

Statistical Discriminant Analysis on the Driving Ability of the Brain-injured¹⁾

Jaehee Kim²⁾ · Jeonga Kim³⁾

Abstract

Brain injured patients who had the driver's license before the injury of the brain were tested with the newly developed tool CPAD by Hanyang Medical School and the National Rehabilitation Center. The CPAD contains many variables to measure the ability of driving. Also for each patient the American standard CBDI score was measured and the result was compared with the CPAD results. Of interest is to classify the patients as pass, border, fail group after the CPAD test. To derive the discriminant functions with the group information based on CBDI, parametric/nonparametric and multivariate/univariate discriminant analysis was performed and discussed.

1. 뇌손상 장애인 운전능력 판별문제

현대 사회에서 자동차의 중요성이 점점 더 커지고 있는 가운데 대중교통의 이용에 제한이 있는 장애인에게는 일상생활에 있어서 더 필수적이다. 그러나 신오수 등(2000)에서 지적한대로, 타인이나 자신에게 신체적, 경제적, 정신적 손상을 입힐 수 있는 사고의 가능성도 함께 가지고 있으므로 안전성에 대한 고려를 해야 한다. 자동차 운전을 위해서는 적절한 반응을 위한 지각, 인지기능과 실행능력들이 필요하다. 또한 뇌손상 장애인은 신체기능, 인지기능, 지각기능은 물론 의사소통, 감정 등의 여러 부분에 문제가 발생하여 운전능력에 영향을 받을 수 있다. 따라서 이전에 운전이 가능했더라도 뇌손상 이후의 운전능력을 다시 평가하여 안전한 운전의 가능성을 전문적으로 판단해주어야 한다.

뇌손상 장애인의 운전능력에 대해서는 다양한 연구가 이루어져 왔다. 뇌손상 장애인의 운전능력을 12가지 인지-지각 평가 도구로 검사하여 이들 도구와 운전능력 간에

1) 본 연구는 2004년 덕성여대 자연과학연구소 연구지원을 받았습니다.

2) 제1저자 : 서울시 도봉구 쌍문동 419 덕성여자대학교 정보통계학과 부교수
E-mail : jaehee@duksung.ac.kr

3) 서울시 도봉구 쌍문동 419 덕성여자대학교 수학 및 통계학과 대학원 석사과정

상관관계가 있다고 알려졌다. Barbara와 Mazer(1998)는 9가지의 지각기능 검사도구와 도로 주행 시험을 비교하여 Motor Free Visual Perception Test(MVPT)와 Trail Making B test를 같이 사용하여 더 높은 운전능력 예측도를 얻을 수 있다고 하였다.

Engum 등(1988)은 뇌손상 장애인을 위한 Bracy's Cognitive Rehabilitation Program을 응용하여 뇌손상 장애인의 운전능력 선별 검사인 Cognitive Behavioral Driver's Inventory(CBDI)를 개발하였고, 도로 주행 시험과 비교하여 이 도구의 민감도가 95.5%, 특이도가 100%라고 보고하였다. 그중에서 뇌손상 장애인을 위한 Bracy's Cognitive Rehabilitation Program을 응용하여 뇌손상 장애인의 운전능력 선별 검사인 Cognitive Behavioral Driver's Inventory(CBDI)를 개발하였다. 국내에서는 이 CBDI를 이용하여 뇌손상 장애인의 운전능력을 평가한 연구에서 유용하지만 허안나 등(2003)에서 설명한 바에 의하면, 국내 실정에 맞지 않는다는 판단 하에 개발이 필요하다고 하였다. 아직 국내에는 뇌손상 장애인의 인지, 시지각 기능을 반영할 수 있는 적절한 운전능력 평가 도구가 없는 실정이기 때문이다. 이에 국립재활원에서는 한양대학교 의공학교실과 함께 뇌손상 장애인의 운전을 위한 인지-지각능력 평가도구(Cognitive-Perceptual Assessment for Driving : CPAD)를 개발하였다. 이 CPAD를 이용하여 2003년 4월부터 국립재활원의 운전클리닉에 참여하여 운전능력 평가를 받은 뇌손상 장애인 100명을 대상으로 CBDI 판정 결과로부터 얻은 그룹정보를 이용하여 합격(pass), 경계(border), 불합격(fail)의 세 그룹으로 나누는 그룹판별을 하고자 한다.

