

스코어카드를 활용한 개인 및 지역정보화 수준 측정 사례연구

정형철¹ · 조영임²

¹수원대학교 통계정보학과, ²수원대학교 컴퓨터학과

(2008년 5월 접수, 2008년 9월 채택)

요약

정보화 지수 등 사회의 현상을 특정 지수로 계량화하기 위한 연구에서 설문조사가 많이 활용된다. 그런데, 설문지로 개인의 능력과 같은 정성적 부분을 측정하게 되면, 설문에 응하는 사람의 주관적 판단이 개입될 수밖에 없다. 그러므로 설문지에 표현된 개인의 주관적인 점수를 비교적 객관적 점수로 변경하는 방법이 필요하다. 본 연구에서는 설문조사로 개인의 정보화 수준을 측정하여 이를 기초로 지역의 정보화 수준을 산출하고자 할 때, 개인의 주관적 설문 점수를 스코어카드를 사용하여 비교적 객관적 점수로 변경하는 과정을 소개한다. 스코어카드 작성 시 로짓모형을 사용하였다. 본 연구를 통해 지역의 인구통계학적 분포로 지역의 정보격차를 비교측정 할 수 있는 단초를 제공하였다고 할 수 있다.

주요용어: 지역정보화, 평가지수, 스코어카드, 로짓모형, t-PIS.

1. 서론

전자정부의 실질적 성과에 대한 평가 및 이에 기초한 새로운 추진계획을 수립하고자 여러 국가에서는 다양한 전자정부 활용지수를 개발해 왔다. 예를 들어 192개 전체 유엔 회원국을 대상으로 전자정부 준비단계를 평가하는 UN의 웹 측정 지수, 세계 22개국(한국 제외)의 정부 웹사이트를 평가하는 Accenture의 전자정부 서비스 수준 평가지수, 전 세계 198개 정부 웹사이트를 통해 온라인 서비스 제공 수준을 측정하는 브라운 대학의 전자정부 평가 지수 등이 그것이다 (행정안전부, 2008). 그런데, 언급된 평가 지수들은 관련 기관 연구원이 해당 국가의 웹사이트를 접속한 후 웹사이트 내에서 계량적으로 측정할 수 있는 부분을 중심으로 전자정부 서비스 수준을 측정하고 있기 때문에, 해당 국가의 기업, 개인의 정보화 수준을 반영한다고 하기에는 다소 무리스러운 면이 존재한다. 전자정부 수준 진단을 위해서는 공급자인 정부 뿐 아니라 실수요자인 기업 및 개인의 정보화 수준까지를 아울러 측정할 수 있어야 한다.

국내에서도 우리나라 정보화 정책성과를 국제적으로 비교할 수 있는 기초 계량자료로 국가정보화지수 등을 측정해 왔다. 여기에는 PC 보급률, 인터넷 이용자, 무선초고속인터넷 가입자, TV 보급률 등을 측정한 후 이를 표준점수로 환산한 다음, 가중치를 곱하여 부분별 지수를 산출한 후, 다시 이들 값을 종합하여 종합지수를 작성하고 있다 (한국정보사회진흥원, 2007). 그런데, 이 역시 국제전기통신연합(ITU)의 디지털기회지수 산정 방식과 거의 동일한 방법으로 소프트웨어 측면이 아닌 하드웨어 측면만을 강조하는 측정방식이라고 할 수 있다. 따라서, 정부에서는 전자정부 정책수립을 위해 정부, 기업 및 개인의 전자정부 활용도를 파악할 수 있는 종합지수를 개발하고자 계획 중이다 (행정안전부, 2008).

¹(445-743) 경기도 화성시 봉담읍 외우리 2-2, 수원대학교 통계정보학과, 조교수. E-mail: jhc@suwon.ac.kr

²교신저자: (445-743) 경기도 화성시 봉담읍 외우리 2-2, 수원대학교 컴퓨터학과, 조교수.

E-mail: ycho@suwon.ac.kr

현재, 국민의 정보화 수준을 측정하는 대표적인 지수로 정보격차지수를 들 수 있다 (정보통신부, 2005; 한국정보사회진흥원, 2007). 정보격차지수는 접근지수에 0.3, 역량지수에 0.2, 활용 지수에 0.5의 가중치를 부여하여 100점 만점의 점수를 산출하는 다차원 개념의 계량지표이다. 여기서, 다차원 개념이란 다양한 부분의 통계정보를 활용한다는 의미이다. 그런데, 컴퓨터 보유율과 같이 비교적 객관적인 수치를 산출하는 접근지수에 반해, 역량지수는 설문조사에 의해 컴퓨터 사용능력(9개 문항)과 인터넷 사용능력(8개 문항)을 측정함으로써 비교적 계량화하기 어려운 지수라 할 수 있다. 즉, 역량지수는 개인의 설문결과에 크게 의존하는 단점을 지니고 있다. 그러므로 비교적 계량적인 다른 지수에 반해 역량지수는 객관화 등의 문제로 가장 낮은 가중치가 부여되고 있다.

본 연구에서는 지역정보화와 관련되어 개인 및 지역의 정보화 수준을 측정하는 문제를 다루었다. 우리나라에서 지역정보화 정책이 본격적으로 등장한 것은 1990년 ‘정보사회 종합대책’이 발표되면서부터이다 (한상진과 박재홍, 1999). 초기에는 지역정보화를 지역 내 정보통신 하드웨어 기반의 정비를 의미했으나, 지역 내 행정, 산업, 생활 등의 총체적 정보화 활동으로 확대되었으며, 최근에는 유비쿼터스 환경에서 지역에 기반을 둔 지방자치에 위한 종합정보추진전략이라는 포괄적 의미로 인식되고 있다. 조영임과 정형철 (2006)은 이를 u-지역정보화라 칭하였다. 지역정보화를 공급과 수요 측면으로 구분할 때, 공급 측면에서는 정보시스템 등 하드웨어적 측면이 강조되지만, 수요 측면에서는 정보마인드 진작을 위한 교육, 사용자의 활용 및 만족, 정보 수용, 접근성 및 친화적 인터페이스 등이 강조되고 있다. 그런데, 지금까지 대부분의 정보정책이 공급 측면과 수요측면의 교육 계몽 등에서는 비교적 적극적으로 추진되어 왔지만, 교육 계몽을 제외한 다른 수요 측면은 진척이 더딘 상태라 할 수 있다 (한국정보사회진흥원, 2007; 한세억과 조찬형, 1997). 그러므로 본 연구에서는 수요자 측면에서 지역정보화 수준측정에 관심을 두고, 개인의 정보화 수준을 측정하는 문제를 다루었다. 또한 정형철과 조영임 (2008) 연구의 연속선상에서 개인의 정보화 수준을 객관적인 지수로 변경하는 문제를 다루었다.

