

KTX 단기수요 예측을 위한 통행행태 분석

김한수^a, 윤동희^b, 이성덕^{1,c}

^a한국철도공사 연구원, ^b한국철도공사 대전충남본부, ^c충북대학교 정보통계학과

요약

본 연구는 KTX의 단기수요예측 방향을 설정하기 위한 통행행태 분석이 목적이다. 분석결과는 첫째, 이상치 판단기준은 통행량 표준편차의 2배가 적정한 것으로 판단된다. 둘째, ANOVA 분석을 이용하여 요일별 통행량의 동질여부를 분석한 결과 주중(월~목)과 주말(금~일)로 구분되었다. 셋째, 통행빈도, 통행량의 평균, 통행거리를 이용하여 철도역간 O/D에 대해 군집분석을 시행하였다.

주요용어: 단기수요예측, 통행행태분석, KTX, 군집분석, 분산분석.

1. 서론

철도수요 예측은 철도계획 및 철도건설사업의 타당성 분석을 위해 주로 사용되어 왔다. 이러한 계획 및 타당성 분석은 건설 후 30년 동안의 장기수요 예측 결과가 필요하기 때문에 주로 전통적 4단계 모형을 이용하여 추정하고 있다(국토해양부, 2010). 철도수요는 철도운영에 필요한 각종 수송정책을 수립하기 위한 기본자료 이기도 하다. 고속철도와 같은 간선철도를 운영하기 위해서는 장래 철도수요를 기반으로 열차운행계획과 차량도입계획을 수립해야 한다. 그러므로 철도운영에 필요한 수송정책을 수립하기 위해서는 장래 철도수요를 정확히 예측하는 것이 무엇보다 중요하다. 철도수송정책을 수립하기 위해 필요한 철도수요는 철도계획 및 타당성 분석에서 요구하는 장기수요 보다 단기수요가 요구된다. 이러한 철도수요는 예측목적에 따라 두 가지로 구분할 수 있다. 첫째, 현재 운영노선에 대한 철도수요이다. 현재 운영노선의 열차운행계획을 보다 효율적으로 개편하기 위해 활용되며, 익년 철도수요를 예측하는 것이 필요하다. 예측방법으로는 시계열모형이나 신경망모형과 같은 패턴인식 기법을 이용할 수 있다(최태성과 김성호, 2004; Tsai 등, 2005). 둘째, 장래 신설노선에 대한 철도수요이다. 장래 신설노선의 열차운행계획과 차량도입계획을 수립하기 위해 활용되며, 노선 신설 후 5년~10년까지의 철도수요를 예측하는 것이 필요하다. 예측방법으로는 계량경제모형(econometrics model)에 기반한 직접수요모형과 전통적 4단계 모형을 이용할 수 있다(Kanafani, 1983; Wardman 등, 2007; MVA Consultancy, 2007). 열차운행계획의 역별 정차횟수와 열차운행횟수를 결정하기 위해서는 철도역간 기종점(origin-destination, 이후 O/D로 표시) 통행과 구간재차인원이 필요하다. 기존 연구에서는 대부분 전통적 4단계 모형을 이용했기 때문에 일 평균수요를 제시하고 있다(국토해양부, 2010). 일부 연구에서 일 평균수요에 요일별 계수를 적용하여 요일별 수요를 산출하기는 했지만 주중통행과 주말통행의 행태차이를 반영하지는 못하였다(한국철도공사 연구원, 2010). 2010년 KTX의 주말 이용자는 주중 이용자에 비해 34%(표 1의 이상치 제거 후 요일별 평균을 이용하여 산출함) 정도 많다. 이와 같이 주중수요와 주말수요의 차이가 크기 때문에 주중과 주말 요일을 구분하여 철도수요를 예측하는 것도 중요한 이슈이다.

¹ 교신저자: (361-763) 충북 청주시 흥덕구 내수동로 52, 충북대학교 정보통계학과, 교수.
E-mail: sdlee@chungbuk.ac.kr

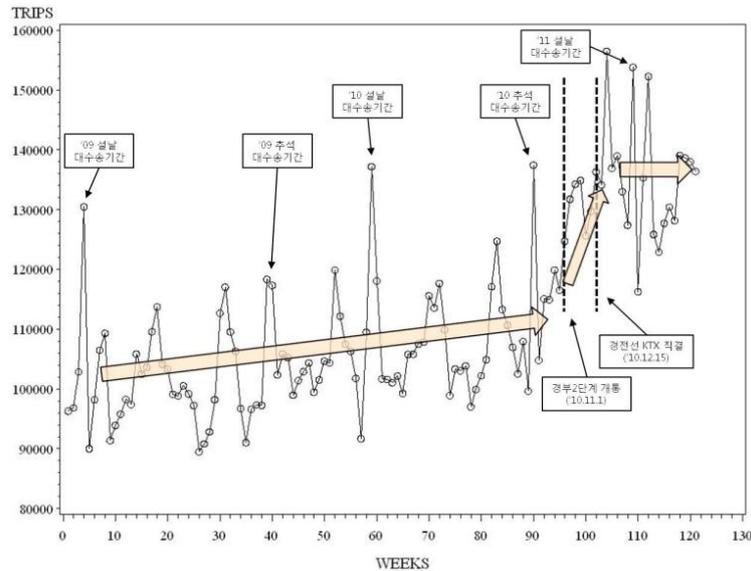


그림 1: KTX 주평균 이용실적 (2009년 1월~2011년 4월)

