

非母數的 統計 프로그램의 開發[†]

(Computer Programs for Nonparametric Tests)

裴 道 善*
張 重 淳**
金 相 復*

Abstract

Computer programs for IBM PC/XT/AT or compatibles, are presented for running 9 nonparametric tests. They include sign test, Wilcoxon signed rank test, Mann-Whitney-Wilcoxon test, Kruskal-Wallis test, Kolmogorov-Smirnov one sample and two sample tests, Kendall and Spearman rank correlation coefficient tests, and Chi square test for contingency table. Each program is written with BASIC language and is combined into a statistical package, 'NONPARA'. It is easily accessible through the menu programs. The algorithms on which each test is based, are also explained and 3 examples are given.

1. 序 論

産業의 急速한 發展에 따른 生産 시스템의 複雜化는 더욱 効率的인 管理技法의 도입을 必要로 하게 되었는데, 이를 위해서는 合理的인 意思決定의 가장 有用한 道具인 各種 統計技法의 活用이 불가피하게 되었다. 國內에서도 이미 實驗計劃, 샘플링檢査, 管理圖法, 工程能力分析 등 많은 統計技法들이 各 産業現場에서 널리 應用되고 있다. 그러나 이러한 技法들은 資料의 母集團이 正規分布를 따른다는 것을 假定하고 있는 것이 대부분이어서, 斷續生産, 少量生産 등으로 인하여 製品 特性值가 正規分布를 따른다고 假定하기 어려운 경우에는 그 소기의 目的을

達成할 수 없게 되는 경우가 많아지게 된다.

이와같이 母集團의 分布가 正規分布를 따른다고 보기 힘들거나, 또는 分布에 對한 假定이 실제와 差異가 날 경우에는 傳統的인 統計技法들에서 使用되는 推定量이나 檢定統計量들은 그 效率이 매우 나빠지게 되는 바, 이를 爲하여는 母集團에 對한 特殊한 分布를 假定하지 않는 非母數的 統計技法의 使用이 有用하리라고 판단된다.

非母數的 統計 技法들은 대부분 順序統計量에 基礎를 두고있어 模型의 分布에 對한 假定에 무관한 長點을 갖고 있다. 그러나 非母數的 統計 技法를 利用하기 爲해서는 많은 計算을 必要로 하고, 또 非母數的 統計量의 各種 數值表 등을

*韓國科學技術院 産業工學科

**亞洲大學校 工科大學 産業工學科

†本 研究는 峨山社會福祉財團 支援에 의하여 이루어진 것임.

손쉽게 利用하기 어려울 뿐만 아니라, 이에 對한 專門的인 知識을 갖고 있는 人力도 不足한 實情이다. 本 研究에서는 이와같은 점에 유의하여 最近 急速하게 發展, 普及되고 있는 퍼스널 컴퓨터에서 非母數的 統計技法을 손쉽게 利用할 수 있도록 하는 非母數 統計 패키지 'NONPARA' 를 開發하였다.^(註)

本 컴퓨터 패키지는 9개의 非母數的 檢定方法에 對한 프로그램을 포함하고 있으며, BASIC 언어와 데이터 베이스 매니지먼트 시스템을 利用

하여 非母數的 統計技法에 對한 專門知識이 없어도 수월하게 使用할 수 있도록 하였으며, 必要한 경우에는 各 使用者의 主프로그램과 連結해 쓸수도 있도록 하였다.

2. 構 成

本 研究에서 開發한 非母數 統計 패키지 'NONPARA' 는 9개의 非母數 檢定 서브루틴을 포함하고 있으며 그 概要는 그림 2-1 과 같다.