정형철과 조영임 (2008)은 수요자 측면에서 개인의 정보화 수준을 측정하기 위해 측정지표를 (1) 역량지표 (2) 활용지표 (3) 수용지표로 선정하였으며, 총 152개의 문항으로 각 지표를 측정하였다. 각 지표의 세부구성에 대해서는 정형철과 조영임 (2008)의 연구를 참고할 수 있다. 그런데, 개인의 정보화 수준 측정이 설문조사 방식으로 진행됨으로 개인의 주관적 판단에 의존한 설문조사 결과가 지수로서 객관성을 지닐 수 있는가 하는 의문이 제기된다. 그러므로 본 연구에서는 이러한 의문을 해결하고자, 설문결과에 전적으로 의존하지 않고, 각 3대 지표에 대한 가중치를 로짓모형을 통해 결정하고, 스코어카드를 통해 설문결과를 필터링시키는 과정으로 새로운 정보화 지수를 유도하는 방법론을 사용하였다. 즉, 3지표를 일정하게 재 범주화하여 각 범주에 새로운 수치를 부여하는 방법으로 1000점 만점의 새로운 개인 정보화 지수를 산출하였는데, 이 새로운 수치를 u-People Information Score(u-PIS)라 칭하였다. 또한 지수 산출 시 스코어카드에 개인의 인구통계학적 속성이 반영되도록 하였다. 이를 통해 개인의 주관적 의사가 반영된 설문에 객관성을 부여하려는 노력을 하였고, 정보화 지수 산출 방법에 인구통계학적 및 지리적 속성을 고려하여야 한다는 점을 강조하였다.

본 논문의 2장에서는 스코어카드를 유도하는 과정을 소개하였다. 3장에서는 개발된 지수의 특징을 분석하였으며 4장에서는 결론과 활용방안을 언급하였다.

2. 스코어카드

2.1. 지표의 특징

개인의 정보화 수준이 지역별 차이가 있으리라는 가정 하에 한국정보사회진흥원 (2006)의 정보화실태조사에 기초하여 전국을 행정교육권역, 주거권역, 상업권역, 산업권역, 농업권역의 5대 권역으로 분리한

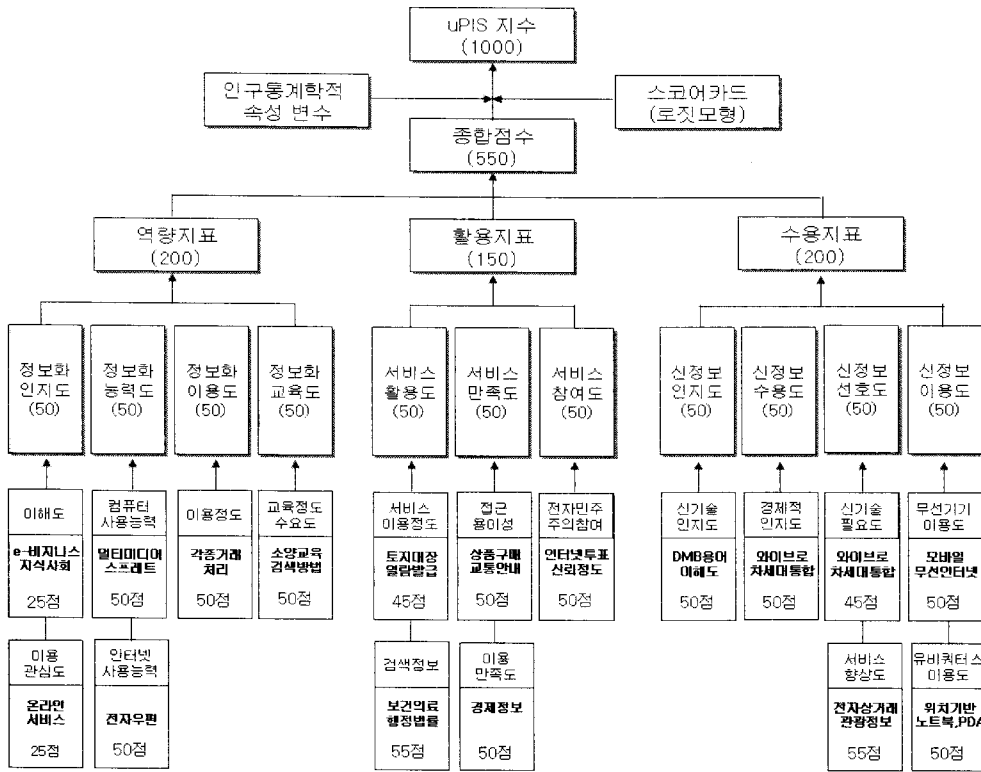


그림 2.1. 개인의 정보화수준을 측정하기 위한 각 지표와 각 지표의 최대값 및 각 지표를 대표하는 대표문항(각 영역별 대표문항의 함은 G/B 분리를 위해 사용되며, 종합점수(SCORE)는 대표문항을 포함하는 각 지표들의 모든 문항들의 합으로 유도됨)

후, 10대 지역에서 설문조사를 실시하여 개인의 정보화 수준을 측정하였다. 실제 조사는 2006년 11월 27일에서 2006년 12월 6일까지 시행되었으며, 총 표본의 크기는 500이다. 표본으로 선정된 500명의 인구통계학적 특징을 살펴보면, 남자는 42.4%, 여자는 57.6%이며, 연령은 30대가 21.6%, 40대가 45.2%로 주로 30-40대가 중심이 되었다. 학력수준으로 고졸이 42.4%, 대졸이상이 38.2%이며, 소득수준으로 300만원 이하가 전체의 67.8%를 차지하였다. 자세한 설문조사과정과 인구통계학적 현황은 조영임과 정형철 (2006)을 참고하길 바란다.

그림 2.1은 개인의 정보화 수준을 측정하기 위해 역량지표, 활용지표 그리고 수용지표가 계산되는 과정을 보여주는데, 각 지표의 종합점수는 550점으로 계산됨을 볼 수 있다 (정형철과 조영임, 2008). 한 예로 역량지표는 정보화인지도, 정보화 능력도, 정보화 이용도, 정보화 교육도라는 4개의 소개념으로 측정되며, 각 소개념들은 다시 10-20여개의 세세문항별로 측정되는데, 역량지표는 최대 200점 만점으로 측정됨을 볼 수 있다. 각 지표의 타당성에 대해서는 정형철과 조영임 (2008)의 결과를 참고하길 바란다.

이제, 산출된 550점의 종합점수는 개인의 주관적 판단이 가미된 점수라는 가정 하에, 각 지역의 특징과 개인의 인구통계학적 속성이 반영된 스코어카드를 통해 1000점 만점의 개인 정보화 지수를 산출하는 과정을 소개하기로 하자. 본 논문의 이하(그림과 표 포함)에서, 550점 만점 종합점수의 영문 표현을 SCORE, 스코어카드로 조정된 1000점 만점의 점수를 u-PIS 또는 ADSCORE로 표기하기로 한다.

