응용통계연구 제 12권 1호, 1999년, pp. 275-293

누적법에 관한 연구 *

백운붕[1)] 이우선[2)]

요 약

다구찌의 누적법은 다구찌 품질공학에서 중요한 통계분석 방법이다. 그러나 이 방법이 복잡하고 비효율적일 뿐만 아니라 실험의 결과가 잘못 해석 될 수 있는 문제점을 가지고 있다. 특히 순서 지어진 범주형에 관한 다요인(multifactor) 실험에서는 이러한 가능성이 큰 것으로 지적되고 있다. 이에 대한 걱정과 비판이 Nair(1986) 그리고 Hamada and Wu(1990)에 의하여 심각하게 제기되어 왔다. 본 논문은 이러한 내용들을 정리하고 이들의 논란과 주장에 대한 평가와 이에대한 최선의 실천방안을 제안하고 있다. 아울러 실제 자료분석을 위하여 필요한 SAS/IML 프로그램을 제시하고 있다.

1. 서론

우리는 산업분야등의 품질관리분야에서 범주형 자료를 자주 접하게 된다. 이 중에서도 특히 순서 지어진 범주형 자료(ordered categorical data)에 대한 분석이 문제로 된다. 이 경우 다구찌(田口玄一, Genichi Taguchi)는 그가 창안한 누적법(累積法, Accumulation Analysis)을 사용한다. 순서 지어진 범주형 자료에 대한 분석에서는 피어슨의 χ^2-검정의 적용이 적절하지 않다고 생각되어 이에 대신되는 방법으로 누적법을 제안하고 발전시켜 온 것이다.

다구찌의 누적법은 제3절에서 서술하는 바와 같이 중요한 위치효과(location effects)를 확인하는데 상당한 검정력을 가지고 있지만 산포효과(dispersion effects)를 따지는 경우에는 좋은 방법이 되지 못한 것으로 지적되고 있다. 더구나 위치효과에서 이러한 산포효과를 분리해 내는데 어려움이 있다. 왜냐하면 대부분 이들 두 효과가 서로 혼재(confounding)해 있기 때문이다.

이러한 관점에서 통계학자들은 다구찌의 누적법이 필요 이상으로 복잡할 뿐만 아니라 효율성도 떨어진다고 지적하고 누적법은 가르쳐서도 안되고 권장해서도 안된다고 부정적인 견해를 표명하고 있다. 그런데 누적법은 아직도 다구찌 방법에서 주요부분을 차지하고 있다. 이것은 통계학자들이 경고한 것처럼 누적법의 오류가 과히 심각하지 않거나 품질개선에 그래도 이바지하고 있다는 반증이 아닐까? 통계학적 이론으로는 받아들일 수 없으나 실제 산업현장에서 이를 배제하고 다른 방법을 이용할 준비가 되어 있지 않은 것이다. 지금도 많은 다구찌 추종자들이 누적법을 사용하고 있다. 많은 통계학자들은 누적법의 사용을 반대하지만, 다구찌방법 사용자들은 이것으로 문제를 훌륭히 해결하고 있다고 주장한

* 본 연구는 1997년 성신여자대학교 연구소 특별사업지원 계획에 따른 기초과학연구소 특별과제인 "품질 향상을 위한 통계적 방법-다구찌 방법의 개선과 적용 방법을 중심으로 -"의 연구결과임.

1) (305-755) 대전시 유성구 어은동 한빛 APT 120-901, 전 고려대학교 통계학과, 교수
2) (136-742) 서울시 성북구 동선동 3가 249-1, 성신여자대학교 통계학과, 교수

다. 다구찌방법으로 훈련된 다구찌 추종자 들에게는“누적법”이 아닌 다른 분석법에는 큰 관심을 갖지 않는다.

그러나, 이들은 누적법에 대한 통계학자들의 견해에 귀를 기울일 필요가 있는 것이다. 또, 통계학자들은 누적법의 단점을 설명하고 이것을 극복하기 위하여 추가적인 통계적 분석이 필요하다는 것을 엔지니어들이 이해하도록 하여야 한다. 그리고 구체적으로 추가적인 통계적 분석방법을 제공해 주는 것이 필요하다. 이 논문은 이러한 설명과 분석방법을 제시하고 있다. 논문의 구성은 2절에 다구찌의 누적법에 익숙하지 않은 사람을 위하여 예를 들어 설명하고 3절에 누적법에 대한 반대논리를 소개한다. 또 4절에 누적법의 올바른 사용법을 제시하고 마지막 5절에서 컴퓨터를 이용한 분석방법을 제시하고 있다.

2. 다구찌의 누적법

다음은 다구찌의 논문(1974)에서 누적법을 설명하는데 사용한 예이다.

(예 2.1) 두 가지의 의약품 A와 B를 사용해서 각각 80명씩에 표준투약(약성분이 없는 가짜약, placebo)을 하고 다른 80명씩에 대해서는 어떤 처치투약(약품A, 약품B)을 하여 표 2.1과 같은 가상적인 관측도수를 얻었다 하자. 여기에서 관측도수는 투약효과를 네 종류의 순서지어진 계급범주로 나누어 관찰하여 기록한 것이다. 여기에서 순서 지어진 계급 범주라는 것은, ‘전혀 효과가 없다(-)’, ‘약간 효과가 있다(+)’, ‘효과가 있다(++)’, ‘현저하게 효과가 있다(+++)’의 4종류이다.

표 2.1: 약품 A, B에 대한 효과 실험

	약품 A						약품 B				
	−	+	++	+++	합계		−	+	++	+++	합계
표준	40	24	10	6	80	표준	40	24	10	6	80
처치	24	40	10	6	80	처치	24	29	16	11	80
차이	−16	16	0	0		차이	−16	5	6	5	
c_i	64/160	64/160	20/160	12/160		c_i	64/160	53/160	26/160	17/160	
	$\chi^2 = 8.00$, $p = 0.047$						$\chi^2 = 7.33$, $p = 0.06$				