본 연구는 현재 운영 중인 KTX의 단기수요를 예측하기 위해 KTX 이용자의 통행행태를 분석하고자 한다. 통행행태 분석은 단기수요 예측방향을 설정하기 위함으로 일일 통행량의 이상치 특성 분석, 동질 요일 선정 및 특성 분석, 동질 O/D 선정 및 특성 분석으로 구분하여 실시한다. 첫째, 일일 통행량의 이상치 특성 분석은 통계적 방법을 이용하며, 둘째, 동질 요일 선정 및 특성 분석은 분산분석, 마지막으로 동질 O/D 선정 및 특성 분석은 군집분석을 이용한다. 본 연구는 2장에서 연구자료 및 분석방법에 대해 기술한다. 3장에서는 통행행태 분석결과를 제시하며, 4장에서는 분석결과를 기반으로 단기수요 예측방향에 대해 논한다. 마지막으로 5장에서 연구결과를 정리하고 향후 연구를 제시한다.

2. 연구자료 및 분석방법

2.1. 연구자료

본 연구는 현재 운영 중인 KTX의 단기수요 예측방향을 설정하기 위한 통행행태 분석이 목적이다. 단기수요 예측은 시계열모형과 같이 과거 이용실적을 기반으로 예측하기 때문에 통행행태 분석 자료는 동질 서비스 기간의 자료를 이용하는 것이 필요하다. 그림 1은 2009년 1월부터 2011년 4월까지 KTX 이용자의 주 평균 이용실적이다. 2010년 11월 1일 경부고속철도 2단계 개통과 2010년 12월 15일 경전선 KTX 직결로 인해 이용실적이 증가하였다. 본 연구에서는 경부고속철도 2단계 개통 전/후의 통행행태에 차이가 있는지도 파악하기 위해 분석기간을 경부고속철도 2단계 개통 전(2010년 6월 20일~2010년 10월 30일)과 경부고속철도 2단계 개통 후(2010년 12월 19일~2011년 4월 30일)로 구분하였다.

경부고속철도 2단계 개통 전의 KTX 정차역은 23개 역으로 이용 가능한 철도역간 O/D는 506개이나, 해당 기간 동안 1인/일 이상 실적이 있는 O/D는 292개이다. 또한 경부고속철도 2단계 개통 후에는 31개 역으로 이용 가능한 철도역간 O/D는 930개이나, 해당 기간 동안 1인/일 이상 실적이 있는 O/D는 471개이다. 분석기간 동안의 열차운행횟수는 경부고속철도 2단계 개통 직전 144회/일(화~목),

181회/일(토·일)에서 2011년 3월 1일 현재 172회/일(화~목), 216회/일(토), 215회/일(일)로 19% 증가하였다. 운임은 2007년 7월 이후 동일하다. 두 개의 분석기간 동안 일부 열차운행횟수가 변경되기는 하였으나 소폭의 조정이며, 동일한 운임이 적용되었기 때문에 각각 동질 서비스 기간이라 할 수 있다.

2.2. 분석방법

본 연구에서는 일일 통행량의 이상치 특성 분석, 동질 요일 선정 및 특성 분석, 동질 O/D 선정 및 특성 분석으로 단기수요 예측방향을 설정하고자 한다. 첫째, 일일 통행량의 이상치 특성 분석은 해당 분석기간의 일일 평균 통행량(μ)의 표준편차(σ) 배수 범위를 초과한 자료를 이상치로 간주한다. 표준편차의 1배 범위는 신뢰구간 68.3%를 초과한 값이며, 2배 범위는 95.4%를 초과한 값, 3배 범위는 99.7%를 초과한 값이다. 둘째, 동질 요일 선정 및 특성 분석을 위해서는 요일별 일일 통행량을 이용하여 분산분석(ANOVA)을 실시한다. 요일 구분에 따른 일일 통행량의 영향을 분석하기 위해 일원배치법을 이용하며, 각 요일 간 차이 여부를 확인하기 위해 최소유의차(Least Significant Difference; LSD)와 튜키(Tukey) 방법을 이용하여 다중비교 결과를 분석한다. 셋째, 동질 O/D 선정 및 특성 분석은 O/D 간 특성을 파악하기 위함으로 O/D 간 특성을 파악할 수 있는 변수들을 이용하여 군집분석을 실시한다.

군집분석에 의한 분류결과를 그림으로 나타낸 것을 덴드로그램(dendrogram) 또는 수목도(tree diagram)라 부른다. 객체를 분류하여 얻어진 덴드로그램이 원래의 거리행렬을 얼마나 잘 적합 시켰는가를 측정하는 측도는 코페네틱 상관계수(cophenetic correlation coefficient)와 결합계수(agglomerative coefficient)가 있다. 객체 r 과 s 의 거리인 d_{rs} 와 덴드로그램에서 객체 r 과 s 가 처음으로 같은 그룹으로 분류될 때의 분류수준을 δ_{rs} 로 하면 이 둘 사이의 상관계수를 코페네틱 상관계수라 한다. 이 값이 1에 가까우면 계층분류법에 의한 덴드로그램 적합이 적정함을 의미한다. 또한 결합계수는 Rousseeuw (1986)이 제안한 것으로 i 번째 관측치가 처음으로 그룹을 형성할 때의 거리를 최종 하나의 그룹이 될 때의 거리로 나눈 값을 $d(i)$ 라 하면, $1 - d(i)$ 들의 평균을 결합계수라 한다 (정강모와 김명근, 2007).