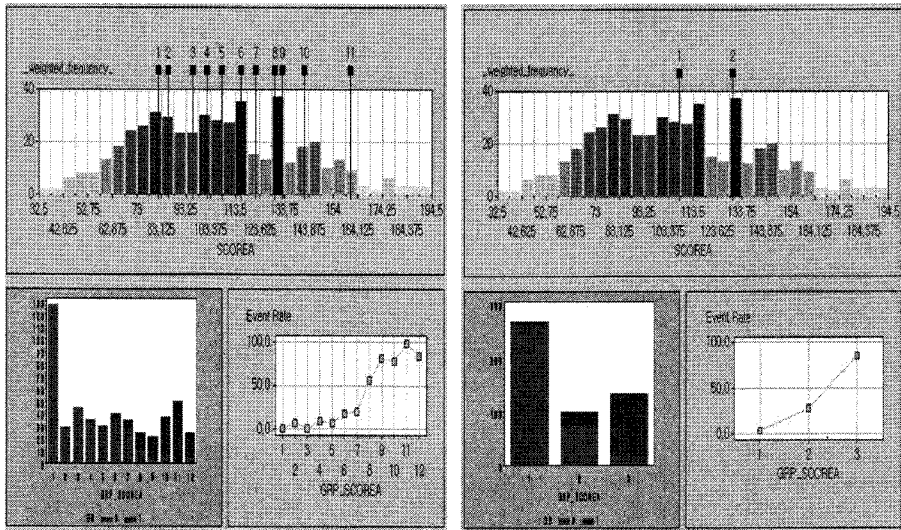


그림 2.2. 역랑지표의 사전 이산화 과정과 사후 이산화 과정

2.2. G/B 분리

스코어카드를 신용평가 기관에서 대출, 신용카드발급, 보험가입 등의 행위에서 기업이 손실을 입지 않도록 소비자의 신용을 측정할 때 주로 사용된다. 초기의 신용평점시스템에는 판별분석 등이 사용되었으나, Fair & Issac사에서 로짓회귀모형으로 스코어카드 방법론을 개발한 이래 본격적으로 스코어카드가 실용화되고 있으며, 국내 소비자금융회사들은 Fair & Issac사에서 점수화한 피코(FICO)점수와 자체의 경험데이터를 결합한 평점시스템을 만들어 고객의 신용리스크를 관리하고 있는 실정이다 (참고, <http://www.fairisac.co.kr>). 일반적으로 신용평가와 관련되어 스코어카드를 만들 때 고객을 Good 그룹과 Bad 그룹으로 분리하여야 하며 이를 G/B 분리라고 칭한다.

개인정보화 지수와 관련되어 Good 그룹은 개인의 정보화 수준이 높은 집단에 속한 사람, Bad 그룹은 그렇지 않은 집단에 속한 사람을 의미한다고 할 수 있다. 그런데, 개인의 정보화 수준과 관련된 사전정보 없이 개인을 정확히 Good과 Bad로 분리한다는 것은 다소 어려운 문제이다. 하지만, 설문조사결과가 개인의 정보화 수준을 반영한다는 가정 하에 설문결과에 기초하여 본 연구에서는 다음과 같은 방법으로 G/B 분리를 시도하였다.

먼저, 본인의 컴퓨터 사용능력과 정보화 수준을 상, 중, 하로 체크하는 문항을 마련하였다. 그리고, 각 지표를 대표하는 대표문항을 크론바흐 알파가 높은 값으로 선발하였다. 그림 2.1은 각 지표에서 선발된 대표문항들을 보여준다. 예를 들어, 그림 2.1에서 컴퓨터 사용능력은 컴퓨터 사용능력을 물어보는 10개의 문항 중 멀티미디어나 소프트웨어 시트 사용 능력과 밀접한 관련이 있음을, 인터넷 사용능력은 10개의 인터넷 사용 능력을 체크하는 문항 중 이메일 사용 능력과 밀접한 관련이 있음을 보여준다. 여기서 크론바흐 알파가 높은 문항들끼리의 단순합을 계산한 후, 이를 두 집단으로 분리하여 컴퓨터 사용능력이나 정보화 역량과의 분할표 분석을 실시하였다. 여기서 각 지표의 크론바흐 알파 상위 문항들의 합이 상위 30%와 하위 70%로 분리될 때, 카이제곱 통계량의 p 값이 최소가 됨을 볼 수 있었다. 그러므로 본 연구에서는 G/B 분리를 위해 크론바흐 알파로 선정된 대표문항들의 합을 목표변수로 잡고, 상위 30% 집단을 Good 그룹, 하위 70% 집단을 Bad 그룹으로 분리하였다.

표 2.1. 스코어카드를 위한 최종 그룹화된 자료의 GOOD 비율(EVENT RATE)

변수	그룹	Cut-off	Event	Non-Event	합계	Event Rate
역량지표	그룹(하)	108미만	6	261	267	2.25
	그룹(중)	130미만	26	74	100	26.00
	그룹(상)	200이하	117	16	133	87.97
활용지표	그룹(하)	76미만	8	197	205	3.90
	그룹(중)	93미만	66	132	198	33.33
	그룹(상)	150이하	75	22	97	77.32
수용지표	그룹(하)	122미만	7	310	317	2.21
	그룹(중)	137미만	40	34	74	54.05
	그룹(상)	200이하	102	7	109	93.58
성별	남자	-	77	135	212	36.32
	여자	-	72	216	288	25.00
연령	20대	-	54	33	87	62.07
	30대	-	28	80	108	25.93
	40대	-	55	171	226	24.34
	50대이상	-	12	67	79	15.19
학력	대졸	-	92	99	191	48.17
	전문대졸	-	15	35	50	30.00
	고졸	-	37	174	211	17.54
	초중졸	-	5	43	48	10.42
거주유형	행정지역	-	14	24	38	36.84
	주거지역	-	115	222	337	34.12
	농업지역	-	10	42	52	19.23
	상업지역	-	8	38	46	17.39
	산업지역	-	2	25	27	7.41
지리유형	산업권역	-	34	37	71	47.89
	행정권역	-	20	32	52	38.46
	농업권역	-	42	79	121	34.71
	주거권역	-	39	139	178	21.91
	상업권역	-	14	64	78	17.95

역량지표, 활용지표, 수용지표를 사용하여 G/B 판정에 대한 의사결정나무분석을 시도하면, 일차적으로 수용지표에서 1차 분류가 일어나며, 다음으로 역량지표에서 2차 분류가 일어난다. 그런데, 활용지표는 G/B 판정에 그다지 중요한 분류 기준이 되지 못함을 의사결정나무분석에서 발견할 수 있었다. 또한, 성별, 연령, 학력, 거주지 유형, 지리 권역의 5대 인구통계학적 변수를 사용하여 G/B 판정에 대한 의사결정나무분석을 시도하면, 가장 먼저 연령에 의한 분류가 발생하는데, 연령이 20대인 집단과 그렇지 않은 집단으로 1차 분류가 된다. 다음으로 학력수준에 의한 분류가 일어나는데, 인구통계학적 변수에서는 연령과 학력이 G/B 판정에 중요한 역할을 함을 사전검토할 수 있었다.

2.3. 이산화 과정

스코어카드에 투입되는 변수는 성별, 연령, 학력, 거주지 유형, 지리 권역의 5가지 인구통계학적 변수와 역량지표, 활용지표, 수용지표의 3개의 연속형 변수이다. 이제, 연속형 변수인 3대 지표를 이산화하는 과정이 필요하다. 이산화과정은 스코어카드에 대한 이해와 적용을 쉽게 하기 위한 것이며, 스코어카드가 특정변수의 특이값에 민감하지 않게 하고 예측력을 높게 하기 위한 목적도 있다 (강현철 등, 2006).

