문에 比率推定은 不偏推定法을 이용하는 편이 좋다.

#### 10.7 推定과 豫測을 위한 統計資料의 利用

### 10.7.1 單純線型回帰 (Simple linear regression)

統計的 데이터로 부터 推定이나 豫測을 하길 원하는 경우 그 이상의 變數간의 關係를 알아볼 必要가 있다. 두 變數에 對한 資料를 2 變量 (bivariate) 데이터라하고 이것이 各축에 한개의 變數를 갖는 圖表上的 點으로 表示된다면 散布圖를 얻을수 있다. 아래 圖表의 例는 여러家口의 月間支出과 收入을 나타내고 있다.



이러한 點들은 一般的으로 上向趨勢인데 우리가 豫상했듯이 高所得家口가 低所得家口보다 支出이 많다는 것을 나타내주고 있다.

두 變數間的 關係에 對한 研究를 通하여 變數에 依存한 한 變數를 推定할 수 있다. 例를들면 수입이 200 \$인 家口의 月支出額을 圖表로 推定하고자 할때 圖表上的 다른 點들의 上向趨勢方向을 따라 이가구를 代表하는 點이 어디에 위치하는가를 假定하여 대충 짐작할 수 있다. 그러나, 統計學에서는 이보다 더 正確히 求

하고자 할 것이다. 그리고 数学的 函数나 직선 관계인 가장 간단한 形態의 公式로 變數를 설명함으로써 그資料를 나타낸다.

우리는 直線關係公式의 見지에서 데이터를 說明하는 技法을 線型 回歸 (Linear regression)라 한다. 그래서 이러한 回歸線上의 點에 基조를 누어 推定할 수 있다. 이렇게하는데는 X에 대한 Y의 變數를 表現하는것이 一般的이다. 다시 말하면 Y는 從屬變數이고 X는 獨立變數이다.

이 직선 관계를 나타내는 방정식은  $y = ax + b$ 의 形態일 것이다. 여기서 b는 절편이 oy축상의 어디에 있는가를 나타내는 常數이고 a는 직선의 기울기를 나타내는 常數이다. 다시 말하면 a는 x의 단위증가량에 대한 y의 증가량을 나타낸다. 직선 관계를 갖고 있는 이런 算式을 만들때마다 圖表上에 나타난 點의 가장 適合한 方程式을 表示하는 2개의 상수 a, b값을 計算할 수 있을 것이다. “보다 適合한”이란말은 좋을수도 있지만 나쁠경우도 있다. 그러나 항상 어떤線보다 더 適合한 直線이 있을 것이다. 例를들면 收入과 支出圖表에서 A를 表示한 直線이 기울기에 關於 限 適合하다. 이 直線은 oy축의 절편에 대해서는 좋지못한것이 명백하다. 반면에 B선은 거의 正確한 位置에서 oy축을 절편하는것처럼 보이지만 그 기울기는 데이터에 適合하게 보이지는 않는다. 명백히 도표상에 나타난점에 대해서 이들보다 더 나은 적합한 직선을 구할 수 있다. 그러나 문제는 모든 가능한 선 中에서 求하는 것이다.

“최적”이 의미하는것이 무엇인지 정의할 수 있는 많은 판단의 기준이 있다. 그러나 一般的으로 使用되는 方法은 最小自



乘法이다. 公式에 의하여 計算된 값과 觀測值간의 差의 自乘合計가 가능한한 最小가 되는 方法으로 x에 대한 y를 나타내주는 公式을 설정할 수 있을 것이다.

우리는 最小자승법을 사용하여, 最適回帰線을 나타내는 a와 b를 數學적으로 計算할 수 있다. 이것은 여러方法으로 쓸 수 있지만 우리는 다음과 같이 使用할 것이다.

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$a = \bar{y} - b\bar{x}$$

여기서 n은 散布圖에서 나타날 수 있는 兩觀測數이다. 이 公式은 다소 複雜하게 보일지 모르지만 事實상  $\sum x$ ,  $\sum y$ ,  $\sum x^2$ ,  $\sum xy$ , n의 값을 計算하면 데이터에 가장 適合한 최적직선에 대한 公式은 쉽게 計算될 것이다.

回帰分析은 需要와 供給曲線, 生産과 消費函數등과 같은 經濟的인 變數間的 關係를 推定하는데 있어서 實際적으로 広範圍하게 使用되어진다. 또한 時系列分析(Time series analysis)을 하는데 크게 이용되고 流失된 데이터를 推定하기 위하여 利用可能한 데이터에 對하여 直線추세선을 一致시키는데, 또한 장래에 대한 推定을 하여 計劃의 方向을 얻는데 가장 重要하다.

實際적인 적용에 대하여 알아보기로 하자. 그리고 어떻게 線型回帰가 時系列分析을 위한 추세선을 그리는데 사용될 수 있는 가를 알아보자. 시멘트生産을 나타내는 다음과 같은 데이터를

이용해 보자.

| 年 度     | 生 産 量 (千 Tons) |
|---------|----------------|
| 1 9 7 2 | 849            |
| 1 9 7 3 | 863            |
| 1 9 7 4 | 928            |
| 1 9 7 5 | 897            |
| 1 9 7 6 | 이용불가           |
| 1 9 7 7 | 944            |
| 1 9 7 8 | 935            |
| 1 9 7 9 | 1,011          |

실습문제로서 1976年の 시멘트(統計資料를 求할 수 없던해)生産량을 推定하는것과 또 1980年과 1981年 生産량이 얼마나 될것인가를 推定해보자. 數理的으로 알기쉽게 하기위하여 '71年을 1로 시작하여 매해에 번호를 부여하고  $\Sigma X$ ,  $\Sigma Y$ ,  $\Sigma xy$ ,  $\Sigma X^2$ 을 計算한다. 또한 現在 計算하는데는 利用되지 않았으나 이節의 뒷節에서 必要하게 될  $\Sigma y^2$ 을 計算해 보자.

| X | y   | X <sup>2</sup> | xy   | y <sup>2</sup> |
|---|-----|----------------|------|----------------|
| 1 | 849 | 1              | 849  | 720801         |
| 2 | 863 | 4              | 1726 | 744769         |
| 3 | 928 | 9              | 2784 | 861184         |
| 4 | 897 | 16             | 3588 | 804609         |
| 6 | 944 | 36             | 5664 | 891136         |

| X    | y    | X <sup>2</sup>     | xy    | y <sup>2</sup>     |
|------|------|--------------------|-------|--------------------|
| 7    | 935  | 49                 | 6545  | 874225             |
| 8    | 1011 | 64                 | 8088  | 1022121            |
| 31   | 6427 | 179                | 29244 | 5,918,845          |
| (Σx) | (Σy) | (Σx <sup>2</sup> ) | (Σxy) | (Σy <sup>2</sup> ) |

관측수 n은 7, 을 공식에 대입하면

$$b = \frac{29,244 - \frac{31 \times 6,427}{7}}{179 - \frac{(31)^2}{7}}$$

$$= \frac{781.57}{41.71}$$

$$= 18.74$$

$$a = \frac{6,427}{7} - 18.74 \left( \frac{31}{7} \right)$$

$$= 918.14 - 82.99$$

$$= 835.15 \text{ (반올림하여 } 835.2)$$

고로 回歸線의 方程式은

$$y = 835.2 + 18.74 X$$

X = 5 (1976年)일때 y의 추정치는  $835.2 + 93.7 = 930$ , X =

9 (1980年)일때  $y = 835.2 + 168.7 = 1,004$ , X = 10 (1981年)

일때  $y = 835.2 + 187.4 = 1,023$

그러므로 線型回歸를 사용하여 1976年の 生産量을 推定하면

930,000 屯이며 1980년에는 生産이 1,004,000 屯이 될것이며

1981년에는 1,023,000 톤일 것이다.

$x = 0$  일때 즉 1971년에 있어서  $y$ 는 835로 觀察될 것이다. 이것은 1971년 835,000 톤이라는 추정된 生産量을 가리키는  $y$ 의 절편이다. 선의 기울기는 18.74이므로 매년 18,740 톤의 시멘트生産이 增加한다고 推定했다.

回帰線에 대한 公式이 決定되어지면 그것은  $y$ 값의 推定量으로 적용되고 이것은 變數  $x$ 의 어떤값에 대한  $y$ 를 推定하는데 사용되어질 수 있다. 回帰公式으로부터 얻어진 값과 觀測值간의 偏差는 誤差로서 정의되어지며 다음과 같은 公式을 사용함으로써  $x$ 에 근거를 두고 예측하는데 關係된 오차의 測定方法을 求할수 있다. 사용한 公式은

$$\sigma_e = \sqrt{\frac{1}{N}(y - y_e)^2}$$

여기서  $\sigma_e$ 는 推定量의 標準誤差이고  $(y - y_e)$ 는  $y$ 의 觀測值와 推定值간의 差이다.

回帰分析은 豫測의 手段으로 重要하고 널리 사용되어지지만 時系列관계에 있어서 使用은 많은 演習을 해야 한다. 장기추세이의 資料의 다양한 이유로써 波動은 있을지도 모른다. 그리고 그러한 波動을 고려하는것은 필수적이다. 波動은 다음 3개의 形態로 나타낼 수 있다.

- (a) 季節變動 - 이것은 時系列에 있어서 단기간(1年以下) 變化에 대하여 적용한다. 月降雨量 月農作物 生産등의 記錄은 季節變動을 나타내는 데이터의 例이다.

- (b) 循環變動 - 時系列資料에 있어서 規則的인 정점과 골짜기를 기대할 수 있는 反復的인 長期變動이다.
- (c) 不規則變動 - 長期趨勢의 부분적 變動이 아니고 季節, 순환변동에 의해 설명할 수 없는 부분에 대한 變動이다. 이變動은 洪水, 가뭄, 파업 등의 비정상적인 조건에 기인할 것이다. 예를들면, 앞예제에서 1975年 시멘트生産이 長期罷業에 의하여 600,000 ton으로 떨어졌다면 罷業의 影響을 받지 않았던 다른해의 生産水準을 豫測하기 위하여 이런 非正常的인 낮은 數値를 利用하기에는 상당히 부적당한 것이다.

이러한 變動의 影響을 除去하기 위하여 많은 技法이 있고 回歸分析을 통하여 長期趨勢를 測定할 수 있다. 예를들면 季節變動의 影響은 단지 年間資料에 관한 趨勢를 취함으로써 除去될 수 있고 循環變動은 循環週期的 길이와 同一한 期間의 資料를 平均한 移動平均法에 의해 除去될 수 있다. 여기서는 이 技法에 대해 상세히 공부하지 않을 것이지만 趨勢線은 어떤 特別한 外部의 影響을 받기쉬운 資料에 기초를 두지 않는다는 것을 強調하기 위한 時系列分析(혹은 어떤 回歸分析 문제에 대해)에 必須的이라는 것을 다시한번 強調하는 것이다.

이제까지 單純線型回歸에 대해서만 論議했지만 물론 모든 資料가 直線形態를 취하는 것은 아니다. 예를들면 우리는 資料를 나타내는 點들이 曲線을 이루고 있음을 알 수 있다. 그러나 이러한 事實이 發生한다 하더라도 線型回歸分析을 利用하는 것이 可

能할지도 모른다. 원데이터 자체가 線型關係를 나타내지 않는다고 할지라도 X에 대한  $\log y$  또는 x에 대한  $\sqrt{y}$ 를 圖表로 작성하면 直線에 가까운線을 求할 수 있다. 物價指數를 다루는 11章 끝에서는 時間의 變化에 따른 物價上昇이 曲線의 形態로 나타난 實例가 있다. 그러나 指數에 대해서 對數(log)를 취하면 時系列은 直線이 된다. 統計資料를 分析할때 直線은 複雜한 計算을 간단하게 해줌으로써 이러한 指數에 대한 線型關係의 形態를 발견하는데 큰 도움이 된다.

#### 10.7.2 相關係數 (Coefficient of correlation)

觀測值 y와 回歸線으로 推定한 값과의 偏差는 推定量의 標準偏差  $\sigma_e$ 에 依해 測定될수 있음을 배웠다. 데이터가 얼마나 잘 回歸線에 適合한지를 나타내는데 매우 有用하며 또다른 測度는 r로 表示한 相關係數이다. 이測度에 대한 公式은 다양한 方法으로 表示할수 있으나 가장 많이 쓰이는 方法은

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2 \cdot n \sum y^2 - (\sum y)^2}}$$

이測度에 대해서는 상세히 공부하지 않을것이지만 r의 最大값은 +1이고 最小값은 -1이다. +1의 相關係數는 모든 觀測值가 完全하게 回歸線上에 놓여있는 即 X와 Y사이의 完全한 양의 相關을 나타내고 (다시말하면 X가 증가함으로써 y가 增加) -1의 相關係數는 完全한 음의 相關 (다시말하면 X가 增加하면 y는 減少)을 나타낸다. 그래서 回歸線은 그래프 위의

왼쪽에서 오른쪽으로 下向한다.  $r = 0$  일때  $X$ 와  $Y$ 사이에는 線型關係가 없고, 相關이 없다.