군집 수 결정방법은 Sarle (1983)이 제안한 방법으로 군집 수와 CCC(Cubic Clustering Criterion)의 산점도를 그려 CCC 값이 3 이상이고 최대인 경우 그 때의 군집 수로 결정한다. 또한 Pseudo Hotelling's T^2 검정 통계량은 두 군집 다변량 평균의 차이를 보는 통계량으로 개체의 군집 간 평균의 차이가 유의하지 않으면 두 군집을 합치고, 유의하면 군집 그대로 유지하는 방법이다. PST2 값이 크다는 것은 군집간 거리가 멀다는 것을 의미하므로 군집을 나누는 것이 좋다. 반대로 값이 적다는 것은 군집 간 거리가 가깝다는 것으로 군집을 합치는 것이 좋다. PST2 값이 주변 값들보다 크면 이 값 바로 전의 군집 개수가 적절하다 (권세혁, 2008).

3. 통행행태 분석

3.1. 일일 통행량 이상치 특성 분석

KTX 일일 통행량의 이상치 특성을 분석하기 위해 2009년 1년간 자료를 이용하였다. 일일 통행량 분포는 평균(mean)이 중앙값(median)보다 크며, 오른쪽 꼬리 분포를 가지고 있다. 이러한 원인은 설날, 추석의 대수송기간과 연휴 또는 징검다리 휴일과 같이 지역간 이동이 집중되는 기간에 KTX 통행이 크게 증가하였기 때문이다. 이런 관측치는 이상치라 할 수 있는데 일반적인 수요예측방법으로 추정하기 쉽지 않다. 또한 이런 이상치를 포함할 경우 모수를 추정함에 있어 편의(bias)가 발생할 수 있기 때문에 이상치 특성을 분석하여 제거하는 것이 필요하다.

이상치 특성 분석을 위해 다음과 같이 전체 일일 통행량 기준과 요일별 통행량 기준으로 구분하여 시행하였다. 첫째, 전체 일일 통행량을 이용하여 통행량의 평균과 표준편차를 산출하고, 표준편차의 배수범위에 해당하는 자료를 이상치로 간주하였다. 분석결과 표준편차의 1배 범위인 신뢰구간

표 1: 경부고속철도 2단계 개통 전/후의 KTX 이용 통행량의 평균 및 표준편차(단위: 인/일)

구분	경부고속철도 2단계 개통 전 (2010년 6월 20일~2010년 10월 30일)						경부고속철도 2단계 개통 후 (2010년 12월 19일~2011년 4월 30일)					
	이상치 제거 전			이상치 제거 후			이상치 제거 전			이상치 제거 후		
	자료 건수	평균	표준 편차	자료 건수	평균	표준 편차	자료 건수	평균	표준 편차	자료 건수	평균	표준 편차
전체 자료	133	110,514	20,981	125	108,522	18,940	133	135,373	24,351	125	134,557	22,831
일	19	120,644	14,572	18	122,353	12,890	19	151,638	19,641	18	154,762	14,568
월	19	103,033	14,225	17	98,619	4,984	19	123,781	9,458	18	122,521	7,922
요일별	19	94,682	15,719	18	91,548	8,001	19	116,493	15,334	18	113,794	10,124
구분	19	96,294	19,345	18	92,264	8,341	19	118,142	17,636	18	114,943	11,112
자료	19	99,915	15,601	18	96,749	7,493	19	125,407	20,783	18	121,393	11,537
목	19	132,724	8,682	18	133,717	7,745	19	161,751	14,358	17	162,181	9,602
금	19	126,304	18,043	18	123,854	14,965	19	150,397	22,372	18	153,841	17,071
토	19			18			19			18		