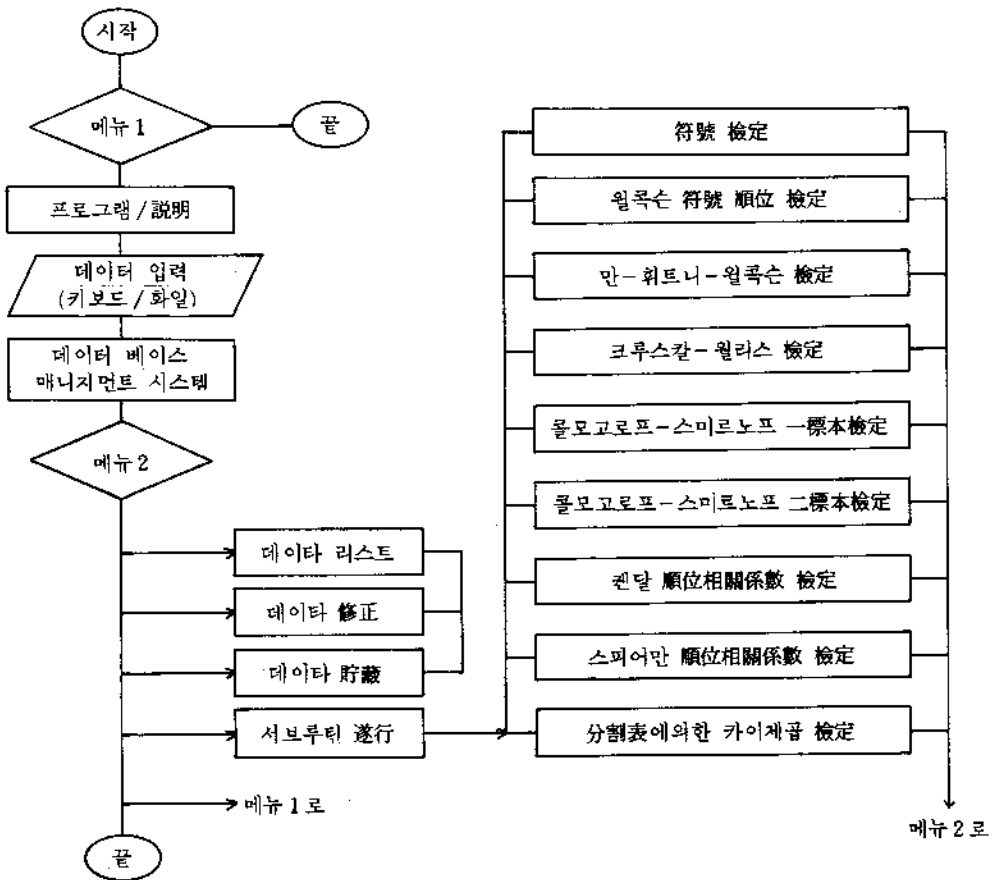


그림 2-1. 非母數 統計 패키지 'NONPARA' 의 構成圖

註) 本 研究에서는 非母數的 統計技法 以外에도, 基礎統計量計算, 分散分析, 回歸分析, 特殊分布函數, 特殊函數, 亂數發生函數 등을 開發하였는 바, Program List는 著者에게 要請하면 받아볼 수 있다.

非母數 檢定을 위해 'NONPARA' 를 實行하는 것은 메뉴 1로 부터 시작된다. 메뉴 1로 부터 實行하고자 하는 서브루틴을 選擇하면 選擇된 非母數 檢定 서브루틴의 目的 및 入力, 出力 변수에 對한 說明이 畫面上에 나타난다. 서브루틴에 對한 說明이 畫面上에 나타난 후에, 分析에 필요한 데이터를 키보드나 화일을 利用하여 入力시킨다. 데이터가 入力되면 데이터 베이스 매니지먼트 시스템에서 入力된 데이터의 리스트나 修正이 可能하다. 새로운 데이터의 入力일 경우나 기존의 데이터를 修正하여 貯藏해 둘 必要가 있는 경우, 메뉴 2에서 '데이터 貯藏' 을 選擇하여 入力화일에 貯藏할 수 있다. 入力 화일의 이름은 d:fn 形態로 입력시킨다. (d는 드라이브, fn은 화일 이름이다). 데이터 入力이 끝나면 '서브루틴 遂行' 을 選擇하여 이미 메뉴 1에서 選擇한 非母數 檢定을 遂行하게 된다.

本 研究에서 開發된 소프트웨어를 使用하기 爲해서는 다음과 같은 컴퓨터 및 주변 기기들이 必要하다.

1. IBM PC XT/AT 또는 이와 호환성이 있는 컴퓨터.
2. 하드 디스크나 360KB (또는 1.2MB) 플로피 디스크 드라이브.
3. 프린터(선택적 기기).

3. 프로그램 說明 및 實行結果

다음은 本 研究에서 開發한 9개의 非母數 檢定 서브루틴에 對한 알고리즘 및 3개의 서브루틴에 對한 例題의 遂行結果이다.

