철도수요예측을 위한 직접수요모형 개발에 관한 연구

Development of A Direct Demand Estimation Model for Forecasting of Railroad Traffic Demand

김효종† 정찬묵*

Hyo-Jong Kim Chan-Mook Jung

ABSTRACT

The Korea Transportation Database (KTDB) is used to obtain data on the origin and destination (OD) of inter-city travel, which are currently used in railroad planning when estimating traffic demand. The KTDB employs the trip assignment method, whereby the total traffic volume researched for inter-city travel in Korea is divided into road, rail and air traffic, etc. However, as regards rail travel, the railroad stations are not identical to the existing zones or the connector has not been established because there are several stations in one zone as such, certain problems with the applicable methods have been identified. Therefore, estimates of the volume of railroad traffic using the KTDB display low reliability compared to other modes of transportation.

In this study, these problems are reviewed and analyzed, and use of the aggregate model method to estimate the direct demand for rail travel is proposed in order to improve the reliability of estimation. In addition, a method of minimizing error in traffic demand estimation for the railroad field is proposed via an analysis of the relationship between the aggregate model and various social-economic indicators including population, distances, numbers of industrial employees, numbers of automobiles, and the extension of roads between cities.

국문요약

현재 철도계획에 이용되고 있는 교통수요예측 지역간기종점(OD)자료는 국가교통데이타베이스(KTDB)를 이용하고 있다. KTDB의 자료는 우리나라 지역간 총통행량을 조사한 후 도로, 철도, 항공등으로 배분하는 방법을 이용하고 하고 있다. 그러나 철도의 경우는 철도역이 기존의 존과 일치하지 않거나 1개 존에 다수의 철도역이 존재하여 Connector가 연결되지 않아 적용 방법상에 문제점이 있는 것으로 나타났다. 이로 인해 KTDB를 이용한 철도부분의 교통예측은 다른 교통수단에 비해 신뢰성이 크게 떨어지고 있다.

본 연구는 이러한 문제점을 검토 분석하고, 신뢰성을 높이기 위해 집계모형(Aggregate Model) 기법을 활용한 철도의 교통수요예측 방법(직접수요추정 : Direct Demand Estimation)을 제안하였다. 본 연구에서는 집계모형에 지역간의 인구, 거리, 산업체종사자수, 자동차대수, 도로연장등 사회경제지표의 연관성을 분석하여 철도분야 수요예측의 오차를 최소화할 수 있는 방법을 연구하였다.

† 정회원, 우송대학교대학원 석사과정, (주)동아기술공사 부사장
E-mail : khjtrans@hanmail.net
TEL : (062)383-0567 FAX : (062)383-0568

* 정회원, 우송대학교,철도건설환경공학과, 교수

1. 서 론

1.1 연구의 목적

본 연구는 우리나라 철도계획에 이용되고 있는 수요예측 현황 및 기반연구등을 검토하고 이들에 대한 문제점을 분석하였다. 특히 2006년 1차 완성된 국가교통DB의 철도분야 지역간 OD는 최근 수요예측의 근간으로 인식되고 있으나 연구결과 여러 가지 문제점을 내포하고 있는 것으로 나타났다.

따라서 본 연구는 이러한 문제점을 개선하기 위한 방안으로 다양한 사회경제 지표를 이용한 직접수요모형(Direct Demand Estimation Model)을 개발하기 위한 방안을 제시하고 모형의 특성을 분석하였다.

본 연구를 통하여 나타난 기존 역 중심에 국가교통DB자료의 활용 제약의 문제점을 넘어 신선계획등에 적용이 가능하며 철도수송 수요예측의 신뢰성을 높일 수 있는 기초연구가 될 것으로 판단된다.

1.2 연구의 방법

본 연구는 기존 국가교통DB등 철도 수요예측 기반에 대해 검토하고 문제점을 분석하였다. 수송수요모형구축을 위한 지역간 자료는 총10개지역(KTDB 존과 동일한 市급이상 존) 데이터(10×10=100개)를 이용 모형화하였다. 이용자료는 한국철도공사의 2006년 지역간 OD와 2006년 KTDB(2008년 발표자료), 2006년 연구대상 지역의 통계자료등을 이용하였다. 독립변수의 분류는 상관분석과 주성분분석등을 실시하였으며 모형은 다중회귀분석 모형을 이용하였다.

2. 철도 수요예측 기반

2.1 국가교통데이터베이스(KTDB)

2.1.1 구축현황

정부에서는 1998년부터 도로, 철도, 항공, 해운등의 교통계획 및 투자의 효율성과 신뢰성을 확보하기 위해 교통체계효율화법을 제정하고 국가기간망계획 및 중기 투자계획등 주요 국가 정책의 합리적 시행을 위해 국가교통조사를 실시하였다. 그동안 총 5차연도에 걸쳐 권역별 여객통행량 및 화물통행량등을 조사하여 데이터베이스를 구축하였다. 또한 매년 보완조사를 통하여 구축자료를 갱신하고 있으며 이 자료는 도로 및 철도, 공항등 계획시 이용토록 하고 있다.