실제로, 우리는 당연히  $r$ 은  $+1$ 과  $-1$ 사이의 어떤 中間값 이라는걸 예상할 수 있고  $r$ 이  $+1$ 이나  $-1$ 에 가까울수록  $X$ 와  $Y$ 간의 關係는 보다더 密接하다. 다른말로 하면 回歸線이 觀측치에 잘 適合하면 할수록, 回歸公式에서 유도될 推定値를 보다더 信賴할 수 있다. 반대로  $r$ 이  $0$ 에 가까울수록 相關關係는 弱하고 豫測과 計劃을 위한 어떤 回歸分析에서 평가할 수 있는 信賴度는 낮을것이다.

이제  $r$ 을 計算한다면, 시멘트生産에 對한 앞例에서

$$\begin{aligned} r &= \frac{7 \times 29244 - 31 \times 6427}{\sqrt{7 \times 179 - (31)^2} \cdot \sqrt{7 \times 5918845 - (6427)^2}} \\ &= \frac{5471}{\sqrt{292} \cdot \sqrt{125586}} \\ &= 0.90 \end{aligned}$$

$r$ 은 기대했듯이  $x$ 가 增加하면  $y$ 가 增加하므로 陽의수이고, 상당히 높다. 回歸線은 데이터에 매우 잘 맞고 상당한 信賴度를 갖고있어서 다른해의 生産量을 推定한다면 實際値에 상당히 접근할 것이다.

이보다 더 接近하는것이 可能하고 推定値에 대한 信賴限界를 計算하는것도 可能하다. 이것은 10.5.2節에서 論議한 標本으로부터 推定値에 대한 信賴限界의 計算과 상당히 유사하다. 그러나 수식은 보다 複雜하게 되어서므로 이책에서는 이러한 計算은 하지 않을것이다.

## 第 11 章 消費者物價指數 및 其他指數

### 11.1 概 要

統計를 나타낼 수 있는 가장 重要한 方法 - 특히 傾向 및 成長率을 表示하는 方法 - 中의 하나는 指數利用에 關한 것이다. 즉 많은 사람들은 指數라는 用語를 消費者物價指數 또는 都売物價指數로 생각하나, 經濟 및 社會統計를 分析하고, 表示하기 爲해서 많은 相異形態의 指數가 있으며, 本章에서는 몇 가지의 指數를 간단히 說明하고, 地方事務所職員과 實查要員들이 그들의 業務遂行過程에서 直面할 可能性이 높은 消費者物價指數에 關하여 重點的으로 說明하고자 한다. 指數와 關聯된 計算은 간단하지만 公式은 複雜하게 보이므로 우선 概念과 計算過程에 關한 것을 論議하고 本章의 後部에서 公式을 소개하고자 한다. 指數란 比率로서 計算된 統計的 測定值이며 아래와 같다.

- (i) 時点間의 變量의 크기 및 變量 Group 內에서의 相對的 移動
- (ii) 서로 相異한 地域間에 있어서 同一時点에서의 變量의 크기 및 變量의 Group 內에서의 相對的 差異.

變量이 하나뿐인 경우에, 指數는 주어진 時点에 있어서 變量의 基準值를 100으로 할때 基点 및 期間에 있어서 變量의 크기를 퍼센트 (Percentage)로 表示한다. 즉 어떤 地方에서 쌀의 基準年度 生産量이 850 톤 (Ton) 이고, 現年度에 1,127 톤이 生産되었다면,  $1,127/850 \times 100 = 113$ 으로 計算하여, 指數로 나타낼 수 있다. 다른말로 表現하면 쌀의 生産이 基準年度에 比하여 現



年度에 33% 増加하였다는 것을 나타낸다. 그리고 이런 생각은 別個의 Group 즉 關聯된 變量(支出, 消費項目)을 포함하여, 綜合指數 作成에까지 擴大된다. 이미 說明한바와 같이, 變量들이 다른 項目에 關聯되어있고 서로다른 單位로 測定되어 있다 하더라도 指數作成이 可能하며 이런 指數는 發生한 全般的인 變化를 測定 하게 된다. 그리고 指數값 計算方法和 指數에 포함된 各項目에 加重值를 주는 方法에 關해서는 11.3 節에서 說明하고자 한다.

## 11.2 指數의 形態

指數는 價格, 實質產出, 消費, 所得 또는 어떤 財貨나 用役의 流量과 貯量等を 指數로 表示할 수 있으나, 세계의 主要形態로 区分하고자 한다.

- (i) 物價指數 (例: 消費者 物價指數)
- (ii) 數量指數 (例: 産業生產指數)
- (iii) 價格指數 (例: 輸出價格指數)

## 11.3 消費者 物價指數

商品의 價格은 時間에 따라서 變化하기 때문에 發生된 變化를 測定하고 그 意味를 파악할 必要가 있다. 단지 한 商品의 경우에는 時點間의 價格變化를 計算하여 퍼센트로 表示하는 것은 어렵지 않다. 예를 들면 重量 500 g 짜리 茶의 값을 每分期 中旬을 基準으로 다음과 같이 調査하였다고 假定하자

|               |         |
|---------------|---------|
| 1979 . 2 월 중순 | \$ 2.12 |
| "    5 월 "    | \$ 2.18 |
| "    8 월 "    | \$ 2.24 |
| "   11 월 "    | \$ 2.40 |
| 1980 . 2 월 "  | \$ 2.52 |

1979년 2월에서 5월간의 增加率은  $2.18-2.12 / 2.12 \times 100 = 2.8\%$ 이며, 이런 增加現象을 나타내는 다른 方法은 1979년 5월의 價格과 比較하여 1.028 배라고 말함으로써 相對價格으로도 나타낼 수 있다. 또 消費者 物價指數로서 1979년 2월 中旬을 100으로 하여 1979년 5월은 指數가 102.8임을 얻을 수 있다. 또 위의 資料는 個別價格指數와 指數로도 아래와 같이 나타낼 수 있다.

| 期 間      | 價 格 (\$) | 個 別 價 格 指 數 | 指 數   |
|----------|----------|-------------|-------|
| 1979年 2月 | 2.12     | 1.000       | 100.0 |
| "    5月  | 2.18     | 1.028       | 102.8 |
| "    8月  | 2.36     | 1.113       | 111.3 |
| "   11月  | 2.40     | 1.132       | 113.2 |
| 1980年 2月 | 2.52     | 1.189       | 118.9 |

위의 例의 모든 價格比較는 1979년 2월을 基準으로 해서 行하여졌다. 그리고 마지막 두난중의 하나에서 1980년 2月末까지 1年동안의 價格增加率이 18.9%이었다는 것을 쉽게 알 수 있고, 指數計算法과 마찬가지로 成長率도 計算할 수 있다. 즉 1979年 5月과 11月間의 個別價格增加率은  $113.2/102.8 =$

1.101로 計算될 수 있고, 그 增加率은 10.1%임을 나타낸다. 그러나 實際에 있어서, 財貨나 用役의 全品目에서 全體費用이 어떻게 變化되었는지를 把握하는데 더욱 關心이 있고, 이것은 各種의 食료품, 化粧品, 주택, 수송도구, 의류등이 包含되고, 또 世界의 거의 모든 國家들은 上記의 目的을 達成하기 위하여 一般 消費者에 依하여 支出된 價格變化를 測定할 수 있도록 消費者 物價指數를 편제, 공표한다. 消費者 物價指數는 個別價格指數와 同一한 方法으로 計算되나 兩者間의 主要差異는 個別價格은 한 商品을 對象으로 하나 消費者 物價指數는 財貨나 用役의 全體品目を 對象으로 한다는 것이다.

#### 11.3.1 消費者 物價指數의 作成

大部分의 物價指數는 社會의 多樣한 支出形態를 精確히 反映하도록 數百個의 다른 商品의 項目을 包含하고 있으나 이 冊에서는 그런 範圍에 까지 擴大하여 說明하기가 困難하므로, 物價指數作成方法을 5個의 品目에 限定하여 說明하고자하며, 同一한 方法으로 實際에 利用되고 있는 指數도 作成됨을 認識하여 주길 바란다. 作成코자하는 5個品目の 價格이 1979年2月부터 1980年2月中의 期間동안 아래표와 같이 變化하였다고 하자

또한 만족할만한 物價指數는 단지 各品目の 한 單位에 대한 價格을 合計함으로써 作成되어 지지않는다는 事實을 理解하는게 重要하다. 이에 관해서, 위의 例를 가지고 說明하면, 쌀값의 10% 引上은 消費者의 全體費用을 約3센트 引上시키는 反面에 가루비누가격의 10% 引上은 消費者의 全體支出을 11% 引上하게 한다.

| 상 품  | 단 위      | 가 격 (\$) |        |        |         |        |
|------|----------|----------|--------|--------|---------|--------|
|      |          | 1979.2   | 1979.5 | 1979.8 | 1979.11 | 1980.2 |
| 쌀    | Kg       | .30      | .30    | .30    | .30     | .34    |
| 차    | 500g (포) | 2.12     | 2.18   | 2.36   | 2.40    | 2.52   |
| 통조림  | 375g (통) | .91      | .94    | .95    | 1.06    | 1.11   |
| 석 유  | 리 터      | .14      | .15    | .18    | .18     | .22    |
| 가루비누 | 1Kg (포)  | 1.07     | 1.09   | 1.09   | 1.14    | 1.14   |

즉 品商單位當價格이 높으면 높을수록, 그 商品價格引上이 全體 指數에 미치는 영향이 크다는 점이다. 物價指數에 새로운 自動車를 包含하고 있는 어떤 國家에서는, 이 새로운 自動車 品目單位當 높은 價格을 갖게 될 것이므로 指數에 큰 영향을 미치게 되어 作成 된 指數는 매우 불만족한 狀態가 될 것이다.

(a) 加重值

위와같은 問題點들은 實際의 支出形態에 一致할 수 있도록 各商品의 品目單位들에 加重值를 부여함으로써 可할 수 있다. 例를 들면 1個月 동안에 家計가 平均的으로 쌀 25 Kg, 차 1포, 돼지고기통조림 4통, 가솔린 25 리터와 가루비누 1포를 購入한 것으로 생각하고, 各品目的 價格은 다음과 같다고 하자.

| 項 目         | 支 出 (\$) |        |        |         |        |
|-------------|----------|--------|--------|---------|--------|
|             | 1979.2   | 1979.5 | 1979.8 | 1979.11 | 1980.2 |
| 쌀 25 Kg     | 7.50     | 7.50   | 7.50   | 7.50    | 8.50   |
| 차 1 포       | 2.12     | 2.18   | 2.36   | 2.40    | 2.52   |
| 돼지고기통조림 4 통 | 3.64     | 3.76   | 3.80   | 4.24    | 4.44   |
| 가솔린 25 리터   | 3.50     | 3.75   | 4.50   | 4.50    | 5.50   |
| 가루비누 1 포    | 1.07     | 1.07   | 1.09   | 1.14    | 1.14   |
| 합 계         | 17.83    | 18.26  | 19.25  | 19.78   | 22.10  |

위 表는 가루비누보다도 쌀에 더 많이 支出하고 있다는 實際的 狀況을 보다 잘 나타내고 있다. 즉 基準時点에서 쌀 값 10%의 上昇은 全體支出費用을 75센트 增加시키게 되나, 가루비누가 격의 10% 上昇은 全體支出費用을 11센트 增加하게 한다. 消費者 物價指數作成의 基本目的은 個個商品의 價格變化가 消費者의 全般的인 購買力에 미치는 效果를 測定하는데 있고, 各商品의 單位에 여러 숫자를 配分하는 것을 加重化한다고 하며, 두번째 例에서 加重化된 品目は 첫번째 例와는 달리 消費者들이 그들의 所得을 支出하는 方法을 反映한다. 各 項目의 全體費用을 利用함으로써, 以前에 차에 대한 個別價格指數를 計算했던 것과 同一 方法으로 消費者物價指數를 計算하면 다음과 같다.

大部分의 統計機關에서 利用되고 있는 作成方法은 同一한 指數 計算時에도 다른 方法을 利用하고 있다. 즉 基準時点에 있어서의 支出額이 加重指數를 計算하는데 利用되기도 하고, 各 項目의

相對的 重要度 ( 쌀은 指數로  $\frac{7.50}{17.83} \times 100$  또는 42.1%의 代表度를 나타낸다 )가 利用되기도 한다.

| 期 間     | 指 數                                      |
|---------|--|
| 1979.2  | $\frac{17.83}{17.83} \times 100 = 100.0$ |
| 1979.5  | $\frac{18.26}{17.83} \times 100 = 102.4$ |
| 1979.8  | $\frac{19.25}{17.83} \times 100 = 108.0$ |
| 1979.11 | $\frac{19.78}{17.83} \times 100 = 110.9$ |
| 1980.2  | $\frac{19.78}{17.83} \times 100 = 123.9$ |

各 項目에 대한 個別價格指數는 相應하는 加重值를 곱하여 計算하며, 이 加重化된 個別價格指數는 指數를 計算하기 위하여 合計된다. 1979年 5月의 物價指數는 아래와 같이 計算된다.

| 品 目 | 加 重 值                                  | 個 別 價 格 指 數                 | 加 重 值 × 個 別 價 格 指 數 |
|-----|--|-----------------------------|---------------------|
| 쌀   | $\frac{7.50}{17.83} \times 100 = 42.1$ | $\frac{.30}{.30} = 1.000$   | 42.10               |
| 차   | $\frac{2.12}{17.83} \times 100 = 11.9$ | $\frac{2.18}{2.12} = 1.028$ | 12.23               |

| 品 目     | 加 重 値                                  | 個 別 価 格 指 数                 | 加 重 値 ×<br>個 別 価 格 指 数         |
|---------|--|-----------------------------|--------------------------------|
| 돼지고기통조림 | $\frac{3.64}{17.83} \times 100 = 20.4$ | $\frac{.94}{.91} = 1.033$   | 21.07                          |
| 가 솔 린   | $\frac{3.50}{17.83} \times 100 = 19.6$ | $\frac{.15}{.14} = 1.071$   | 20.99                          |
| 가루비누    | $\frac{1.07}{17.83} \times 100 = 6.0$  | $\frac{1.07}{1.07} = 1.000$ | 6.00                           |
| 계       | 100                                    |                             | 102.39<br>( 102.4 로<br>반올림했음 ) |

1979年 5月에 指数가 102.4라는 것은 마켓 바스켓(Market basket) 方法에 의한 것과 完全 一致하며, 마찬가지로 다른 時點을 基準으로 이 方法에 의해서 作成된 指数는 同一한 結果를 提供할 것이다.