*차이=처치− 표준, c_i=(표준+처치)/160

표 2.1에서 피어슨의 χ^2-검정결과를 살펴 보면 유의수준 5%하에서 약품 A가 통계적으로 유의한 반면 약품 B는 유의하지 않는 것으로 되어있다. 이때 χ^2-검정의 귀무가설은 계급 범주내 표준과 처치 사이에 차이가 없다는 것이다. 그러나 위의 실험자료를 살펴보면 약품 A에서는 이 차이가 들숙날숙하거나 전연 차이가 없는 것으로 보이나, 약품 B의 경우는 두 번째 계급범주부터는 분명히 처치효과가 있는 것으로 나타나 있다. 이와같이 χ^2-검정결과와 분할표를 관찰해서 얻는 내용이 어긋나는 현상을 다구찌는 다음과 같이 설명하고 있

다. 피어슨의 χ^2-검정방법은 위와같은 실험자료에서 요인효과의 크기를 검정하는데 적당하지 못하다. 위의 예에서 약품 A보다는 약품 B의 효과가 크다는 것이 분명한 것으로 생각되기 때문이다. 따라서 다음에 설명하는 누적법을 사용하면 이러한 문제점을 해소하여 우리의 상식과 일치하는 결과를 얻을 수있다. (예 2.1)에서 약품 A, B에 관한 분할표인 표 2.1를 근거로 다음과 같은 도수를 누적시킨 표 2.2를 작성한다. 우선 순서 지어진 4개 계급 범주에 대해서 (-)열은 계급 제 I조로 열의 명칭만 바꾸고, (-)열과 (+)열의 대응도수를 합한 것을 계급 제 II조로 하고, 이 제 II조와 (++)열의 서로 대응하는 도수를 합한 것으로 계급 제 III조를 구성한다. 마찬가지로 계급 제 IV조를 구성할 수 있으나 이것은 처리별 도수 합계이므로 분석 대상에서는 제외된다.

표 2.2: 표 2.1에서 작성한 누적표

약품 A					약품 B				
A	I	II	III	IV	B	I	II	III	IV
표준	40	64	74	80	표준	40	64	74	80
처리	24	64	74	80	처리	24	53	69	80
합계	64	128	148	160	합계	64	117	143	160
d_i	$\frac{64}{160}$	$\frac{128}{160}$	$\frac{148}{160}$	1	d_i	$\frac{64}{160}$	$\frac{117}{160}$	$\frac{143}{160}$	1

표 2.2의 각 칸의 수치는 기본 관측치 0과 1로 이루어진 것 즉, 0,1의 데이터의 합계라는 것으로 이해하여야 한다. 예를 들어 약품A에서 계급 I조에서 표준처리의 40이라는 것은 표준처리를 받은 80명 중 제 I조에 속하면 1 그렇지 않으면 0으로 한 관측치 80개의 합계가 40이라는 것이다. 이와 같이 생각하여 각 조별로 분산분석을 하고 이것을 각 변인별로 합한다. 그런데 각 조내에서의 분산이 조간에 걸쳐 서로 같지 않다는 것이 문제이다. 이것을 조정 하기 위하여 예를들어 i조는 비율이 d_i인 이항 모집단을 구성한다는 생각에서 분산 $d_i(1-d_i)$를 이용한다. 이 분산으로 i조에서의 제곱합을 나누어 표준화를 실시한다. 즉, i조에 대해서 $W_i = 1/[d_i(1-d_i)]$와 같이 정의한 가중치(weight)를 i조의 각 제곱합에 곱하는 것에 의해서 표준화된다. 이와 같이 수정된 각 조에서의 제곱합을 합하여 자유도 3인 처리제곱합 T를 얻는다. 이것을 누적분석에 의한 통계량이라는 뜻으로 AA(accumulation analysis)-통계량 T라고도 한다. 약품 B의 누적표에서 표 2.3과 같은 분산분석표를 얻는다.

표 2.3: 약품 B에 관한 누적표에 대한 분산분석표

변인	자유도(df)	제곱합(SS)	평균제곱(MS)	F
처리	3	$T = 12.16$	4.05	4.11**
오차	474	467.84	0.99	
계	477	480	**:유의함($\alpha = 0.01$)	

표 2.3에서 F-값을 보면 유의수준 1%하에서도 유의성이 나타나 자료의 내용과 잘 부합된다는 것을 알 수있다. 한편, 약품 A의 경우는 $T = 6.6667$로 되어 χ^2-검정 때와는 달리 5% 유의수준하에서도 유의성이 나타나지 않는다.

이러한 방법은 다음과 같은 다요인(multi-factor)실험의 경우에도 확대 응용 될 수 있다.

(예 2.2) 어떤 주조(molding)공정에서 최적 주조조건을 구하기 위하여 다음과 같은 실험설계를 이용한 실험을 하였다. 여기에서 고려된 인자와 수준은 다음과 같다.

A=재료(資料), 2 수준(현재의 것과 새로운 것)

B=주조(鑄造)의 형식, 2 수준(현재의 형식과 새로운 형식)

C=경화(硬化) 방법, 2 수준(현재의 방법과 새로운 방법)

D=용해(鎔解) 온도, 2 수준(현재의 온도와 현재 온도+30)

표 2.4: 주조 실험의 계획과 데이터

	A	B	$A\times B$	C	e	D	$A\times D$	관측도수			
실험번호	(1)	(2)	(3)	(4)	(5)	(6)	(7)	수	우	양	가
1	1	1	1	1	1	1	1	15	3	1	1
2	1	1	1	2	2	2	2	14	3	2	1
3	1	2	2	1	1	2	2	16	2	2	0
4	1	2	2	2	2	1	1	19	1	0	0
5	2	1	2	1	2	1	2	17	1	0	2
6	2	1	2	2	1	2	1	17	2	1	0
7	2	2	1	1	2	2	1	10	2	4	4
8	2	2	1	2	1	1	2	16	3	1	0
							합계	124	17	11	8

여기에서 관측도수는 각 인자의 수준별 처리조합(실험조건)별로 20개씩의 제품을 조사 하여 수(no flow-holes), 우(a few flow-holes), 양(several flow-holes) 그리고 가(many flow-holes)로 분류하였다. 이것을 정리한 것이 표 2.4이다.

이 자료를 분석하기 위하여 각 인자별로 관측도수를 합집(collapsing)한다. 예를들어, 인자 A의 경우는 표 2.5와 같다.

표 2.5: 인자A의 경우 합집

	수	우	양	가	계	I	II	III
A1	64	9	5	2	80	64	73	78
A2	60	8	6	6	80	60	68	74

표 2.5는 2행 4열의 분할표이다. 이 표로부터 $T = 4.1718$를 얻는다. 이와 같은 분석을 모든 인자에 대하여 실시하여 정리한 것이 표 2.6이다.