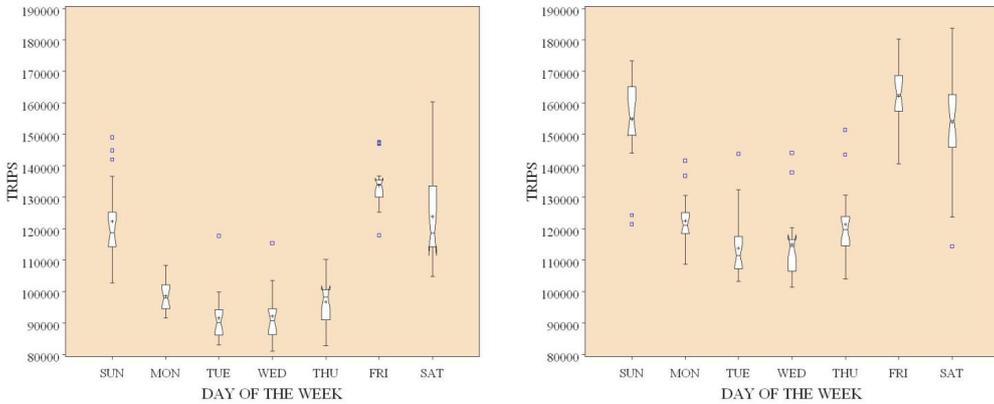
68.3%를 초과한 값은 전체 365건 중 125건이었으며, 2배 범위인 신뢰구간 95.4%를 초과한 값은 12건, 3배 범위인 신뢰구간 99.7%를 초과한 값은 1건이 해당 되었다. 둘째, 전체 일일 통행량을 요일별로 구분하여 요일별 통행량의 평균과 표준편차를 산출하고, 표준편차의 배수범위에 해당하는 자료를 이상치로 간주하였다. 분석결과 표준편차의 1배 범위인 신뢰구간 68.3%를 초과한 값은 전체 365건 중 69건이었으며, 2배 범위인 신뢰구간 95.4%를 초과한 값은 16건, 3배 범위인 신뢰구간 99.7%를 초과한 값은 6건이 해당 되었다.

이러한 분석결과를 정리하면 첫째, 전체 통행량의 표준편차 1배를 이상치 판단기준으로 선정할 경우 요일별 통행량을 기준할 때 보다 많은 건수가 이상치에 해당된다. 이는 요일별 통행량으로 구분할 경우 정상치임에도 불구하고 일부 자료가 이상치로 판단되는 경우가 발생한다. 둘째, 전체 통행량의 표준편차 2배 또는 3배를 이상치 판단기준으로 선정할 경우 요일별 통행량을 기준할 때 보다 적은 건수가 이상치에 해당된다. 이는 전체 일일 통행량의 표준편차가 요일별 통행량의 표준편차 보다 크기 때문이다. 이는 요일별 통행량으로 구분할 경우 이상치 임에도 불구하고 일일 통행량의 표준편차가 크기 때문에 2배 또는 3배를 이상치 판단기준으로 선정하면 일부 자료가 정상치로 판단되는 경우가 발생한다. 그러므로 이상치 판단기준은 요일별 통행량을 기준으로 정하는 것이 바람직하다. 요일별 통행량 표준편차의 2배를 기준할 경우 설날, 추석 대수송기간과 연휴 또는 징검다리 휴일이 포함되며, 표준편차의 3배를 기준할 경우 설날, 추석 대수송기간의 일부만 포함된다. 이와 같은 분석결과를 토대로 이상치 판단기준은 요일별 통행량 표준편차의 2배 기준이 적절할 것으로 판단된다.

경부고속철도 2단계 개통 전/후 자료에 이상치 판단기준을 적용한 결과 표 1과 같다. 주중 요일은 상한 이상치들이 많았던 반면, 주말 요일은 하한 이상치들이 많았다. 경부고속철도 2단계 개통 전 토요일은 다른 주말 요일과 다르게 상한 이상치가 더 많았다. 이러한 이상치 제거로 인해 표준편차가 상당히 감소하였다. 주중 요일인 월~목은 통행량 평균과 표준편차가 유사한 반면, 주말 요일인 금~토는 통행량 평균은 유사하지만 표준편차 차이는 큰 것으로 나타났다.

3.2. 동질 요일 선정 및 특성 분석

열차운행횟수는 철도수요 특성에 따라 요일로 구분하여 설정할 수 있다. 그러나 모든 요일을 구분하여 열차운행횟수를 결정하는 것은 비효과적이다. 그러므로 요일별 통행량을 이용하여 동질 요일을 선정하는 것이 필요하다. 철도수요 특성이 동질한 요일이면 열차운행횟수를 동일하게 정할 수 있다. 동질 요일을 선정하기 위해 분산분석을 이용하여 요일 구분에 따른 통행량 차이를 분석하였다. 앞에서 분석한 이상치 제거 후의 요일별 통행량 Box-Plot은 그림 2와 같다. 경부고속철도 2단계 개통 전/후로