위에 보인 方法이 마켓-바스켓 接近法에 의한 것보다 난해하고, 불편하다는 사실은 간단하게 제시된 예로 부터 알 수 있을 것이다. 그러나 實際에 있어서 特히 各 品目들에 대하여 세밀한 指数를 使用하지 않더라도 必要한 것에, 相應하는 個別價格指數를 포함으로써 적절한 加重値를 얻을 수 있다는 點이며, 두번째 方法에 의한 公式이 마켓-바스켓 方法에 의한 公式보다도 훨씬 簡單하다는 事實을 알게 될 것이다.

(b) 品目の 範圍 및 分類

앞에서 본바와 같이, 消費者 物價指數는 단지 5個의 品  
目に 限定되는 것이 아니라 수 많은 品目を 포함하고 있으며,  
消費者物價指數에 包含되고 있는 모든 費目の 範圍에는 食料品,  
衣類 및 家口商品뿐만 아니라 다른 品目들 例를들면 서어비스,  
전세 및 住宅維持費等 수 많은 項目을 包含하고 있다. 한 指數  
로 이 모든 項目들을 結合하는 것이 사람들에게 生計費指數로 알  
려지고 있으나, 잠시 후에 說明하는 바와같이 實際에 있어서 正確  
한 用語가 아니므로 이런 用語使用은 自制되어야 한다.

財貨와 用役價格의 全般的인 上昇을 記錄하는 외에 一般的으로  
消費者物價指數는 많은 項目그룹(Group)으로 나누어질 수 있고, 그  
런 그룹간의 모든 項目에 대한 價格變化를 나타내며, 그런 그룹  
分類가 全世界的으로 一致하는 것은 아니나 아래와 같이 나눌 수  
있다.

食 料 品

담배 및 酒類

住宅 및 家口用品類

衣類 및 신발류

運輸部門

其 他

各 그룹은 보다 적은 소 그룹으로 나눌 수 있다. 즉 食品類  
그룹은 육류, 해류, 곡물류, 과일 및 채소류등과 같이 보다 적은  
소 그룹으로 分類할 수 있고, 소 그룹은 몇개의 개개 項目을



包含한다.

이미 周知하는 바와같이, 消費者物價指數는 어떤 一定 期間의 社會支出形態를 反映하기 위해서 作成되고 있는데, 大部分의 경우에 支出의 平均的인 形態는 標本家口를 選定하여 期間中 모든 標本家口の 支出을 記錄함으로써 測定한다. 그리고 이런 類의 調査를 家計支出調査 또는 家計所得支出調査라 칭하고, HES 또는 HIES로 약칭하기도 하는데 이 教材에서는 HES를 使用하고자 한다.

HES 調査結果는 範圍에 包含되어 있고, 加重値가 부여된 項目들을 測定하기 때문에 消費者物價指數가 作成되기 前에 利用하는 것이 必要하나, 어떤 特殊狀況에서는 家計支出調査가 實施됨이 없이 家計支出의 代表的인 形態를 測定할 수 있다. 商店이나 市場에서 購入할 수 있는 多様하고 많은 品目들 모두가 實際적으로 消費者物價指數에 包含될 수 없다. 그 代身 代表品目이 選定되고 있는데 이는 指數作成 對象에 包含되지 않는 品目の 價格變化가 對象에 包含된 品目の 價格變化와 거의 같도록 選定해야 할것이다. 즉 스파게티 한 통은 HES에 包含되지 않을 수도 있으나 한 品目の 價格變化가 다른 品目の 價格에 영향을 미친다고 생각할때 스파게티의 加重値는 유사품인 볶은 콩 한 통에 配分할 수 있다. 그러나 이러한 方法으로 對象品目の 構成을 決定하는데에는 상당히 신중한 判斷을 하여야 한다. 이러한 過程이 너무 지나치면 그 對象에는 상당히 적은 數의 品目만이 남게 될 것이다.

그렇게 되면 어떤 한 품目の 價格變化는 指數에 과도한 影響을 미치게 되므로 包含되지 않은 品目 價格의 增加나 減少는 測定하지 못한다. 또 다른 側面에서 너무 많은 項目들은 指數를 過大하게 만드는 結果를 낳게 되고, 指數作成에 必要的한 資料蒐集 및 指數計算에 過大한 노력을 必要로 한다.

(c) 品目明細

한 品目を 包含하기로 決定을 하였을때 實際로 어떤 製品이 調査되어야 하는지 明白하게 定義하는 것이 必要하다. 例를 들면 어떤 期間에 값이 싼 250 g의 차의 값을 記錄하고, 또 어떤 期間에는 質이 좋은 차 500g의 값을 記錄하는 경우에는 指數에 차의 品目を 包含하는 것이 有用하지 않으며, 단지 實際 價格變化가 完全히 同一한 品目の 價格을 記錄하고 있다는 點을 明白히 하기 위해서 每期의 價格資料를 蒐集하여야 한다. 그러므로 正確한 品目分料가 必要하다. 즉 商標, 品質, 크기단위 및 포장형태등이 完全하게 分類되어야 하며, 可能한 한 每期 指數作成에 있어서 同一品目이 調査되어야 한다. 또 價格을 調査하기 위해서 商店 및 市場等을 訪問하며, 實際的인 實查業務를 行하고 있는 職員들이 變更되고 있기 때문에, 指數作成과 關聯된 書類가 매우 仔細하게 함으로써, 새로 任命된 實查要員이 올바르게 모든 過程를 理解할 수 있고, 모든 경우에 分類된 製品을 完全히 確認할 수 있도록 明確히 하여야 한다.

各 品目이 注意깊게 定義되어 있지 않으면, 차, 설탕, 쌀, 담배 등과 같이 아주 쉽고, 간단한 경우에도 신중한 고려를 기울이지

않는다면 다른 品目보다는 複雜하고, 不正確한 調査를 낳게 되는  
보이지 않는 많은 問題點들이 있는데, 빵이라는 品目은 자주 잘  
못 調査될 수 있는 製品의 例이다. 즉 一部 國家에서 빵 製造  
業者는 아주 오랫동안 빵 한개의 價格을 一定하게 維持하는 것  
이 常例이나 製造費用의 變化를 보상하기 위하여 實際로는 빵의  
크기를 變更시키고 있다. 그러므로 價格蒐集時에 빵의 무게를 測  
定하는게 必要하다. 왜냐하면 어떤 時點에서 顧客이 250 센트로  
빵 500g을 購入할 수 있었으나, 다른 時點에서는 빵 460g을  
250 센트로 購入할 수 있다면 實際로는 보이지 않는 價格引上이  
發生한 것이므로 指數에 反映되어야 한다.

#### (d) 質的·變化

品目分類에서 가장 重要하고, 複雜한 問題點은 천이나 편  
물등과 같은 品目들에 關해서 經驗을 가져야 한다는 것이다. 그  
理由는 商品種類의 多樣성과 生産된 商品의 質이 多樣할 뿐만  
아니라, 商店에 在庫로 있는 製品들이 유행과 기호에 따라서 계  
속적으로 變化되고 있기 때문에 간혹 어떤 期間에 있어서는 價  
格比較가 不可能한 경우가 발생할 수 있기 때문이다. 이런 問題  
點들을 解決하기 위한 方法으로는 統計當局이 천이나 편물등에  
대한 식별전문가를 고용하는 경우등이 있다.

指數에 列挙된 製品이 더 이상 利用할 수 없는 것으로 밝혀  
졌을때 專門家들은 相異製品에 대하여 적정가격을 推計하고, 質의  
差異를 고려하기 위해서 그들의 知識을 利用할 것이다. 물론 모  
든 國家가 이렇게 정교한 技法을 利用할 수는 없으며, 實際에 있어서

發生하고 있는 質的 變化를 만족할 程度로 고려한다는 것은 극히 어려운 것으로 생각된다. 또한 一定 期間동안 商品價格을 正確하게 調査할 수 없다면 그러한 類의 品目은 指數作成에서 除外시켜야 할 것이다. 이런 조치는 消費者들의 支出慣習의 代表性을 저하 시킬수도 있으나, 잘못 測定된 價格變動이 계속적으로 指數에 記載될 수 있는 不正確한 價格測定을 代替하는 것이 바람직하다.

그리고 物價指數의 신뢰성은 어떤 다른 要素보다도 價格測定의 正確性에 依存하고 있으므로, 實查要員이 調査한 價格이 列挙된 品目に 正確하게 關聯되도록 하는 것이 중요하다.

#### (e) 消費者 物價指數에서 除外된 品目

一部 品目들은 消費者物價指數의 範圍밖에 있다고 생각되거나, 價格測定 및 量的 單位를 定義하기 어렵기 때문에 指數에서 完全히 除外되고 있는데, 이런 品目들에는 直接稅, 國家基金 또는 罰金支払, 貯蓄, 貨幣贈與, 노름비, 第2次製品, 藝術品, 教會 및 其他 寄贈品과 住宅이 建築되고 있는 垡地도 大部分의 國家에서 除外되고 있으며, 住宅購入 및 建築費用을 除外하고 있는데, 除外理由는 이런 財貨購入을 消費支出이 아닌 資本支出로 간주하고 있기 때문이다.

#### (f) 調査對象處

各 品目を 完全하게 分類한 다음에 價格이 蒐集되어야 할 곳(市場, 商店)을 決定하는 것이 必要하고, 一般的으로 指數는

特別地域 및 都市에 관한 것이므로 實際적으로 지역내 小売商이 調査對象이 되어야만 한다. 또 選定된 小売商들은 可能한 모든 商店들을 代表할 수 있어야 한다. 즉 어떤 品目の 大部分이 수퍼마켓에서 購入되고, 구멍가게에서는 아주 小量이 購入되고 있다면 指數에 使用할 價格은 수퍼마켓의 價格을 利用하는 것이 必要하다.

### 11.3.2 消費者物價指數의 管理

消費者物價指數中에서 하나의 重要한 問題는 消費者物價指數가 消費者가 購入하는 品目の 價格變化를 反映하도록 最近의 指數로 維持하여야 한다는 점이다. 그러므로 消費者物價指數의 主要 用度를 列挙하고 이에 關하여 論議코자 한다.

- (i) 經濟狀況의 分析 및 評價 즉 인플레이率決定, 經濟成長問題 處理, 需給狀態觀察
- (ii) 價格變化가 雇傭人과 消費者에게 미치는 影響의 差異의 測定 즉 生活水準 및 購買力測定
- (iii) 賃金決定: 消費者의 購買力을 維持하도록 價格指數의 上昇에 따라 俸給 및 賃金を 引上한다. 이를 賃金 Indexation 이라 한다.
- (iv) 物價디플레이터計算( 이 章의 後部에서 說明 )

價格變化가 사람들의 生活形態에 미치는 效果에는 特別한 關係가 있다는 것을 위의 說明에서도 찾아볼 수 있다. 그러므로 定義하기가 어렵고, 곤란하다 할지라도 消費者物價指數를 生計費變化

로 간주하고자 할것이다. 그러나 不幸하게도 特定品目の 財貨와 用役의 價格變化만을 測定하기 때문에, 消費者 物價指數를 生計費로 간주하는 것은 實際적으로는 난해하고, 거의 不可能하다.

基準時点에서 支出形態의 代表性을 나타 내도록 하여야 하나, 消費者들의 消費習慣의 變化가 격심하기 때문에 消費者物價指數에 包含되고 있는 品目들은 점차로 現 支出形態의 代表性을 줄어들게 한다. 이런 理由로 生計費의 概念을 反映한다고 볼 수 없다. 이전에 言及한 두가지의 指數, 即 消費者物價指數와 小売物價指數는 實際로 測定되는 것을 나타내는 正確한 表現이며 實際로도 同一하다. 아마도 消費者物價指數라는 用語가 훨씬 좋은 用語이고, 보다 더 폭넓게 使用하고 있으나, 根本적으로 消費者들은 그들의 所得을 小売水準에서 消費하고 있으므로 小売物價指數라는 用語로 代替할 수 있다.