표 2.6: 표 2.4에서 인자별 AA-통계량을 분산분석표로 종합한 것

변인	자유도	제곱합	평균제곱	F
A	3	4.1718	1.3906	1.47
B	3	0.6809	0.2270	0.24
$A \times B$	3	13.9680	4.6560	4.93**
C	3	11.8684	3.9561	4.19**
D	3	8.9482	2.9827	3.16**
$A \times D$	3	1.2072	0.4024	0.43
e	3	8.2368	2.7456	2.91*
오차	456	430.9188	0.9450	
전체	477	480.0000	*:유의함($\alpha = 0.05$)	

3. 누적법에 대한 비판

Nair는 그의 논문(1986b)을 쓰게 된 동기를 다음과 같이 말하고 있다. "미국에서 다구찌 방법이 널리 수용되고 있으며 이러한 결과로 누적법이 널리 사용되고 있다. 따라서 이 방법의 장단점을 따져서 부적절 하다면 이에 대신되는 단순하고도 효율적인 방법을 제시할 필요가 있다." 흔히 누적법(累積法)이라고 하면 다구찌 특유의 방법이라고 할 수 있는 누적도수표를 작성하여 분산분석을 하고 F-검정과 같은 방법에 의해서 처리효과를 판정하는 방법을 말한다. 그런데 순서 지어진 범주형 분할표에서 누적 도수표를 작성하는것 자체는 전통적 통계학에서도 사용하는 방법으로서 이러한 누적표의 작성이 문제가 되는 것은 아니다. 그러나 다구찌의 누적법에 의한 통계분석의 기본절차로서 분산분석표를 작성하여 F값을 계산하는 것에 대한 비판의 내용을 다음과 같이 요약할 수 있다.

(1) 분산분석표에서 처리제곱합(AA-통계량)만이 빈도 관측치에서 계산된 것이다.

(2) AA-통계량 T는 분산분석표에 제시된 자유도를 가진 χ^2-분포에 따르지도 않을뿐만 아니라 분산비F도 F-분포를 한다고 생각할 수 없다.

(3) 전체제곱합 SS_T는 미리 설계된 표본의 크기에 의해서 정해지는 하나의 상수이다. 그리고 오차제곱합 SS_E는 이 상수에서 통계량 T를 뺀 것이므로, 통계량 T와 서로 독립이 아니다. 또, 오차에 관한 평균제곱 $MS_E = SS_E/(\text{자유도})$는 일반적으로 1에 가까운 값을 갖는다.

Nair(1986a)는 벨 연구소의 "연구보고서 18"로 발표된 논문이다. 이 논문이 다음 해에 미국 통계학회지(JASA)에 발표되어 Nair의 1987년도 논문으로 알려져 있다. Nair(1987)는 다구찌의 누적법을 비판하는 논문이 아니다. 이 논문은 누적법의 본질을 수학적으로 규명하려는 연구 결과이다. 그는 이 논문에서 분할표에서 검정의 일반 부류인 소위 누적카이제곱-형(CCS-type)검정을 고찰하는 논문이라고 하고 있다. CCS-형 검정의 성질을 연구하는 주

무기로 T-통계량을 직교 성분의 가중합계로 표현하는 방법을 사용하고 있다. Nair(1987)논문을 살펴보면 통계량 T를 고유치 문제를 풀어서 $T = \sum_{1}^{K-1} \gamma_j Z_j^2$ 와 같이 분할하는 것이 자세히 설명되어 있다. 여기에서 γ_j는 고유치이고 Z_j는 이에 대응하는 고유벡터를 계수 벡터로 하여 계산한 값이다. Nair(1986b)에서는 산포효과(dispersion effects)의 추정의 중요성을 강조하고 있다. 아울러 다구찌의 누적법은 산포효과의 검출에는 무력하다는 점을 지적하고 있다.

Nair(1986b)에서는 앞에서의 논문(Nair, 1986a)과는 달리 다구찌의 누적법에 대한 결점을 지적하고 대안을 제시하고 있다. 그리고 이 논문에 대한 토론이 이루어지고 있다. 그는 여기에서도 다구찌의 AA-통계량의 특성을 직교성분으로의 분할에 의해서 다음과 같이 설명하고 있다.

$I \times K$-분할표의 각 열의 합계도수가 동일할때(등확률일때) AA-통계량 T의 직교성분으로의 분할은 Nair(1987)에서 논의되고있다. 이때 고유벡터는 체비세프의 다항식 (Chebychev polynomials)의 형태를 갖는다. 즉, j번째 고유벡터 q_j는 j차 다항식의 계수이고 j 번째 고유치는 $\gamma_j = K/[j(j+1)]$와 같다. 등확률이 아닌 경우에도, 시뮬레이션의 결과에 의하면 직교성분 Z_1^2와 Z_2^2이 각각 위치변동과 산포변동에 대응하고, 특히 Z_1^2은 위치변동에 대해서 좋은 검정력을 가지고 있는 것으로 평가된다. Z_1^2의 계수는 $\gamma_1 = K/2$이고 $\sum \gamma_j = K-1$이므로 Z_1^2이 다른 직교성분에 비하여 월등하게 크다면 T값의 대부분을 $\gamma_1 Z_1^2$이 차지하게 되어 결국 다구찌의 T는 위치효과에 대한 검출력이 크게 되는 것이다.

Hamada and Wu(1990)에서 AA-통계량은 가짜 요인효과(spurious factor effects)를 잘못하여 진짜로 오인하여 검출해 내고 인자의 중요성의 순위(order of factor importance)를 사실과 다르게 바꿀 수 있다고 주장하고 있다. 더욱이 순서 지어진 범주형 자료에서 산포효과를 검출해 내는데는 본질적으로 어려운 문제가 있다는 것이다. Z_2^2는 오직 주변빈도의 함수이므로 산포효과와 인자의 각 수준에서의 분포의 상이한 위치에 대한 설명을 동시에 할 수는 없을 뿐만 아니라 AA-통계량을 사용하는 경우보다 심각한 문제를 가져온다는 것이다. 이것은 오직 위치효과가 없을 때 단일인자를 다루는 경우에만 산포효과에 대한 검정을 하는 것이라고 설명할 수 있다는 것이다. 다만 Z_1^2의 위치 효과에 대해서는 Nair(1986)의 주장대로 검출력이 좋다는 것을 인정한다. 그러나, Hamada and Wu(1990)는 다구찌의 누적법은 권장할 수 없다는 결론을 내리고 있다.