2. 뇌손상 장애인 운전능력에 대한 데이터

2003년 4월부터 국립재활원의 운전클리닉에 참여하여 운전능력 평가를 받은 뇌손상 장애인 100명을 대상으로 하였다. 대상군 모두는 발병 이전에 운전면허가 있었고, 실제 운전 경험이 있었다. 수용성 실어증(receptive aphasia)이 있어 지시 따르기를 못하는 경우나, 혼자 앉기가 불가능하거나, 운동 실조증이 있는 경우는 제외하였다. 대상군은 남자 95명, 여자 5명이었고, 연령별대로는 50대가 가장 많았다. 마비 부위는 좌측 편마비 58명, 우측 편마비 42명으로 좌측 편마비자가 많았고, 원인 질환은 뇌경색이 50명, 뇌출혈이 46명으로 많았다. 각 환자에 대해 마비부위, 학력, 장애등급, 운전면허 종류, 운전경험, 사고경험, 과거병력 등이 측정되었고 특히 운전능력에 관해 측정된 변수는 <표 1.2>에서 보여주며 여기서 task는 장애인 운전능력을 판단하기 위해 개발된 과제를 말한다.

CPAD 검사로 구해진 10가지 변수를 살펴보면 각 변수를 구성하는 검사항목은 정답수, 반응시간 등으로 그 값의 성격과 단위가 서로 달라 결과값을 일률적으로 처리할 수 없어 각 변수의 결과값을 평균이 50, 표준편차가 10이 되도록 표준화하여 표준화한 변수에는 원래의 변수이름 앞에 z를 붙이기로 한다. 환자의 검사점수가 50일 경우 이는 이 군에서 평균적 수행능력을 보인 것을 의미하고 그 보다 높으면 평균보다 더 좋은 수행능력을, 낮으면 평균보다 못한 수행능력을 보이는 것으로 의미하도록 하였다. 또한 운전자의 전반적인 운전능력에 관한 10개의 변수로부터 표준화와 주성분을 이용해 이미 생성된 변수 td11을 운전능력점수라고 부르기로 한다.

설명변수의 수는 10개이며, 외적기준은 3개(pass(1), border(2), fail(3))이다. 각 그룹의 도수분포는 <표 1.1>에 나타나있다. 또한 각 변수에 대해서 전체표본과 각 그룹별

로 기술통계량을 통해 기본적인 특성을 파악하였으나 결과표는 생략하기로 한다.

<표 1.1> CBDI 판별에 의한 그룹 빈도표

CBDI	외적기준	명수	비율
1	pass(합격)	36	0.36
2	border(경계)	27	0.27
3	fail(불합격)	37	0.37

<표 1.2> 뇌손상 장애인 운전능력에 대한 측정 변수

변수명	표준화변수명	변수 설명
cbdots		CBDI 검사후 점수
cbdipf		CBDI 검사에 의한 판정결과1:pass, 2:border, 3:fail
ont		주행시험(on road test) 총점
onpf		on-road 합격여부
t1	zt1	task1의 맞은 횟수
t3	zt3	task3의 맞은 횟수
t4	zt4	task4의 맞은 횟수
t5	zt5	task5의 맞은 횟수
t6	zt6	task6의 맞은 횟수
t7	zt7	task7의 맞은 횟수
t1stm	zt1stm	1단계 맞은 경우-평균시간(msec)
t5tm	zt5tm	맞은 경우 -평균시간(msec)
t81m	zt81m	task8의 1단계 진행시간(msec)
t82m	zt82m	task8의 2단계 진행시간(msec)

3. 뇌손상 장애인의 운전능력 판별분석

판별(discrimination) 및 분류(classification)분석은 집단에 대한 정보로부터 집단을 구별할 수 있는 판별함수(discriminant function) 또는 판별규칙(discriminant rule)을 만들고, 새로운 개체에 대해 어느 집단에 속하는지를 판별하여 분류하는 다변량 기법으로 집단에 대한 정보를 이용한 탐색적인 통계 기법이다. 판별분석의 주요목적은 소속 집단에 속하는 관측값들을 분류할 때 오분류 확률을 최소화할 수 있는 독립변수들의 선형함수 혹은 이차함수를 찾는 것이라고 할 수 있다. 본 연구에서는 과거 운전 경력이 있는 뇌손상 장애인에 대해 컴퓨터를 이용한 운전능력검사 후 CBDI에 의한 판정결과와 CPAD 측정변수들을 이용하여 합격, 경계(판정보류), 불합격에 대한 판정을 내릴 수 있는 판별함수를 구하고자 한다.