또 하나의 重要點이 이 段階에서 紹介되는 것이 필요하다는 點이다. 지금까지 基準加重指數만이 論議되었으나, 잠시 후 指數公式에 關한 節에서 보는 바와같이 理論적으로 現時點의 加重値를 利用하여 指數를 計算하거나, 基準時點과 現時點의 消費形態에 따라 加重化된 어떤 平均値를 挾하여 指數를 計算할 수도 있다. 한편 이것이 理論上으로는 妥當하고 實際적으로도 어떤 다른 指數算定에 利用되고 있으나, 消費者物價指數作成에 利用하기는 곤란하다. 왜냐하면 이런 指數는 每 指數의 作成時點에서 計算되어야 할 加重値를 必要로 하고, 또 加重値를 正確히 計算하기 위해선 每 分期마다 家計支出調査를 行할 必要가 있으나, 財政上

또는 人力上의 問題때문에 一部 國家에서만 行하여 지고 있다.

(a) 指數의 連繫方法

主要 家計支出調査는 每 5 年마다 實施하고, 그 調査結果에 따라서 調査對象의 品目範圍와 加重值를 最近值로 바꾸는 것이 一般的이다. 또한 그 調査가 實施된 時점이 새로운 基準時点으로 選定되고 그 時点を 基準으로 指數를 100으로 바꾸게 된다. 이 경우에 舊指數와 新指數를 連繫하는 方法이 必要하게 되는데, 이것은 連繫法에 의하여 遂行된다. 그러므로 舊指數가 新指數와 同等한 값으로 바뀌어진다. 그 例를 보면 아래 표와 같다.

| 分 期        | 舊 指 數                       | 新 指 數                       | 連 繫 指 數                     |
|------------|-----------------------------|-----------------------------|-----------------------------|
|            | ( 기준 : 4/4 분기 )<br>1974:100 | ( 기준 : 4/4 분기 )<br>1975:100 | ( 기준 : 4/4 분기 )<br>1979:100 |
| 1979 1 / 4 | 139.6                       |                             | 96.7                        |
| " 2 / 4    | 140.1                       |                             | 97.2                        |
| " 3 / 4    | 142.7                       |                             | 99.10                       |
| " 4 / 4    | 144.1                       | 100.0                       | 100.0                       |
| 1980 1 / 4 |                             | 102.6                       | 102.6                       |
| " 2 / 4    |                             | 102.4                       | 102.4                       |

위 표는 各各의 구지수에 100.0/144.1을 곱하여 新 基準에 合당한 指數로 전환한 것이다.

一部 國家에서는 빈번하게 加重形態를 最近의 것으로 바꾸기

위하여 消費者物價指數에 充分한 資源을 配分하고 있으나, 그런 것은 部分的으로 자주 實施되는 家計支出調査 및 消費形態의 變化를 確認하기 위하여 輸入, 生産 및 小売業과 같은 統計資料를 利用함으로써 達成할 수 있으나, 빈번한 變便은 統計利用者들에게 혼란을 야기시키며, 또한 舊기준 시점에 尙하도록 指數變便을 위한 技法이 使用되어야 하기 때문에 每期마다 새로운 基準時点으로 變更하는 것은 바람직 하지 못하다. 이것은 一定 期間以上의 어떤 추세를 보다더 明白히 理解하게 하는데, 이런 指數를 連鎖(Chain-linked) 指數라 한다.

(b) 基準變更

다른 指數와 比較하고, 資料分析에 도움을 줄 수 있도록 基準을 變更하여야 하는데, 그 計算方法은 基準이 變更된 時点的 指數로 前 時点的 指數를 나누어서 100을 곱하면 되고, 또 이 計算法은 連鎖(Splicing) 指數로 만들때 使用한 方法과 同一하나, 다만, 어떤 二重加重이나 指數에 對한 變更이 없이 새로운 基準時点이 使用되었다는 점이 다르다.

그러면 1979年 4/4分期 基準의 新·舊지수를 連鎖했던 前의 例를 利用하여 基準變更에 關하여 說明해 보기로 하자 이제 基準時点이 1980年の 2/4分期로 變更되었다면, 1980年の 2/4分期의 指數는 102.4에서 100으로 變更되어야 하고, 모든 다른 指數들도 同一比率에 의해서 調整되어야 한다. 그러므로 1979年 1/4分期의 指數는  $96.9 \times \frac{100.0}{102.4} = 94.6$ 으로 變更된다.



(c) 品目代替 및 帰属法

消費者物価指数中에서 重要な것 중의 하나가 品目代替인데 간혹 最初의 品目範圍에 包含되 있는 어떤 品目들은 在庫가 없거나 어떤 다른 包装( 例를들면, 다른 크기의 통 및 상자 ) 을 使用하고 있기 때문에 同一 製品에 대한 直接的인 価格比較가 不可能한 경우가 생겨난다. 위와같은 경우의 發生時 가장 一般的인 対処方法은 現在 利用可能한 類似品目으로, 品目範圍内에 있는 品目 및 品目を 代替하는 것과 새로운 製品價格을 基準時点의 價格에 帰属하게 하는 것이다.

價格의 帰属은 여러가지 方法으로 行할 수 있으나, 여기서는 두가지 方法을 보고자 한다. 첫째 方法은 指數의 作成에 利用한 品目は 完全히 同一하나, 容器의 크기가 變化하거나, 製品이 새로운 基準에 의하여 販賣되므로, 量의 單位가 다를 경우 이 方法이 利用된다. 이 경우에 基準時点의 價格이 關聯된 크기나 量에 根拠하여 간단하게 帰属이 行하여 진다. 따라서 基準時点에서 쌀값이 파운드당 13 Cent 이고, 價格變化가 없다고 가정하면, 미터單位로 轉換後의 價格은  $13 \times 2.205 = 28.7 \text{ Cent / Kg}$  이다. 以後의 指數計算에 있어서 쌀의 個別價格指數는 28.7 Cent 로 나눔으로써 얻을 수 있다.

둘째, 값을 評價하여야 할 品目に 어떤 根本的인 變化가 있거나, 同一 品目이더라도 基準時点과 現時点間의 直接的인 價格比較가 不可能할 정도로 製品의 質과 特性에 重要な 變化가 있을 때 이 方法이 利用된다. 볶은 콩과 스파게티의 경우를 利用하여

볶은 콩은 基準時점에 32 Cent 이었으나, 볶은 콩을 購入할 수 없게 되어 指數의 代替가 必要하기 直前の 價格이 45 Cent 였다고 假定하자. 그러면 볶은 콩의 最終在庫時 새로운 品目인 스와게티의 價格을 把握하는 것이 必要하며, 그렇게 함으로써 두 品目の 價格이 直接的으로 同時に 比較될 수 있다. 즉 스와게티의 價格이 56 Cent 라면 스와게티 價格으로 轉換한 基準時点的 價格은  $56 \times \frac{32}{45} = 39.8$  Cent 가 된다.

實際에 있어서는 이 方法이 同時に 두 製品에 대한 比較價格을 얻기가 可能的한 것은 아니다. 例를들면 볶은 콩의 價格이 1979年 5월에 45 Cent 이고, 스와게티의 價格이 1979年 8월에 56 Cent 라면, 이 경우 適切한 相應 價格을 얻을 수 없기 때문에 어떤 共通 價格을 얻도록 最善의 努力을 하여야 한다. 그러나 이것이 不可能한 것으로 判斷될 경우 共通 價格을 얻은 것 처럼 計算할 수 있고, 또 이로 인한 不正確度는 價格調査 時점에 큰 差異가 없다면 容認될 수 있다. 品目代替의 過程이 國家마다 相異하나, 그 根本的인 原理는 同一하다.

指數를 管理하기 위한 以外の 節次的인 問題가 대두되고 있는 바 例를들면 2個以上の 品目に 二重加重을 한다거나, 새로운 品目を 追加시킬때 新品目에 對한 加重值를 부여하고 對象에 이미 包含되어 있는 品目에 對한 加重值는 除去해 버리는 경우이다.

그러나 이 冊에서는 이런 問題의 解決方案이 論議되지 않으므로, 物價指數의 作成에 關한 指針書에서 이에 對한 對策을 收錄하여야 할 것이다.

#### 11.4 其他 物価指数

또한 消費者物価指数外에 여러 國家에서 많은 價格指數를 作成하고 있는데, 이런 指數中 가장 重要한 것은 經濟企劃家들이 價格變化에 依한 全體의 輸出額變化 및 實際的인 輸出量의 變化는 얼마 程度인지를 測定하는 手段인 輸出價格指數이며, 또한 貿易統計는 通常的으로 每期當의 商品輸出額 및 量에 關한 明細를 包含하고 있기 때문에, 現時點에서 加重値를 안다는 것은 不可能하므로 輸出(輸入) 價格指數의 경우에는 消費者物価指數와 같이 基準加重體系가 利用되지 않는다.

또 다른 重要 經濟指數는 消費者物価指數보다 広範圍한 品目을 包含하고 있는 都売物価指數인데 간혹 價格上昇이 最初의 都売段階에서 發生하여 消費段階의 消費者에게 미치게되므로 都売物価指數는 消費者物価指數보다도 經濟內的 變化狀態를 迅速히 알리 주기 때문에 經濟企劃에 있어서 特히 重要하다.

또한 物価指數는 經濟分析 特히 國民計定과 關聯하여 利用되는 디플레이터(Deflator)로 使用되나 이 教材에서는 國民計定統計에 關해서는 간단히 論述하고자 한다. 國民計定은 주어진 기간(1년)내에 國家의 總生産量 即 GDP라는 用語를 使用하여 國家의 全財質와 用役의 價值를 提供한다. 이를 基礎로 하여 經濟企劃家와 分析家들은 總生産에서 純生産을 분리하고자 하기 때문에 價格變化의 效果를 除去하는 것이 必要하게 되는데 이를위한 가장 一般的인 方法은 不變價格으로 GDP를 算出하는 것이다. 다시 말해서 現 年度의 總生産額이 採択된 基準時의 價格에 따라 價值를 評價할 수 있도록 計算하는 것이다.

또 이것은 期間中에 發生한 價格變化에 따라서 줄이거나 디플레이팅(Deflating)함으로써 現 價格을 推計할 수도 있으며, 이런 目的을 위하여 消費者物價指數가 利用될 수도 있으나 現實적으로 滿足스럽지 못하다.

그러므로 우리는 國家에서 모든 財貨 및 用役に 대한 物價推移를 測定하는 指數를 가지고자 하며, 디플레이터도 消費者物價指數보다 더 많은 品目들이 包含되어 있는 指數에 基礎해야 할 것이다. 實際에 있어서는 이러한 目的을 위해서 여러가지 物價指數가 必要하게 되므로 特別한 指數가 作成되거나 都壳物價指數와 같은 보다 폭 넓은 基礎指數가 使用된다. 여기에서는 國民計定에서 디플레이터使用에 관한 것만이 論議되었나, 實際에 있어서는 이와 같은 同一한 方法이 어떤 값에 대한 統計時系列을 위해서도 使用될 수 있고, 그 目的은 정상가격을 불변가격으로 전환시키는 것인데 이렇게 함으로써 實際적인 趨勢를 나타내 주게 된다.

#### 11.5 數量指數

數量指數는 한 期間과 다른 期間사이의 生産 및 消費 등의 實際量에 대한 相對적인 變化를 測定하며 그 原理는 物價指數에 관한 說明과 비슷하나 數量指數에 關하여 간단히 說明하면 아래와 같다.

한 商品에 대한 簡單한 數量指數는 現時點과 基準時點間의 量에 대한 比率을 取함으로써 計算되는데 이러한 比率을 量的比率

이라 한다. 그 방법은 한 상품에 대한 價格指數의 計算方法과 同一하다.

몇개의 品目에 대한 數量指數의 作成에 있어서 噸, 세계공메타 個數等의 單位들로 測定된 品目들을 어떻게 취합하느냐에 대한 問題가 생기게 되며 여기에서 해야할일은 數量에 價格을 곱하여 貨幣價值로 轉換시키는 일이다.

또 基準加重化된 數量指數를 求하기 위해서 基準時點의 價格에 調査된 量을 곱한다. 이에 관한 簡單한 例를 보면 아래와 같다.

| 品目單位      | 單位當<br>基準時<br>價格 | 量      |        |        | 量 × 基準價格 |          |          |
|-----------|------------------|--------|--------|--------|----------|----------|----------|
|           |                  | 1977   | 1978   | 1979   | 1977     | 1978     | 1979     |
|           | \$               |        |        |        | \$ 1,000 | \$ 1,000 | \$ 1,000 |
| 쌀 Ton     | 287.50           | 9,285  | 10,600 | 14,190 | 2,664.8  | 3,042.2  | 4,072.5  |
| 밀가루 Ton   | 234.50           | 5,167  | 6,426  | 6,705  | 1,209.1  | 1,503.7  | 1,569.0  |
| 우유 1,000ℓ | 187.50           | 3,240  | 3,316  | 3,095  | 607.5    | 621.8    | 580.3    |
| 차 Carton  | 41.25            | 11,219 | 13,460 | 16,118 | 462.8    | 555.2    | 664.9    |
| 계         |                  |        |        |        | 4,944.2  | 5,722.9  | 6,886.7  |

基準年度(1977年)의 指數는 100이다. 즉  $\frac{4,944.2}{4,944.2} \times 100 = 100.0$ , 1978年의 指數는  $\frac{5,722.9}{4,944.2} \times 100 = 115.7$  이고 1979年의 指數는  $\frac{6,886.7}{4,944.2} \times 100 = 139.3$  이다. 物價指數와 마찬가지로 數量指數는 基準時加重指數 또는 現時點에 근거한 加重指數中

의 하나로 可能하나, 後者의 경우에는 基準時点의 價格 代身에 現時点의 價格이 計算에 使用된다. 基準時와 現時点에 根拠한 數量指數의 平均값을 求해야 하는데 이에 關한 公式은 다음의 節에 說明된다.