2.1.2 구축체계

(1) 세부기준

국가교통데이터베이스는 전국을 16개 대존(광역시), 166개 중존(시군구), 249개 소존(소규모 시급 구)으로 나누어 교통존(Traffic Zone)별 통행발생량과 분포, 수단선택 그리고 배정통행량을 DB화하여 제공하고 있다. KTDB상의 통행발생 및 분포는 군집분석(총 5개군)을 통해 지역특성을 일반화하여 인구와 승용차통행시간, 자동차등록대수를 이용한 모형식을 이용 산정하였다.

$$\ln TRIP_{ij} = a + b_1 \ln T_{ijp} + b_2 \ln POP_{ij} + b_3 \ln CAR_{ij} + b_4 D_1 + b_5 D_2 \qquad \text{(수식 1)}$$

여기서, $TRIP_{ij}$ = 존 i와 j간의 여객 통행량(통행/일) (도로, 철도, 항공등 전체통행량)

T_{ijp} = 존 i와 j간의 승용차 통행시간(분)

POP_{ij} = 존 i와 j의 인구의 곱(인)

CAR_{ij} = 존 i와 j의 1인당 자동차등록대수의 곱(대)

D_1 = 존 i와 j간의 통행시간(T_{ijp})에 대한 더미변수

D_2 = 존 i와 j의 인구의 곱(POP_{ij})에 대한 더미변수

a, b_1, b_2, b_3, b_4, b_5 = 파라미터

<표 1> 국가교통DB 해당지역별 통행량산정을 위한 군집분류

구분	군집1	군집2	군집3	군집4	군집5
지역	서울	부산, 대구, 인천, 광주, 대전	경기도지역	울산 및 23개 지방거점도시	나머지 지역

(2) 수단선택 절차

국가교통DB의 지역간 통행량 산정방법은 원칙적으로 전통적4단계 추정법을 이용하고 있다. 즉 총 249개존의 발생통행량을 산정하여 해당지역별 분포를 군집별 모형을 이용 산정하였다. 산정된 분포량을 육상교통인 도로와 철도 그리고 항공 및 해운등으로 수단을 분류하고 이들 수단에 각각 통행량을 배정하는 절차로 이루어졌다. 이중에서 수단선택의 과정에서 철도분야의 지역간 통행량이 나누어지게 된다.

수단선택은 통행분포 단계에서 추정된 교차통행량을 수단별 통행 교차량으로 바꾸는 모형으로 통행자의 선호형태를 고려하여 육상교통과 철도등으로 나누는 단계이다.

일반적으로 수단선택 모형은 전환곡선법, 회귀분석모형, 로짓모형, 프로빗모형등 대부분 확률 모형이 이용되며 KTDB에서도 이들 모형을 이용하였다.

확률선택모형이란 효용극대화 이론을 이론적 배경으로 하고 있는데 통행 비용과 통행 사간을 일반변수로 하고 확률효용함수는 수단 특성 상수로 모형화 된다. 즉 특정수단을 선택할 확률(k)는

$$P(k) = \frac{e^{(uk)}}{\sum_{i=0}^{n} e^{(ui)}} \qquad \text{(수식 2)}$$

여기서, uk는 수단 k의 효용

ui 는 수단 i의 효용

n은 수단의 수

이때 효용 함수는 다음과 같다.

$$U_{ijm} = \alpha_1 (T_{time})_{ijm} + \alpha_2 (T_{cost})_{ijm} + (DUM)_m \qquad \text{(수식 3)}$$

여기서 U_{ijm} 은 수단 m의 교통존 i 와 j간의 효용함수

$(T_{time})_{ijm}$은 수단 m의 교통존 i 와 j간의 총통행시간

$(T_{cost})_{ijm}$은 수단 m의 교통존 i 와 j간의 총통행비용

$(DUM)_m$은 수단 m의 더미변수

여기서 수단 특성 상수의 파라미터 값은 다음 값을 사용하였다.

<표 2> 수단선택 모형의 파마리터 값

구 분	T_{time}	T_{cost}	DUM
승용차	-0.00254	-0.0000243	-
버 스		-0.0000733	-1.326
철 도		-0.000115	-1.797

자료 : 2006년 “국가교통DB 구축 사업 보고서”, 2007. 한국교통연구원

2.2 철도분야 지역간 통행량

2009년 KTDB 자료를 보면 2007년 지역간 1일 총 수단통행량은 13,983천통행/일로 2006년 13,842천통행/일에 비해 1.0% 증가하였다. 승용차 통행은 1일 8,373천통행/일로 전체 수단통행량의 59.9%, 버스는 3,608천통행/일로 25.8%, 철도는 1,937천통행/일로 13.8%를 분담하는 것으로 나타났다.

<표 3> 수단별 통행량(2007년)

구 분	승용차	버스	철도	항공	해운	계
통행/일	8,372,684	3,608,079	1,936,567	46,780	18,621	13,982,731
분담비(%)	59.9	25.8	13.8	0.3	0.1	100.0

자료 : 2008년 “국가교통수요조사 및 DB구축사업”, 2009.4, 한국교통연구원

장래 통행분포 예측은 구축된 발생량과 도착량을 이용 다음 (수식 4)의 2중제약 프라타모형(Two-Dimensional Fratar Model)을 이용하여 산정하였다.

$$T_{ij} = A_i B_j O_i D_j t_{ij}$$ (수식 4)

여기서 t_{ij} : 지역간 여객 기종점통행량(A,B지역간 O/D)

O_i : 장래목표연도 존별 발생량

D_j : 장래목표연도 존별 도착량

장래 수단별 통행량을 살펴보면, 2007년 승용차가 8,373천통행/일로 지역간 통행의 60.0%를 분담하였으나 2036년 9,180천통행/일로 58.6%를 분담할 것으로 예측되었다. 버스의 경우, 2007년 25.8%인 3,797천통행/일을 분담하는 것으로 분석되었으며, 2036년의 경우 2,267천통행/일로 14.5%를 분담하는 것으로 예측되었으며 항공은 2006년 0.3%인 47천통행/일을 분담하는 것으로 분석되었으며 2036년의 경우 144천통행/일로 0.9%를 분담하는 것으로 예측됨