4. 누적법의 올바른 사용법

다구찌는 전기통신 전공자이고 이 방면의 연구소에서 연구업무에도 종사했던 사람이다. 그는 2차대전후, 새로운 통계학, 그 중에서도 실험설계방법을 접하고 자기 분야와 여러 가지로 유사한 점이 많다는 것에 놀랐다. 그 후 다구찌는 통계학 공부, 그 중에서도 실험설계방법에 빠져 들었다. 그리고 인도에서의 직교배열에 관한 연구 성과에 매료되었다. 이렇게 해서 다구찌는 실험설계방법에 대한 공부에 몰두하게 된 것이다. 다구찌는 엔지니어로서 번쩍이는 아이디어와 실험설계방법이라는 새로운 도구를 이용해서 품질개선에 효과적인 성과를 얻으려고 노력한 사람이다. 이렇게 해서 그가 개발한 방법들이 순수 통계학자들

의 눈에는 기발하고 참신하기도 하나 이론적 뒤받침이 부족할 뿐만 아니라 때로는 필요이상으로 복잡한 것으로 생각되었다. 그러나 생산현장에서는 이러한 다구찌의 방법으로 능률적이고 효과적으로 품질개선을 달성할 수 있었던 것이다. 이러한 점에서 다구찌의 방법이 미국의 엔지니어들에게 실천가능하고 효과적인 방법인 신선한 아이디어로 받아 들여지고 환영받게 된 것이다.

누적법은 이러한 배경에서 순서 지어진 범주형 자료를 분석하는 왕도(王道)로서 등장한 것이다. 그러나 통계학에서도 이러한 문제를 해결하기 위한 나름대로의 방법들이 연구되어 왔다. 이러한 입장에서 보면 다구찌의 누적법은 불합리한 것이고, 이러한 누적법의 보급은 전통적인 통계적 방법을 오염(汚染)시키는 것이라고 생각 될 수 있다. 이러한 이유로 다구찌의 누적법이 통계적 방법의 정도(正道)로서 수용되지 않고 있는 것이 사실이다. 그러나 통계학자들이 제시하는 방법들은 엔지니어들이 선택해서 이용할 수 있도록 정비되어 있지 못하다. 반면에 다구찌의 누적법은 그 응용방법이 엔지니어들에게 쉽게 실천할 수 있도록 해설되어 있어 친숙한 방법으로 이해되어 왔을 뿐만 아니라 대부분의 경우 큰 오류가 없기 때문에 널리 사용 되고 있는 것이다. 엔지니어들은 순서지어진 범주형 자료의 분석에 자신감을 가지고 있다. 이러한 문제를 해결하기 위하여 제어인자를 잡아내고 직교배열표를 이용한 실험결과에 대해서 누적법으로 문제를 해결하고 있는 것이다.

물론 누적법을 쓰고있는 이러한 문제들도 다른 전통적 통계적 방법으로 해결될 수 있다. 그러나 다구찌방법의 사용자들에게는 누적법이 가장 이용하기 편리한 방법이다. 다구찌방법들 중에서 누적법만은 사용하지 말라고 한다면 '다구찌 방법'의 기본 체계에는 커다란 공동(空洞)이 생기고 결국 무너지고 말 것이다. 그리고 결국 산업계에서 통계적방법의 사용이 크게 위축되고 말 것이다. 엔지니어들에게 품질향상을 위한 실험에 자신을 가지고 통계적 방법을 응용할 수 있는 계기를 마련해 주어야 한다. 다구찌의 누적법은 실용적으로 우수한 것이므로 이것을 잘 보완하여 사용하는 것이 가장 현실적인 해결책이다. 이것이 본 연구에서 얻은 결론이다.

순서형 범주 데이터에 대해서, 결국 다음과 같은 처리 절차에 따를 것을 권장한다.

(1) 다구찌의 누적법에 따라서 AA-통계량을 계산한다.

(2) 직교 성분으로 분할하여 Z_1^2, Z_2^2을 검토한다.

이 두 가지는 제6절에서 제시하는 프로그램 'Accumu'에 의해서 쉽게 계산할 수 있다. 사실 대부분의 경우 이것으로 순서 지어진 범주형 데이터를 분석하여 충분히 그 목적을 달성할 수 있다. 또 예2.2와 같은 경우는

(3) 범주에 점수 수열을 대응시켜 관측치를 가중하는방법으로 SN비를 계산하는 다구찌 방법을 적용해서 분석할 수 있다. 예2.2에 대한 이러한 분석이 Park(1996)에 제시되어 있다.

(1)과 (2)의 방법으로 충분하지 못할 경우는 비-누적법의 적용이 합당한 지를 검토하고 이 방법으로 분석해 본다. 비-누적법이라고 하는 것은 순서형 범주 도수를 누적시켜 고찰하는 방법이 아니라는 뜻이다. 그러므로 물론 다구찌식 누적법도 아니다. 순서 지어진 계수 분류치에 관한 분할표의 경우 누적법이 아닌 방법으로도 분석할 수 있는 것이다. Hammda and Wu(1990)에서는 다음 제5절에서의 예5.1과 관련해서 다음과 같은 방법을 대안으로 제시하고 있다.

(4) 범주의 점수화(fixed scoring categories)

(5) 평균 반응치 모형(mean response model) (SAS/STAT p.478)

(6) 로지스틱 회귀(SAS/STAT p.481)

(7) 비례 오즈 모형(proportional odds model) 등이 권장할 만한 방법이라고 하고 있다. 이들 중에서도 범주의 점수화와 평균 반응치 모형은 방법이 단순하고 또 점수(score)가 적절히 매겨진다면 아주 효과적인 방법이라는 것이다.

5. 예에 의한 설명과 계산

제 6절에 작성되어 있는 SAS/IML 프로그램(파일 이름: Accumu.sas)을 이용한다. 앞으로 이것을 단순히 'Accumu'라고 불러 이용할 것이다. 그리고 여기에서 분석대상은 제2절에 제시되어 있는 예로 부터 시작하자.