3.1 다변량 분산분석과 그룹간 거리

다변량 판별분석에서 선형판별함수를 이용하기위해서 기본 가정은 첫째, 각 설명변수들이 다변량 정규분포를 이루어야 하고, 둘째, 각 설명변수들의 공분산행렬(covariance matrix)이 동일해야 한다. 표본의 크기가 클 경우에는 선형판별함수보다 2차판별함수를 사용하는 것이 더 좋다고 알려져 있다. Velilla와 Barrio(1994)는 데이터가 정규분포를 따르지 않을 경우 변환식을 통해 근사적으로 정규분포를 따르게 만든 후 정규분포를 따를 경우의 판별함수를 이용하여 그룹을 판별하는 방법을 제안하였다.

장애인 운전능력 판별에 의미있는 변수들에 대해 일변량 분산분석 결과 모두 매우 유의하며 (p -값 <0.01) 이들 변수들에 대해 다변량 분산분석 결과 Wilks Lambda=0.3479 (근사적 p -값 <0.0001)로 그룹간의 평균벡터에 대한 차이가 매우 유의함을 알 수 있다.

그룹간의 거리를 구하기 위해 합동공분산을 이용한 마할라노비스거리(Mahalanobis distance)는 <표 3.1>과 같으며 이는 분류된 집단 상호중심까지의 거리이다. 그룹2는 경계그룹으로 판정보류그룹인데 합격 그룹인 그룹1과의 거리나 불합격 그룹인 그룹3과의 거리가 비슷함을 알 수 있다.

<표 3.1> 각 집단 중심까지 마할라노비스 거리

CBDIF	1(pass)	2(border)	3(fail)
1(pass)	0	2.94905	7.68862
2(border)	2.94905	0	2.62275
3(fail)	7.68862	2.62275	0

3.2 그룹간 공분산행렬에 대한 동일성 검정

다변량 정규분포를 이용해 판별함수를 구할 경우 공분산행렬의 동일성여부에 따라 선형/2차 판별함수가 결정되므로 공분산행렬이 동일한가에 대한 가설검정을 수행하고자한다. 검정결과 <표 3.2>에서 보면 세 그룹간 공분산행렬이 유의수준 1%, 5%에서 동일하다고 할 수 없음을 알 수 있다. 그러므로 각 집단의 공분산행렬이 같다고 할 수 없으므로, 각 집단의 공분산행렬이 판별분석에 고려되어야한다. 이 절에서는 각 집단의 공분산행렬이 고려된 이차판별함수를 구하고자한다.

<표 3.2> 공분산행렬의 동일성 검정 결과

카이제곱통계량	자유도	p-값
172.1061	90	<0.0001

3.3 선형판별함수

다변량 정규분포를 가정하고 집단간 공분산행렬이 같다고 가정한 후 선형판별함수를 구하고자한다. 선형판별함수는 이해하기 쉬운 함수로 판별함수의 계수를 통해 그룹 판별에 비중있는 역할을 하는 변수를 알 수 있게 된다.

첫 번째 선형판별함수는

$$G_1(\mathbf{X}) = 84.37 + 0.28zt1 + 0.52zt3 + 0.47zt4 + 0.31zt5 + 0.48zt6 + 0.01zt7 + 0.76zt1stm + 0.17zt5m - 0.19zt81m + 0.27zt82m \quad (3.1)$$

이고, 두 번째 선형판별함수

$$G_2(\mathbf{X}) = 70 + 0.31zt + 0.53zt3 + 0.34zt4 + 0.33zt5 + 0.50zt6 - 0.03zt7 + 0.63zt1stm + 0.18ztm - 0.09zt81m + 0.08zt82m \quad (3.2)$$

이며, 마지막으로 세 번째 선형판별함수는

$$G_3(\mathbf{X}) = -56.87 + 0.30zt1 + 0.46zt3 + 0.32zt4 + 0.29zt5 + 0.48zt6 - 0.04zt7 + 0.62zt1stm + 0.08ztm - 0.10zt81m + 0.02zt82m \quad (3.3)$$

와 같이 구해진다.