(a) 경부고속철도 2단계 개통 전

(b) 경부고속철도 2단계 개통 후

그림 2: 경부고속철도 2단계 개통 전/후의 고속철도 이용 통행량 Box-Plot

표 2: 경부고속철도 2단계 개통 전/후의 요일별 차이여부

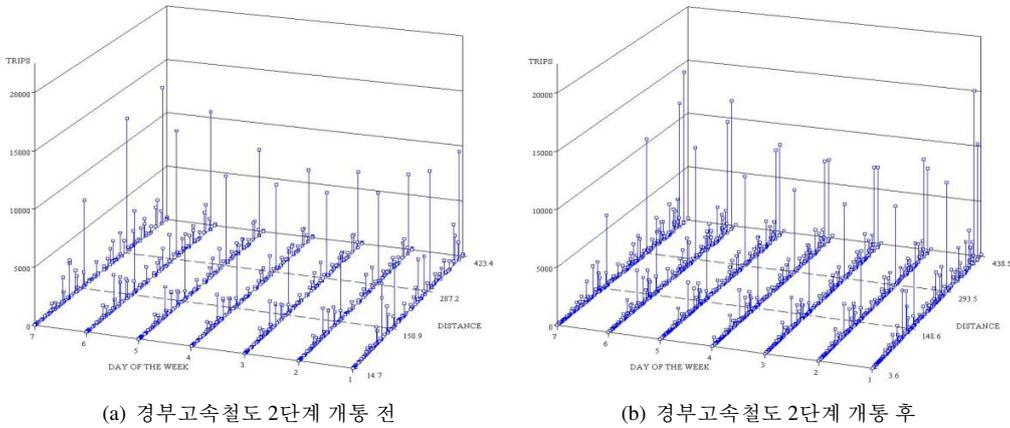
(a) 경부고속철도 2단계 개통 전								(b) 경부고속철도 2단계 개통 후							
	일	월	화	수	목	금	토		일	월	화	수	목	금	토
일	-						차이 없음	일	-						차이 없음
월		-	차이 없음	차이 없음	차이 없음			월		-	차이 없음	차이 없음	차이 없음		
화		차이 없음	-	차이 없음	차이 없음			화		차이 없음	-	차이 없음	차이 없음		
수		차이 없음	차이 없음	-	차이 없음			수		차이 없음	차이 없음	-	차이 없음		
목		차이 없음	차이 없음	차이 없음	-			목		차이 없음	차이 없음	차이 없음	-		
금						-		금	차이 없음					-	차이 없음
토	차이 없음						-	토	차이 없음					차이 없음	-

주: 진하게 표시된 부분은 최소유의차와 튜키방법 공통결과인 반면, 진하게 표시되지 않은 차이없음은 튜키방법에 의한 결과임

구분하였으며, 요일 순서는 일요일에서 토요일로 구분되어 있다.

경부고속철도 2단계 개통 전 자료를 이용한 분산분석 결과는 검정통계량 F 값이 57.97로 유의수준 0.01에서 요일별 통행량에 차이가 있음이 나타났다. 또한 총 변동 중 요일 구분이 통행량에 미치는 영향을 설명하는 R^2 가 0.7465로 요일 구분이 통행량 차이에 영향이 있음을 알 수 있다. 최소유의차와 튜키방법을 이용한 다중비교 결과는 표 2의 왼쪽 표와 같다. 경부고속철도 2단계 개통 전에는 주중 요일인 월~목, 주말 요일인 토~일 간에 차이가 없는 것으로 나타났으나, 다른 요일 간은 차이가 있는 것으로 나타났다. 즉, 주중, 금, 토~일로 구분되는 것으로 나타났다.

경부고속철도 2단계 개통 후 자료를 이용한 분산분석 결과는 검정통계량 F 값이 54.20로 유의수준 0.01에서 요일별 통행량에 차이가 있음이 나타났다. 또한 총 변동 중 요일 구분이 통행량에 미치는 영향을 설명하는 R^2 가 0.7338로 요일 구분이 통행량 차이에 영향이 있음을 알 수 있다. 최소유의차와 튜키방법을 이용한 다중비교 결과는 표 2의 오른쪽 표와 같다. 경부고속철도 2단계 개통 후에는 주중 요



(a) 경부고속철도 2단계 개통 전

(b) 경부고속철도 2단계 개통 후

그림 3: 경부고속철도 2단계 개통 전/후의 통행거리대별 요일별 통행량 분포

일인 월~목, 주말 요일인 금~일간의 차이가 없음이 나타났다. 즉, 주중과 주말로 구분되는 것으로 나타났다.

3.3. 동질 O/D 선정 및 특성 분석

KTX 통행은 대도시권 통행이나 다른 교통수단의 지역간 통행분포와 차이가 있다. 경부고속철도 2단계 개통 전/후를 구분하여 통행거리별 통행량 분포(Trip Length Distribution; TLD)를 살펴보면 그림 3과 같다. 일반적인 통행거리대별 통행량 분포는 통행거리가 증가함에 따라 우하향 하는 통행행태를 나타내는 반면, KTX 통행은 주요 O/D(서울-부산, 서울-동대구, 서울-대전 등)의 통행량 비중이 큰 반면 다른 O/D는 상대적으로 작다. 또한 통행거리가 증가함에 따라 우하향 하는 통행행태를 나타내지 않는다. 이러한 특징은 KTX가 다른 교통수단에 비해 중장거리 통행에서 경쟁력이 높은 반면, 단거리 통행에서는 다른 교통수단에 비해 통행비용이 높고, 접근통행 저항으로 인해 경쟁력이 낮기 때문이다.

