

## 統計資料의 分析計劃과 資料處理

李 東 宇 \*

目 次	
I. 實驗計劃의 必要性	III. Package Programme의 利用
II. 資料處理의 順序	참고문헌

### I. 實驗計劃의 必要性

統計資料의 分析計劃設定은 統計資料가 수집된 以後에 作成하는 것이 아니고, 統計資料의 수집 以前에 設定되는 것이다.

즉, 統計資料를 수집하기 前에 統計資料를 수집하기 위한 研究計劃 또는 實驗計劃이 作成되어 있어야 研究目的에 合理的인 統計資料를 수집할 수 있고, 또 研究計劃을 設定할 때 研究의 目標과 具體的인 測定目的이 있으므로 分析計劃은 實은 研究計劃作成時에 作成되어 있는 것이다.

統計資料가 수집된 以後에 作成하는 分析計劃은 다만 研究計劃수립시 設定된 分析計劃의 범위내에서 資料의 特性, 分布 또는 범위 등을 考慮한 具體的 處理를 위한 計劃에 지나지 않는다. 따라서 統計資料의 合理的인 分析을 위하여는 研究計劃(또는 實驗計劃)의 先行이 必要하다.

實驗計劃은 研究에서 규명 또는 해결하려는 目的을 達成하기 위한 계획, 구조 및 方法을 組織化한 것이다. 여기서 계획이란 것은 研究의 全般的인 內容과 順序, 時期 等에 對한 內容이고, 構造는 研究에서 규명되고 또는 利用되는 變數의 具體的인 선정, 활용, 설명 및 규경에 對한 內容이고, 方法은 資料의 수집과 分析에 對한 計劃內容이다.

一般的으로 實驗計劃의 設計에는 다음의 세가지 事項을 考慮하여야 한다.

첫째, 構造設置를 위한 變數의 選定

變數는 同一項目의 관찰 또는 測定값의 調査 또는 觀察對象이 달라짐에 따라 變化가능한 상태 또는 特性을 뜻한다.

한 研究에서는 母集團의 性格, 實驗內容 또는 調査內容 및 研究目的에 따라 여러 種類의 變數로 나누어 選定할 수 있다.

여러 變數의 選定에 있어서 研究目的을 最終的으로 說明할 수 變數는 종속變수이다.

종속變수를 說明하는 變數는 獨立變數라고 하며, 이 獨立變數의 選定은 研究目的과 實驗의 性格에 따라 定한 說明形式 범위내에서 定한다.

獨立變數는 다시 實驗의 效果를 얻기 위한 變數가 있고(이를 實驗變數라고 한다) 實驗에 참여되는 對象, 即 調査單位의 性格에 따라 종속變수를 說明하는 變數도 獨立變數인바, 이 變數中 母集團 設定時 調査單位의 特性에 對한 變數는 基本獨立變數이고, 이 基本獨立變數中에서도 時系烈上 實驗變數가 效果를 미치기 以前에 作用할 수 있는 變數를 中間變數라고 한다.

이런 여러 獨立變數는 또 다른 觀點에서 分類할 때 두 가지 變數로 나눌 수 있다.

하나의 變數는 作用變數(Active Variable)인데 이

\*. 延世大學校醫科大學

變數는 종속변수에 영향을 最大限으로 미치게 하기 위하여 設定된 獨立變數이다. 이 作用變數 以外的 變數는 研究者가 獨立變數에 영향을 미치지 않으나 直接 間接으로 영향을 미치는 變數가 있게 마련이다. 이런 變數를 歸國變數(Attributable Variable)라고 한다.

研究設計時 變數의 選定 및 設定은 이미 分析의 方向과 內容을 제시하고 있다.

둘째, 研究對象 抽出에 있어서 確率性의 介在.

一般的으로 標本에 依한 研究 또는 調査對象의 選定에 있어서는 確率性의 介在가 必要하다.

標本抽出의 確率性의 介在는 確率事象에 依한 標本の 抽出에서 標本間의 同質性을 設定하기 위한 方法이며, 研究者의 期待效果를 노리는 實驗上 또는 觀察上의 Bias를 除去하기 위한 것이다.