3.1 符號 檢定

符號檢定은 雙體比較의 한 方法으로 랜덤샘플 $(X_1, Y_1), (X_2, Y_2), \dots, (X_m, Y_m)$ 의 n 개 쌍으로 이루어진 데이터로부터 각 쌍의 차이의 中央값이 零인가를 檢定하는 것이다. 符號檢定을 위한 프로그램의 알고리즘은 다음과 같다.

1) 각 쌍의 차이 $D_i = X_i - Y_i$ 의 符號를 구한다.

2) $X_i = Y_i$ 가 되는 쌍은 버리고 $X_i \neq Y_i$ 인 쌍의 갯수 K 를 구한다.

3) D_i 가 陽인 것과 陰인 것의 갯수를 각각 구한 후, 작은 것을 M 이라 한다.

4) K 가 25보다 작거나 같은 경우, 二項分布로부터 같은 부호를 갖는 쌍의 수가 관측된 M 값보다 작을 확률 $p = \sum_{i=1}^M \binom{K}{i} / 2^K$ 를 구하며, K 가 25보다 큰 경우는 二項分布 대신 正規分布로 近似시켜서 p 를 구한다. 이때 평균은 $\frac{K}{2}$, 표준편차는 $\frac{\sqrt{K}}{2}$ 가 된다.

檢定統計量은 M 이 되며 충분히 작은 M 이어질 때, 귀무가설을 棄却하게 된다. 이때 유의수준이 PROB로 주어진다. (단 PROB는 단측검정을 위한 값이다). 符號檢定 프로그램을 遂行하기 위한 入力변수와 出力변수는 다음과 같다.

- 입력변수: 1) n : 관측쌍의 갯수
 2) $(X_i, Y_i), i=1, \dots, n$: 관측치
- 출력변수: 1) K : $X_i \neq Y_i$ 인 쌍의 갯수
 2) M : 양의 차이와 음의 차이의 쌍의 갯수중 작은 것
 3) PROB: p -value (단측)

단 p -value는 歸無假說을 棄却할 수 있는 가장 작은 有意水準의 값이다.

例 1. 다음은 2개의 生産라인에서 生産된 製品들을 랜덤샘플해서 측정한 데이터이다. 키보드를 利用하여 15쌍의 데이터를 入力시킨 후 '리스트' 시킨 結果는 아래와 같다.

NUMBER OF OBSERVATIONS : N = 15

A (1) = .15	B (1) = .16
A (2) = .21	B (2) = .15
A (3) = .16	B (3) = .18
A (4) = .15	B (4) = .15
A (5) = .14	B (5) = .2

A (6) = .18	B (6) = .2
A (7) = .16	B (7) = .17
A (8) = .12	B (8) = .14
A (9) = .15	B (9) = .13
A (10) = .17	B (10) = .24
A (11) = .12	B (11) = .21
A (12) = .15	B (12) = .14
A (13) = .14	B (13) = .17
A (14) = .16	B (14) = .12
A (15) = .18	B (15) = .2

메뉴 2로부터 '서브루틴 遂行'을 選擇하여符號檢定 서브루틴을 遂行하면 아래와 같은 結果를 얻는다.

OUTPUT RESULTS

K : THE NUMBER OF PAIRS = 14

M : THE NUMBER OF +, - DIFFERENCES (SMALLER) = 4

PROB : COMPUTED PROBABILITY (ONE-TAILED) = 8.978271 E - 02

위의 結果로 부터 유의 수준 $\alpha=0.05$ 인 경우 PROB 값이 α 보다 크므로 두 生産라인에서 生産되는 製品의 品質에 차이가 없다는 귀무가설을 棄却할 수 없다.

3. 2 윌콕슨 符號 順位 檢定

符號檢定이 관측된 쌍의 값의 차이에 대한 符號만을 利用하여 차이의 중앙값에 대한 檢定을 하는데 반해서, 윌콕슨 符號 順位檢定은 차이의 크기도 利用하며, 차이의 중앙값이 어떤 값 M_0 와 같은가를 檢定하는데 쓰인다. 이 檢定프로그램에 대한 알고리즘은 다음과 같다.

1) 각 관측된 쌍에 대한 차이(d_i)를 구한다.

2) 1)에서 구한 d_i 에 대해 절댓값 $|d_i|$ 를 利用하여 順位를 정한다. 이때 $d_i=0$ 인것은 버리고 $d_i=d_j (i \neq j)$ 인 것에 대해서는 順位의 평균값을 각 順位로 한다.