열차운행횟수를 산정하기 위해서는 구간재차인원을 산출할 수 있는 철도역간 O/D를 예측하는 것이 필요하다. 그러나 모든 O/D에 대해 개별적인 모형을 적용하는 것보다 동질 O/D를 파악하여 해당 동질 O/D별로 모형을 적용하는 것이 효과적이다. 그러므로 본 연구에서는 기존 KTX 이용실적으로 산출 가능한 설명변수를 이용하여 동질 O/D를 분류하고, 이들 특성을 분석하고자 한다. 설명변수는 통행빈도, 통행량 평균과 표준편차, 통행거리, 운임을 1차로 선정하였다. 통행빈도는 분석기간 동안 해당 O/D의 통행 발생일수이며, 통행량 평균과 표준편차는 분석기간 동안 해당 O/D의 통행량의 평균과 표준편차, 통행거리와 운임은 해당 O/D의 통행거리, 운임이다. 선정된 설명변수 간의 상관분석을 실시한 결과는 표 3과 같으며, 통행량 평균과 표준편차, 통행거리와 운임이 높은 양의 상관관계를 나타내기 때문에 최종 설명변수는 통행빈도, 통행량 평균, 통행거리를 선정하여 이들 변수를 이용하여 군집분석을 실시하였다. 군집분석을 위한 결합법은 표 4와 같이 최단결합법, 최장결합법, 군평균결합법, 최소분산결합법을 적용하였으나, 군집분석 적합도 검정결과 코페네틱 상관계수와 결합계수가 공통적으로 높은 최장결합법과 군평균결합법이 적절할 것으로 판단된다. 이러한 분석결과는 경부고속철도 2단계 개통 전/후 동일한 결과를 나타냈다.

경부고속철도 2단계 개통 전 자료에 대한 군집분석 결과 선정된 군집수는 최장결합법으로 분석할 경우 5개 군집, 군평균결합법은 8개 군집으로 구분된다. 군집 수 선정기준은 CCC와 PST2를 이용하였다. 이들 군집들의 특성은 표 5와 같다. 우선 최장결합법에 의한 결과를 살펴보면, 군집 5는 통행량 평

표 3: 동질 O/D 선정을 위한 설명변수 간의 상관분석 결과

구분	통행빈도	통행량 평균	통행량 표준편차	통행거리	운임
통행빈도	1				
통행량 평균	0.1478 (0.0114)	1			
통행량 표준편차	0.1600 (0.0062)	0.9722 (< .0001)	1		
통행거리	0.3057 (< .0001)	0.2452 (< .0001)	0.2879 (< .0001)	1	
운임	0.3122 (< .0001)	0.3073 (< .0001)	0.3392 (< .0001)	0.9737 (< .0001)	1

표 4: 군집분석 적합도 검정 결과

구분	경부고속철도 2단계 개통 전 (2009년 1월 1일~2010년 12월 14일)		경부고속철도 2단계 개통 후 (2010년 12월 15일~2011년 3월 31일)	
	코페네틱 상관계수	결합계수	코페네틱 상관계수	결합계수
최단결합법	0.7687	0.9764	0.7540	0.9825
최장결합법	0.8114	0.9918	0.8648	0.9940
군평균결합법	0.8880	0.9885	0.8890	0.9920
최소분산결합법	0.5633	0.9966	0.4566	0.9977

표 5: 경부고속철도 2단계 개통 전 군집별 특성(단위: 인/일, km)

구분	경부고속철도 2단계 개통 전					경부고속철도 2단계 개통 후				
	O/D 개수	통행빈도	통행량 평균	통행량 표준편차	통행거리	O/D 개수	통행빈도	통행량 평균	통행량 표준편차	통행거리
전체	292	111.8	371.8	95.6	176.1	292	111.8	371.8	95.6	176.1
군집1	162	122.3	340.4	95.2	250.8	34	122.8	563.9	168.6	372.7
군집2	94	116.4	72.3	22.9	71.3	193	123.5	221.2	61.1	163.1
군집3	17	18.5	1.3	0.5	101.1	29	94.1	2.6	2.0	98.0
군집4	13	63.1	1.8	1.1	50.6	17	18.5	1.3	0.5	101.1
군집5	6	125.0	7765.7	1703.9	287.1	12	61.7	1.7	1.1	53.0
군집6						2	125.0	5549.5	865.8	159.8
군집7						4	125.0	8873.8	2123.0	350.8
군집8						1	80.0	2.7	1.9	363.3

군이 7,765.7인/일로 서울-부산, 서울-동대구, 서울-대전 간 O/D가 해당되며, 군집 1, 2는 중간 정도의 통행량이나 군집 1이 군집 2보다 장거리 O/D로 구성되어 있다. 군집 3, 4는 통행량 평균이 2 이하로 매우 낮은 통행량을 나타내며, 통행빈도와 통행거리에 따라 차이를 보이고 있다. 군집 3, 4는 전체 O/D의 10%를 차지하고 있다. 군평균결합법에 의한 결과는 최장결합법 보다 많은 군집으로 세분되었다. 서울-부산, 서울-동대구(군집 7), 서울-대전(군집 6)이 두 개 군집으로 구분되었으며, 군집 3, 4, 5, 7은 통행량 평균이 3 이하로 매우 낮은 통행량을 나타내며, 통행빈도와 통행거리에 따라 차이를 보이고 있다. 이들 군집은 전체의 20%를 차지하고 있다. 군집 구분특성을 정리하면, O/D 통행량 다소에 따라 크게 구분되며 장거리, 단거리로 또 다시 구분되는 특징이 나타난다. 통행량 평균이 클수록 표준편차 또한 크게 나타나는데 이러한 원인은 통행량이 큰 O/D는 주중과 주말 통행량 차이가 크기 때문이다. 그러므로 이들 군집은 수요를 예측할 때 주중과 주말 차이를 반영할 수 있는 예측기법 적용이 필요하다.