표 2.2. 스코어카드

변수	범주	회귀계수	보정추정	POD	평점	범위	비율
성별	남성	0.0000	0.5082	4.429375	2	2	0.2
	여성	-0.5082	0.0000	4.429375	0		
연령	20대	8.6492	12.5754	4.429375	56	56	5.6
	30대	0.0000	3.9262	4.429375	17		
	40대	-1.0297	2.8965	4.429375	13		
	50대이상	-3.9262	0.0000	4.429375	0		
학력	대졸	24.9596	24.9596	4.429375	111	111	11.1
	초대졸	23.8768	23.8768	4.429375	106		
	고졸	23.0090	23.0090	4.429375	102		
	중졸이하	0.0000	0.0000	4.429375	0		
거주 유형	행정	3.3681	16.3752	4.429375	73	73	7.3
	주거	1.7045	14.7116	4.429375	65		
	상업	0.0000	13.0071	4.429375	58		
	산업	-13.0071	0.0000	4.429375	0		
	농업	1.1801	14.1872	4.429375	63		
지리 권역	행정	-0.4117	14.2701	4.429375	63	65	6.5
	주거	-4.0080	10.6738	4.429375	47		
	상업	-14.6818	0.0000	4.429375	0		
	산업	0.0000	14.6818	4.429375	65		
	농업	-1.2473	13.4345	4.429375	60		
역량 지표	하위	-64.3952	0.0000	4.429375	0	285	28.6
	중위	-27.3283	37.0669	4.429375	164		
	상위	0.0000	64.3952	4.429375	285		
활용 지표	하위	-27.3083	0.0000	4.429375	0	121	12.1
	중위	-2.6522	24.6561	4.429375	109		
	상위	0.0000	27.3083	4.429375	121		
수용 지표	하위	-64.7223	0.0000	4.429375	0	287	28.7
	중위	-38.1776	26.5447	4.429375	118		
	상위	0.0000	64.7223	4.429375	287		
총합						999	100

SAS E-miner를 사용하여 3대 지표를 각각 상, 중, 하 3개의 그룹으로 분리하였는데, 정보지수(information value)가 커지는 방향으로 그룹화를 시도하였다. 그림 2.2는 역량지표에 대해 SAS E-miner 도구를 사용하여 초기 12개의 그룹을 3개의 그룹으로 분리하는 과정의 일부를 보여준다. 다른 변수에도 동일한 방법을 적용하여 연속형 자료를 이산화하였다. 이산화 결과, 8대 변수에 대한 Good 그룹 비율은 표 2.1과 같이 주어진다. 표 2.1에서 event rate는 정보화인(Good 그룹)의 비율을 나타낸다. 연속형 변수인 3대 지표는 cut-off에 의해 상, 중, 하로 분리되었음을 볼 수 있다.

2.4. 스코어카드

이항반응 로짓회귀모형을 사용하여 스코어카드를 계산하였다. 표 2.2의 회귀계수추정치는 정보화인을 $Y = 1$, 비정보화인을 $Y = 0$ 으로 놓고 계산한 로짓회귀모형의 추정된 회귀계수를 의미한다. 여기서, 보정은 회귀계수를 수정한 것이며, 평점은 POD(point of double odds)를 활용하여 스코어의 총합이 1000점이 되도록 조정된 값이다 (강현철 등, 2006; Scallan, 1999). 즉, 보정된 추정치와 평점은 다음과

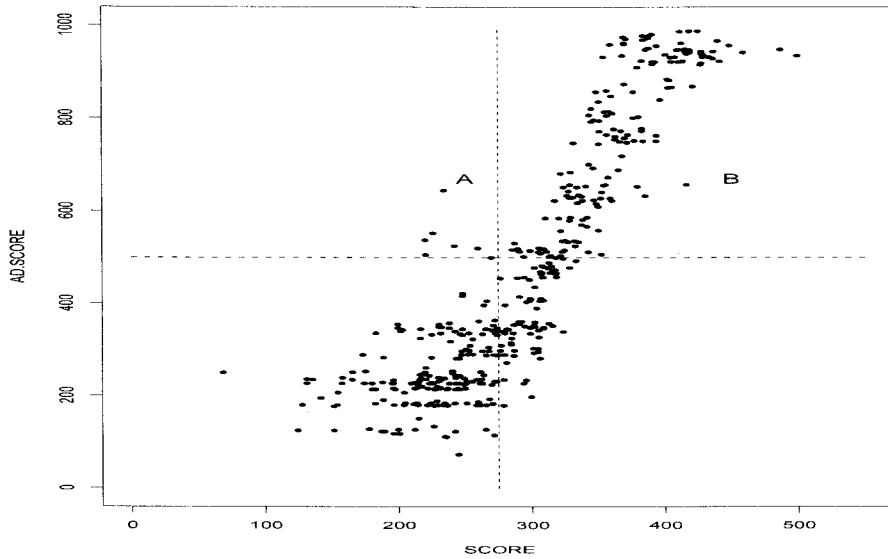


그림 2.3. 설문조사 결과 점수(X축이며 550 만점의 종합점수(SCORE))와 스코어카드로 변환된 점수(Y축이며 1000점 만점인 u-PIS(ADSCORE))

같이 계산된다.

보정된 추정치 = 회귀계수 추정치 - 가장 작은 회귀계수 추정치평균,

$$\text{평점} = \text{보정된 추정치} \left[\frac{\text{POD}(4.429375)}{\log(2)} \right],$$

여기서, 변수의 각 범주와 평점만을 따로 정리한 표를 스코어카드라 한다. 또한 표에서 비율은 각 변수의 영향력이나 중요도 등을 의미한다고 할 수 있다. 정보화점수(u-PIS)를 산출하는 방법은 다음과 같다. 예를 들어, 어떤 사람의 인구통계학적 속성이 남성(2)이고, 40대(13)이며, 초대졸(106)이고, 주거지역(65)에 거주하는데 조사대상 지역이 상업권역(0)에 속하며, 설문조사에서 역량지표는 상(285), 활용지표는 중(109), 수용지표는 하(0)의 집단에 속하는 결과가 나왔다면, 이 사람의 개인정보화 점수는 각 범주에 해당하는 평점을 합한 580점으로 결정되게 된다. 이와같이 각 개인의 3대 지표의 설문점수 SCORE를 그대로 사용하지 않고, 설문에 응한 개인의 인구통계학적 속성을 반영하여 새로운 변형점수를 유도하는 것을, 스코어카드를 사용한 점수산출 과정이라 하며, 이값이 u-PIS이다.

스코어카드의 범위를 중심으로 각 변수의 중요도를 해석하여 보자. 먼저, 범위값이 큰 변수는 수용지표와 역량지표이다. 범위값을 변수의 중요도라고 본다면, 이러한 결과는 앞의 의사결정나무분석 결과와도 일치한다. 여기서 수용지표는 정보화 기술을 받아들이 자세가 되어 있는가를 집중적으로 측정하는 지표인데, 스코어카드는 수용지표 값이 높은 사람들의 개인정보화 점수를 더 강조하는 특징이 있음을 볼 수 있다. 단, 스코어카드는 설문지 상에서 수용지표 점수가 137점부터 150점인 사람(상인 그룹)을 모두 같은 집단으로 보고 동일한 점수를 부여한다는 점이다. 즉, 스코어카드는 개인이 응답한 설문결과에 비례하여 정보화 점수를 산출하는 것이 아니라, 일정 점수구간은 개인의 주관적 편차에 의해 나타난 값일 수 있으므로 이들을 동일 집단으로 고려한다는 점을 주지할 필요가 있다. 그리고, 인구통계학적 속성에서