3) d_i 의 부호에 따라 각 順位값의 합을 구하여 둘중 작은것을 T라고 한다.

$$4) \mu = \frac{K(K+1)}{4}, \sigma = \sqrt{\frac{K(K+1)(2K+1)}{24}}$$

를 利用하여 평균과 표준편차를 구하고 $Z = \frac{T-\mu}{\sigma}$ 값과 이에 대응하는 確率값을 구한다.

윌콕슨 符號 順位檢定에 使用되는 입력변수와 출력변수는 다음과 같다.

- 입력변수 : 1) M : 각 변수의 관측치 수
 2) $(X_i, Y_i), i=1, 2, \dots, M$: 관측치
 출력변수 : 1) K : $d_i \neq 0$ 인 쌍의 수
 2) T : $d_i > 0$ 인 데이터의 順位합과 $d_i < 0$ 인 데이터의 順位합중에서 작은 값
 3) $Z = \frac{T-\mu}{\sigma}$ 값
 4) PROB : p-value

3. 3 만-윌트니-윌콕슨 檢定

만-윌트니-윌콕슨 檢定은 獨立인 두 데이터 그룹이 같은 母集團으로부터 抽出된 것인가를 檢定하는데 使用된다. 만-윌트니-윌콕슨 檢定 서브루틴에서는 데이터의 입력시 데이터 수가 작은 그룹의 데이터를 먼저 입력시키고 나머지 데이터를 입력시킨다. 이 檢定 프로그램에 使用된 알고리즘은 다음과 같다.

입력된 각 그룹의 데이터는 작은 값부터 順位를 매기며 같은 값을 갖는 관측자들에 대해서는 順位의 평균값을 順位로 한다. 만일 n_1, n_2 를 각 데이터 그룹의 데이터수, R_2 를 데이터 수가 많은 그룹의 順位의 합이라고 하면, $u' = n_1 \cdot n_2 + \frac{n_2(n_2+1)}{2} - R_2$ 로 계산되며 檢定統計量은

$$u = \begin{cases} n_1 n_2 - u' & \text{만약 } u' < u \\ u' & \text{만약 } u' \geq u \end{cases} \text{가 된다.}$$

이때 $Z = \frac{u - \frac{n_1 n_2}{2}}{S}$ 가 近似的으로 標準正規

分布를 따른다는 관계를 利用하여 檢定을 遂行하게 된다. 이때 S는 표준편차이며 다음과 같이 구한다.

$$S = \begin{cases} \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}, & T = 0 \text{ 일때} \\ \sqrt{\frac{n_1 n_2}{N(N-1)} \left(\frac{N^3 - N}{12} - T \right)}, & T > 0 \text{ 일때} \end{cases}$$

단 T는 동일한 관측치가 발생할 경우 수정항이며, $N = n_1 + n_2$ 이다. 만-휘트니-윌콕슨 檢定 프로그램을 遂行하기 위한 입력변수 및 출력변수는 다음과 같다.

입력변수 : 1) n_1, n_2 : 각 그룹의 데이터 수
(단 $n_1 < n_2$)

2) $X_i, i=1, \dots, n_1, n_1+1, \dots, n_1+n_2$: 관측치

출력변수 : 1) u : 檢定統計量 값

2) Z : 표준정규분포에서의 Z 값,

$$\frac{u - \frac{n_1 n_2}{2}}{S}$$

3) PROB : p-value

만일 n_2 가 20보다 작은 경우 $Z = 0$ 로 주어진다.

3.4 크루스칼-윌리스 檢定

크루스칼-윌리스 檢定은 K개의 母集團의 分布函數가 同一한가를 檢定하는 것이다. 만일 R_i 를 i번째 母集團에서 抽出한 랜덤샘플의 개수라고 할 때, 전체 觀測值 수는 $N (= \sum_{i=1}^K R_i)$ 이 되며, 각 관측치에 대해 작은것부터 크기순으로 順位를 매긴 후, 각 母集團의 샘플의 順位 합 T_i 와 $\bar{T}_i = (T_i / R_i)$ 를 구한다. 이때 檢定統計量 $H = \frac{12}{N(N+1)} \sum_{i=1}^K R_i \left(\bar{T}_i - \frac{N+1}{2} \right)^2 - 1$ 이 自由度 $K-1$ 인 카이제곱 分布를 따르는 관계를 利用하여 檢定을 하게된다. 만일 관측치의 수가 5보다 작을때는 베타分布를 利用하여 檢定을 하게된다.