(예 2.1)의 약품 B에대해서:

'Accumu'의 모듈 DECOMPO끝의 FINISH다음에

```
DATA={40 24 10  6,
      24 29 16 11};
RUN DECOMPO;
```

와 같이 입력하고 Submit하면 다음 결과를 얻는다.

```
          DATA
              40         24         10          6
              24         29         16         11

   P_CHISQ       TAGUCHI        GD      T_CHISQ       DEF
  7.3269017    12.160222   1.4318237   8.4928204    2.09523
```

피어슨의 카이제곱 값은 7.3269이고 다구치의 AA-통계량 T값은 12.1602이고 T_CHISQ = TAGUCHI/GD = 8.49 는 자유도 2.10의 카이제곱 값을 나타낸다.

Z1	Z2	EIG			
5.8094356	1.4781168	1.906924	0.722186	0.37089	-8.57E-17

Z1, Z2는 각각 직교 성분 Z_1^2와 Z_2^2를 나타내며, EIG는 고유치 $\gamma_1, \gamma_2, \gamma_3$, 그리고 γ_4값을 나타낸다.

약품 A에 대해서도 DATA를 입력하여 이에 상응하는 분석결과를 얻을 수 있다.

(예 2.2)에 대한 분석

'Accumu'의 /** OA-8 **/의 끝의 /** DATA **/다음에 다음과 같이 입력한다.

```
Y={15 3 1 1, 14 3 2 1, 16 2 2 0, 19 1 0 0,
  17 1 0 2,  17 2 1 0, 10 2 4 4, 16 3 1 0};
RP=Y;
AX=D1'*Y; BX=D2'*Y; ABX=D3'*Y; CX=D4'*Y; ERX=D5'*Y; DX=D6'*Y; ADX=D7'*Y;
```

여기에서, 예를들어 DATA=AX; RUN DECOMPO;와 같이 입력하면 다음과 같이 출력된다.

```
DATA
  64        9        5        2
  60        8        6        6

 P_CHISQ      TAGUCHI       GD        T_CHISQ      DEF
2.2787649    4.1718343   1.6906755   2.4675547   1.7744387

    Z1         Z2        EIG
1.8044455 0.4695633    2.161149  0.576947  0.261904 9.996E-18
```

나머지 인자들에 대해서 계산결과를 종합한 것이 다음 표 5.1이다.

표 5.1: 인자별 AA-통계량과 분할

Source	DF	T	Z_1^2	Z_2^2
A	3	4.1718	1.8044	0.1718
B	3	0.6809	0.2078	0.1523
$A\times B$	3	13.9680	6.1701	1.0513
C	3	11.8684	5.3618	0.4495
D	3	8.9482	3.7089	1.2058
$A\times D$	3	1.2072	0.5233	0.1100
e	3	8.2368	3.3046	1.8589

Z_1^2는 어느 것이나 Z_2^2에 비하여 월등하게 크며 AA-T값에 비례하고 있다. 이와 같은 결과는 AA-T값들이 위치 효과를 나타내고 있는 것이라고 판단할 수 있고, 따라서 더 이상의 분석을 시도할 필요가 없을 것으로 생각된다.

표5.1에서 교호작용 $A\times B$와 주효과 C와 D가 중요 인자로 인지되고 최적 인자수준을 찾아낼 수 있게 된다. 범주 데이터에 점수 수열의 가중치를 주어 SN비를 계산하는 것에 의한 분석에서도 같은 결론을 얻는다(Park,1996).

다음 예는 Taguchi(1987, 189쪽)에 수록되어 있고 Hamada and Wu(1990)에서 아크-용접(Arc-Welding) 실험으로 소개되고 비판의 대상으로 논의되고 있는 것이다.

(예 5.1) 두 개의 강철판 사이의 아크-용접을 위한 최적조건을 알아내기 위한 실험이다. 여기에 9개의 제어 인자가 취해졌다. 이것을 A, B, C, D, E, F, G, H, I로 표시한다. 그리고, 교호작용 $A \times G$, $A \times H$, $G \times H$, $A \times C$도 알아보기 위한 실험설계이다. 실험계획은 $L_{16}(2^{15})$-형 직교배열표를 이용하였고 이들 각 조건에서의 반응 관측치로는 관찰된 겉모양(Appearance), X선 검사 결과, 용접 부분의 강도 그리고 작업성(workability)에 관한 것이다. 다음 표에는 작업성에 관한 3개 범주, 즉 쉽다(easy), 보통이다(normal), 어렵다(difficult)의 어느 범주에 속하는가를 판단하여 속하는 범주에 1 그렇지않으면 0으로 기록한 것이 제시되어 있다.

표 5.2: $L_{16}(2^{15})$를 이용한 실험계획과 작업성에 관한 데이터

	직교배열표															관찰치		
	A	G	$A\times G$	H	$A\times H$	$G\times H$	B	D	E	F	I	e	e	$A\times C$	C	작업성		
열번호	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
실험번호																E	N	D
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	0	1	0
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	0	1	0
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	1	0	0
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	0	0
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1	0	0	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	1	0	0
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2	0	0	1
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	0	1	0
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1	0	1	0
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1	0	1	0
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2	0	1	0
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	0	0
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2	0	0	1
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	0	1	0
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	0	1	0

각 실험조건(처리조합)에 대한 작업성이 3범주 중 어느 하나로 판정된 특징이 있다. 따라서 자료가 극단적으로 제한적이어서 결핍자료(the paucity of data) 라는 특징을 가지고 있다. 그러나 다구찌는 여기에서도 다른 경우와 마찬가지로 누적법을 이용하여 분석하고 있다. 예2.2의 경우와 마찬가지로 인자별로 정리한 2행 3열의 분할표를 작성하고, 이것에서 계산한 AA-통계량 T를 계산하고 분산분석표를 작성하고 있다.