셋째, 統計學 理論의 利用性

研究計劃에 있어 目的하는 종속변수의 모든 해석형식 또는 獨立變數의 종속변수의 설명형식을 해결할 수 있는 理論根據는 現在 상당히 發展되어 있다. 경우에 따라서는 요즈음 地域社會에 있어서의 實驗的 研究는 실로 複雜하여 合理的인 分析上의 實驗計劃을 設定하기 어려운 때도 있으며, 또한 變數의 測定값이 一定한 形式으로 測定되는 것이 아니고 命名의 척도, 順序의 척도, 간격적 척도 및 비척도 등의 여러 性格으로 測定되므로 測定값의 척도內容에 따른 分析上의 方法利用 한계성도 있다.

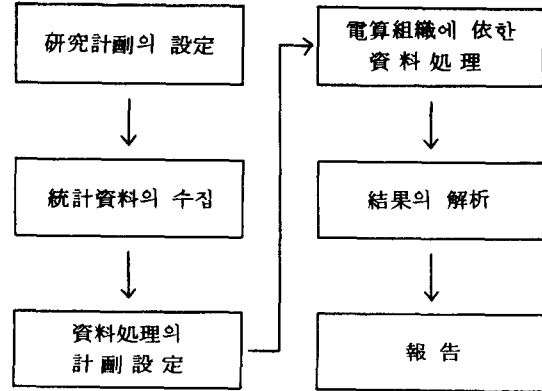
종속변수의 測定값이 間隔척도以上으로 測定될 경우, 獨立變數가 命名의 척도以上으로 測定되면 獨立變數間의 效果는 一次元的인 分散分析, 또는 두개의 獨立變數가 存在할 때에는 2次元的인 分散分析 方法에 따라 效果說明이 가능하고 셋 이상의 獨立變數가 存在할 때에는 各 變數間의 效果뿐만 아니라 相互作用까지도 分析가능하다.

分散分析은 여러 統計學 解析형식의 한 方法에 불과하고 現在까지 發展된 理論의 利用性은 變數의 성격과 分析의 구조에 따라 분리 利用할 수 있다.

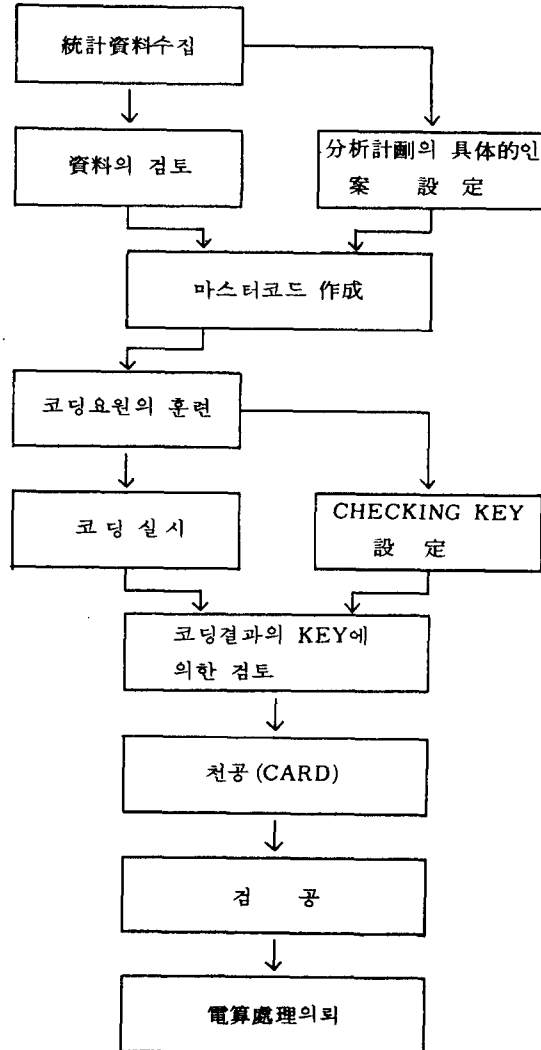
## II. 資料處理의 順序

統計資料의 機械的 處理를 위한 順序는 다음과 같은 몇개의 FLOW 그림으로 表現할 수 있다.