本 서브루틴을 使用하기 위한 입력변수와 출력변수는 다음과 같다.

입력변수 : 1) K : 샘플을 抽出한 母集團의 수

2) $NI(i), i=1, 2, \dots, K$: 각 母集團에

서 抽出한 샘플의 수

3) X : 관측치 (X는 길이가 $\sum_{i=1}^K NI(i)$ 인 벡터이다).

출력변수 : 1) STAT(1) : 관측치에 같은값이 있을 경우, 수정항을 包含시켜 計算한 檢定統計量 값

2) STAT(2) : STAT(1)에 대응하는 $1 - (p\text{-value})$ 값

3) STAT(3) : 관측치에 같은 값이 있을 경우, 수정항을 包含시키지 않고 計算한 檢定統計量 값

4) STAT(4) : STAT(3)에 대응하는 $1 - (p\text{-value})$ 값

단 p-value는 귀무가설을 棄却할 수 있는 가장 작은 유의수준의 값이다.

3.5 콜모고로프-스미르노프 一標本 檢定

콜모고로프-스미르노프 一標本 檢定은, 만일 $F_0(X)$ 를 假定된 알려진 分布函數라고 할때, 未知의 分布函數 $F(X)$ 를 갖는 母集團에서 랜덤 샘플을 취하여 모든 X에 대해 $F(X) = F_0(X)$ 인 관계가 成立 하는가를 알아보는 檢定이다. 따라서 $S_n(X)$ 를 샘플 累積分布函數라고 할때 $F_0(X)$ 와 $S_n(X)$ 의 차이가 귀무가설 $H_0 : F(X) = F_0(X)$ 의 棄却與否를 결정하게 된다. 주어진 자료 $\{X_i, i=1, 2, \dots, n\}$ 으로 부터 샘플 累積分布函數 $S_n(X)$ 를 구한 후, $D_n = \max_{1 \leq k \leq n} |S_n(X_{(k)}) - F(X_{(k)})|$ 를 計算한다.

이때 $X_{(k)}$ 는 K번째 順位統計量이다. 檢定統計量은 D_n 이며, 이 서브루틴에서 제공하는 $F_0(X)$ 는 正規分布函數, 指數分布函數, 코시分布函數, 矩形分布函數이다.

本 서브루틴을 遂行하기 위한 입력변수와 출력변수는 아래와 같다.

입력변수 :

변 수	正規分布	指數分布	코시分布	矩形分布
IFCOD	1	2	3	4
u	평균	평균	중앙값	왼쪽끝점
s	표준편차	표준편차	u-s가제일 사분위수가되는 s 값	오른쪽끝점

출력변수 : 1) $D_n = \max_{1 \leq k \leq n} |S_n(X_{(k)}) - F(X_{(k)})|$

2) PROB : p-value

例 2. 다음은 PLA-DECK 이라는 製品을 生産하는 工場에서 WDW에 대해 測定한 20개 데이터를 콜모로고프-스미르노프-二標本 檢定 서브루틴을 이용하여 檢정한 結果이다.

키보드를 利用하여 20개의 데이터를 入力시킨 후, 메뉴 2의 '데이터 리스트' 기능으로 출력해 본 結果는 다음과 같다.

IFCOD = 1 (NORMAL)

U = 2.78 S = .43

- | | |
|-----------------|-----------------|
| X (1) = 2.8222 | X (11) = 2.8395 |
| X (2) = 2.9073 | X (12) = 2.7157 |
| X (3) = 2.7493 | X (13) = 2.8453 |
| X (4) = 2.6392 | X (14) = 2.7507 |
| X (5) = 2.7492 | X (15) = 2.8744 |
| X (6) = 3.1392 | X (16) = 2.8213 |
| X (7) = 2.6876 | X (17) = 2.6261 |
| X (8) = 2.7262 | X (18) = 2.7415 |
| X (9) = 2.6508 | X (19) = 2.7012 |
| X (10) = 3.0723 | X (20) = 2.5239 |

주어진 20개의 데이터가 평균이 2.78 이고 표준편차가 0.43인 正規分布를 하는 母集團에서 抽出된 것인가를 檢定하기 위한 것이다. 메뉴 2에서 '서브루틴 遂行'을 選擇하여 出力해 보면 다음과 같다.