여기에서는 우선 AA-통계량과 직교성분 Z_1^2과 Z_2^2를 계산하여 고찰하고 Hamada and Wu(1990)에 따라서 범주의 점수화에 의한 분석, 평균 반응치 모형의 적용, 그리고 비례 오즈 모형을 적용한 분석을 하고 고찰한다. 이와 같은 것을 실행하기 위하여

'Accumu'의 /** OA-16 **/의 끝의 /** DATA **/다음에 다음과 같이 반응치를 입력한다.

```
Y={0 1 0, 0 1 0, 0 1 0, 1 0 0, 1 0 0, 0 0 1, 1 0 0, 0 0 1,
   0 1 0, 0 1 0, 0 1 0, 0 1 0, 1 0 0, 0 0 1, 0 1 0, 0 1 0};
RP=Y;
AX=D1'*Y; BX=D7'*Y; CX=D15'*Y; DX=D8'*Y; EX=D9'*Y; FX=D10'*Y; GX=D2'*Y;
HX=D4'*Y; IX=D11'*Y; ACX=D14'*Y; AGX=D3'*Y; AHX=D5'*Y; GHX=D6'*Y;
Z1X=D12'*Y; Z2X=D13'*Y;
```

여기에서, 예를 들어 PRINT, AX;와 같이 입력하면 다음과 같이 출력된다.

```
AX
3    3    2
1    6    1
```

인자 A에 대한 분석을 위해서는 DATA=AX; RUN DECOMPO;와 같이 입력하여 실행시키면 된다. 모든 인자에 대한 분석결과에서 주목되는 것은 D, F, G의 3개 인자이다. 실험설계에서 G는 교호작용 $D \times F$와 교락되어 있다. 이들 3개 인자에 AA-통계량을 구해 본다.

```
DATA= DX;          /* D에 대한 2x3 분할표 */
RUN DECOMPO;
DATA=FX;           /* F에 대한 2x3 분할표 */
RUN DECOMPO;
DATA=GX;           /* G에 대한 2x3 분할표 */
RUN DECOMPO;
```

다음은 이들 출력 결과를 종합하여 표로 작성한 것이다.

표 5.3: D, F. G에 대한 결과

인자	AA-T	Z_1^2	Z_2^2	γ_1	γ_2
D	5.0256	3.7043	0.4069	1.2774	0.7226
F	9.0256	7.0070	0.1041	1.2774	0.7226
G	5.0256	0.2302	6.5476	1.2774	0.7226

여기에서 각 인자에 대한 Z_1^2와 Z_2^2를 비교해 보면 인자 D, F에서는 Z_1^2이 상대적으로 크고 G에서는 Z_2^2가 상대적으로 크다. 따라서 인자 G는 산포효과를 나타내는 것으로 판단할 것이다.

이 예에 대하여 비-누적법에 의한 분석을 시도해 본다. 우선 절차의 편리를 위하여 행렬 M의 열순서를 이에 할당된 인자의 알파벳 순으로 재배열하여 행렬 MM를 구성한다. 그리고 SAS/CATMOD를 이용할 수 있는 데이터 세트 MRES를 창출한다.

```
MM=C1 || C7 || C15 || C8 || C9 || C10 || C2 || C4 || C11 || C14
      || C3 || C5 || C6 || C12 || C13;
```

SAS/CATMOD를 실행할 수 있는 데이터 세트 MRES를 다음과 같이 창출한다.

```
MM=MM // MM // MM;
Y1=RP[,1]; Y2=RP[,2]; Y3=RP[,3];
W=Y1 // Y2 // Y3;
REP=J(16,1,1) // J(16,1,2) // J(16,1,3);
MD=MM || REP || W;
VNAME={A B C D E F G H I AC AG AH GH Z1 Z2 REP W};
CREATE MRES FROM MD [COLNAME=VNAME];
APPEND FROM MD;
CLOSE MRES;
```

또 하나의 데이터 세트 SCORE를 다음과 같이 창출한다. 이 것은 범주 데이터의 점수화에 의한 비교분석을 위한 것이다.

```
FF=AX // BX // CX // DX // EX // FX // GX // HX // IX //
   ACX // AGX // AHX // GHX // Z1X // Z2X;
PRINT, FF;
S1={0,1,2}; S2={0,1,5}; C={1,-1};
C=BLOCK(C,C,C,C,C,C,C,C,C,C,C,C,C,C,C);
FS1=FF*S1; FF1=FS1'*C; FS2=FF*S2;
FF2=FS2'*C; FM=FF1' || FF2'; FM=ABS(FM);
VNAME={CT1 CT2};
CREATE SCORE FROM FM [COLNAME=VNAME];
APPEND FROM FM;
CLOSE SCORE;
```

이제 다시 IML프로그램 끝의 QUIT을 빠져나가 일반적인 SAS/PROC 절차를 실행한다. PROC PRINT DATA=SCORE; RUN; 데이터 세트 SCORE를 프린트해 보는 것이다. 점수 수열 (0, 1, 5)를 택한 경우가 우리의 검토 대상으로 된다. 이유는 작업이 어렵다는 것이 큰 비용부담과 관계되기 때문이다. 이 결과에서는 인자 D, F, 그리고 G의 효과가 있다고 판단한다. 이들 대비의 분포상태를 고찰하기 위하여 다음 절차를 취한다.

```
PROC UNIVARIATE PLOT; VAR CT2; RUN;
```

다음 두 프로그램(점수열이 0 1 2 일 때와 0 1 5의 경우)은 평균 반응치모형(MEAN RESPONSE MODEL)으로 분석한 것이다.

```
PROC CATMOD DATA=MRES;
WEIGHT W;
RESPONSE 0 1 2;
MODEL REP=A B C D E F G H I AC AG AH GH/ADDCELL=0.01; QUIT;
PROC CATMOD DATA=MRES;
WEIGHT W;
RESPONSE 0 1 5;
MODEL REP=A B C D E F G H I AC AG AH GH/ADDCELL=0.01; QUIT;
```

점수열 (0,1,2)를 사용했을 때는 D, F 두 인자가 중요하고 (0,1,5)를 사용했을 때는 G도 중요 한 인자로 등장한다. 다음 표5.4는 분산분석표의 일부이다.

표 5.4: ANALYSIS-OF-VARIANCE TABLE

		(0,1,2)		(0,1,5)	
Source	DF	Chi-Square	Prob	Chi-Square	Prob
D	1	41.16	0.00	46.40	0.00
F	1	80.28	0.00	60.99	0.00
G	1	1.57	0.21	24.26	0.00
RESIDUAL	2	21.60	0.00	15.42	0.00

다음은 누적 로짓분석을 한 것이다.

```
PROC CATMOD DATA=MRES;
WEIGHT W; RESPONSE CLOGITS;
MODEL REP=_RESPONSE_ A B C D E F G H I AC AG AH GH/ADDCELL=0.25;
QUIT;
```

표 5.5의 결과를 보면 어느 인자도 유의성이 없다. 다만 인자 D와 특히 인자 F는 중요인자 로 부각된다. 이 예의 경우 요인실험 전체를 CATMOD/CLOGITS로 처리한 결과에서는 신통한 결과를 얻지 못한다. 다만 인자 D와 F 2개 인자가 중요 인자로 부상하고 있다. 이 예와 같은 다원 분할표(multi-way contingency table)의 실험자료에서는 각 인자별 로 하나 하나의 인자로 합집(collapsing)한 다음에 인자별로 로짓 분석을 생각할 수 있다. 여기에서 합집한다는 뜻은, 예를 들어 인자 A의 경우 A의 2수준과 3개 범주의 두 방향으로 분류 정리한 2×3 분할표로 작성한 것을 말한다. 이 예의 경우 MODEL REP=_RESPONSE_ D /ADDCELL=0.25;와 같이 하면 인자 D로 합집 한 다음에 로짓분석을 한다는 것을 알 수 있다.