분류함수의 능력을 판단하는 방법으로 오분류(misclassification)의 확률로서 오류율(error rate)과 정확한 분류율(correct classification)을 이용한다. 재대입(resubstitution)분류, 표본분할(partitioning the sample), 교차타당성(cross validation) 방법 등을 이용하여 오류율을 계산할 수 있는데 본 연구에서는 재대입 분류에 의한 오류율 계산을 이용하였다. 앞에서 구한 선형판별함수를 이용하여 재대입분류에 의해 <표 3.3>의 오분류표를 얻었으며, 정확한 분류율은 $(30 + 11 + 29)/100 = 0.7$ 이며 30%의 오분류율을 보인다.

3.4 2차판별함수와 오분류율

세 그룹에 대해 각각의 공분산행렬을 고려하여 표본으로부터의 추정량을 이용하면 다음의 2차형식의 판별함수

$$Q_i(\mathbf{X}) = \ln p_i - \frac{1}{2} \ln |S_i| - \frac{1}{2} \bar{\mathbf{X}}_i' S_i^{-1} \bar{\mathbf{X}}_i - \frac{1}{2} \mathbf{X}' S_i^{-1} \mathbf{X} + \bar{\mathbf{X}}_i' S_i^{-1} \mathbf{X} \quad (3.4)$$

를 구할 수 있다. 여기서 $i = 1, 2, 3$ 에 대해 p_i 는 i 번째 그룹에 대한 사전확률(prior),

\mathbf{X} 는 관측벡터, $\bar{\mathbf{X}}_i$ 는 i 번째 그룹의 표본평균벡터, \mathbf{S}_i 는 i 번째 그룹의 표본공분산 행렬, $\bar{\mathbf{X}}$ 는 전체 표본평균벡터를 나타낸다. 2차 판별함수에 의한 오분류표 <표 3.4>에서 보면, 정확한 분류율은 $(34 + 19 + 30)/100 = 0.83$ 이고 오분류율은 $1 - 0.83 = 0.17$ 으로 17% 오분류율을 보인다.

<표3.3> 운전능력 관련 변수들에 대한 선형판별함수의 함수의 오분류표

CBDI 판정	분류결과			행합
	1(pass)	2(border)	3(fail)	
1(pass)	30	6	0	36
2(border)	10	11	6	27
3(fail)	3	5	29	37
열합	43	22	35	100
PRIORS	0.36	0.27	0.37	

<표3.4> 운전능력 관련 변수들에 대한 이차판별함수의 함수의 오분류표

CBDI 판정	분류결과			행합
	1(pass)	2(border)	3(fail)	
1(pass)	34	2	0	36
2(border)	6	19	2	27
3(fail)	4	4	30	37
열합	43	25	32	100
PRIORS	0.36	0.27	0.37	

3.5 정준판별분석

정준판별분석(canonical discriminant analysis)은 주성분분석과 정준상관계수를 결합한 차원축소기법이라고 할 수 있다. 그룹 간에 가능한 가장 큰 다중 상관계수를 갖게 되도록 변수들의 선형조합을 구하게 되며, 이와 같이 얻어진 선형조합을 첫 번째 정준변수(the first canonical variable)라 하고 여기서 가장 큰 다중 상관계수를 첫 번째 정준상관계수(the first canonical correlation)라 한다. 두 번째 정준변수는 첫 번째 정준변수와는 서로 독립이며 그 다음으로 큰 다중 상관계수를 갖게 되도록 변수들의 선형조합을 구한다.

첫 번째 정준변수는 그룹간의 차이를 최대한 반영하는 변수가 되며 이러한 이유로 정준변수는 판별함수(discriminant function)라고 부르기도 한다.

각 객체에 대한 판별점수는 비표준정준계수와 대응하는 판별변수와의 선형결합을 통해 얻어지며 구해진 정준판별함수들은 다음과 같다:

$$\begin{aligned} \text{정준판별함수1} = & -0.00997zt1 + 0.0193zt3 + 0.057zt4 + 0.0089zt5 - 0.0004zt6 \\ & - 0.0106zt7 + 0.0476zt1stm + 0.0287zt5m - 0.0340zt81m \\ & + 0.0909zt82m \end{aligned} \quad (3.5)$$

$$\begin{aligned} \text{정준판별함수2} = & 0.0247zt1 + 0.0596zt3 - 0.547zt4 + 0.0402zt5 + 0.0213zt6 \\ & - 0.0612zt7 - 0.0576zt1stm + 0.0700zt5m - 0.0539zt81m \\ & - 0.0644zt82m. \end{aligned}$$