<표 4> 장래 목표연도별 수단별 통행량 비교

구 분		승용차	버스	철도	항공	계
2007년	통행/일	8,372,684	3,608,079	1,936,567	46,780	13,964,110
	분담비(%)	60.0	25.8	13.9	0.3	100.0
2011년	통행/일	8,830,924	3,847,190	2,090,586	63,281	14,831,981
	분담비(%)	59.5	25.9	14.1	0.4	100.0
2016년	통행/일	9,071,289	3,972,314	2,161,714	74,190	15,279,508
	분담비(%)	59.4	26.0	14.1	0.5	100.0
2021년	통행/일	9,194,006	4,047,099	2,258,224	87,573	15,586,901
	분담비(%)	59.0	26.0	14.5	0.6	100.0
2026년	통행/일	9,284,293	4,103,705	2,285,888	103,347	15,777,234
	분담비(%)	58.8	26.0	14.5	0.7	100.0
2031년	통행/일	9,304,860	4,126,041	2,291,601	122,035	15,844,537
	분담비(%)	58.7	26.0	14.5	0.8	100.0
2036년	통행/일	9,179,727	4,082,275	2,267,035	144,232	15,673,268
	분담비(%)	58.6	26.0	14.5	0.9	100.0

주 : 2007년도 총 통행량은 13,982,731통행/일이며, 이 중에서 해운통행은 18,621통행/일을 차지함
자료 : 2008년 “국가교통수요조사 및 DB구축사업”, 2009.4, 한국교통연구원

2.3 철도 수송수요 적용 문제점

국가교통DB는 모든 교통수단(도로, 철도, 항공, 해운등)의 총 통행량을 집계하여 지역특성함수를 반영한 수단선택 모형을 이용 배분하고 있다. 그러나 철도의 경우 역간 이용자 자료가 매년 집계되어 발표하고 있어 해당모형 적용을 통한 존 배분은 무의미한 것으로 판단된다.

또한 철도의 경우 통행비용 함수에 대한 기초적인 연구가 미미하여 적용이 어려운 실정으로 노선배정시 신뢰성이 저하되고 오차가 크게 발생하고 있다. 또한 KTDB의 경우 1개의 존에 다수의 철도역이 존재할 경우 존에서 나오는 connector가 모든 철도역에 연결되지 않아 통행배정시 관측지점이 적게 나타나며 아울러 존과 철도역간의 연결문제로 존에서 나오는 connector의 연결이 철도역과 비합리적으로 이루어져 있어 통행배정에 문제점 발생한다. 즉, 도로배정 connector와 철도배정 connector가 다를 경우 배정오류가 발생하고 있다.

또한 대중교통 통행배정에 널리 활용하고 있는 최적전략(optimal strategy)은 Spiessety al.(1989)에 의해 개발된 방법으로 EMME/2의 대표적인 대중교통 통행배정기법을 적용하였으나 이 방법은 최적전략을 찾는 단계와 최적전략으로 선택된 구간에 통행수요를 배정(loading)하는 2단계로 이루어져 있다. 이로 인해 최적전략은 출발지에서 선택가능한 노선정보를 알고 있을 때 가장 최소비용이 소요되는 노선을 선택하고 최적전략으로 경로가 결정되면, 통행수요를 해당노선에 배정하게 되며, 여기서 각 노드를 통과하는 노선들의 배차간격(headway)에 비례하여 수요를 배정한다.

그러나 철도의 경우 실측자료를 가지고 배정한 결과, 배정되지 않는 노드(node)들이 다수 존재하는 것으로 나타났다. (2009년 국가교통수요조사 및 DB구축사업 보고서)

3. 직접수요추정모형

3.1 수요추정 모형 방법

교통수요 추정 방법은 전통적 4단계추정법(Tradtional 4-step Method)과 직접수요추정방법(Direct Demand Estimation Method), 개별형태 수요추정방법(Disaggrigate Behavioral Method) 그리고 개략적 수요추정방법(Sketch Planning Method)등으로 나눌수 있다. KTDB에서 이용한 4단계 추정법은 통행발생과, 통행분포, 수단분담, 통행배정의 단계로 각 교통존(Traffic zone)을 중심으로 해당 단계별 모형으로 이용 수요를 예측하는 방법이다.

이와는 상대적인 개념으로 통행발생 및 통행분포, 수단선택등의 단계를 필요에 따라 통합하여 예측하는 방법을 직접수요 추정모형이라 한다 이 모형은 집계모형(集計模型 :Aggregate Model)의 일종으로 이론의 배경은 경제학의 소비자 행태이론에 기초한다. 직접수요추정 모형은 대상지역간 인구와 거리, 시간등 통행비용등 다양한 지표를 이용 설명변수들간의 관계를 규명하는 이론적 배경을 가지고 있다.

(1) Quandt-Baumol 모형(전도성 모형)

이 모형은 미국 Boston과 Washington을 연결하는 교통축 연구사업에서 도시간 통행수요의 추정과정에서 개발되었다.

$$T_{kij} = \alpha_0 P_i^{\alpha 1} P_j^{\alpha 2} Y_i^{\alpha 3} Y_j^{\alpha 4} M_j^{\alpha 5} M_j^{\alpha 6} N_{ij}^{\alpha 7} f_1(H) f_2(C) f_3(D) \quad \text{(수식 5)}$$

여기서

$$f_1(H) = (H_{ij}^b)^{\beta 0} (H_{kij}^r)^{\beta 1}$$

$$f_2(C) = (C_{ij}^b)^{\gamma 0} (C_{kij}^r)^{\gamma 1}$$

$$f_c(D) = (D_{ij}^b)^{\delta 0} (D_{kij}^r)^{\delta 1}$$

T_{kih} = i 지역에서 j지역으로의 수단 k의 통행량

P_i = 지역 i의 인구

Y_i = 지역 i의 평균소득

M_i = i지역에 대한 특성 index

N_{ij} = 지역i,j간 운행수단의 종류(수)

H_{ij}^b = 지역i,j간 최소통행시간

C_{ij}^b = 지역i,j간 최소통행비용

D_{ij}^b = 지역i,j간 최소운행빈도

H_{kij}^r = 지역i,j간 최소통행비용에 대한 k번째 수단의 상대적 통행시간

C_{kij}^r = 지역i,j간 최소통행비용에 대한 k번째 수단의 상대적 통행비용

D_{kij}^r = 지역i,j간 최소통행비용에 대한 k번째 수단의 상대적 운행빈도

$\alpha_0, \alpha_1 \cdots \alpha_m$ = 모형의 파라메타

파라메타 α_0부터 α_6의 부호는 양의 형태로 나타나며, 통행빈도를 제외한 통행시간이나 비용과 같은 비효욕의 변수는 음의 형태로 나타난다.