표 5.5: ANALYSIS-OF-VARIANCE TABLE

Source	DF	Chi-Square	Prob
INTERCEPT	1	0.06	0.8052
RESPONSE	1	10.62	0.0011
A	1	0.06	0.8052
B	1	0.06	0.8052
C	1	0.54	0.4606
D	1	1.50	0.2213
E	1	0.06	0.8052
F	1	2.88	0.0894
G	1	0.06	0.8052
H	1	0.06	0.8052
I	1	0.54	0.4606
$A \times C$	1	0.06	0.8052
$A \times G$	1	0.06	0.8052
$A \times H$	1	0.06	0.8052
$G \times H$	1	0.06	0.8052
RESIDUAL	17	6.41	0.9899

다원 분할표에 대한 다구찌의 누적법도 이와 같이 인자별로 합집하는 과정을 거쳐서 실행되었다. 그러나 이와 같은 방법에 문제가 있다는 것은 통계학자들 사이에 잘 알려진 사실이다. 위에서 9개 요인의 실험 결과를 CLOGITS로 분석한 것과 각 인자별로 합집한 다음에 CLOGITS로 분석한 것을 비교하여 보면 물론 분석결과에 차이가 있다. 전자는 16×3로 주어지는 조건별 관측수 열에 대해서 로짓값을 구한 다음에 분석한 것이고 후자는 인자별로 각각 합집한 2×3 분할표에 대해서 로짓값을 구한 것을 분석한 것이므 로 이와같은 차이가 생기는 것은 당연하다고 할 수 있다. 이 예에서는 인자 G의 선형효과 즉 위치효과와 이와 교락되어 있는 교호작용 $D \times F$을 인정할 수 없다. 그런데 표5.3에서 G에 대한 다구찌 통계량의 직교성분을 보면 Z_1^2의 값에 비하여 Z_2^2의 값이 월등하게 크다. 그러나 이것만으로는 요인 G는 산포효과가 있다고 판정할 수 없다는 것이다. 이것은 제어인자가 9개인 요인 실험, 즉 $L_{16}(2^{15})$를 이용한 실험 결과를 인자 G에 관한 2×3 분할표로 합집한 것을 분석한 결과에서 오는 가짜효과일 가능성이 있다는 것이다. Hamada and Wu(1990)는 이점을 지적하고 우려하고 있는 것이다. 결국 다요인 실험결과에서 한 인자에 대해서 계산된 AA-통계량은 그 인자에 대한 합집한것에서 계산된 것이므로 분포의 혼합(mixture)된 것을 비교하는 것이라는 것이다. 그러므로 AA-통계량은 가짜 효과를 검출하는 특성(취약점)을 가지고 있다는 것이다.