### 11.6 金額指數

金額指數는 매우 簡單하고 한가지 뿐이다. 基本的으로 金額指數는 基準時点에서 어떤 品目들의 價値에 대한 한 時点의 總價値를 말한다.

따라서 金額指數는 期間中の 價格變化, 數量變化 및 兩者의 結合에 起因한 指數의 變化를 나타내는 것이 아니라 全體的인 變化를 나타내며, 事實上 金額指數는 物量指數와 物價指數를 作成하는 하나의 方法이며 이것은 주로 보다 進展된 指數作成 業務에 利用된다. 特히 現時点에 근거한 加重物量指數로 부터 基準物價指數를 作成하기 위해 利用된다.

### 11.7 指數의 公式

指數公式이 얼핏보기에 좀 어려울것 같아 지금까지 提示하지 않았으나 이 章에 提示된 指數公式은 實질 簡單하고, 이미 說明한 것을 簡單한 數式으로 表現한 것에 지나지 않는다.

公式을 說明함에 있어서 使用될 符號는 다음과 같다. 價格은 P로, 數量은 q로 表示되며, 指數에 關聯된 期間과 品目は 添字로 表示된다. 即 基準時点是 添字 "0"로 現時点是 添字 "n"

으로, 特定 品目は " i "로 表示할 것이다. 그리고 品目範圍內에 있는 特定品目を 考慮하여 基準時價格은  $P_{0i}$  및 基準時數量은  $q_{0i}$ 로 表示한다. 그러므로 基準時點에서의 이 品目에 대한 支出은  $P_{0i} q_{0i}$ 이며, 모든 品目에 대한 基準期間의 全體支出은  $\sum P_{0i} q_{0i}$ 로 表示된다.

또한 各種指數가 一般的으로 特別한 名稱으로 불리워지고 있다는 것을 알게 될 것이다. 즉 基準時點에 根拠한 加重指數는 Laspeyres 指數로, 現時點에 根拠한 加重指數는 Paashe 指數로, 또 Laspeyres 指數와 Paashe 指數를 幾何 平均한 것을 Fisher 指數로 부르고 있다. 이런 用語들이 매우 広範하게 使用되고 있기 때문에 本章의 說明中에 가끔 이런 用語들을 使用하게 될 것이다.

Laspeyres 指數 (가장 잘 알려져 있는 消費者物價指數)는 아래와 같다.

$$P_L = \frac{\sum P_{ni} q_{0i}}{\sum P_{0i} q_{0i}} \times 100$$

위의 公式은 다음과 같이 說明될 수 있다. 各 品目の 基準數量에 現時點의 價格을 곱하여 合計한 다음 100을 곱하여 그 값을 基準時의 總金額으로 나눈다. 一部 國家에서는 Laspeyres 指數로 變更하여 使用하고 있는데, 이는 現時點과 그 以前時點에 대한 價格比를 算出하는데 利用된다. 公式은 다음과 같다.

$$P_L = \frac{\sum \left( \frac{P_{ni}}{P_{(n-1)i}} \times P_{(n-1)i} q_{0i} \right)}{\sum P_{0i} q_{0i}} \times 100$$

이를 변형시켜 소위 “加重平均 價格比較” 方法으로 알려진 Laspeyres 指數를 나타내주는 一般的으로 利用되는 方法이 있는데 이 경우의 公式을 다음과 같이 簡單하게 된다.

$$P_L = \sum W_i \frac{P_{ni}}{P_{oi}}$$

이 算式에서  $W$ 는 퍼센트(%)로 表示된 加重值 또는 相對的 重要度를 나타내며, 따라서 “ $W_i$ ”는  $i$ 번째 品目에 割當된 加重值이다.

지금까지 添字 “ $i$ ”는 어떤 特定品目を 表示하기 위해서 公式에 利用되고 있다. 이것을 쓰는것은 理論적으로 正當하나 公式은 屢번히 添字를 생략하고 사용되고 있는데 이는 理解를 明確하고 容易하게 하므로 다음 公式부터는 添字 “ $i$ ”를 省略하고자 한다.

Laspeyres 와 Paashe 指數公式을 表記하면는 아래와 같이 나타낼 수 있다.

Laspeyres의 價格指數

$$P_L = \frac{\sum P_n q_o}{\sum P_o q_o} \times 100 \quad \text{또는} \quad \sum W \frac{P_n}{P_o}$$

Laspeyres의 數量指數

$$Q_L = \frac{\sum P_o q_n}{\sum P_o q_o} \times 100$$

Paashe의 價格指數

$$P_P = \frac{\sum P_n q_n}{\sum P_o q_n} \times 100$$



## Paashe 의 数量指數

$$Q_P = \frac{\sum P_n q_n}{\sum P_n q_n} \times 100$$

大文字 "P" 와 "Q" 는 價格 및 數量指數를 나타내고, 添字 "L" 과, "P" 는 各各 Laspeyres 및 paashe 指數식을 나타낸다.

나머지 두지수를 幾何平均하여 얻어진 指數 즉 Fisher 指數에 關於하여 以前에 언급한바 있는데 或者는 이보다 더 複雜한 指數를 編制하는지에 關於해서 疑念을 품을 수 있다. 그러나 Laspeyres 指數와 Paashe 指數가 모두 欠點을 內包하고 있으므로 다른 指數를 作成하여야 할 充分한 理由가 있다. 卽 基準時의 消費形態가 變하지 않은 것으로 가정하고 있기 때문에 基準加重價格指數는 發生하는 價格變化를 過大評價하는 傾向이 있다. 實際上에 있어서도 一般的인 現象은 價格上昇이 아주 적거나 거의 없는 경우의 商品을 더욱더 많이 購入하고, 價格上昇이 매우 큰 商品은 거의 購入되지 않은 傾向이 있다. 卽 어떤 要因 때문에 돼지고기의 價格은 크게 上昇하고, 닭고기의 價格은 不變이라면 많은 사람들은 닭고기를 더 많이 그리고 돼지고기를 더적게 購入함으로써 消費패턴에 대한 全般的인 效果는 크게 變하지 않을 것이다. 그러나 社會의 消費패턴이 價格變化의 效果를 신속하게 反映시킬 수 있는한 Laspeyres 指數는 以前의 消費패턴에 따라 繼續 作成될 수 있다. 위의 例에서 보다높은 돼지고기의 價格上昇을 指數水準을 增加하게 하는 實際的인 要因이 될 것이나 消費習慣의 갑작스런

變化가 모든 時点에서의 價格變化를 誘發시킬 것으로는 생각되지 않는다. 基準加重指數가 價格變化의 效果를 過大評價할 수 있다는 事實은 一般的으로 認定되고 있다. Paashe 價格指數는 現時点的 消費패턴을 根拠로하고 있기 때문에 實際의 價格變化를 過小評價하는 傾向이 있는데 이런 問題点들은 數量指數에서도 나타난다.

그러므로 統計學者들은 이런 問題点를 解決하기 위해서 複雜한 形態의 指數를 開發하게 된다. 基準時点에 근거한 指數가 價格變化를 過大評價하고 現時点에 根拠한 價格指數가 價格變化를 過小評價한다면 明確한 解決策은 이들 두 指數에 대하여 어떤 平均을 利用하는 것이다. 第10章에서 論議한 바와같이 變化率에 適用할 수 있는 最善의 平均方法은 Fisher 指數式과 같이 幾何平均법을 想起하여 주길 바라며, 公式은 아래와 같이 表記된다.

Fisher의 價格指數

$$P_F = \sqrt{\frac{\sum P_n q_o}{\sum P_o q_o} \times \frac{\sum P_n q_n}{\sum P_o q_n}} \times 100$$

Fisher의 數量指數

$$Q_F = \sqrt{\frac{\sum P_n q_n}{\sum P_o q_o} \times \frac{\sum P_n q_n}{\sum P_n q_n}} \times 100$$

結局 金額指數 (V) 에 대한 公式은  $\frac{\sum P_n q_n}{\sum P_o q_o} \times 100$  으로 簡單하며 이것은 實際적으로 Fisher의 價格指數 및 數量指數와 同一하다.

또 이 章을 通하여 指數는 基準時的 100 과 關聯되어 表現되어 있다는 것을 알았을 것이나, 實際에 있어서는 이것이 大部分 國家에서 使用되고 있는것은 단지 傳統일뿐이며, 指數는 變化率을

測定하는 것이기 때문에 基準値가 100 일 必要는 없고 어떤 경우에는 1000 또는 1이 使用될 수도 있다. 위와같이 1,000 또는 1이 基準値로 使用된다면 誤解를 払拭할 수 있도록 明確한 說明이 주어져야 한다.

### 11.8 指数公表 및 解釈

前節에서 指数는 品目이나 品目集團의 全般的인 變化를 測定하기 위한 것임을 보았다. 따라서 平均增加率을 計算할 때에는 算術平均이 아닌 幾何平均이 利用된다는 點을 周知하였을 것이다.

이제 우리는 이에 對해 좀더 研究해보고 指數의 解析上 뒤따르는 몇가지의 簡單한 原則을 討議해야 할 것이다.

記述되어야 할 概念은 事實上 單純하나 흔히 誤用되고 있음을 놀라운 사실이 아닐 수 없다.

어떤 時点에서 다른 時点으로 指數가 增加하게 되면 同期間 동안의 增加值를 當初의 指數로 나누고 이에 100을 곱하여 算出할 수 있다. 이 章에서는 이미 다섯 品目に 關하여 作成한바 있는 消費者物價指數를 利用하여 分期의 增加率을 計算하면 다음과 같다.

| 期 間    | 指 数   | 增 加 率                                |
|--------|-------|--------------------------------------|
| 1979.2 | 100.0 |                                      |
| 1979.5 | 102.4 | $\frac{2.4}{100.0} \times 100 = 2.4$ |

| 期 間     | 指 数   | 增 加 率                                  |
|---------|-------|--|
| 1979.8  | 108.0 | $\frac{5.6}{102.4} \times 100 = 5.5$   |
| 1979.11 | 110.9 | $\frac{2.9}{108.0} \times 100 = 2.7$   |
| 1980.2  | 123.9 | $\frac{13.0}{110.9} \times 100 = 11.7$ |

위 表는 1979. 2月부터 1980. 2月까지 1年 동안의 指數는 100에서 123.9로 卽 23.9%가 增加되었음을 나타내고, 이런 增加率は 分期增加率의 合計와는 一致되지 않는다. 卽 分期增加率  $2.4 + 5.5 + 2.7 + 11.7 = 22.3$ 은 年間增加率과 一致되지 않는다. 또한 變化率을 計算할때에 分期의 增加率은 다음과 같은 法으로 年間値를 計算할 수 있다. 卽 年末 現在의  $100.0 \times 1.024 \times 1.055 \times 1.027 \times 1.117 = 123.9$ 이고, 年間增加는 23.9%이며, 이 값은 이미 위에서 計算된 값이다.

또 이것은 가끔 指數增加가 퍼어센트나 "index points"로 나타낼 수 있음을 알게 됨으로써 좀더 다른 觀點에서 이를 檢討해 볼수 있을 것이다. 따라서 위의 경우에 1979年 11月과 1980年 2月사이에 13.0%가 增加 되었다고 말할 수 있는데, 이런것은 同一한 資料를 두가지 方法으로 說明할 수 있음을 意味한다. 그리고 基準時点以後 첫번째 기간중의 指數포인트(Points) 增加는 퍼어센트(%)로 表示한 增加와 同一하나, 期間이 經過함에

따라서 指數水準이 보다 높게 움직이는 것처럼, 指數포인트 (Point)로 測定된 增加值는 퍼센트로 表示된 增加率보다 점점 크게 되므로 指數變化가 "포인트"로 引用될때 指數利用者들은 指數增加가 퍼센트값으로 나타냈을때 그 增加가 더욱더 큰 것으로 믿게된다. 또 利用者들의 혼란을 避하기 위해서 統計刊行物の 發刊에 있어서 指數에 關한 資料를 포인트(Points)로 測定된 測定値가 아닌 퍼센트값으로 發表하는 것이 바람직하다는 意見이 있다.