경부고속철도 2단계 개통 후 자료에 대한 군집분석 결과 선정된 군집수는 최장결합법으로 분석할 경우 5개 군집, 군평균결합법은 7개 군집으로 구분된다. 이들 군집들의 특성은 표 6과 같다. 우선 최

표 6: 경부고속철도 2단계 개통 후 군집별 특성(단위: 인/일, km)

구분	최장결합법					균평균결합법				
	O/D 개수	통행빈도	통행량 평균	통행량 표준편차	통행거리	O/D 개수	통행빈도	통행량 평균	통행량 표준편차	통행거리
전체	471	108.7	286.1	74.9	184.0	471	108.7	286.1	74.9	184.0
군집1	72	37.9	3.5	2.9	107.6	46	24.5	3.6	3.3	69.5
군집2	235	119.6	101.3	29.4	118.5	241	123.0	238.3	71.4	272.2
군집3	158	124.1	394.1	114.3	312.1	134	122.8	213.6	57.6	75.4
군집4	2	125.0	5867.4	803.8	159.8	32	78.1	2.4	1.8	85.6
군집5	4	125.0	9175.4	2108.0	354.6	12	58.2	4.9	3.0	272.3
군집6						2	125.0	5867.4	803.8	159.8
군집7						4	125.0	9175.4	2108.0	354.6

장결합법에 의한 결과를 살펴보면, 군집 5는 서울-부산, 서울-동대구 간 O/D가 해당되며, 군집 4은 서울-대전 간 O/D가 해당된다. 군집 2, 3은 중간 정도의 통행량이나 군집 3이 군집 2보다 장거리 O/D로 구성되어 있다. 군집 1은 통행량 평균이 4 이하로 매우 낮은 통행량을 나타낸다. 이들 O/D는 전체 O/D의 15%를 차지하고 있다. 균평균결합법에 의한 결과는 최장결합법 보다 많은 군집으로 세분되었다. 군집 6, 7은 최장결합법의 군집 5, 6과 같다. 군집 2, 3은 중간 정도의 통행량이나 군집 2가 군집 3보다 장거리 O/D로 구성되어 있다. 군집 1, 4, 5는 통행량 평균이 4 이하로 매우 낮으나, 통행빈도와 통행거리에 의해 구분되었다. 경부고속철도 2단계 개통 후의 군집 구분은 1단계와 비슷한 결과를 나타내고 있다. O/D 통행량 다소에 따라 크게 구분되며 장거리, 단거리로 또 다시 구분되는 특성이 나타난다. 그러나 통행량이 작은 O/D를 세 개 군집으로 구분하여 군집 수가 증가하였다.

4. 단기 수요예측 방향

본 연구에서는 KTX 단기수요 예측을 위해 통행행태를 분석하였다. 그 결과 단기수요 예측을 위해 다음과 같은 방향을 설정할 수 있었다. 첫째, 이상치 제거와 결측치 보완이 필요하다. KTX 이용 통행량은 오른쪽 꼬리 분포를 가지는데, 이는 하한 이상치 보다 상한 이상치가 많기 때문이다. 이러한 이상치는 대수송기간, 연휴 등으로 지역간 이동이 집중되는 기간에 발생한다. 이들 이상치를 포함하여 장래 수요수요를 예측할 경우 추정치에 편의가 발생할 수 있으므로 이상치를 제거하고, 제거된 결측치를 대체(imputation)하는 것이 필요하다. 이상치를 제거하는 방법은 요일별 통행량 표준편차의 2배 범위인 신뢰구간 95.4%를 초과한 값을 이상치로 판단하는 것이 적합할 것으로 판단된다. 이와 같이 적용할 경우 설날, 추석의 대수송기간과 연휴 또는 징검다리 휴일이 포함된다.

둘째, 주중과 주말 통행량 차이를 반영하여 예측하는 것이 필요하다. 분산분석을 이용하여 요일 간 통행량의 차이여부를 검정한 결과 경부고속철도 2단계 개통 전에는 주중 요일인 월~목, 주말 요일인 토~일 간의 차이가 없었다. 개통 후에는 주중 요일인 월~목, 주말 요일인 금~일 간 차이가 없었다. 총 변동 중 요일 구분이 통행량 차이에 미치는 영향을 설명하는 R^2 가 0.7465, 0.7338로 영향이 있음을 알 수 있었다. 이와 같이 주중과 주말 통행량에 큰 차이가 있는 이유는 주중 통행은 업무 및 출퇴근 통행이 주를 이루는 반면 주말 통행은 여가통행, 친지방문 통행, 주말가정방문 통행이 주를 이루기 때문이다. 즉, 주중 통행과 주말 통행의 통행목적에 큰 차이를 보이기 때문이며, 이로 인해 주말 통행의 변동이 주중 통행에 비해 큰 특징이 있다. 주중과 주말 통행차이를 반영하는 방법은 주중과 주말 통행을 구분하여 예측하거나, 시계열 자료를 주기(cycle), 추세(trend), 계절(seasonality)변동으로 구분하여 예측하는 방법이 있다.