정준판별함수1에서는 zt_{82m} (task8의 2단계 진행시간)이 상대적으로 가장 큰 공헌을 하고 있고, 그 다음은 zt_4 (task4의 맞은 횟수), 그리고 zt_{81m} (task8의 1단계 진행시간), zt_7 (task7의 맞은 횟수)은 음의 값을 취하고 있다. 한편 정준판별함수2에서는 zt_5m (맞은 경우-평균시간)이 가장 크고 둘째로 음의 값인 zt_{82m} 를 고려할 때 뚜렷한 대조관계를 이루는 함수로 생각된다. 정준변수에 해당하는 고유값은 1.4421, 0.1632로 정준판별함수1의 비중이 90%가 된다. <표 3.7>은 각 그룹별 정준변수의 중심점을 보여준다. 새로운 개체에 대해서는 정준판별함수를 계산한 후 그룹 중심점과 거리가 가까운 그룹으로 판별하게 된다.

<표 3.5> 각 그룹에 대한 정준변수 중심점

CBDI 판정	정준판별함수1	정준판별함수2
1(pass)	1.417622690	-0.242062679
2(border)	-0.033086107	0.676902127
3(fail)	-1.355164648	-0.258435162

3.6 단계적 판별분석

분류모형에 포함된 독립변수의 수가 많으면 그 중에는 집단을 구분하는데 중요한 독립변수가 있는가 하면 그렇지 못한 변수가 포함될 수 있다. 이러한 경우에는 분류력이 유의한 독립변수들을 선택하여 모형을 단순화하는 측면과 판별에 기여하는 변수 선택 측면에서 도움을 주는 분석으로 볼 수 있다. 그러므로 판별분석에서 유의한 변수를 선택하는 방법 중 본 연구에서는 단계적 변수 선택(stepwise selection)을 이용하고자 한다. 여기서 단계별 변수선택방법을 사용하기 위해 그룹간 공통 공분산행렬을 가지며 변수벡터는 다변량 정규분포를 따른다고 가정하며, 각 단계에서 부분 F-검정에 의해 변수가 선택되고 제거되는 과정을 반복하게 된다.

이러한 방법에 의해 선택된 변수는 zt_3 , zt_4 , zt_{1stm} , zt_5m , zt_{81m} , zt_{82m} 으로 <표 3.6>에서 선택과정에서의 통계량들을 보여준다. Wilks Lambda는 그 단계의 변수까지 포함된 벡터에 대한 다변량 분산분석 통계량을 나타내고, 그 단계의 변수까지 포함한 정준변수와 그룹을 나타내는 변수와의 정준상관계수는 적어도 2개 이상의 그룹을 잘 분리해 낼수록 1에 가까운 값을 갖게 되는데 <표 3.6>에서는 0.35로 경계그룹이 합격 그룹과 불합격그룹에 걸쳐있기 때문으로 여겨진다.

선택된 변수들로만 선형판별함수와 이차판별함수를 구해 분류해 본 결과 오분류표 <표 3.7>, <표 3.8>을 얻었으며 전체적인 오분류율은 30%, 20%로 변수를 모두 포함했을 때의 오분류율과 거의 같다.

<표 3.6> 단계적 판별분석 결과 선택된 변수

단계	들어온 변수	부분 R^2	부분 F-통계량	부분 F-검정 p-값	Wilks Lambda	정준상관계수 제곱
1	zt82m	0.4316	36.82	<.0001	0.56844643	0.21577679
2	zt4	0.1414	7.91	0.0007	0.48804723	0.25720827
3	zt1stm	0.1189	6.41	0.0024	0.43001135	0.29374015
4	zt5m	0.0676	3.41	0.0373	0.40094210	0.32068121
5	zt81m	0.0491	2.40	0.0962	0.38125009	0.33478139
6	zt3	0.0421	2.02	0.1382	0.36519831	0.35108665

<표3.7> 선택된 변수들에 대한 선형판별함수의 함수의 오분류표

CBDI 판정	분류결과			행합
	1(pass)	2(border)	3(fail)	
1(pass)	31	5	0	36
2(border)	9	11	7	27
3(fail)	3	6	28	37
열합	43	22	35	100
PRIORS	0.36	0.27	0.37	

<표3.8> 선택된 변수들에 대한 이차판별함수의 함수의 오분류표

CBDI 판정	분류결과			행합
	1(pass)	2(border)	3(fail)	
1(pass)	35	1	0	36
2(border)	6	17	4	27
3(fail)	4	5	28	37
열합	45	23	32	100
PRIORS	0.36	0.27	0.37	

4. 운전능력점수에 의한 일변량 판별분석

4.1 정규분포를 가정한 판별분석

운전능력에 대한 총체적 점수로 볼 수 있는 운전능력점수는 운전능력평가 항목 모두를 고려한 일변량 변수인 td11만을 이용한 일변량 판별분석을 하고자한다.