(2) McLynn 모형

이 모형은 두지역간 총 통행수요가 아래와 같은 형태로 주어지고, 기본적인 가정은 논리적인 제약조건 또는 소비자이론에 의거한 수요탄력성에 대한 가정에 전제한 것이다.

$$T_{ij} = k_{ij} \sum_{q} [C_{qij}]^{g}$$

이러한 가정의 조건을 만족하는 McLynn 모형의 일반적 형태를 보면 다음과 같다.

$$T_k = \frac{\alpha_k \prod_{m,l} V_{ml}^{b_{mlk}}}{\sum_{n} \alpha_n \prod_{m,l} V_{ml}^{b_{mln}}}$$

여기서 지역 i,j간 교통수단 k에 대한 통행수요의 모형은 다음과 같다.

$$T_k = \frac{K_{ij} \sum_{n} [C_{nij}]^{g} C_{kij}}{\sum_{n} [C_{nij}]}$$

위 두식을 정리하여 간편화하면 다음과 같다.

$$C_{nij} = \alpha_n \prod_{m,l} V_{ml}^{b_{mlk}} \qquad \text{(수식 6)}$$

여기서 C_{ij} = 지역i,j간 q수단의 비용

T_k = K수단통행량

K_{ij} = 지역i,j간 수단K의 총 통행량

$V_{ml}^{b_{mlk}}$ = 지역내 총 통행량

α_n, g = 모형의 파라메터

(3) 수단경쟁모형

수단경쟁모형은 형태와 모형식의 모형에 도입되는 변수는 동일하지만 모형식 내에서 동시적으로 수단간의 서비스특성 변수를 고려함으로써 수단간 교차관계를 규명한다는 점에서 전도성모형과 차이점이 있다.

수단경쟁모형의 대표적인 모형으로 Kraft-SARC모형이 있다.

$$T_{kij} = \alpha_k (Y_i)^{\beta_1} (P_i)^{\beta_2} (E_j)^{\beta_3} \prod_{l} (V_{kij})^{\delta kl} \prod_{q(\neq k)} \prod_{l} (V_{qijl})^{\gamma_{ql}} \qquad \text{(수식 7)}$$

여기서 $Y_{i,} P_i$ = 지역i의 소득과 인구의 중앙값

E_j = 지역j의 고용수준

V_{qijl} = 지역i,j간 수단 q의 l 서비스특성

$\alpha_k, \beta_1, \beta_2, \beta_3, \delta_{kl}, \gamma_{ql}$ = 특성 파라메타

3.2 이용자료

3.2.1 변수규모 선정

본 연구에서는 모형예측을 위해 총 10개지역간 주요역을 선정하여 모형화를 시도하였다. 해당역의 선정은 역간 거리의 규모를 고려하여 목포와 서울간 총 10개역, 100개 데이터 (10×10 Matrix)를 이용하였다. 아울러 역 선정시 KTDB의 존(Zone) 체계와 일치할 수 있도록 하였다. 해당역의 규모는 최도한 시(市)이상 역을 대상으로 선정하였다.

3.2.2 KTDB 배분편차

본 연구 기준연도는 2006년 자료(2008년 발표)를 이용하였고 지역간 여객 통행량(매표소 자료집계)과 KTDB자료를 비교하였다.

편차분석결과 서울에서 대전구간은 편차가 46.1%로 크게 나타났으며 기타 구간은 약 14.1%로 나타났다.

<표 5> 국가교통DB 분석대상 존간 여객발생량(2006년) (단위 : 통행/일)

OD	서울시	천안시	대전시	논산시	익산시	김제시	정읍시	광주시	나주시	목포시
서울시	-	7,413	14,807	1,239	1,787	432	824	2,786	311	986
천안시	7,964	-	2,955	154	165	35	52	155	19	64
대전시	14,814	2,547	-	875	496	71	112	296	31	149
논산시	1,246	147	975	-	78	14	37	123	12	34
익산시	1,780	153	480	86	-	24	527	349	33	109
김제시	450	32	78	17	42	-	55	109	9	38
정읍시	898	51	116	38	507	49	-	94	15	66
광주시	2,648	123	276	114	346	107	102	-	50	607
나주시	323	20	41	19	45	12	19	36	-	66
목포시	1,049	61	173	40	128	43	84	702	54	-

자료 : 2008년 국가교통DB 지역간 OD 재분류(본연구)

<표 6> 분석대상존간 여객 수송실적(2006년) (단위 : 통행/일)

OD	서울시	천안시	대전시	논산시	익산시	김제시	정읍시	광주시	나주시	목포시
서울시	-	3,469	8,610	654	1,715	352	534	2,570	229	1,072
천안시	3,346	-	1,491	151	148	34	38	62	13	46
대전시	8,882	1,425	-	529	484	82	127	286	29	198
논산시	598	139	481	-	115	23	28	219	8	33
익산시	1,886	152	467	118	-	90	359	139	25	115
김제시	339	36	78	22	89	-	61	68	8	30
정읍시	557	38	122	28	339	69	-	44	9	45
광주시	2,572	50	274	80	275	88	72	-	16	777
나주시	253	13	27	10	26	9	9	12	-	64
목포시	1,082	45	194	33	120	29	45	789	64	-

3.2.3 독립변수

기존 대부분의 모형구조는 지역간 사회경제 지표등을 이용한 중력모형 (지역간 지표 곱)형태를 띄고 있다. 따라서 본 연구에서도 동일하게 각 지역간 독립변수의 곱을 이용한 Matrix를 구축하였다. 다음 <표 7>는 본 연구에서 검토한 독립변수의 내용이다. 다만 거리는 지역간 거리를 이용하였다.