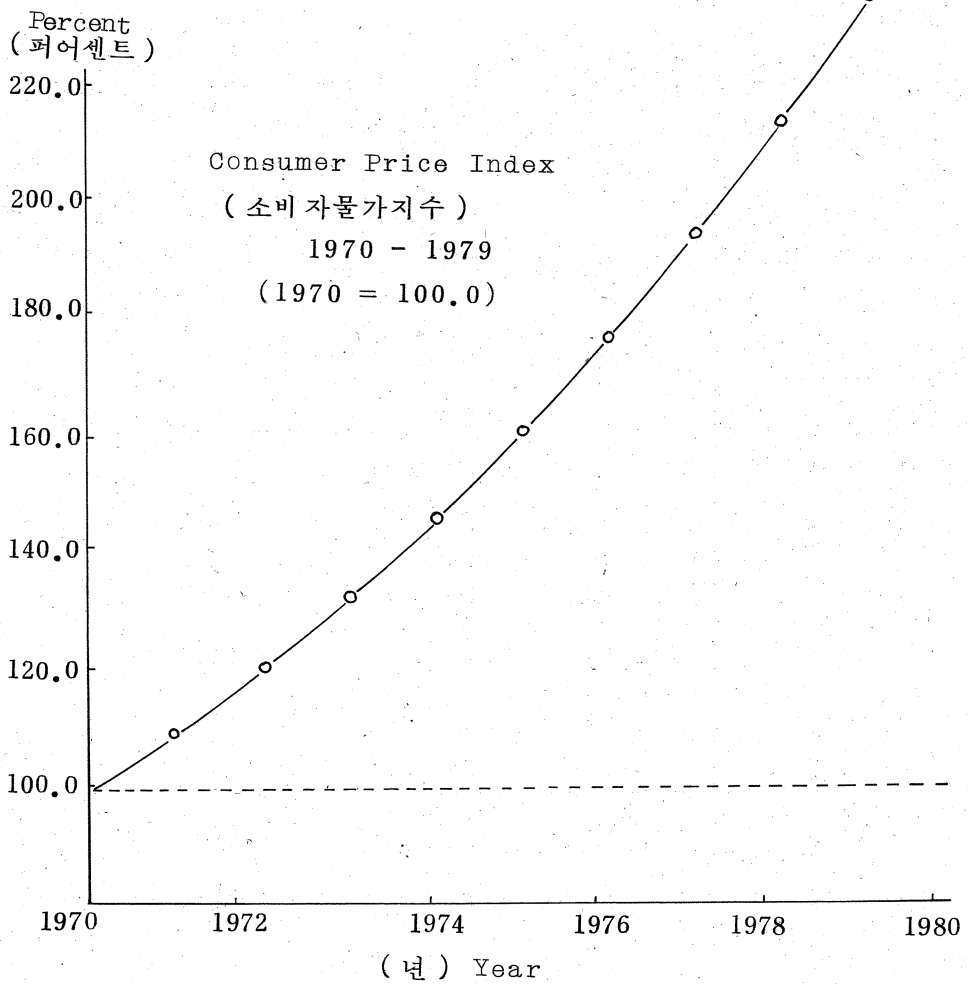
또 이런 現象은 指數資料가 그림표 形態로 表現되었을때 아마도 前述한 說明은 더욱 明白해 질 것이다.

年間인플레이率이 10% (即 前年度의 消費者物價水準에 比하여 現年度의 消費者物價指數가 10% 增加)인 한 國家를 簡單한 例로 들어 說明해 보기로 하자.

年度別 指數는 아래와 같다.

| 基 準 時 | 100.0                       |
|-------|-----------------------------|
| 1 年 後 | $100.0 \times 1.10 = 110.0$ |
| 2 年 後 | $110.0 \times 1.10 = 121.0$ |
| 3 年 後 | $121.0 \times 1.10 = 133.1$ |
| 4 年 後 | $133.1 \times 1.10 = 146.4$ |
| 5 年 後 | $146.4 \times 1.10 = 161.1$ |
| 6 年 後 | $161.1 \times 1.10 = 177.2$ |
| 7 年 後 | $177.2 \times 1.10 = 194.9$ |
| 8 年 後 | $194.9 \times 1.10 = 214.4$ |
| 9 年 後 | $214.4 \times 1.10 = 235.8$ |

이 결과를 그래프에 옮긴다면 右向의 直線이 아니라 年度가 經過함에 따라서 急傾斜의 曲線으로 나타난다는 것을 보게 될 것이다. 이것은 Oy 軸에는 指數의 實質水準( 퍼어센트가 아닌 指數 포인트의 增加 )을 記錄함으로써, 그래프의 原点으로 부터의 수직 거리는 指數水準이 높게 移動하는 것과 마찬가지로 같은 比率로 增加하기 때문에 보다 크게 나타난다. 위의 資料에 依한 그래프는 아래와 같다.



이와같이 繼續增加하고 있는 傾斜를 가진 그래프는 계속 增加하는 인플레이션을 보이고 있는 것으로 判斷되며 이럴경우 그 判斷이 잘못되었다 하더라도 그리 놀라운 일은 아니다.

이런 것은 分明히 資料의 表現이 잘못 된것이 아니라 解釋上의 잘못때문이므로, 잘못된 解釋의 可能性을 避하기 위해서 다른 方法으로 資料를 그래프에 표시하는 것이 必要하게 된다.

結論적으로 이런 問題點을 解決하기 위한 最善의 方法은, 第9章에서 言及한 바와 같이 指數를  $O_x$  軸에는 定常值로  $O_y$  軸은 代數值로 表示된 半代數그래프용지에 表示함으로써 原點으로부터  $O_y$  軸上에서 100과 110間的 거리가 200과 220 또는 300과 330사이의 거리와 同一한 것 처럼, 結果적으로 同一한 比率의 變化를 나타낸다. 이제 前의 그래프를 半代數表用紙에 다시 表示한다면 右上向 直線을 얻게 되며, 이것은 一定한 인플레이션 率을 그림표로 나타낼 수 있는 最善의 方法이다. 一般的으로 指數를 그래프에 表示할때에 絶對값 보다는 퍼센트로, 通常의 그래프 用紙보다는 半代數用紙에 表示하는 것이 훨씬더 合理的이다.

## 第 12 章 統計制度와 基準

### 12.1 概 要

이 教育教材를 通하여 研究한 바와같이 지금까지 統計蒐集, 資料處理, 資料公表 및 分析에 關聯된 여러가지 問題들을 論議하였다. 마지막 章에서는 效率적인 統計業務를 위하여 마련될 統計制度의 形態에 關하여 概觀한 다음에 統計의 質과 比較性を 維持하기 위해서 必要한 統計基準에 對하여 說明하고자 한다. 위의 問題를 論議함과 同時에 特히 標準定義, 分類體系 및 標準方法等에 關하여서도 言及하고자 한다. 또 이章에서는 實查要員 및 地方事務所職員이 業務와 關聯하여 直面할 수 있는 主要定義의 一部를 提示하고, 이들 實務적인 側面에서 論議하며, 아울러 主要한 國際分類體系의 一部를 說明하고자 한다.

마지막으로 一般的인 用語로 統計體系가 資料를 提供할 것으로 期待되는 資料의 範圍에 對하여 考察할 것이다.

### 12.2 統計體系

지금까지 이 教材에서는 한 機關이 한 國家에서 作成되는 統計에 對하여 全적인 責任을 맡고 있는 것처럼 政府統計要員 및 統計機關에 關하여 論議하였는데, 이것은 이 教材에서 說明해야 할 概念을 보다 쉽게 나타내도록 하는 簡單한 接近方法이나, 現實적으로 統計體系는 훨씬 더 複雜하다.

여러 政府部처나 統計作成 機關들은 統計의 一部業務에 參與하고



있으며 이를 위해 效率的으로 活動하고 있는데 이는 統計業務에 參與하고 있는 모든 相異機關들의 業務가 調整되고, 단일 體制로 統合되는 것이 必要하다.

만약 統計調整 및 統合이 遂行되지 않는다면 한 部処의 業務가 他 部処의 業務와 重複되고, 各 機關에서 發表된 資料간의 混亂이 惹起됨으로써 매우 浪費的인 節次를 招來하게 될 것이다.

第1章에서 統計部処에 從事하고 있는 職員들은 어떤 作業班의 一員으로 存在하고 있다는 點에 關하여 論議한바있고, 統計制度가 實査要員으로부터 本部事務所의 指導員에 이르는 各 그룹 모두의 기여 없이는 機能을 效率的으로 發揮할 수 없다는 點을 明白히 하였다. 實際的으로 이런 팀-워크(Team work)는 한 政府機關外에 까지 擴大되어서, 國家의 經濟計劃 및 政策決定을 行하는데 必要한 統計資料를 蒐集, 處理, 分析 및 發刊하고 있는 各 分野에 從事하고 있는 모든 從事자들이 包含되어야 한다.

統計制度의 企劃 및 管理는 任意的으로 行할수 없는 龐大한 業務이다. 우리가 아는바와 같이 이러한 制度는 여러가지 方法으로 組織될 수 있다. 즉 인도네시아의 業務體系는 日本 및 오스트리아의 業務體系와 아주 다르다. 그러나 組織形態가 어떻게 되어 있던간에, 모든 政府部処内の 모든 活動을 統合할 수 있는 體系의인 接近方法이 必要하게 되는데, 이는 政府統計人의 責任이다.

本章에서의 論議는 좀 學究的인 面에서 다루어지고, 實査過程이나 地方事務所의 業務와는 直接的인 關聯이 없음을 알게 된다. 그러나 統計要員數는 매우 적지만 重要한 部分인 全體統計組織에 보다 좋

은 아이디어를 提供할 것이므로 이러한 事項에 대해서는 簡單하게  
나마 托의할 가치가 있다고 생각된다.

國家統計體系의 主要機能은 完全한 基準에 의해서 國家의 經濟  
社會의 構造 및 活動에 關聯된 統計를 蒐集, 編纂, 發刊하는 것이므  
로, 이 體系의 第1의 業務는 어떤 統計가 必要하고, 또 必要한  
統計를 어떻게 가장 잘 生産할 것인가 하는 問題를 決定하는 것  
인데, 이 業務遂行과 關聯하여 生産되고 있는 統計에 대한 定期的  
인 檢討가 必要하다. 또 傳統的인 統計發刊物은 向上되고, 補完  
되어, 發刊物間의 相互連繫가 이루어질 수 있도록 檢討와 特別研  
究가 必要하다.

統計의 統合調整을 위하여 政府機關은 여러 統計 利用者와 協議  
하여 一般的인 計劃을 樹立해야 하며 이러한 計劃은 一般的인 定  
義, 分類, 作成技法, 適切한 調查範圍 및 調查關聯期間, 等を 提示해  
줌으로써 各級統計 作成機關을 調整하는데 目標을 두어야 할 것  
이다. 또한 現在 作成되고 있는 統計相互間의 差異와 이들이 作  
成될 우선순위가 提示되어야 할 것이며 各機關이 作成한 統計資料  
가 不必要하게 重複되지 않고 現在와 앞으로 必要하지 않게 될  
統計에 대해서는 豫算이 浪費되지 않도록 計劃되어야 한다.

간혹 國家統計體系가 集中化 되어지고 있느냐 分散化되어지고 있  
느냐에 關하여 論議가 있는데, 이러한 用語들을 地域的인 觀點에  
서 使用할 수 있다. 즉 中央統計機關과 地方事務所로 잘 發達  
된 組織을 갖추고 있는 分散型體系와 모든 統計業務가 中央에 集  
中되어 있는 集中型體系로 区分 可能하다. 이런 意味에서 實際

的으로 資料蒐集에 責任을 지고 있는 職員이 応答者와 密接하게 연결되고 있기 때문에 大部分의 統計制度는 상당히 分散型的인 面을 보이고 있으며, 이런 現象들은 地方事務所의 發展을 고무시켜 주고 있다.

그러나 一般的으로 우리가 統計制度의 集中化에 대해 論議하게 되는 경우에는 "責任과 統制"가 集中되었느냐 또는 分散되었느냐를 말하게 되는데 即 統計活動에 對한 責任이 全的으로 하나의 커다란 國家統計當局에 있느냐 아니면 政府 各 機關이 各己의 特別 關心 分野에 對하여 統計作成 責任을 分擔하고 있느냐로 생각할 수 있다.

어떤 國家에서 發展되고 있는 制度의 形態는 統計蒐集과 利用에 큰 影響을 미치기 때문에, 네가지의 主要한 統計制度의 形態를 살펴보고자 한다.

#### (a) 中央統制와 調整이 없는 分散型體系

이런 意味에서의 統計體系는 國家水準에 있는 相異한 各機關이 그들의 業務分野에서 資料蒐集, 編纂 및 發刊에 責任을 지고 있으며, 各 機關들은 中央調整機關의 干渉없이 完全히 그들의 業務를 獨創的으로 行한다. 이런 類의 體系는 國家統計開發의 初期段階에서 나타나며, 이 體系는 一般的으로 가장 不滿足한 것으로 評價된다. 이러한 統計體系에서는 보다 신속한 처리가 가능하기는 하나, 많은 短点들 때문에 크게 不利하다.

이러한 短点에는 어떤 면에서 各級機關이 同一한 分野에 대해 統計를 作成하는 탓으로 인한 努力의 重複과, 어떤 기관에 直接

적으로 關心이 없는 다른면에 있어서의 差異와 그리고 여러가지 定義나 分類體系를 利用함으로써 생기는 모순점이 內包되어 있다.