표 2.3. 구간별 정보화인과 비정보화인 비율 및 K-S 통계량

구간	범위	빈도			G/B 비율		전체		누적비율		K-S
		비정보인	정보인	전체	비정보인	정보인	비정보인	정보인	비정보인	정보인	
1	0-50	2	0	2	100.0	0.0	0.6	0.0	0.6	0.0	0.6
2	50-100	19	0	19	100.0	0.0	5.4	0.0	6.0	0.0	6.0
3	100-150	40	0	40	100.0	0.0	11.4	0.0	17.4	0.0	17.4
4	150-200	85	0	85	100.0	0.0	24.2	0.0	41.6	0.0	41.6
5	200-250	32	0	32	100.0	0.0	9.1	0.0	50.7	0.0	50.7
6	250-300	52	0	52	100.0	0.0	14.8	0.0	65.5	0.0	65.5
7	300-350	24	0	24	100.0	0.0	6.8	0.0	72.3	0.0	72.3
8	350-400	12	0	12	100.0	0.0	3.4	0.0	75.7	0.0	75.7
9	400-450	27	0	27	100.0	0.0	7.7	0.0	83.4	0.0	83.4
10	450-500	35	2	37	94.6	5.4	10.0	1.3	93.4	1.3	92.1
11	500-550	6	5	11	54.5	45.5	1.7	3.4	95.1	4.7	90.4
12	550-600	10	12	22	45.5	54.5	2.8	8.1	97.9	12.8	85.1
13	600-650	6	9	15	40.0	60.0	1.7	6.0	99.6	18.8	80.8
14	650-700	1	4	5	20.0	80.0	0.3	2.7	99.9	21.5	78.4
15	700-750	0	19	19	0.0	100.0	0.0	12.8	100.0	34.3	65.7
16	750-800	0	16	16	0.0	100.0	0.0	10.7	100.0	45.0	55.0
17	800-850	0	10	10	0.0	100.0	0.0	6.7	100.0	51.7	48.3
18	850-900	0	37	37	0.0	100.0	0.0	24.8	100.0	76.5	23.5
19	900이상	0	35	35	0.0	100.0	0.0	23.5	100.0	100.0	0.0
전체		351	149	500	70.2	29.8	100	100	K-S 최대값		92.1

는 학력과 나이가 중요한 역할을 한다는 점이다. 특히, 20대와 20대가 아닌 사람과 스코어에는 큰 차이가 있음을 볼 수 있다. 반면 학력에서 고졸과 초대졸, 대졸 간에는 큰 차이가 발생하지 않는다. 거주유형에 있어서는 산업지역 거주자와 기타지역 거주자와 차이가 있으며, 지리적으로는 상업권역과 기타권역 간 차이가 있다. 여기서 농업권역의 정보화 스코어 비교적 높게 나타나는 이유는 설문조사가 실시된 지역이 순수 농업지역이라기보다 정보화마을로 선정된 지역이기 때문이다.

2.5. 타당성평가

그림 2.3은 설문조사에 응한 개인의 정보화 점수(3대 지표만을 고려한 설문결과 점수)와 스코어카드를 사용해서 인구통계학적 특징까지 반영한 u-PIS 점수와의 관계를 보여준다. 두 점수사이의 상관관계는 0.893으로 당연히 두 점수 간 일치성이 높다. 그림을 통해 u-PIS의 몇 가지 특징을 살펴보자. 그림 2.3 A의 근방은 원점수(SCORE)는 낮으나 u-PIS가 높은 사람들이며, B 영역들은 그 반대이다. 예를 들어 B에 속한 사람들은, 전반적으로 정보화 수준이 낮은 속성을 지닌 집단에 분류된 사람 중 본인 스스로는 그 집단에 속한 사람과 비교하여 상대적으로 정보화 수준이 높다고 자신을 판단하는 사람이라고 할 수 있다. 그런데, u-PIS는 해당 개인의 주관적 판단을 어느 정도 반영하되, 그가 속한 집단이 전반적으로 정보화 수준이 낮으므로 정보화 지수값을 감소시키는 방향으로 움직인다. 이는 u-PIS가 설문지 결과에 인구통계학적 집단의 속성을 반영하여 점수를 재조정하고 있기 때문이다.

이제, 개발된 스코어카드의 타당성을 평가하여보자. 스코어카드를 설문조사에 응한 개인들을 정보화 수준에 따라 순위화하도록 설계되며, 전체적으로 Bad 그룹에 속한 개인이 Good 그룹에 속한 개인보다 낮은 평점을 지니도록 하였다. 따라서 Good 그룹과 Bad 그룹에 속한 개인에 대한 스코어 분포가 잘 분리되는

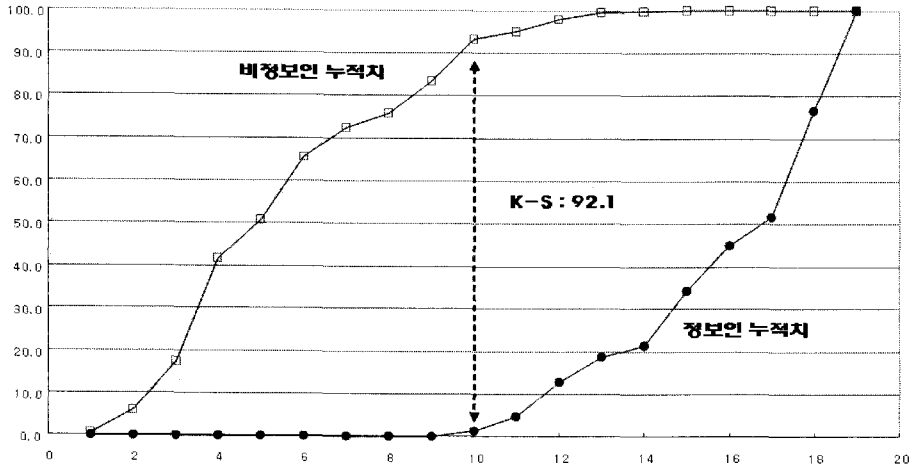


그림 2.4. 정보화인과 비정보화인의 누적비율

것이 바람직하며, 이러한 분리의 정도를 측정하기 위해 K-S(Kolmogorov-Smirnov) 통계량과 ROC 곡선이 스코어카드의 타당성 평가에 사용된다 (강현철 등, 2006).

표 2.3은 u-PIS 점수를 50점 간격으로 분리한 구간에 대한 Good 그룹(정보화인)과 Bad 그룹(비정보화인)의 빈도 및 G/B 비율, 전체의 G/B비율(전체) 그리고 누적비율을 보여준다. 여기서 K-S 값은 누적비율의 차이를 나타내는데 이중 가장 큰 값인 10구간의 92.1이 K-S 통계량이다. 그림 2.4는 G/B 그룹의 누적비율을 그림으로 표시한 것이다. 본 연구에서는 제시하지 않았지만, 표 2.3으로부터 정보화인과 비정보화인으로 분류된 개체에 대한 민감도(sensitivity)와 특이도(specificity) 값으로 ROC 곡선을 유도할 수 있는데, ROC 곡선 아래의 면적값(c-통계량)은 0.985로 K-S 통계량과 마찬가지로 상당히 높게 나타난다. Elizabeth (2003)에 의하면, K-S 통계량 값이 지나치게 좋다고 할 수 있으나, 본 연구에서는 설문결과에 의존하여 G/B 분리를 하였으므로 이러한 결과는 당연하다고 판단된다. 결과적으로 개발된 표 2.2의 스코어카드는 상당히 우수하다고 할 수 있는데, K-S 통계량을 기준으로 특정 개인을 정보화인과 비정보화인으로 분류하는 경계점으로 u-PIS 점수 500점을 선택할 수 있다. 즉, 스코어카드 기준으로 500점을 기준으로 500점 아래면 비정보화인, 500점 이상이면 정보화인으로 분류할 수 있다.

3. u-PIS 현황분석

u-PIS는 인구통계학적 속성과 3대 지표의 특징이 복잡하게 연관된 점수이다. 이제, u-PIS 점수를 분석함으로써 인구통계학적 연관성이 어떻게 표현되는가를 살펴보기로 한다.