<표 7> 본윤구 분석대상 존의 독립변수 지표

구 분	서울시	천안시	대전시	논산시	익산시	김제시	정읍시	광주시	나주시	목포시
인 구 (천명)	10,356.2	531.2	1,475,9	132.8	317.9	101.3	127.4	1,415,9	97.5	242.8
세 대 수 (천가구)	3,979.9	200.3	518.0	51.7	112.6	40.6	50.0	495.9	41.2	90.5
가구당인구 (인/가구)	2.6	2.7	2.85	2.57	2.82	2.49	2.55	2.85	2.37	2.68
면 적 (㎢)	605.33	636.2	539.79	554.98	507.06	545.02	692.93	501.31	604.05	47.95
인구밀도 (인/㎢)	17,008.7	835.0	2,734.3	239.3	626.9	185.8	183.8	2,824.5	161.4	5,062.7
경제인구 (천명)	10,181.2	414.9	1,147.0	110.6	253.6	86.3	105.1	1,108.5	82.6	191.4
GRDP (십억)	193,775.9	-	18,693.434	-	-	-	-	19,049.400	-	-
사업체수 (개수)	733,759	33,616	89,851	9,091	19,708	6,147	7,935	95,656	5,856	19,379
종사자수 (천인)	3,894..7	182.2	412.9	34.6	84.2	25.1	29.1	439.4	24.8	69.3
도로연장 (km)	8,093.0	975.7	2,237.9	500.2	918.1	656.8	589.1	2,136.9	695.3	247.6
도로율 (%)	21.89	-	6.10	3.38	4.92	3.71	3.57	6.19	3.69	10.99
자동차수 (천대)	2,856.8	194.2	521.2	44.19	105.2	34..8	38.9	449.9	34.0	71.1
영업용 자동차대수	194,238	-	10,303	681	-	603	1,013	9,718	507	1,740
영업차연간 수송인(백만)	2,675.6	-	267.8	20.9	-	8.6	8.7	0.26	4,3	0.24

3.3 독립변수 선정

3.3.1 독립변수의 상관성

본 논문에서 철도수송 수요와 관련이 있을 것으로 판단되는 총 15개의 사회 경제 지표를 선정하였다.

이 사회 경제지표중 변수간 직접 연관이 있는 변수, 즉, 도로율과 도로와 같이 동일 지표로 판단되는 함수는 제외하고 총 9개의 지표를 선정하여 상관메트릭스 분석을 실시하였다.

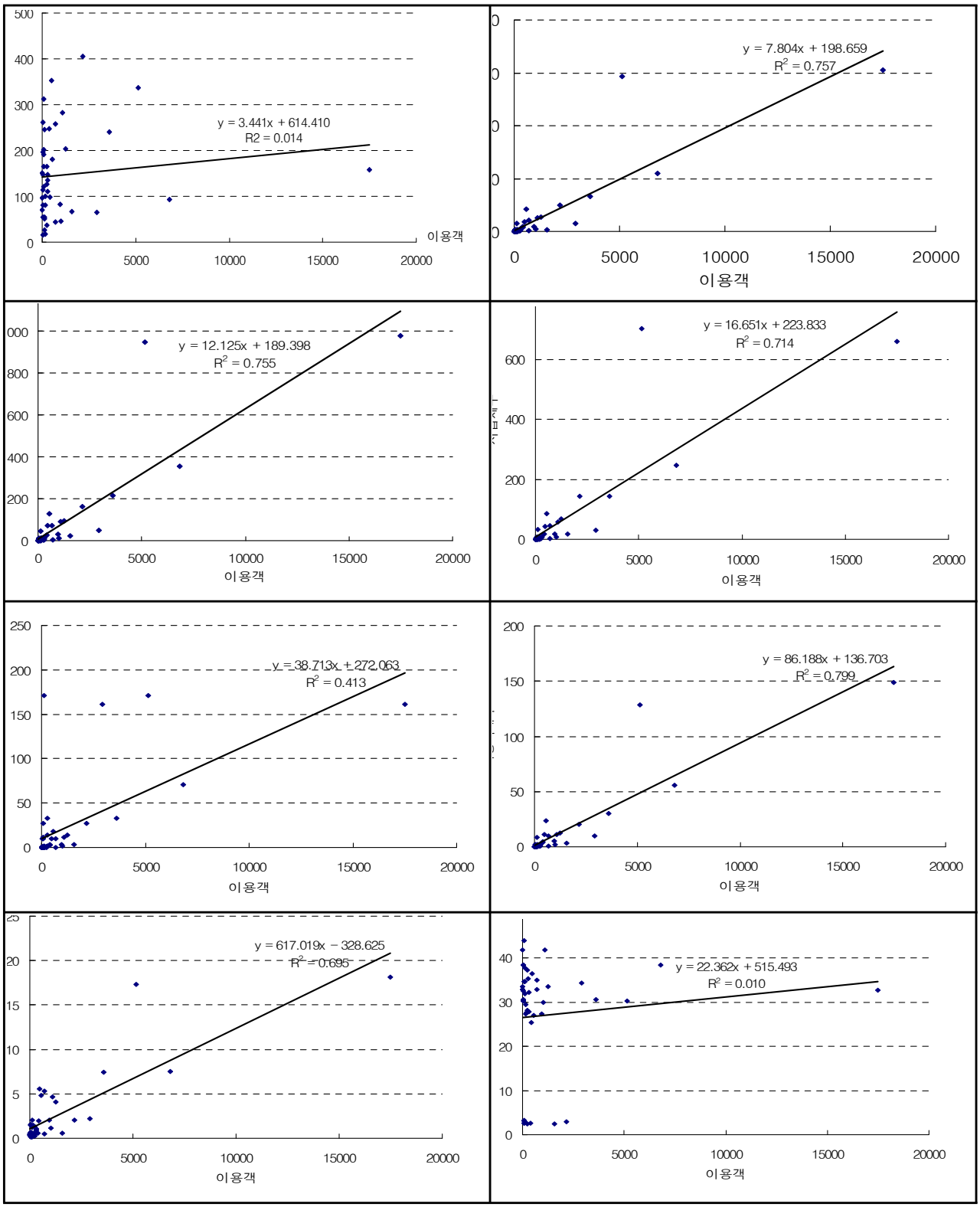
분석결과는 다음과 같이 변수들 간에 서로 상관성이 높은 것으로 나타났다. 특히 일부 자료의 경우 상관도가 1.0으로 직접적인 관련이 있어 데이터간 주성분분석이 필요한 것으로 판단된다