셋째, 동질 O/D 군집의 특성에 적합한 모형을 적용하는 것이 필요하다. 군집별 특성을 살펴보면 통

행량이 작은 군집은 요일 구분이 큰 영향을 미치지 않는다. 그러나 통행량이 큰 군집은 주기변동 뿐만 아니라 추세 및 계절변동이 존재한다. 이러한 특성을 반영하여 동질 O/D 군집에 따라 예측모형을 개별적으로 적용하는 것이 필요하다. 특히 유가 및 국내경기 변동에 영향을 받는 O/D 군집 또한 통행량이 큰 군집들로 이들 O/D 군집들을 대상으로 사회경제적 요인에 의한 영향여부를 분석하는 것이 요구된다.

5. 결론

KTX 단기수요 예측을 위해 KTX 통행행태의 분석결과는 다음과 같다. 첫째, 이상치 특성을 분석한 결과 설날, 추석의 대수송기간과 연휴 또는 징검다리 휴일 등에서 상한 이상치가 나타났으며, 이들 이상치는 모수를 추정함에 있어 편의를 발생시키기 때문에 제거하는 것이 바람직하다. 둘째, 동질 요일 분석결과 주중(월~목)과 주말(금~일)로 구분되며, 요일 구분이 통행량에 미치는 영향이 있음을 통계적으로 검정하였게적으실째, 동질 O/D 분석결과 통행량 다소에 따라 크게 구분되며, 장거리, 단거리로 또 다시 구분되는 특성이 나타났게적으또한 군집별 주중 및 주말 통행량 변동특성이 다르기 때문에 예측모형을 적용함에 있어 군집으로 구분하여 적용함이 필요함을 파악할 수 있었다. 이러한 분석결과 단기수요예측을 위한 방향을 다음과 같이 설정하였다. 첫째, 이상치 제거와 결측치 보완이 필요하다. 둘째, 주중과 주말 통행량 차이를 반영하여 예측하는 것이 필요하다. 셋째, 동질 O/D 군집의 특성에 적합한 모형을 적용하는 것이 필요하다.

본 연구의 후속연구로는 첫째, 군집 특성에 따라 시계열 예측모형 개발이 필요하다. 둘째, 모형의 정산 및 검증을 위해 시계열 예측에 필요한 과거 데이터량과 갱신간격을 결정하는 연구가 필요하다. 마지막으로 대수송기간과 연휴기간의 통행량 예측방법을 개발하는 것이 필요하다. 대수송기간 등의 수요예측을 위해서는 고속도로 TCS(Toll Collection System) 자료를 같이 이용하는 것도 필요할 것으로 판단된다.

참고 문헌

- 국토해양부 (2010). <국가철도망구축계획 수립 연구>, 국토해양부.
- 권세혁 (2008). <다변량 데이터 분석과 활용>, 자유아카데미.
- 정강모, 김명근 (2007). <R 기반 다변량 분석>, 교우사.
- 최태성, 김성호 (2004). 철도여객수요예측을 위한 Holt-Winters모형의 초기값 설정방법 비교, <한국철도학회논문집>, 7, 9-13.
- 한국철도공사 연구원 (2010). <KTX 중장기 수송수요 예측 연구>, 한국철도공사.
- Kanafani, A. (1983). *Transportation Demand Analysis*, McGraw-Hill, New York.
- MVA Consultancy (2007). *Rail Passenger Demand Forecasting Research*.
- Rousseeuw, P. J. (1986). A Visual Display for Hierarchical Classification, In *Data Analysis and Informatics 4*, edited by E. Diday, Y. Escoufier, L. Lebart, J. Pages, Y. Schektman, and R. Tomassone, North-Holland, Amsterdam, 743-748.
- Sarle, W. S. (1983). *Cubic Clustering Criterion*, SAS Technical Report A-108, Cary, NC: SAS Institute Inc.
- Tsai, T., Lee, C. and Wei, C. (2005). Design of dynamic neural networks to forecast short-term railway passenger demand, *Journal of the Eastern Asia Society for Transportation Studies*, 6, 1651-1666.
- Wardman, M., Lythgoe, W. and Whelan, G. (2007). Rail passenger demand forecasting: Cross-sectional models revisited, *Research in Transportation Economics*, 20, 119-152.

Travel Behavior Analysis for Short-Term KTX Passenger Demand Forecasting

Hansoo Kim^a, Donghee Yun^b, Sungduk Lee^{1,c}

^aKorail Research Institute, Korea Railroad Corporation

^bDeajeon-Chungnam HQ, Korea Railroad Corporation

^cDepartment of Information and Statistics, Chungbuk National University

Abstract

This study analyzes the travel behavior for short-term demand forecasting model of KTX. This research suggests the following. First, the outlier criteria is considered to appropriate twice the standard deviation of the traffic. Second, the result of a homogeneity test using ANOVA analysis has been divided into weekdays(Mon~Thu) and weekends(Fri~Sun). Third, a cluster analysis for O/D pairs using trip frequency, traffic averages and the distance between stations was performed.

Keywords: Short-term demand forecasting, travel behavior analysis, KTX, cluster analysis, ANOVA.

¹ Corresponding author: Professor, Department of Information and Statistics, Chungbuk National University, Cheongju, Chungbuk 261-763, Korea. E-mail: sdlee@chungbuk.ac.kr