3.1. 분산분석

표 3.1은 설문결과와 종합점수(SCORE)와 u-PIS에 대한 인구통계학 변수의 분산분석 결과이다. 모든 인자에 대해 설문지 상의 종합점수 SCORE에서 통계적 유의성을 발견할 수 있는데, u-PIS에서는 그 차이가 더 강하게 나타난다는 점이다. 이는 u-PIS가 인구통계학적 특징에 따라 점수를 각기 다르게 부여받기 때문에 나타나는 매우 당연한 결과이다. 그런데, 직업변수는 스코어카드에 반영되지 않으므로써, 표 3.1에서 직업의 F값 변화가 거의 없다는 점을 볼 수 있다. 이는 직업 내에서 다른 인구통계학적 속

표 3.1. 종합점수(SCORE)와 u-PIS에 대한 분산분석(직업은 u-PIS 계산에 사용되지 않았음)

변수	SCORE(550점만점)		u-PIS(1000점만점)		F 변화량
	F값	유의확률	F값	유의확률	
성별	7.63	0.0059	12.50	0.0004	4.87
연령	22.74	0.0001	27.90	0.0001	5.16
학력	16.06	0.0001	29.66	0.0001	13.60
직업*	14.44	0.0001	14.66	0.0001	0.22
거주유형	6.19	0.0001	6.68	0.0001	0.49
지리권역	7.44	0.0001	14.95	0.0001	7.51

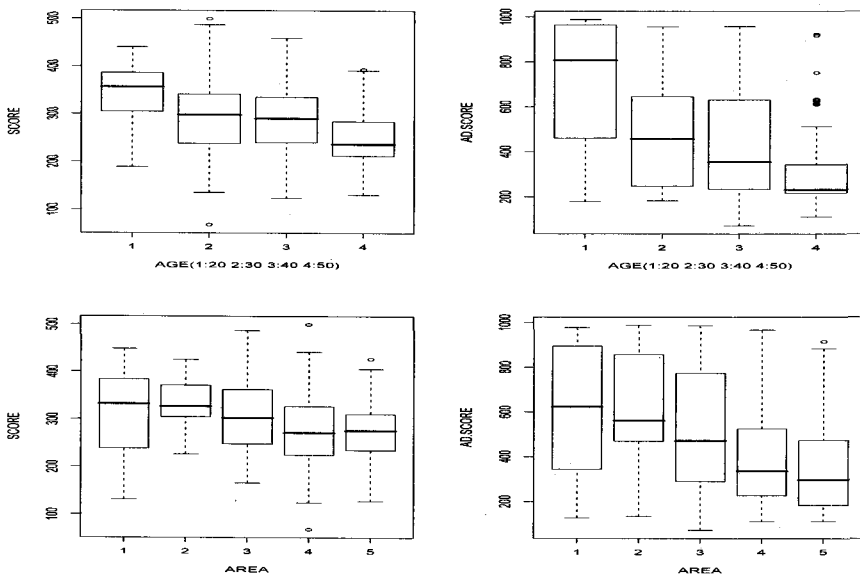


그림 3.1. 연령별(위그림), 지리권역별(아래그림), SCORE(왼쪽그림)와 u-PIS(오른쪽그림)의 상자그림(1: 산업권역, 2: 행정권역, 3: 농업권역, 4: 주거권역, 5: 상업권역)

성이 비교적 균일하게 분포되고 있기도 하지만, 스코어카드에 투입되지 않은 변수에서는 SCORE와 u-PIS 간에 큰 변화가 발생하지 않는다는 점을 나타냄으로 스코어카드에 투입되는 변수 선정의 중요성을 반증한다고 하겠다. 이제, 상자그림을 통해 인구통계학적 요인별 평가지표의 합인 SCORE와 u-PIS를 비교하여 보자.

그림 3.1은 연령별, 5대 지리권역별(행정, 주거, 상업, 산업, 농업 순) 상자그림이다. 상자그림에서 보면, 두 점수의 범위가 다르다는 점을 감안하더라도, SCORE보다 u-PIS 점수의 집단 내 분산이 더 클 수 있다. 먼저, 연령별 특징을 살펴보자. 연령별 SCORE는 30대와 40대에서 비교적 차이가 보이지 않았으며, 연령의 증가에 따라 SCORE의 중앙값이 계단식 감소하는 경향이 있으나, u-PIS에서는 지수적 감소가 일어남을 볼 수 있다. 또한, u-PIS는 20대와 다른 연령대와의 차이를 보다 분명하게 보여 준다. 단, 20대에서 u-PIS의 범위가 상당히 넓은데, 이는 20대의 인구통계학적 다른 요인의 조합이 매우 다양하다는 것을 의미한다고 하겠다. 즉, 20대의 평균적 정보화 수준은 높으나, 20대 내에서도 인구통계학적 여러 요인과 지리적 요인에 의해 정보격차가 매우 크게 발생하고 있음을 보여준다. 반면 50대

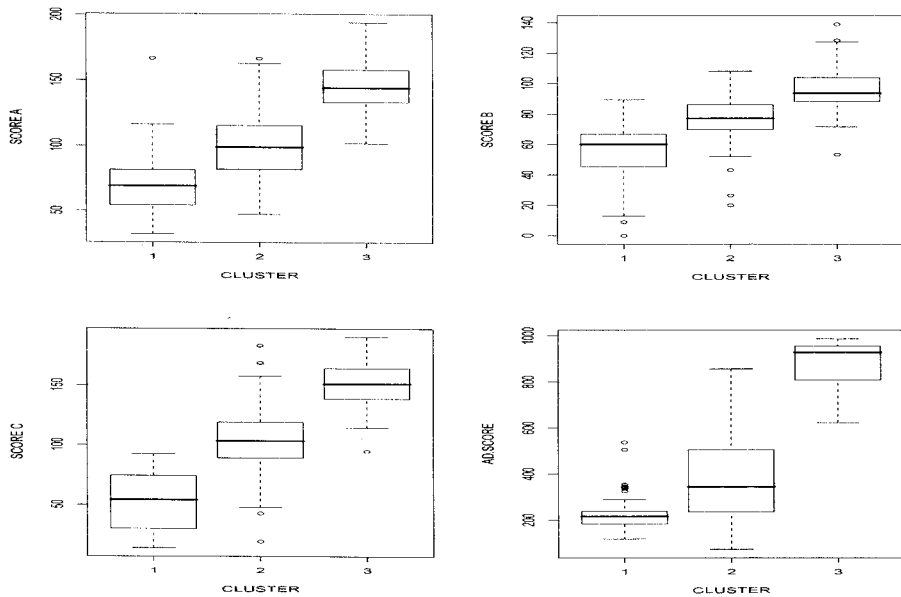


그림 3.2. 3개 군집에 대한 역량지표(SCOREA), 활용지표(SCOREB), 수용지표(SCOREC), u-PIS(ADSCORE)의 상자그림

표 3.2. 정보화 상종하 3그룹 별 각 지표의 평균

정보화 수준		표본크기	역량지표	활용지표	수용지표	종합점수	u-PIS
군집1	하	66	70.235	54.826	53.803	178.864	226.030
군집2	중	315	98.886	77.429	104.752	281.067	381.467
군집3	상	119	146.445	97.036	150.744	394.225	880.605
최대값			200	150	200	550	1000