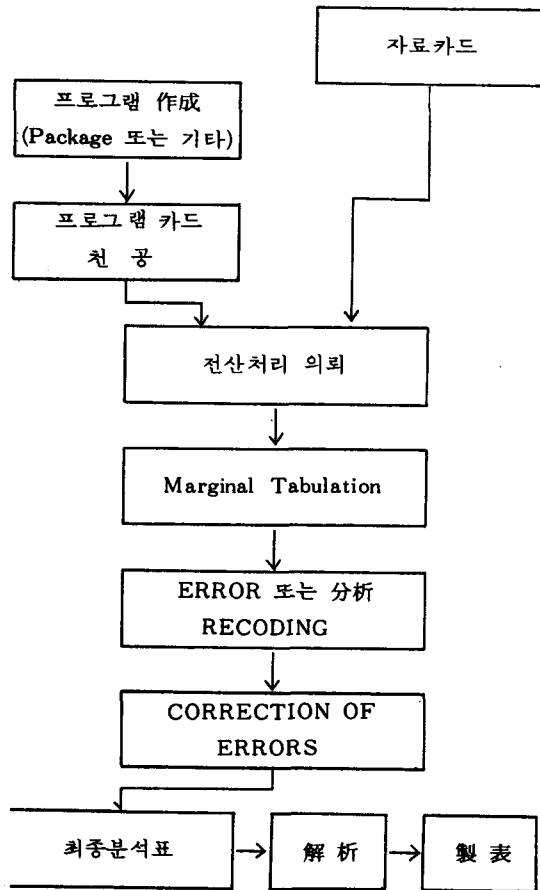
가) 全般的인 順序



나) 수집된 統計資料의 資料處理를 위한 具體的인 順序



다) 電算處理 順序



Ⅲ. Package Programme의 利用

統計資料의 統計學的 分析은 現在 이미 몇 개의 Package Programme이 作成되어 있다. 따라서 各 研究의 統計學的 分析을 위하여 새로 Programme을 作成하지 않아도 이 Package Programme을 利用하면 쉽게 資料를 處理할 수 있다.

現在 우리나라에서 利用 가능한 統計學分析 Package Programme은 SPSS와 BMDP가 있다. 이 중에서도 SPSS 6.0 및 7.0의 利用이 쉬워서 누구나 Parameter Card만 作成해서 自己의 Data Card와 同時에 電算所에 가지고 가면 쉽게 處理할 수 있다.

SPSS의 6.0에 依한 統計學 分析方法은,  
 DESCRIPTIVE STATISTICS ,  
 FREQUENCY TABLE ,  
 CONTINGENCY TABLE ,

MEAN DIFFERENCE TEST ,  
 BIVARIATE CORRELATION ANALYSIS ,  
 PARTIAL CORRELATION ,  
 MULTIPLE REGRESSION ANALYSIS ,  
 LINEAR MODELS  
 ANALYSIS OF VARIANCE ,  
 DISCRIMINANT ANALYSIS ,  
 FACTOR ANALYSIS ,  
 CANONICAL CORRELATION ANALYSIS ,  
 SCALOGRAM ANALYSIS ,

등이고, SPSS 7.0에 새로 追加된 分析的 方法은,  
 GENERALIZED AND THREE STAGE LEAST  
 SQUARES ,  
 JORESKOG'S FACTOR ANALYSIS,  
 MULTIVARIATE ANALYSIS OF VARIANCE ,  
 ANALYSIS OF MULTIPLE RESPONSE  
 ANALYSIS ,  
 NON-UNEAR REGRESSION ,  
 NON-PARAMETRIC TESTS,  
 DIGITAL PLOTTING ,  
 RELIABILITY ,  
 SPECTRAL ANALYSIS ,  
 SUMMARY TABLES ,  
 SURVIVAL ANALYSIS ,  
 TETRACHORIC CORRELATIONS , 등이다.

참 고 문 헌

1. Frankel, L. R. : Statistics and People-The statistitan's Res responsibilities, Jour. of the American Statistical Association, 71 : 358, March, 1976.
2. Meier, Paul : Statistics and Medical Experimentation, Biometrics, 31 : June, 1975. 511-529,
3. Seal, Hilary : Multivariate Statistical Analysis for Biologist, Methuen and Co. Ltd, 1964.
4. Dixon, W. (editor) : BMD, Biomedical Computer Programs, University of California Press, 1971
5. Norman. H. N, et al : Statistical Package for the Social Sciences, 2nd Edition, Mcgraw Hill, 1970.