OUTPUT RESULTS

IFCOD = 1 (NORMAL)

U = 2.78 S = .43

D = .3116602

PROB = 4.108369E - 02

따라서 $D_n = \max_{1 \leq k \leq 20} |S(X_{(k)}) - F(X_{(k)})| = 0.3117$

이며 p-value가 0.041 이므로, 유의수준이 5% 이면 귀무가설을 棄却할 수 있다.

3.6 콜모고로프-스미르노프 二標本 檢定

콜모고로프-스미르노프 二標本 檢定은 각각 獨立인 두 랜덤샘플 X_1, X_2, \dots, X_m 과 Y_1, Y_2, \dots, Y_n 의 모집단의 分布函數를 $F_1(X)$ 와 $F_2(X)$ 라고 할 때, 모든 X에 대해 $F_1(X) = F_2(X)$ 가 성립 하는가를 檢定하는데 쓰인다. $S_m(X)$ 와 $S_n(Y)$ 를 각각 觀측된 X 값과 Y 값의 샘플 累積分布 函數라고 할 때, X, Y에 대한 두 샘플 累積分布函數 값의 差異의 絶대값의 最大값인 $D_{m,n} = \max_{x,y} |S_m(X) - S_n(Y)|$ 를 이용하여 檢定한다.

本 서브루틴을 使用하기 위한 入力변수와 出力변수는 아래와 같다.

입력변수 : 1) m, n : 두 랜덤샘플의 觀측치 수
2) $X_i, Y_j, i=1, 2, \dots, m, j=1, 2, \dots, n$
: 觀측치 값.

출력변수 : 1) Z : $\sqrt{\frac{m \cdot n}{m+n}} D_{m,n}$

2) PROB : $1 - \lim_{n,m \rightarrow \infty} \text{Prob}$

$\left\{ \sqrt{\frac{m \cdot n}{m+n}} D_{m,n} < Z \right\}$

3.7 켄달 順位 相關係數 檢定

켄달 順位 相關係數를 利用하여 두 변수간의 相關關係가 있는가를 檢定하는데 使用된다. 檢定統計量은 $\tau = \frac{S}{\frac{1}{2}n(n-1)}$ 가 되며, 觀측치중

같은 값을 갖는 경우가 생기면

$\tau = \frac{S}{\sqrt{\frac{1}{2}n(n-1) - T_a} \cdot \sqrt{\frac{1}{2}n(n-1) - T_b}}$ 로

주어진다. 이때 S는 표준편차로서 $\sqrt{\frac{2(2n+5)}{9n(n-1)}}$ 로 계산되며, T_a, T_b 는 수정항이다. 이 서브루

틴의 檢定은 $Z = \frac{T}{S}$ 가 近似的으로 標準正規分布을 따른다는 것을 利用하여 檢定한다.

켄달 順位 相關係數 檢定을 위한 입력변수와 출력변수는 다음과 같다.

- 입력변수: 1) n : 관측치 수
 2) NR : 順位를 매긴 입력 데이터인 가를 나타내는 변수(順位를 매긴 데이터의 경우 NR=1이며, 아닌 경우는 NR=0이다)

3) A, B : 관측치 벡터

- 출력변수: 1) TAU : 켄달 順位相關係數
 2) SD : 표준편차
 3) Z : 標準 正規分布에서의 Z 값
 4) PROB : p-value

3.8 스피어만 順位 相關係數 檢定

스피어만 順位 相關係數 檢定은 스피어만 順位 相關係數를 利用하여 주어진 데이터로 부터 두 변수간의 相關關係가 있는지 없는지를 알아 보는데 쓰인다. 스피어만 順位相關係數는 $r_s = 1 - \frac{6D}{n(n^2-1)}$ 로 계산되며, $t = r_s \cdot \sqrt{\frac{n-2}{1-r_s^2}}$ 가 자유도 $n-2$ 인 t-分布를 따른다는 것을 利用하여 檢定을 遂行한다. 이때 $D = \sum_{i=1}^n (R(X_i) - R(Y_i))^2$, $R(X_i)$, $R(Y_i)$ 는 각각 X_i 와 Y_i 의 順位이다.

스피어만 順位 相關係數 檢定을 위한 입력변수와 출력변수는 다음과 같다.