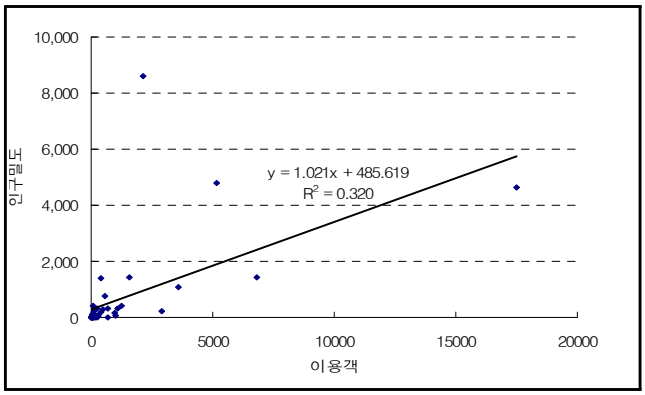
다음은 독립변수의 상관분석 Matrix 결과이다.

<표 8> 독립변수의 상관분석 매트릭스(Correlation Matrix)

구 분	인 구	인구밀도	자동차	도로연장	경제인구	사업체수	종사자	면 적	지역간 거리
인 구	1	0.658	0.998	0.958	1.000	0.998	0.714	0.104	0.307
인구밀도	0.659	1	0.644	0.566	0.660	0.638	0.441	-0.251	0.511
자동차수	0.998	0.644	1	0.960	0.997	0.992	0.720	0.112	0.288
도로연장	0.958	0.566	0.960	1	0.960	0.952	0.684	0.215	0.370
경제인구	1.000	0.660	0.997	0.960	1	0.998	0.711	0.104	0.312
사업체수	0.998	0.683	0.992	0.952	0.998	1	0.710	0.089	0.328
종사자총수	0.714	0.441	0.720	0.684	0.711	0.710	1	0.148	0.270
면 적	0.104	-0.251	0.112	0.215	0.104	0.089	0.148	1	-0.158
지역간거리	0.307	0.511	0.288	0.370	0.312	0.328	0.270	-0.158	1

다음은 철도역간 여객수송수요와 해당지역 사회경제 지표 관계 모형이다.





<그림1> 사회경제지표별 이용수요 모형

3.3.2 주성분분석(Principal Component Analysis)

본 연구와 같은 회귀분석(regression analysis)이 상관관계를 갖는 것은 측정치의 중복성(measurement overlap)이다. 또 변수간의 상관 관계를 분석하거나 한 변인의 측정치로부터 다른 변인의 값을 예측하는 것뿐만 아니라 각각의 변인들의 공통으로 가지고 있는 변량의 크기로써 측정된 값이 얼마나 중복되어 있는가를 계산할 필요가 있다. 즉, 주성분 분석에서는 (1)여러 변인간의 상호관계로부터 공통 변량을 구하고, (2)측정치의 중복성을 찾아내어, (3)몇 개의 기본 변인, 즉 몇 개의 변인 군을 산출해 내는데 사용되는 기법이다. 다시 말해 측정값들의 상호관계를 나타내는 변량들을 좀더 기본적인 몇 개의 가상 변수(hypothetical variables)로 재배열하는 것이다. 본 연구에서는 독립변수들이 어떠한 형태(pattern)로 구성되었는가를 알아내기 위하여 주성분분석을 실시하였으며 이를 근거로 대표적 독립변수를 추출하였다.

주성분분석결과 즉 Varimax법에 의해 회전설계된 각 요인의 적재치는 다음 <표 9>과 같다. 여기서 알 수 있듯이 지역간 철도 수송수요는 3개의 주성분군(group)을 형성하였다.