우선 td11에 대해 일변량 분산분석 결과 $F=35.27$ (p -값<0.0001)으로 그룹간 평균 차이가 유의하다. 또한 그룹간 분산에 대한 동일성 검정결과 $\chi^2 = 32.420$ (p -값<0.0001)으로 그룹간 분산 차이가 유의하므로 분산의 이질성을 고려할 필요가 있다.

확률밀도함수추정을 이용해 비모수적 판별분석과 더불어 선형, 이차판별함수를 구해 오분류율을 구하고자 한다. 우선 td11에 대한 기초통계량은 <표4.1>과 같이 나타난다.

등분산을 가정하고 구한 선형판별함수는

$$\begin{aligned} G_1(x) &= -45.219 + 1.604 \cdot td11 \\ G_2(x) &= -38.769 + 1.477 \cdot td11 \\ G_3(x) &= -28.836 + 1.274 \cdot td11 \end{aligned} \quad (4.1)$$

와 같이 구해진다. <표 4.2>는 선형판별함수에 의한 오분류표로 32% 오분류율을 보인다. <표 4.3>은 분산의 이질성(heterogeneity)을 반영하여 2차판별함수에 의한 오분류표로 37% 오분류율을 보여 분산의 특성을 고려했음에도 불구하고 오분류율은 오히려 증가했다. 그룹별 분산을 다르지만 선형적 경향이 강하므로 이와 같은 현상이 나타난 것으로 보인다.

<표4.1> 운전능력점수 td11에 대한 기초통계량

CBDI 판정	N	평균	분산	표준편차
전체	100	49.653	57.884	7.608
1	36	55.093	9.077	3.012
2	27	50.718	22.767	4.771
3	37	43.744	66.380	8.147

<표4.2> td11의 선형판별함수에 대한 오분류표

CBDI 판정	분류결과			행합
	1(pass)	2(border)	3(fail)	
1(pass)	34	2	0	36
2(border)	15	4	8	27
3(fail)	6	2	30	37
열합	55	8	38	100
PRIORS	0.36	0.27	0.37	

<표 4.3> td11의 2차판별함수에 대한 오분류표

CBDI 판정	분류결과			행합
	1(pass)	2(border)	3(fail)	
1(pass)	32	4	0	36
2(border)	14	7	6	27
3(fail)	4	10	24	37
열합	50	21	30	100
PRIORS	0.36	0.27	0.37	

4.2 비모수적 판별분석

그룹에 대한 확률밀도함수에 대한 비모수적 추정에 근거하여 비모수적 판별분석을 할 수 있다. Silverman(1986)은 확률밀도함수 추정에 대해 자세히 다루고 있으며 커널 함수(kernel function)를 이용한 방법이나 k-이웃점 방법(k-nearest-neighbor method)

등을 사용하여 각 그룹에 대한 비모수적 확률밀도함수를 추정할 수 있다. 이렇게 추정된 확률밀도함수는 그룹 분류의 기준으로 사용할 수 있게 된다. 커널함수로는 균일(uniform), 정규(normal), Epanechnikov, biweight, triweight 커널함수 등을 사용할 수 있다.

본 연구에서는 td11 변수에 대해 Epanechnikov 커널함수를 이용한 비모수적 판별 분석을 하고자한다.

x 에서 그룹 t 에 대해 커널함수를 이용한 확률밀도함수는

$$f_t(x) = \frac{1}{n_t} \sum_y K_t(x-y) \quad (4.2)$$

와 같이 추정된다. 여기서 그룹 t 에 속한 모든 y 에 대해 계산하게 된다. 여기서 사용된 커널함수로 Epanechnikov 커널함수는 다음과 같다.

$$K_t(x) = \begin{cases} \left(1 - \frac{1}{r^2} \frac{a^2}{V^2}\right), & \text{if } \frac{a^2}{V^2} \leq r^2 \\ 0 & \text{elsewhere} \end{cases} \quad (4.3)$$