6. 컴퓨터 프로그램

```
/* Accumu.sas */
 PROC IML WORKSIZE=50;
 RESET NOLOG;
 START DECOMPO;
 B={1 0 0 0 0 0 0 0 0 0,
    1 1 0 0 0 0 0 0 0 0,
    1 1 1 0 0 0 0 0 0 0,
    1 1 1 1 0 0 0 0 0 0,
    1 1 1 1 1 0 0 0 0 0,
    1 1 1 1 1 1 0 0 0 0,
    1 1 1 1 1 1 1 0 0 0,
    1 1 1 1 1 1 1 1 0 0,
    1 1 1 1 1 1 1 1 1 0};
 PRINT, DATA;
 K=NCOL(DATA); NR=NROW(DATA);   /* 분할표 DATA의 열과 행의 크기 */
 RT=DATA[,+];                                       /* 행별 합계치 */
 CT=DATA[+,]; N=SUM(RT);                    /* 열별 합계치와 총계 */
 KC=K-1;
 B=B[1:KC,1:K];                                  /* 결정된 행렬 B */
 CCT=CT*B';                                /* 열별 합계치의 누적값 */
 CC=CCT/(N-CCT);
 S=0; DO I=1 TO KC-1; DO J=I+1 TO KC;
 S=S+CC[I]/CC[J]; END; END;
 G=1+(2/(K-1))*S; GD=G;
 DEF=(NR-1)*(K-1)/G;                         /* T_CHISQ의 자유도 */
 CW=CCT#(N-CCT);
 W=N*N/CW; W=DIAG(W);                        /* 가중치의 대각행렬 */
 DA=CCT/N; DA=DA';
 DA=DA || DA || DA || DA || DA || DA || DA || DA || DA || DA;
 DAM=DA[,1:K];
 A=B-DAM;                                        /* 구성된 행렬 A  */
 AWA=A'*W*A;
 H=AWA;
 R=DIAG(RT); RI=INV(R);
 YY=DATA[1,]; DO I=2 TO NR; YY=BLOCK(YY,DATA[I,]); END;
 Y=YY';
 HH=BLOCK(H,H,H,H,H,H,H,H,H,H); AWA=HH[1:NR*K,1:NR*K];
 TA=Y'*AWA*Y*RI;
 TAGUCHI=SUM(TA);                        /* 다구찌의 AA-통계량 T*/
 T_CHISQ=TAGUCHI/GD;
 CT=CT'; LAM=CT/N;
 L=DIAG(LAM);
 S=H*L;
 F=ROOT(L);
 FHF=F'*H*F;
```

```
CALL EIGEN(G,W,FHF);
EIG=G';
G=DIAG(G);
V=W'*F;
Q=INV(V);                              /* S=H*L의 고유벡터 행렬 */
G=G[1:KC,1:KC];                        /* 0이 아닌 고유치로 된 G */
Q=Q[,1:KC];                        /* G에 대응하는 고유벡터 행렬 */
SDM=SQRT(RI);
U=Q'*DATA'*SDM;
US=U#U; P_CHISQ=SUM(US);                     /* Pearson의 chi-sq */
PRINT, P_CHISQ TAGUCHI  GD T_CHISQ DEF;
Z1=US[1,+]; Z2=US[2,+];                          /* T의 직교성분 */
UU=U[,1]; DO I=2 TO NR; UU=BLOCK(UU,U[,I]); END;
EI=BLOCK(G,G,G,G,G,G,G,G,G,G); G=EI[1:NR*KC,1:NR*KC]; U=UU;
T=U'*G*U;                                            /* T의 분할 */
TAGUCHI=SUM(T);
PRINT, Z1 Z2 EIG;
FINISH;

/** OA-8 **/
M={1  1  1  1  1  1  1,
   1  1  1  2  2  2  2,
   1  2  2  1  1  2  2,
   1  2  2  2  2  1  1,
   2  1  2  1  2  1  2,
   2  1  2  2  1  2  1,
   2  2  1  1  2  2  1,
   2  2  1  2  1  1  2};
C1=M[,1]; C2=M[,2]; C3=M[,3]; C4=M[,4]; C5=M[,5]; C6=M[,6]; C7=M[,7];
D1=DESIGN(C1); D2=DESIGN(C2); D3=DESIGN(C3); D4=DESIGN(C4);
D5=DESIGN(C5); D6=DESIGN(C6); D7=DESIGN(C7);
/***** DATA *****/

/** OA-16 **/
M={1  1 1  1 1  1 1  1 1  1 1  1 1  1 1,
   1  1 1  1 1  1 1  2 2  2 2  2 2  2 2,
   1  1 1  2 2  2 2  1 1  1 1  2 2  2 2,
   1  1 1  2 2  2 2  2 2  2 2  1 1  1 1,
   1  2 2  1 1  2 2  1 1  2 2  1 1  2 2,
   1  2 2  1 1  2 2  2 2  1 1  2 2  1 1,
   1  2 2  2 2  1 1  1 1  2 2  2 2  1 1,
   1  2 2  2 2  1 1  2 2  1 1  1 1  2 2,
   2  1 2  1 2  1 2  1 2  1 2  1 2  1 2,
   2  1 2  1 2  1 2  2 1  2 1  2 1  2 1,
   2  1 2  2 1  2 1  1 2  1 2  2 1  2 1,
   2  1 2  2 1  2 1  2 1  2 1  1 2  1 2,
   2  2 1  1 2  2 1  1 2  2 1  1 2  2 1,
   2  2 1  1 2  2 1  2 1  1 2  2 1  1 2,
```

```
   2 2 1 2 1 1 2 1 2 2 1 2 1 1 2,
   2 2 1 2 1 1 2 2 1 1 2 1 2 2 1};
C1=M[,1]; C2=M[,2]; C3=M[,3]; C4=M[,4]; C5=M[,5]; C6=M[,6]; C7=M[,7];
C8=M[,8]; C9=M[,9]; C10=M[,10]; C11=M[,11]; C12=M[,12]; C13=M[,13];
C14=M[,14]; C15=M[,15];
D1=DESIGN(C1); D2=DESIGN(C2); D3=DESIGN(C3); D4=DESIGN(C4); D5=DESIGN(C5);
D6=DESIGN(C6); D7=DESIGN(C7); D8=DESIGN(C8); D9=DESIGN(C9); D10=DESIGN(C10);
D11=DESIGN(C11); D12=DESIGN(C12); D13=DESIGN(C13); D14=DESIGN(C14); D15=DESIGN(C15);
/***** DATA *****/
QUIT;
```

참고문헌

[1] 박성현 (1990). 〈응용실험계획법〉. 영지문화사.

[2] 백운봉(1997). 〈품질을 좋게 하는 방법〉. 자유아카데미.

[3] 백운봉 (1989). 累積法. 〈응용통계〉, 4. 1-19. 고대 통계연구소.

[4] Hamada, M. and Wu, C.F.J. (1990). A Critical Look at Accumulation Analysis and Related Methods. *Technometrics*. Vol. 32. 119-162.

[5] McCullagh, P. (1980). Regression Models for Ordinal Data (with discussion). *Journal of Royal Statistical Society, Ser. B*. 109-142.

[6] Nair, V.N. (1986a). Components of Cumulative Chi-squared-Type Tests for Ordered Alternatives in Contingency Tables. *Statistical Research Report 18*, Murray Hill, NJ: AT&T Bell Laboratories.

[7] Nair, V.N. (1986b). Testing in Industrial Experiments With Ordered Categorical Data. *Technometrics*. Vol. 28. 283-311.

[8] Nair, V.N. (1987). Chi-squared-Type Tests for Ordered Alternative in Contingency Tables. *Journal of American Ststistical Association*. Vol. 82. 283-291.

[9] Park, S.H. (1996). *Robust Design and Analysis for Quality Engineering*. Chapman & Hall.

[10] SAS Institute Inc. (1990). *SAS/STAT User Guide*, Version 6, Fourth Ed.

[11] Taguchi, G. (1974). A New Statistical Analysis for Clinical Data, The Accumulating Analysis, in Contrast With the Chi-Square Test. *Saishin Igaku* (最新醫學). Vol. 29. 806-813.

[12] Taguchi, G. (1987). System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs, White Plains, NY: UNIPUB/Kraus.

[1998년 11월 접수, 1999년 3월 최종수정]

On Accumulation Analysis

Uhn Boong Paik [1)] Woo Sun Lee [2)]

ABSTRACT

Taguchi promoted "Accumulation Analysis" as a central method of his analysis strategy. Many statistician, however, believe that accumulation analysis is not a useful technique and that this is unnecessarily complicated, as well as inefficient, and consequently should not be taught or recommended. Nair(1986b) proposed using the first two components of the accumulation analysis statistic seperately to detect location and dispersion effects, respectively. Hamada and Wu(1990) also consider other alternatives - the method of scoring categories, a mean response model, and a proportional odds model. Typical Taguchi users should understand these alternatives. In this article, we prepare SAS/IML MODULE program for accumulation analysis and for decomposition of orthogonal components. We also demonstrate the computing for the other alternatives with examples.

1) 120-901, Hanbit Apt., Eoeun-Dong Yusung-Ku, Taejun city
2) Department of Statistics, Sungshin Women's University, Seoul, Korea