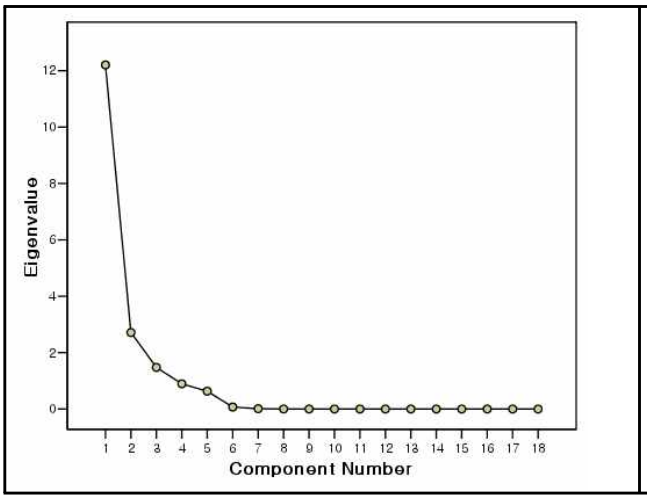
제1주성분은 자동차대수, 인구, 경제인구, 사업체수, 도로연장이며 이때 Eigen Value는 12.203이고 기여율(% of variance)는 67.796%이었다. 제2주성분은 지역간 거리로 설명력의 변동은 15.072%로 높게 나왔으며 누적기여율은 82.868%였다. 제3주성분은 두 지역 면적의 곱으로 기여율은 8.200%이었고 전체 누적기여율은 91.068%로 나타났다.

또한 전체적으로 높은 누적치를 보이고 있고 모두 정(+)의 계수를 나타내고 있어 각각의 성분들이 특성치가 될 수 있음을 보여주고 있다.

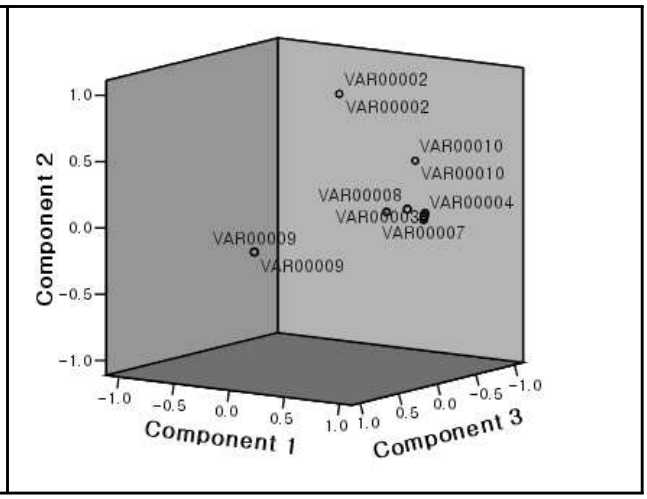
<표 9> 주성분분석결과

구 분		Componet		
		1	2	3
Factor	자동차대수	0.968	0.114	0.001
	인 구	0.986	0.134	-0.008
	경제인구	0.985	0.139	-0.007
	사업체수	0.981	0.158	-0.026
	도로연장	0.937	0.209	0.140
	종사자수	0.750	0.168	0.142
	인구밀도	0.619	0.462	-0.415
	거 리	0.170	0.970	-0.075
	면 적	0.124	-0.079	0.960
Initial Eigen Value	Total	12.203	2.713	1.476
	% of variance	67.796	15.072	8.200
	Cumulative(%)	67.796	82.868	91.068

Extraction Method: Principal Component Analysis
Rotation Method : Varimax with Kaiser Normalization
a. Rrotation converged in 4 iteration.



<그림2> 주성분분석 Scree Plot 결과



<그림> 주성분분석 Component Plot in Rotated Space

3.4 수요예측 모형 개발

본 연구에서는 앞에서 제시한 바와 같이 독립변수들간의 상관분석 매트릭스 중복성 때문에 주성분 분석을 실시하여 총 3개의 독립변수를 선정하였다.

다만 상관분석 매트릭스에서 나타난바와 같이 제1주성분 요소인 인구와 자동차대수, 경제인구, 사업체수, 도로연장은 상관성이 매우 높아(0.958~0.998) 어떠한 독립변수를 모형에 포함하여도 높은 설명력을 보일 것으로 예측되었다. 따라서 본 연구에서는 이들 5개의 모형중 대표적인 지표인 인구를 제1주성분 함수로 선정하여 철도 지역간 여객수송 수요 모형으로 제시하였다.

다음은 본 연구에서 제시한 철도 수요 예측 모형 예이다.

<표 11> 연구대상 노선 수요예측 모형

구 분	모 형 식	R^2
모형식 1 ($Trip1_{ij}$)	$953.107+90.961V_{ij}-4.734L_{ij}-6.435A_{ij}$	0.882
모형식 2 ($Trip2_{ij}$)	$1010.575+8.293P_{ij}-5.023L_{ij}-4.4768A_{ij}$	0.782
모형식 3 ($Trip3_{ij}$)	$1028.473+12.921Cp_{ij}-5.195L_{ij}-5.016A_{ij}$	0.782
모형식 4 ($Trip4_{ij}$)	$944.712+17.839Co_{ij}-5.369L_{ij}-1.652A_{ij}$	0.744
모형식 5 ($Trip5_{ij}$)	$1467.809+712.468R_{ij}-7.766L_{ij}-32.42A_{ij}$	0.757

여기서 $Trip_{ij}$는 두 지역간 철도수송 수요 (통행/일)

V_{ij}는 두 지역간 자동차대수의 곱 (만대)

P_{ij}는 두 지역간 인구의 곱 (만명)

CP_{ij}는 두 지역간 경제인구의 곱 (만명)

Co_{ij}는 두 지역간 사업체수의 곱 (천)

R_{ij}는 두 지역간 도로연장의 곱 (km)

L_{ij}는 두 지역간 거리 (km)

A_{ij}는 두 지역 면접의 곱 (km^2)

본 연구에서 제시한 모형 분석결과 두 지역이 인구와 자동차, 경제인구, 도로연장등 1주성분은 양(+)의 함수이고, 두 지역간 거리와 면적은 음(-)의 함수로 나타났다. 인구등과 거리는 일반적인 기존 모형의 부호를 보이고 있으나 해당지역의 면적은 크면 오히려 철도수송수요가 적다는 것을 알수 있다.