그 結果 機關間的 資料比較를 어렵게 하며, 또한 國家全體의인 統計蒐集을 不可能하게 하며 더욱더 政府의 統計構造 및 需要를 把握하는데 있어 政府를 補助할 수 있는 機關이 存在하지 않는다는 點이다. 結局 政府는 이런 短點들을 分明히 알고 있어야 하며 이로 因해 이러한 體系로 부터 아래 (b)에 記述되는 體系로 轉換해야 할 것이다.

(b) 中央統制 및 調整機能을 가지고 있는 分散型體系

이러한 體系는 美國, 日本 및 韓國 등에서 活用되고 있으며, 中央統計機關自身の 統計作成活動은 매우 小規模의으로 遂行하나, 統計調整(各 機關의 統計重複作成除去) 및 分類와 定義의 基準設定(各 機關에서 生産된 統計가 比較 및 聚合이 可能하도록)의 業務에 重點을 둔다. 通常의으로 이런 機關은 一般統計 刊行物の 발간 및 國民計定推計의 作成 그리고 여러 機關에서 作成한 統計를 蒐集하여 國家的인 次元에서 總括하는 責任을 맡고 있으며, 또 中央機關의 權利와 責任이 他機關의 그것들과 均衡을 이룬다면 훌륭한 統計體系라 볼 수 있다. 한편 中央機關이 效率的으로 調整任務를 遂行할 수 있도록 充分한 權限을 부여하여야 하지만 他機關이 그들의 合法的인 統計需要에 對処하기 위한 統計를 作成하는데에 支障을 줄만큼 과중한 統制를 가해서는 안된다.

(c) 主要統計作成機關이 있으나 統計業務中 一部 重要分野는 獨立機關의 所管으로 되어있는 部分的 中央集中型 이 形態의 制度는 이미 說明한 바와같이 調整任務를 行하는 中央機關이 있고, 또한 一般統計 및 어떤 主要統計部門에 대하여 責任을 맡고 있는 形態로서 (b)와 (c)의 中間形態이다. 이에 는 어떤 다른 機關의 關心部門이 아닌 人口센서스와 特殊專門知識이 必要하며, 統計當局內에서 가장 集中化될 수 있는 標本調査等이 속 하는데, 이러한 形態의 體系는 말레지아, 인도네시아에서 使用되고 있다.

(d) 오스트렐리아와 캐나다와 같이 實際로 모든 政府統計의 作成에 있어서 하나의 機關이 責任을 맡고 있는 中央集中型 體系 이 體系는 많은 長點을 가지고 있다. 다시말해서 높은 程度의 프로그램 및 방법의 조정, 노력중복의 회피 및 範圍, 定義, 分類等의 矛盾除去, 專門化된 人力의 效率的 利用과 컴퓨터 및 其他 處理設備의 利用의 經濟性 등을 들 수 있다. 그러나 이 體系의 效率性을 감소시킬 수 있는 短點도 있다. 特히 이러한 體系는 많은 部處들의 需要 및 關心을 過小評價할 수도 있으며, 統計當局의 우선순위에 대한 判斷이 各 部處들의 우선순위에 대한 判斷과 一致하지 않음으로써 部處가 必要로 하는 統計를 適時에 獲得하지 못하는 경우가 있다. 以外에도 統計當局은 各 部處의 實際業務에 대한 專門知識이 不足하다. 即 例를 들면 農業關係職員은 農夫의 사고방식 및 행위에 대하여 統計當局의 職員보다 더

잘알고 있을수 있다는 것이다.

이런 要因때문에, 中央集中統計體系下에서는 統計機關과 他部처가 密接한 關係를 維持하는 것이 不可避하게 되므로 統計開發과 關聯된 폭넓은 問題들에 關하여 方向을 設定할 수 있도록 主要政府部처 또는 統計利用者들로 構成되는 委員會가 設置되어야 한다. 그러나 統計機關이 效果的으로 機能을 발휘하기 위해선 독자적으로 業務를 遂行할 수 있고 正當한 自治機構로서 認定받는 경우 뿐이므로 統計委員會가 統計當局의 事實的인 統計活動에 關여하지 않아야 할 것이다. 왜냐하면 지나친 외부간섭은 統計人에게 不信感을 야기시킬 수 있고, 또한 統計要員이 政治的인 影響力에 의하여 左右될 것으로 본다면 応答者들이 正確한 資料를 提供하지 않게 되어 生産統計의 質을 저하시키게 된다. 그러므로 이런 特性을 고려하는 것이 매우 重要하며, 아울러 高位統計計劃家들이 계속적으로 關心을 기울이고 있는 以外의 여러 統計體系가 있으나 이러한 體系들은 本教材의 範圍를 넘는다. 따라서 一部 主要考慮事項에 대해서만 言及했을 뿐이다.

## 12.3 統計基準

### 12.3.1 概 要

이 章의 첫머리에서 檢討한 바와같이 統計基準은 統計의 質的水準의 維持와 統計資料의 比較를 위하여 必要하나, 統計의 關聯 分野 및 寄與하고 있는 特殊目的에 따라서 基準이 다르기도 한다. 統計基準은 : (1) 雇傭, 家口 및 事業體等과 같은 用

語에 대한 標準定義, (2) 經濟活動, 商品 및 疾病等과 같은 標準分類 (3) 産業生産指數 作成方法과 標本誤差의 計算方法과 같은 標準方法論의 3 個部門으로 나누어 보고자 한다. 이 교재를 통하여 매우 폭넓은 분야에 影響을 미치는 政府作成統計의 경우를 살펴 보았으며, 事實上 政府發刊統計는 人間活動의 거의 모든 部門을 包含하고 있다는 事實도 알게 되었고 어떤 分野에 對한 統計라도 或者에 依해서 가끔 作成되고 있다고 생각되어진다. 統計가 해결하고자 하는 많은 문제점들은 매우 複雜하고, 많은 다른 요인에 대한 주의깊은 분석을 必要로 한다. 즉 政府가 個人課稅水準을 引上할 것인지, 아닌지를 결정중에 있다면 全般的인 要因을 檢討하는 것이 必要하며, 또한 政府는 追加수입이 얼마정도 增加할 것인지에 對하여 고려하는 것보다는, 個人과세인상이 經濟全體에 어떤 影響을 미칠 것인가를 檢討해야 한다. 이런 分析을 시도하는 경우에, 管理者와 計劃家는 經濟의 어떤 部門이 한 部門에 어떻게 反應하는가를 檢討하기 위해서 經濟理論을 利用하며, 그런 경우 그 經濟理論에 適合하고, 利用可能한 統計를 扞하게 된다. 이런 方法에 의하여, 서로 相異한 行爲의 效果를 豫測할 수 있으며, 이렇게 함으로써 가장 效率的인 方法을 選扞할 수 있게 된다.

### 12.3.2 標準定義

이런 類의 分析이 가능하기 때문에 여러 統計刊行物은 그들간에 모순이 없고 經濟理論에 合當한 定義와 概念을 利用해야 하며, 貿易統計, 業生産統計 및 金融統計等에 利用된 定義

들은 그들 相互間이나 基礎理論에 關聯을 갖고 있어야 할 것이다.

이런 理由 때문에, 統計人들은 各種 統計資料가 一관성을 維持할 수 있도록 하기 위해서 많은 時間을 消費하고 있으며, 이런 作業은 매우 複雜한 節次이므로 여기에서는 簡單하게 紹介할까 한다.

事實上 大部分의 統計刊行物에 利用될 定義와 概念에 體系를 確立하여 이들이 다른 刊行物에도 利用될 수 있고 定義가 妥當性이 있으며 各種資料間의 比較가 可能하도록 해야 할 것이다.

국민계정의 理論과 實際를 研究하고 있는 자들은 이와같은 것이 얼마만큼 重要한지를 쉽게 評價할 수 있을 것이다.

이런 標準定義와 分類는 統計制度發展에 基本的인 것이므로 本教材의 最終章에서 主要課題로 다루지 않고 實查指導나 實查活動過程에 必要한 가장 重要한 定義의 一部分을 여기에 列挙하고자 하며 統計가 發展된 先進諸國에서 直接 利用되고 있거나 地域事情에 알맞도록 修正한 形態로 使用되고 있는 國際分類體系에 關하여 言及하고자 한다.

다음 定義의 一部는 環境에 따라서 ( 예를 들면 우리가 첫번째 定義하고자 하는 " 家口 " 라는 用語는 社會構造上의 差異로 國家에 따라 여러가지 意味를 갖게 된다 ) 國家마다 다를수 있으나 모든 定義는 國際的으로 認定된 基準에 따라 一關性을 維持하고 있다.

이 節에는 各國에서 重要하다고 생각하는 一般的인 定義의 範圍와 實際로 使用하고 있는 定義의 正確한 意味를 나타낼수 있



도록 各国別로 마련 되어야 하며 最少限 다음 用語에 對한 定義  
는 여기에 包含시켜 다루어야 할 것이다.

- 家 口
- 住 宅
- 企 業 體
- 事 業 體
- 農 業 事 業 體
- 勞 動 力
- 經 濟 活 動 人 口
- 失 業
- 職 業 狀 態
- 產 業
- 職 業

### 12.3.3 國際分類體系

위에 주어진 定義目錄에서 産業이란 事業體(即 雇用主)  
에 의해서 수행되는 業務 및 活動形態를 말하고, 職業이란 各各  
의 勞働者(被雇用者)에 의해서 수행된 일의 形態를 말한다. 또  
한 各各의 境遇에 있어 數百, 數千가지의 活動이 있음을 알게 될  
것이며 統計가 事實 그대로를 正確하게 反映하고 있는 경우 産  
業 및 職業에 關한 어떠한 質問에 對해서도 答辯資料를 提示할  
수 있는 分類體系를 갖추어야 할 것이다.

전세계의 各国마다 完全히 獨自的으로 그 國家의 需要를 充足시

키기 위한 分類體系를 開發할 수도 있으나, 이런 方法은 두가지 점에서 결점을 갖고 있다. 첫째는 적절하고, 包括적인 分類體系를 確立하는데에는 莫大한 費用이 소모된다는 점이고, 둘째는 作業에 관한 모든 接近方法이 相異하기 때문에 分類體系에도 크게 差異가 있으며 이에따라 국제비교가 어렵거나 불가능하게 되어 各國은 다른 국가의 分類체계와 관련하여 그들의 경험및 문제점을 평가할수 없다는 점이다. 그래서 이러한 두가지의 어려운 問題를 解決하기 위하여 U.N은 많은 國家의 統計機關과 協力하여 産業 및 職業分類가 널리 普及되고 利用되도록 標準分類體系를 數年間に 걸쳐 發展시켜 왔다. 이러한 標準分類體系는 完全無欠하다거나 各國에서 修正하지 않고 直接利用하는 것이 바람직하다고는 볼 수 없으나 各國이 그들 事情에 맞도록 이들을 修正, 採択할 수 있도록 土台를 마련한 것이라 볼 수 있겠다. 또한 이 分類體系는 열거된 各分類에 숫자로된 코오드(3個以上 數字)를 부여하며, 이들 숫자들은 分類階級을 이루게 된다. 다시말해서 첫번째 숫자는 最高 10個까지 分類可能한 大分類를 나타내며, 두번째 숫자는 各各의 大分類를 몇개의 그룹으로 나눈 中分類를 意味하고 이렇게 계속 進行하여 最終數字는 同體系에서 分類可能한 最終의 細分類를 나타내도록 하고 있다.

産業을 分類하기 위해서 利用되는 分類體系는 國際産業標準分類라고 하며 영어 첫글자를 따서 ISIC라 稱한다. 이러한 分類體系는 生産過程이나 投入原資材, 또는 生産製品等に 있어서 類似한 活動을 한 그룹으로 하여 分類하는 4單位 數字를 利用한 體

系이다.

ISIC 分類의 첫번째 숫자는 대분류로서 産業을 10 個의 群으로 分類한 것이고, 두번째 숫자는 大分類를 中分類로 細分한 것이며, 세번째 숫자는 소분류로, 네번째 숫자는 세세분류로 세분되고 있다. 이 分類가 어떻게 構成되고 있나를 알아보기 위하여 아래에 ISIC의 10 個의 大分類와 이에 相応하는 코드번호를 列挙하고자 한다.

1. 農業, 수렵, 임업 및 어업
2. 鑛業 및 채석업
3. 製造業
4. 電氣, 가스 및 水道事業
5. 建設業
6. 都·小売業, 음식 및 숙박업
7. 觀光, 倉庫 및 通信業
8. 金融, 保險, 不動産 및 용역업
9. 社会 및 個人서어비스업
10. 分類不能産業

따라서 小売商店은 6으로 시작되는 ISIC의 코드들, 양조장은 3으로 시작되는 코드들 갖게되며 以外 産業에도 이와같이 코드들을 부여하는 것이다. 그러나 모든 産業이 이렇게 쉽게 分類될 수 있는 것은 아니며, 특히 여러가지 活動을 行하고 있는 事業體는 分類上相異한 어려움이 있다. 그러므로 이런 경우에는 主要 活動을 決定할 基準이 設定되어야 하며 그 活動에 알맞는 코드

가 부여 되어야 할 것이다.