- 입력변수: 1) n : 각 변수의 관측치 수
 2) NR : 켄달 순위 상관계수의 경우와 같음
 3) A, B : 관측치 벡터

출력변수: 1) RS : 스피어만 순위상관계수

2) T : $r_s \cdot \sqrt{\frac{n-2}{1-r_s^2}}$

3) NDF : 자유도(n-2)

4) PROB : p-value

만일 n이 10보다 작으면 T = 0가 된다.

3-9 分割表에 의한 카이제곱 檢定

分割表로 주어진 데이터로 부터 변수간의 獨立性 여부를 검사하는데 쓰인다. 이 서브루틴에 서는 A_{ij} , $i=1, \dots, n, j=1, \dots, m$ 의 데이터를 갖는 分割表로 부터 자유도와 카이제곱 값을 다음과 같은 방법으로 구한다.

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(A_{ij} - E_{ij})^2}{E_{ij}}$$

이때 $E_{ij} = \frac{T_i T_j}{GT}$ 이며 T_i, T_j 는 각각 i번째 列과 j번째 行의 合이며 GT는 전체 合이다. 그리고 자유도는 $(n-1) \cdot (m-1)$ 로 계산된다. 이 서브루틴의 입력변수와 출력변수는 다음과 같다.

- 입력변수: 1) A : 관측치 行列
 2) n (行), m (列) : A 行列의 크기
 출력변수: 1) CS : 카이제곱 값
 2) NDF : 자유도
 3) PROB : p-value

例3. 다음은 5개 地域에 사는 住民을 대상으로한 說問調査에서 각 地域에 따라 應答한 사람과 應答하지 않은 사람수에 대한 데이터이다. 이 데이터로부터 지역에 따른 說問調査의 應答 정도가 다른가를 알아보려고 한다. 入力된 데이터의 리스트는 다음과 같다.

NUMBER OF ROWS IN A : N = 2
 NUMBER OF COLUMNS IN A : M = 5

A (1, 1) = 108	A (2, 1) = 334
A (1, 2) = 94	A (2, 2) = 151
A (1, 3) = 62	A (2, 3) = 112
A (1, 4) = 67	A (2, 4) = 53
A (1, 5) = 14	A (2, 5) = 6

위의 데이터로부터 카이제곱 檢定을 하면 아래의 結果를 얻게된다.

OUTPUT RESULTS

NDF = 4

CS = 56.88561

PROB = 2.6E - 06

따라서 유의수준 $\alpha = 0.005$ 일때 p-value 가

0.005보다 작으므로 귀무가설을 棄却할 수 있다. 즉, 사는 地域에 따라 說問調査에 응답하는 정도가 다르다고 볼 수 있다.

4. 結 論

本 研究에서 開發한 非母數的 統計 팩키지는 非母數 統計技法중 가장 많이 利用되고 있는 9개의 檢定節次를 包含하고 있다. 本 팩키지는 퍼스널 컴퓨터용으로 開發된 것이므로 少量生産, 斷續生産, 잦은 設計變更등으로 인한 工程의 不安靜化등으로 製品特性值의 正規性을 保障할 수 없는 경우나, 資料의 不連續性으로 인해 正規分

布를 假定할 수 없는 여러 社會科學의 研究에 널리 使用될 수 있으리라 思料된다.

最近 퍼스널 컴퓨터의 發展에 따라, 퍼스널 컴퓨터用 統計 팩키지가 開發, 使用되고 있는 바, 國內에서도 그 需要가 增加하고 있다. 이와같은 팩키지들을 利用하는 데에는 퍼스널 컴퓨터와 統計에 대한 基本的인 知識을 必要로 하고 있으나 이를 더욱 發展시켜 統計에 대한 知識이 적은 사람도 쓸 수 있는 專門家體系(Expert System)의 형태로 發展시키는 것이 有用하리라 생각된다. 이는 非母數的 統計方法 뿐만아니라, 다른 統計的 分析方法들에 관한 팩키지 開發에 있어 指向해야 할 방향이라고 思料된다.

References

1. Daniel, W. W., *Applied Nonparametric Statistics*, Houghton Mifflin Co., 1978.
2. *IMSL Library Reference Manual*, Vol. 2, IMSL Inc., 1980.
3. Lehmann, E. L., *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, Inc., 1975.
4. Marija, J. N., *SPSS/PC+ for the IBM PC/XT/AT*, SPSS Inc., 1986.
5. *SAS Introductory Guide for Personal Computer*, Version 6, SAS Institute Inc., 1985.