그룹 t 에 대한 사전확률(prior probability) q_t 를 고려하여 그룹 t 에 속한 개체에 대한 사후확률(posterior probability)은

$$p(t|x) = \frac{q_t f_t(x)}{f(x)} \quad (4.4)$$

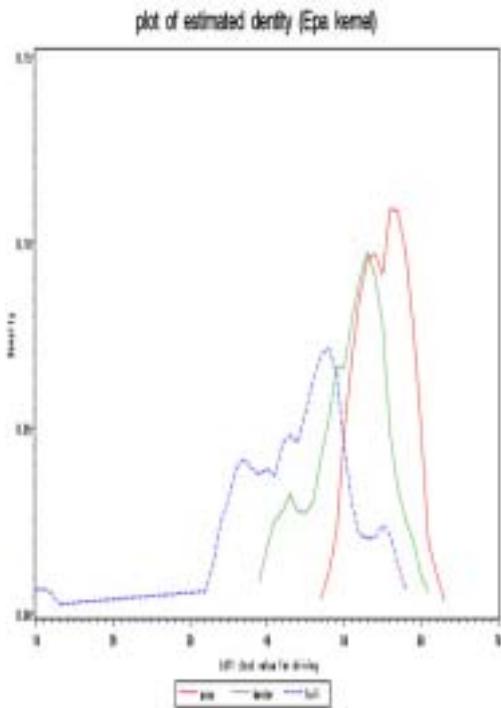
이다. 여기서 $f(x) = \sum_u q_u f_u(x)$ 로 추정된 확률밀도함수이다.

Epanechnikov 커널함수를 이용하여 확률밀도함수와 사후확률을 추정한 후 분류표는 <표 4.4>이고 정확한 분류율은 $(34 + 0 + 31)/100 = 0.65$ 이며 35%의 오분류율을 보인다.

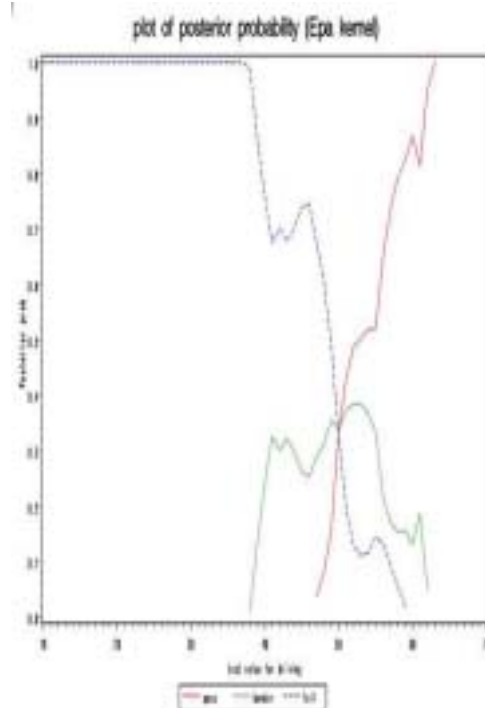
비모수적 방법으로 Epanechnikov 커널함수를 사용하여 그룹별 확률밀도함수를 추정하여 분류한 결과 <그림 4.1>에서 보면 그룹2(경계)의 확률분포가 그룹1(합격)과 그룹3(불합격)에 걸쳐있으며 불합격그룹의 분산이 합격그룹의 분산보다 큼을 알 수 있다. 또한 비모수적 방법으로 그룹별 사후확률분포를 추정한 결과 <그림 4.2>에서 보면 1(pass), 2(border), 3(fail) 세 그룹간 확률분포의 차이를 볼 수 있으며 td11 점수가 낮을수록 불합격, 높을수록 합격 확률이 큼을 볼 수 있다. <그림 4.1>과 <그림 4.2>는 SAS에서 얻은 그래프로, 가장 오른쪽에 위치한 가장 진한선은 합격그룹, 가장 왼쪽에 위치한 점선은 불합격그룹 그리고 중간에 위치한 선은 경계그룹을 나타낸다.