이상(그림에서 60대를 따로 제시하지는 않았지만 특히 60대 이상)의 u-PIS는 원 점수에 비해 분산이 더 좁아, 60대들의 다른 인구통계학적 특성이 서로 유사하며, 이에 따라 60대 연령층에서는 서로의 정보화 수준이 다소 비슷하게 나타나고 있다고 유추할 수 있다. 이제, 지역정보화와 관련지어, 한국정보사회진흥원에서 분류한 지리권역별 특징을 상자그림으로 살펴보자. 지리권역별 SCORE와 u-PIS의 평균적 추이는 큰 차이가 없으나, 앞에서 언급한 바처럼 u-PIS는 지리권역별 분산을 다소 크게 함을 볼 수 있다. 특정 지역의 정보화 수준은 그 지역의 인구통계학적 요인에 크게 영향을 받는데, u-PIS의 분산이 크다는 것은 해당 지역의 인구통계학적 속성이 매우 다양하고 복잡하다는 사실을 나타낸다고 할 수 있다. 또한, 설문조사에 응한 지역주민의 설문으로만 정보화 수준을 측정하는 것에 비해, 실제 특정 지역의 정보화 수준을 측정한다는 것은 상당히 어렵고 힘든 일임을 반증한다고 하겠다. 그러므로 하드웨어 중심에서 벗어나 수요자 중심의 지역정보화를 실현하기 위해서는 해당 지역이 어떤 인구적 속성을 지니고 있는가를 파악하고, 인구통계학적 속성에 따라 지역정보화 전략을 수립해야 함을 본 u-PIS는 제시하고 있다. 또한, 한 지역의 수요자 중심 정보화 수준은 그 지역의 인구통계학적 속성에 의존한다는 사실에 따라, 지역정보화는 ‘인구통계학적 속성을 반영하는 지역적 특성이 강한 행정정보시스템’으로 개발되어야 한다는 점을 제안하고자 한다.

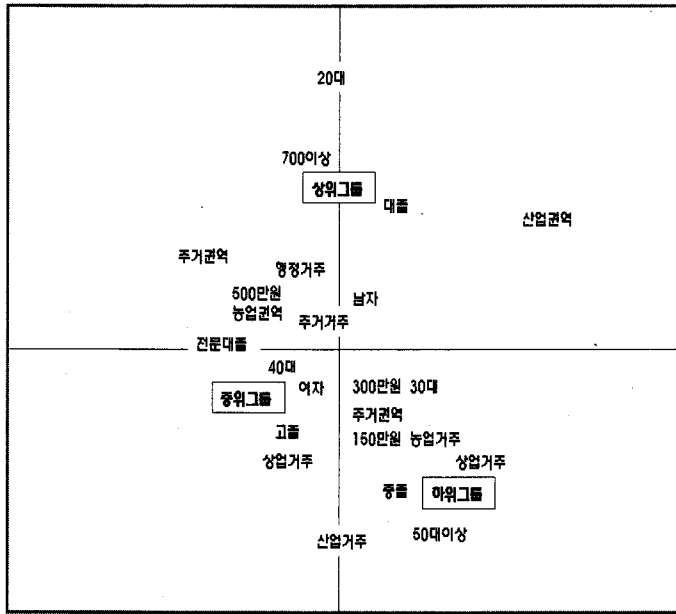


그림 3.3. 인구통계학적 요인과 정보화 그룹과의 다중대응분석

3.2. 군집분석 및 다중대응분석

인구통계학적 요인과 지리권역적 특징을 u-PIS 점수와 연관지어 살펴보고자 다중대응분석을 실시하였다. 다중대응분석을 실시하기 이전에 u-PIS 점수를 범주화 하였는데, 이를 위해 3대 지표와 u-PIS 값을 사용하여 K-평균 군집분석을 실시하였다. 2차원 주성분 그림에서 주축을 중심으로 아래에 존재하는 군집, 중간에 존재하는 대군집 그리고 주축의 위에 존재하는 군집이 존재함을 감지할 수 있었으므로 군집의 개수를 3으로 설정하였다. 이는 개인의 정보화 수준이 상인 그룹, 중인 그룹 그리고 하인 그룹이라고 설명할 수 있으므로 3개의 군집으로 지역주민을 분리하는 것은 상식에 부합되는 결과라 할 수 있다.

그림 3.2와 표 3.2는 3개 군집에 대한 각 지표의 상자그림과 평균을 보여준다. 군집분석 결과 정보화 지표가 상인 그룹, 중인 그룹, 하인 그룹으로 구분이 되는데, u-PIS가 하인 그룹과 상인 그룹은 그룹 내 범위가 비교적 좁으나, 중인 그룹은 범위가 다소 넓은 것을 볼 수 있다. 이러한 군집분석의 목적은 인구통계학적 요인 및 지리적 특성과 정보화 상중하 그룹과의 다중대응분석을 위한 사전단계이다.

이제, 군집과 인구통계학적 요인 및 지리권역별로 대응분석을 실시하였다. 그림 3.3은 인구통계학적 여러 요인과 3대 군집과의 다중대응분석, 그림 3.4는 지리권역(10대 권역)과 3대 군집과의 다중대응분석이다. 대응분석을 통하여 정보화 세 그룹과 해당 인구통계학적 요인 및 지리권역별 대응관계를 살펴보기로 하자.

인구통계학적 요인과 관련되어, 정보화 상위그룹의 특징은 연령별로는 20대, 소득수준은 700만원 이상, 학력으로는 대졸이상과 관련이 있음을 볼 수 있다. 정보화 중위그룹은 그룹 내에서 u-PIS 점수 편차가 크지만, 연령별로는 40대, 성별로는 여성, 학력수준으로는 초대졸이나 고졸과 대응관계를 이룸을 볼 수 있다. 정보화 하위그룹은 연령별로는 50대 이상, 주거지역은 상업지역이나 농업지역, 소득수준도 비교적 낮음을 볼 수 있다.

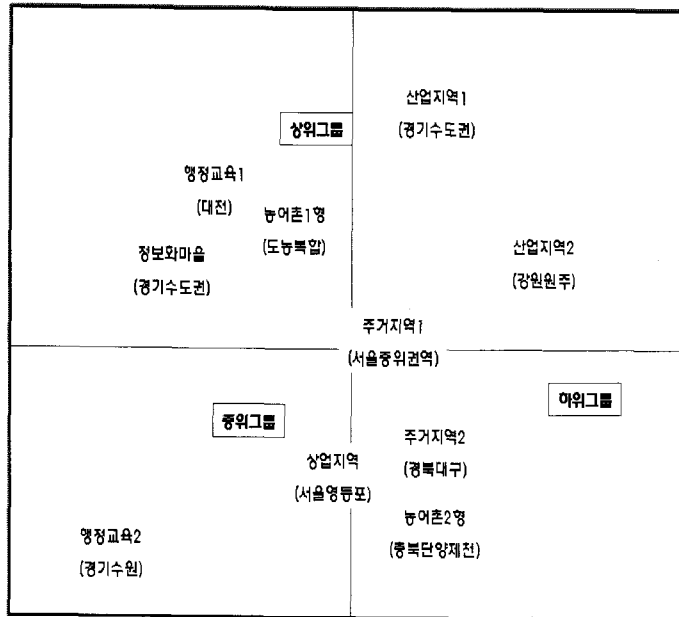


그림 3.4. 10대 지리권역별 다중대응분석

이제, 10대 지리권역별 대응인 그림 3.4를 살펴보자. 정보화 상위그룹은 대전과 경기 수도권, 중위그룹은 서울 상업지역(영등포구)과 대응관계가 있으며, 하위그룹과 가까운 대응을 보이는 권역은 잘 보이지 않는다. 서울 주거권역으로 분리된 지역은 전체 그림의 가운데에 위치함으로 정보화 상, 중, 하의 모든 성격을 다 지닌 지역임을 관찰 할 수 있다.

본 연구에서 사용한 표본의 크기가 500에 불과하기에 설명력이 다소 낮음에도 불구하고 대응분석의 결과가 어느 정도 상식적인 결과를 유도하였다고 볼 수 있다.