職業을 分類하기 위한 國際分類體系로서는 國際標準職業分類 即 ISCO가 있으며, 이 分類는 國際勞動機構 (ILO)에 依해서 開發 되었으며, ISIC 分類보다 최근의 것이다. 그리고 이 分類는 ILO에 依해서 發刊되었으며, 國際比較를 위하여 3자리 數字가 標準임에도 불구하고 職業을 分類하는데 5단위 數字를 利用하고 있다. ISCO 分類의 3자리 숫자는 主要그룹 (Major groups : 첫 數字), 소그룹 (Minor groups : 둘째 數字), 및 단위그룹 (Unit groups ; 세째 數字)로 構成된다. 一見 職業分類는 比較的 簡單하며, 標準産業分類體系 보다 쉽게 發展시켜 나갈수 있을 것으로 생각되나, 實際에 있어서는 잘 이루어지지 않고 있다. 그理由로서 우리는 遂行되고 있는 여러 活動 뿐 아니라 業務에 必要한 어느程度의 知識 (即 必要한 資格과 技術)에 對해서 考慮하지 않으면 안되기 때문이다. 이러한 體系는 現在 遂行하고 있거나 수행할 資格을 가지고 있는 여러가지 業務에 對한 分類가 可能하도록 해 주고 있으며, 教育制度가 여러가지 業務를 遂行하기 위하여 資格을 갖춘 人力을 배출하도록 組織되어야 하는 바와 같이 정부가 人當需給計劃을 樹立함에 있어서 매우 중요한 것이며 職業分類는 資格과 技術에 關한 情報를 包含하고 있기 때문에 人力計劃에 크게 도움이 된다. 現行 ISCO의 分類體系는 아래와 같이 8個의 主要그룹 (Major groups)으로 構成된다.

| ISCO의 첫자리 숫자 | 분 류 직 業                 |
|--------------|-------------------------|
| 0 / 1        | 교수 및 전문직 종사자            |
| 2            | 행정 및 관리부문 종사자           |
| 3            | 사무 및 관련 종사자             |
| 4            | 판매종사자                   |
| 5            | 서어비스 종사자                |
| 6            | 농업, 어업, 수렵, 숙박업 및 관련종사자 |
| 7 / 8 / 9    | 생산 및 관련종사자와 운전수 및 노무자   |
| ×            | 달리 분류되지 않는 종사자          |

ISIC과 ISCO의 두 분류체계는 인구 및住宅센서스와 고용에 관한 統計의 資料를 處理하는데 가장 많이 利用될 것으로 본다. 그러나 이 두가지의 分類體系에 따라 個人의 고용 특성을 코딩하는 경우에는 混動을 일으키기 쉬우므로 이 두가지의 分類用語에 對하여 분명히 理解할 수 있도록 念頭해 두어야 할 것이다.

위의 두 分類體系外에 一般的으로 使用되는 國際分類에 關하여 간단히 說明하고자 한다.

- 國際疾病分類(ICD) : 주로 사망원인을 코딩하는데에 利用되며, 資料의 國際比較를 可能케 함.

- 國際標準貿易分類(SITC) : 매우 重要的한 商品分類로서 輸出入商品을 分類하는데 使用되고 있으며, 다른 分類體系와 마찬가지로 統計의 國際比較를 용이하게 하고 있으며, 貿易統計는 國家間의 거래를 把握하는데 利用되고 있으므로 貿易統計에 關한한 매우

重要한 體系이다. 또한 共通의 分類體系가 利用되는 경우에는 統計分析이 훨씬 쉬워진다.

## 12.4 統計蒐集의 範圍

### 12.4.1 概要

本章에서 지금까지 우리는 一般的으로 직면하게 될 여러 形態의 統計制度和 統計가 一貫性이 있고 統合調整될 수 있도록 各種 統計에 있어서의 共通的인 定義와 分類를 利用하는 即 다시 말해서 統計가 우리나라의 全般的인 양상과 活用實態를 나타내는데 適合한, 하나의 重要한 形態인 統計基準에 對하여 討議 하여 왔다. 이제 우리는 統計體系下에서 一般的으로 어떤 統計가 蒐集되고, 作成되는지 간략하게 檢討하면서 本章을 마치고자 한다. 우선 政府가 蒐集하고 利用하는 모든 資料를 두 種類 即 사람에 關한 資料와 그들이 살아가고, 일하는 方法等에 關한 資料로서 "社會的資料" (Social data)와 물리적환경에 關한 모든 자료를 포함하고 있는 "환경적인 자료" (Environmental data)로 분류할 수 있고, 사회적 자료 (Social data)는 다시 생산, 분배, 소비, 투자등과 關連된 "경제적인 자료" (Economic data)와 이외의 모든 정보를 包含하고 있는 "非經濟社會的資料"로 区分된다. 그러므로 어떤 統計的인 質問은 經濟活動, 非經濟社會活動 및 物理的環境과 같은 주요한 3개의 分類中 어느 하나에 속하게 되며 상당수의 統計表와 統計資料는 이들 세분류와 연關되며, 또한 重複되는 점이 없지 않은바 各 部門을 차례로 살펴보고, 이들간의 關係를 밝혀 보고자 한다.

以外에 各種 統計資料가 상호간에 어떻게 연관되고 있는지를 說明해 주는 도표를 利用하여 나타내 보고자 하며 이들 3가지 分類의 統計에 關하여 言及하고자 한다. 이를 나타낸 도표는 본장의 마지막 페이지에 첨부되어 있다.

#### 12.4.2 經濟統計

統計資料의 3개의 分類中에서 經濟統計는 모든 活動을 화폐가치라는 동일한 共通의 測定手段으로 나타내기 때문에 關聯分類中에서 가장 쉽게 觀察될 수 있으며, 大體的으로 달러 또는 以外の 화폐단위로 표시된 統計를 經濟統計라 말할 수 있다. 다시말해서 保健用役에 對한 支出을 貨幣額으로 表示한 경우는 經濟統計라 볼 수 있으나, 治療인수와 治療형태로 나타낸 統計는 經濟統計라 할 수 없다.

經濟統計의 理論的인 構造는 金세기 초에 유럽과 美國의 經濟學者의 研究에 의해서 開發되었으며 UN 指針書에는 國民計定體系 (System of national accounts)라 記錄되어 있는데 흔히 SNA라고 불려지고 있다.

簡單히 말해서 会社나 企業이 그들의 모든 活動을 일단의 계정에 기록하듯이 SNA는 한 國家의 各種의 關聯經濟統計를 計定에 記錄하여 나타내고 있다. 사실 이것은 各種活動이 測定되고 있는 方法과 이러한 活動이 다른 活動과 어떻게 聯關되고 있는가를 具體的으로 다루는 複雜한 體系이다.

圖表를 보게 되면 經濟統計는 다음과 같이 2개 그룹으로 나누

어 지게 됨을 알 수 있다. 첫째는 外國과 關係되는 모든 去來  
即 海外計定(The rest of the world)이며, 둘째는 國內의 모든  
活動 即 國內經濟(The domestic economy)이다.

海外計定 밑에는 輸入, 輸出, 國際收支(한나라에서의 貨幣의 流出  
入) 및 海外投資統計가 있으며 外國에 對한 支出과 外國으로 부  
터의 受取를 취급하는 經濟統計는 이 部門에 包含될 것이다.

아시아의 여러 國家에서는 國內經濟를 2개의 주요 그룹으로 分  
類하고 있는데 그 하나는 企業活動, 政府, 雇傭등을 다루는 公式  
經濟이며 다른 하나는 廣範圍한 活動을 다루는 非公式部門이다. 그  
런데 여기에는 自身の 食糧을 생산하는 住民, 住宅과 기타 生活必  
須品, 非公式마을 企業및 会社, 가구활동 및 其他 事項들이 包含되고 있다.

一部國家에서 非公式部門은 매우 重要하며, 全體國民生産의 半以上  
을 점할 수 있으나, 非公式部門으로 부터의 統計蒐集問題가 公式  
的인 企業으로 부터의 統計蒐集方法과는 매우 상이하기 때문에 그  
자체가 獨立部門으로 놓여 있다.

公式部門의 領域아래에서는 統計를 3部門으로 区分하여 蒐集,  
作成하고 있는데, 첫째는 財政에 關聯되며 조세, 金融, 通貨供給, 保  
險, 公共財政 및 家口의 收支를 包含하고, 둘째는 生産에 關聯되며  
農産物, 鑛産物, 製造商品, 都·小売去來額, 輸送費와 警察, 公務員, 国  
防, 保健, 教育 등의 政府의 用役提供에 對한 統計를 蒐集, 作成하며,  
또 이 두부문에는 상이한 특성치를 결합할 수 있도록 貨幣額으  
로 評價된 統計가 包含되고 있다.

그리고, 사람수, 쌀생산량 등의 양에 關한 資料는 세번째의 部門인



其他 經濟統計에 속하고, 이 資料는 雇傭 및 勞動, 산출량, 생산성 등에 關聯된 資料를 包含한다. 마지막 장에서 볼 수 있는 바와 같이 價格과 양을 연계시킨 것은 物價統計인데 이것은 圖表에서 알아 볼 수 있다. 여기에서는 양을 價値에 關聯시킨 物價와 價値를 量에 關聯시킨 物價를 利用하게 된다.

#### 12.4.3 非經濟社会統計

이 部門에는 사람과 그들의 活動에 關聯되어 있는 資料를 包含하나, 經濟活動과 關聯되어 있는 資料는 除外되고, 關聯된 各種 統計를 包含시킬 수 있는 體系를 開發함에 있어서 아래와 같은 問題에 直面하게 될 것이다. 非經濟社会資料는 經濟統計와는 달라. 完全한 體系로 利用할 수 있는 共通測定單位가 없으며, 大部分의 統計表는 사람을 單位로 취하고 있으나, 以外에 統計는 事件, 용역, 家口, 犯罪等を 다루고 있다. 따라서 貨幣額으로 表示된 모든 經濟活動의 国民總生産은 集計가 可能한데 非經濟資料에서는 經濟資料에서와 같이 總量을 구하는 것이 어려울 뿐 아니라, 범죄수, 도로사고 및 병원중사자들을 合計하여 全般的인 狀態를 구하는 것은 무의미하다.

이런 이유 때문에 SNA에 알맞은 非經濟資料를 分類할 수 있는 體系가 認定되고 있지 않다. 따라서 UN은 SSDS로 불리어지고 있는 社会 人口統計體系를 開發중에 있으나, 現在까지는 國際的인 承認을 받지 못하고 있으며 確固하게 組織된 體系를 갖추지 못하고 있다. 마지막 페이지의 圖表에서 非經濟社会資料는 두 部門 即 첫째는 人口數, 出生 및 死亡者數, 国内外的 移住 및

가족형성과 가구에 관한 統計를 內包하는 소위 "人口" (Population)라 불려지고 있으며 둘째는 所得分配, 住宅, 社会福祉, 教育, 雇傭 및 其他의 所得活動과 여가를 나타내는 "国民福祉統計" (Well-being of the people)로 名称이 주어지고 있다. 이 統計의 많은 部門은 SNA의 一部와 關聯되어 있다. 다시말해서 雇傭 및 所得活動에 관한 資料는 經濟統計와 非經濟統計의 양부문에 나타낼 수 있다. 即, 教育事項은 非經濟統計인 반면 이런 용역을 提供하는 費用은 經濟統計分野의 정부 서비스란에 包含될 것이다. 이런 關係는 여러 다른 部門에서도 볼 수 있다.

#### 12.4.4 環境統計資料

이 部門에서는 사람과 그들의 活動과는 달리 物理的 세계에 關聯된 情報를 다루고 있다. 강우량과 氣溫等에 관한 기상학적자료, 토지이용, 공해 및 에너지자원에 관한 資料가 이에 包含된다. 單一化된 統計體系를 갖추고자 하는 경우에는 여러 測定單位를 使用한 統計表 問題에 부딪치게 된다. 그런데 UN은 環境統計體系(SES)를 開發하기 시작했으나, 現在로서는 초기단계에 있고 國際적으로 권고될 程度까지는 못되고 있다.

분명히 環境統計와 經濟統計 및 非經濟統計間에 어떤 關聯性이 있다. 즉 토지 이용에 관한 資料는 農業生産에, 또 에너지자원에 관한 資料는 에너지의 生産, 消費等에 關聯이 있다.

#### 12.4.5 統計活動의 構造

一般的으로 政府統計活動에 包含되는 主要業務와 各種

統計資料間의 어떤 關聯性을 보여주는 圖表에 關하여 이미 言及한 바 있다. 그런데 이 圖表는 우리에 의해서나 國家統計業務에 종사하는 本부사무소, 地方事務所 및 實查現場의 우리 職員들에 의해 遂行되고 있는 各種 統計活動의 構成形態를 보여주고 있으므로 要點을 나타내는데 있어 이 보다 좋은 方法은 없을 것이다.

統計資料間的 어떤 關聯性을 보여주는 圖表에 關하여 이미 言及한 바 있다. 그런데 이 圖表는 우리에 의해서나 國家統計業務에 종사하는 本부사무소, 地方事務所 및 實查現場의 우리 職員들에 의해 遂行되고 있는 各種 統計活動의 構成形態를 보여주고 있으므로 요점을 나타내는데 있어 이 보다 좋은 方法은 없을 것이다.

公共統計의 主要部門