<표 4.4> Epanechnikov 커널함수를 이용한 경우 오분류표

CDBI 판정	분류결과			행합
	1(pass)	2(border)	3(fail)	
1(pass)	34	1	1	36
2(border)	16	0	11	27
3(fail)	6	0	31	37
열합	56	1	43	100
PRIORS	0.36	0.27	0.37	



<그림 4.1> Epanechnikov 커널함수를 이용한 확률밀도함수 그래프



<그림 4.2> Epanechnikov 커널함수를 이용한 그룹별 사후확률분포 그래프

5. 결론

뇌손상 장애인들의 안전한 자동차 운전을 뒷받침 해주기 위해 국립재활병원팀은 한양대학교 의과대학 의공학실과 함께 뇌손상 장애인의 인지-지각 기능을 평가 할 수 있는 새로운 검사 도구인 CPAD를 개발하였다. 해당 연구팀은 이 도구의 유용성을 알아보고자 외국에서 개발된 CDBI 검사와 같이 측정하였으며 CDBI와의 관련성을 이용

하여 운전능력에 대한 판정도구로 이용하고자 한다. 이와 관련하여 본 논문에서는 운전능력 판정법에 대한 연구로 통계적 판별분석을 시도하게 되었다.

운전능력 측정 변수들 중 통계적으로 의미있게 선택된 10개의 변수를 선택하여 합격, 경계, 불합격 그룹으로 구별하는 판별분석을 수행하였다. 우선 모수적인 방법으로는 선형판별함수, 2차판별함수, 정준상관분석, 단계적 판별분석을 시도하였고, 비모수적인 방법으로는 운전능력점수에 대해 Epanechnikov 커널함수를 이용하여 확률밀도 함수 추정을 이용한 판별분석을 하였다.

집단별 공분산행렬의 동일성 검정 결과 공분산행렬이 같다고 할 수 없었으며, 2차 판별함수를 이용한 경우의 오분류율이 17%로 가장 작게 나타났으며 운전능력점수인 일변량 변수를 사용한 경우보다 여러 변수의 다차원 정보를 포함한 다변량 변수를 사용한 경우의 오분류율이 훨씬 작게 나왔다. 합격그룹과 불합격그룹에 걸쳐있는 경계 그룹(판정보류그룹) 때문에 오분류율을 줄이는데 한계가 있으며 경계그룹을 제외한 후 판별분석을 시도할 수 있겠다.

이러한 결과를 살펴보았을 때 CPAD 검사도구를 이용하여 뇌손상 장애인의 운전능력판별이 가능하고 2차판별함수를 이용한 판별분석이 가장 나은 방법으로 제안될 수 있다. 그러나 CBDI 결과를 참고하여 판정한 것이므로 이 부분이 없는 경우를 고려하여, 개발된 변수들만의 자체 판정 기준도 연구해야 할 필요가 있다.

참고문헌

1. 신오수, 장순자, 김완호, 이범석, 홍병진, 김재형, 김병식(2000) 국립재활원의 장애인 자동차 운전훈련 프로그램 적용 결과분석. 대한재활의학회지 24: 618-623
2. 허안나, 박시운, 이범석, 신오수, 이은선, 나인수, 김병식(2003) Cognitive Behavioral Driver's Inventory를 이용한 뇌졸중 환자의 운전능력 평가. 대한재활의학회지 27: 7-12
3. 김기영, 전명식(1994) 다변량 통계자료분석, 자유아카데미.
4. Barbara L., Mazer B. L.(1998) Predicting Ability to Drive after Stroke. *Archives of Physical and Medical Rehabilitation*. 79, 743-750.
5. Engum E.S., Lambert E.W., Womac J., Pendergrass T.M.(1988) Norms and Decision making Rules for the Cognitive Behavioral Driver's Inventory. *Cognitive Rehabilitation*. 6: 12-18
6. Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.
7. Jerwood, D. Price, D. J. and Georgiakodis, F. A.(1991). Problems of Diagnosis in Severe Head Injuries using Discriminant Analysis Techniques. *Statistics in Medicine*, 199-217.
8. Khatree, R. and Naik, D. N. (1995) *Applied Multivariate Statistics with SAS Software*, SAS Institute, North Carolina.
9. Rencher, A. C. (1998) *Multivariate Statistical Inference and Applications*, Wiley, New York.

10. Rencher C.A.(2002) *Methods of Multivariate Analysis*, Wiley, New York.
11. Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.
12. Velilla, S. and Barrio, J. A. (1994) A Discriminant Rule under Transformation, *Technometrics*, 36, 348-353.

[2004년 10월 접수, 2005년 1월 채택]