4. 결론 및 시사점

범국가적 차원에서 유비쿼터스 시대에 부합되게 개인의 정보화 능력을 평가하고, 정보화 점수를 측정하는 객관적인 방법을 개발하고자 노력하고 있다. 지금까지 기업, 교육기관, 지역 주민 등에 대한 정보화 수준평가 지표 및 모형들이 많이 연구되어 왔으나, 유비쿼터스 시대에 맞게 지역정보화를 달성하기 위한 수혜자 입장의 정보화 능력 평가 방법론은 부재할 뿐 아니라, 지역주민을 대상으로 하는 객관적인 평가 지표 역시 부족한 편이라고 할 수 있다 (정형철과 조영임, 2008). 이에 따라 본 연구에서는 평가지표의 객관성 확보 방안으로 정형철과 조영임 (2008)에서 연구한 지역주민의 정보화 수준 측정 설문에 대해, 3대 지표를 스코어카드로 변경하는 방법과 스코어카드 점수의 특성을 분석하는 내용을 다루었다.

본 연구의 핵심은 개인의 정보화 수준에 기초하여 지역정보화 수준을 측정할 때, 개인의 설문 결과를 그대로 점수화 하지 않고, 적당한 범위로 분리한 후 동일 그룹에게는 동일 점수를 부여하는 스코어카드를 사용하여 개인정보화 수준을 측정하였다는 점에 있다. 또한 스코어카드에 개인의 인구통계학적 속성과 지리적 속성을 투입하여 비교적 객관적 지수를 산출하고자 노력하였다는 점이다. 물론, 스코어카드를 사용했을 때 각 지표의 점수를 분리하는 경계값에 따라 스코어가 상당히 큰 차이가 날 수 있는 단점이

존재한다. 하지만, 설문지 상에서 ‘아주만족’과 ‘만족’ 그리고 ‘보통’ 등의 판단이 개인적 주관 및 인식의 정도에 따라 엄격하게 수량화할 수 없다는 점을 생각한다면, 일정한 점수대를 같은 그룹으로 변경하여 동일한 점수를 부여하는 방법이 보다 합리적이지 않는가하는 생각이 든다.

현재, 차세대 전자지방정부 추진 시 10대 고려사항에 다양한 전자정부 서비스 및 행정, 법률, 보건 관련 콘텐츠 개발의 중요성과 더불어, 개인정보화 역량강화 및 특성화를 위한 지자체 정보화 전략 개발이 포함되어 있다 (조영임과 정형철, 2006). 개인정보화 역량강화 및 지자체 정보화전략은 해당 지역의 인구통계학적 특성 및 환경요인에 의존하며, 이를 위해 지역주민을 대상으로 하는 개인정보화 수준을 측정하는 문제가 제기되고 있는 현실에 기초하면, 본 연구에서 제안한 스코어카드를 사용한 u-PIS는 개인의 정보화 수준을 측정 및 지역정보화 수준 측정에 하나의 단초를 제공한다고 하겠다.

향후 u-PIS의 재개발 과정 주기는 신용평가기관에서 실시하는 스코어카드 변경 주기와 비슷하게 3-5년 정도라고 생각된다. 그 이유는 IT 기술의 급격한 변화로 지역주민 수요가 변화되고 있으며, 그로 인해 정보화 전략도 다변화되어야 하고, 궁극적으로 유비쿼터스 시대로의 변화에 따라 지역사회가 고도화 단계에 진입하기 때문이다. 따라서 기존의 u-PIS 측정 지표 역시 변경되어야 하며, 스코어카드 역시 일정 주기에 따라 변경되어야 한다.

본 연구에서 언급한 u-PIS를 에이전트 중심의 지능형 시스템으로 전환하는 과정을 추후 연구로 다루고 있는데, u-PIS가 지능형 시스템 상에서 구현되는 것을 u-PISS라 하였다. 에이전트는 task agent engine, knowledge, communication 등 3가지 주요한 컴포넌트에 의해 자동성, 사용자 대리성, 지능성을 갖고 다양한 틀을 이용하여 에이전트 자신의 목적을 수행하게 된다 (Cho, 2007). 추후 개발될 u-PISS를 통해 개인 맞춤형 정보화 전략이 여러 관련 실행과 함께 시스템에서 제공될 수 있게 되면, 국가나 지방자치단체의 정보화전략 수립에 많은 도움이 되리라 예상된다.

참고문헌

- 강현철, 한상태, 최종후, 이성건, 김은석, 엄익현, 김미경 (2006). <고객관계관리(CRM)을 위한 데이터마이닝 방법론>, 자유아카데미, 서울.
- 정보통신부 (2005). <정보격차지수 조사>, 한국정보사회진흥원, 서울.
- 정형철, 조영임 (2008). 지역정보화 수준 평가를 위한 문항분석: 크론바흐 알파 계수에 의한 중요변수 선택, *Journal of the Korean Data Analysis Society*, 10, 1073-1087.
- 조영임, 정형철 (2006). <지역정보화의 효율적 추진을 위한 지역주민의 정보화능력 평가지표 연구>, 한국정보사회진흥원, 서울.
- 한국정보사회진흥원 (2006). <정보화통계집>, 한국정보사회진흥원, 서울.
- 한국정보사회진흥원 (2007). <국가정보화백서>, 한국정보사회진흥원, 서울.
- 한상진, 박재홍 (1999). 지역정보화 계획의 흐름과 지역별 특성, <사회과학연구>, 17, 87-100.
- 한세익, 조찬형 (1997). <지역정보화 정책 이론과 실제>, 한국정보문화센터, 서울.
- 행정안전부 (2008). <전자정부 활용지수 개발 추진계획>, 행정안전부.
- Cho, Y. I. (2007). Intelligent multi agent application system in AI system, *Intelligent Journal of Artificial Life and Robotics*, Plenary talk, Japan.
- Elizabeth, M. (2003). <신용평점의 이해와 활용>, 이파로스, 김민정 역, 서울.
- Scallan, G. (1999). *Bulding Better Scorecards*, Austin, TX: Austin Logistics.

A Case Study on the Information Capacity Assessment Index of Local Residents Using Score Card

Hyeong Chul Jeong¹ · Young Im Cho²

¹Dept. of Applied Statistics, University of Suwon; ²Dept. of Computer Science, University of Suwon

(Received May 2008; accepted September 2008)

Abstract

In this study, we consider to assess the information ability of individuals related to local information. Questionnaire surveys are frequently used for studies on the measuring of information index and such specified indexes of society development. However, if questionnaire papers measured individuals' abilities and such parts of quality, the subjective judgments of respondents cannot but be intervened. Therefore a method which will enable subjective points to be comparatively objectified is needed. In this study, when an attempt is being made to measure individuals' information capacity through questionnaire surveys as the process for calculating the standards of regional information levels. Cases have been looked at for times when, by using a scorecard, individuals' subjective survey result have been be changed to be comparatively objectified scores and such process. The logit model was used in the preparation of the scorecard. In this study, we have presented a clue which enables the comparative measurement about regional information gaps by using characteristics of regional population statistics.

Keywords: Information capacity, assessment index, scorecard, logit model, u-PIS.

¹Assistant Professor, Dept. of Applied Statistics, University of Suwon, Suwon, 445-743, Korea.
E-mail: jhc@suwon.ac.kr

²Corresponding Author: Assistant Professor, Dept. of Computer Science, University of Suwon, Suwon, 445-743, Korea. E-mail: ycho@suwon.ac.kr