3.5 모형진단

모형진단 결과 제시한 모형은 R^2 및 F검정결과 모든 모형은 유의한 것으로 나타났다. 다음은 모형검증 결과이다.

<표 12> 모형식 해석 결과

구 분	R	R Square	Std. Error of the Estimate	Sum of Squares	df	Mean Square	F-test	Sig.
$Trip1_{ij}$	0.907	0.882	1,246.7	2.94E+08	3	9.81E+07	63.103	0.000
$Trip2_{ij}$	0.884	0.782	1,379.4	2.80E+08	3	9.33E+07	49.045	0.000
$Trip3_{ij}$	0.884	0.782	1,380.9	2.80E+08	3	9.32E+07	48.912	0.000
$Trip4_{ij}$	0.862	0.744	1,496.5	2.66E+08	3	8.87E+07	39.619	0.000
$Trip5_{ij}$	0.870	0.757	1,456.06	2.71E+08	3	9.00E-07	42.621	0.000

<표 13> 모형검증 결과

구 분		Unstandardized Coefficients		Standardized Coefficients	T-test	유의확률 (sig.)
		B	Std. error	Beta		
$Trip1_{ij}$	Constant	953.107	566.350	-	1.683	0.100
	P_{ij}	90.961	6.728	0.944	15.520	0.000
	L_{ij}	-4.734	2.074	-0.160	-2.282	0.208
	A_{ij}	-6.435	14.907	-0.029	-0.432	0.668
$Trip2_{ij}$	Constant	1010.575	627.779	-	1.610	0.115
	P_{ij}	8.293	0.696	-0.022	-0.289	0.774
	L_{ij}	-5.032	2.309	-0.170	-2.179	0.035
	A_{ij}	-4.768	16.483	-0.022	-0.289	0.774
$Trip3_{ij}$	Constant	1028.473	628.698	-	1.636	0.110
	P_{ij}	12.924	1.087	0.926	11.892	0.000
	L_{ij}	-5.195	2..316	-0.176	-2.243	0.030
	A_{ij}	-5.016	16.504	-0.023	-0.304	0.763
$Trip4_{ij}$	Constant	994.712	681.282	-	1.460	0.152
	P_{ij}	17.839	1.668	0.906	10.693	0.000
	L_{ij}	-5.369	2.523	-0.182	-2.128	0.039
	A_{ij}	-1.652	17.848	-0.007	-0.093	0.927
$Trip5_{ij}$	Constant	1467.809	670.976	-	2.188	0.034
	P_{ij}	712.468	64.221	0.963	11.094	0.000
	L_{ij}	-7.766	2.534	-0.263	-3.064	0.004
	A_{ij}	-32.421	17.986	-0.147	-1.803	0.079

4. 결론

본 연구는 기존 국가교통DB의 철도분야 여객수송수요 구축현황과 문제점등 수요예측 기반에 대하여 검토하였다. 검토결과 존간 여객발생량은 육상공로와 철도, 항공등 총 수요를 집계하여 4단계추정법 중 수단선택모형을 이용 철도로 배분 산정하고 있는 것으로 나타났다.

그러나 철도는 역기반 시설이므로 기존 도로네트워크와 상이 connector가 연결되지 않은 존이 발생하는등 예측의 신뢰성에 문제가 있는 것으로 나타났다. 특히 신설계획시 이용에 한계가 있는 것으로 판단된다.

따라서 본 연구는 이러한 문제점을 개선하고 수요예측의 정밀도를 높이기 위해 다양한 사회경제지표를 독립변수로 직접수요추정모형을 개발 제시하였다. 본 연구에서는 상관매트릭스 및 주성분분석을 통해 독립변수를 계층 그룹화 하였다. 모형분석결과 대표적인 지표로 인구, 경제인구, 사업체수, 자동차대수와 지역간거리, 그리고 해당지역의 면적이 포함되었으며 인구 및 사업체수, 자동차대수등은 양(+)의 함수, 거리와 면적은 음(-)의 함수로 나타났다.

본 연구의 향후 과제는 전국 지역간 통행실측 자료(역간OD자료)를 이용하여 모형을 확대 해석하여야 하며 이를 통해 철도 신설계획등에 일반화 할 수 있을 것으로 본다. 또한 거리등을 계층화 하는등 다양한 지역특성을 반영하여 모형을 구축할 필요가 있을 것으로 판단된다. 아울러 모든 지역에 적용가능한 일반화된 모형개발을 위해 추가적인 독립변수등의 확장이 필요한 것으로 판단된다. 특히 철도수송수요와 직접관련이 있을 것으로 판단되는 역의 수와 운행회수등도 함수로 포함하는 것이 필요할 것으로 판단된다.

참고문헌

1. 한국교통연구원 “2008년 국가교통수요조사 및 DB구축사업 보고서”, 2009.4
2. 한국철도공사 “철도통계연보” 각연도(2003～2009)
3. 성수련외 “RP와 SP의 결합추정에 의한 교통수단선택 모델 연구”, 대한토목학회 학술대회, 2006
4. 장준석외 “철도승객의 노선선택을 고려한 철도망 설계모형 연구”, 한국철도학회논문집 제11권 제56호, 2008.12
5. 박상명외 “호남고속전철 이용자의 환승 교통수단 선택 행태분석”, 대한토목학회논문집 제15권 제60호, 2005. 11
6. 김익희외 “문경선 운영재개에 따른 이용수요예측 연구”, 한국철도학회 학술대회, 2008년(추계)
7. 한국철도기술연구원, 국가고속철도망 전략수립 연구, 2009.2
8. 박창호외 “교통공학개론”, 영지문화사
9. 서울시,천안시,대전시,논산시,익산시,김제시,정읍시,광주시,나주시,목포시 홈페이지, 통계자료, 연도별