



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

김 영 준 교수지도
기술경영학석사 학위논문

택배 물동량 추정에 영향을 미치는
요인에 관한 연구

A Study on Factors Influencing
Forecasting of Short-term Trade
Volume in Parcel Service Industry

고려대학교 기술경영전문대학원

기술경영학과

남 기 찬

2017년 12월 16일

技術經營學碩士學位論文

택배 물동량 추정에 영향을 미치는
요인에 관한 연구

A Study on Factors Influencing
Forecasting of Short-term Trade
Volume in Parcel Service Industry

高麗大學校 技術經營專門大學院

技術經營學科

南基讚

2017年 12月 16日



金 永 竣 教 授 指 導
技術經營學碩士學位論文

택배 물동량 추정에 영향을 미치는
요인에 관한 연구

A Study on Factors Influencing
Forecasting of Short-term Trade
Volume in Parcel Service Industry

이 論文을 技術經營學碩士學位 論文으로 提出함.

2017年 12月 16日

高麗大學校 技術經營專門大學院

技術經營學科

南 基 讚 (印)



남 기 찬 의 技術經營學碩士學位論文

審 査 를 完 了 함 .

2017年 12月 16日

委 員 長 _____ (印)

委 員 _____ (印)

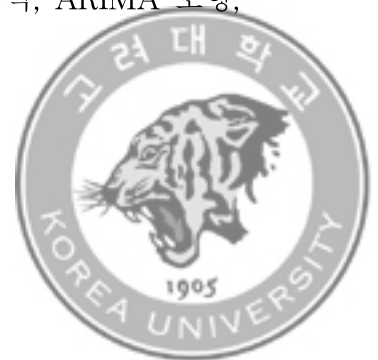
委 員 _____ (印)



국문 초록

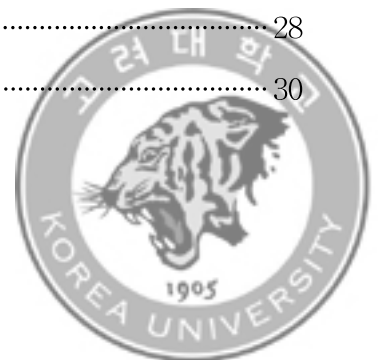
물류 서비스 시장의 성장에 따른 대 고객 서비스를 위한 예측 필요성이 강조되고 있는 상황이다. 기존 물류 서비스 업체도 시장점유율을 유지하고 견고히 하기 위하여 경쟁업체와 차별화되는 경쟁력을 갖추어야 하지만, 단순화된 집화-분류-배송의 프로세스로 경쟁업체와의 차별화된 품질을 제공하는 것이 어렵다. 본 연구에서는 국내 최대 종합물류기업 C사의 택배 지점들 중 수도권을 중심으로 한 택배 물동량 수요예측을 진행하여 물동량 예측에 영향을 미치는 요인을 분석하고 여러 요인이 반영된 日 단위 수요 예측에 대한 정확도 높은 예측모형을 제시하고자 한다. 궁극적으로 물류 고유의 문제를 해결하고 고객의 트렌드 파악과 자원의 효율적인 활용에 좋은 매개체로 물동량 예측 방법을 제시하여 물류 자원 운영, 전략, 마케팅 등 다양한 기업 활동의 의사결정 과정에 도움을 주어 기업의 장기 수익 개선에 기여할 수 있도록 하고자 한다. C사의 경우 예측 모델에 대한 반영을 각 터미널 단위로 반영하지 않고 지역적 위치와 日 단위 Sub 터미널의 집합을 묶는 군집화를 실시하였다. 선행연구 및 인터뷰를 통해 도출된 기준별 변수를 요인 분석을 통해 주요 요인을 선정하고 회귀 예측 모형을 통해 시물레이션 하여 결과 값을 토대로 가장 정확도가 높은 예측 모형을 선정하는 데 중점을 두었다. 본 연구는 물동량의 예측을 日 단위 예측이 활성화되기 전에 연구를 진행하였다는데 큰 의의가 있으며 앞으로 물동량의 단기 예측을 수행하는데 이론적으로나 실무적으로 유의미한 시사점을 제공했기를 기대해 본다.

주제어 : 물동량 예측, 요인분석, 시계열 분석, 다중회귀분석, ARIMA 모형, 절대평균편차, 절대평균오차비율, 오차제곱평균



목 차

목 차	i
표 목차	iii
그림 목차	v
제 1 장 서 론	1
제 1 절 연구의 배경 및 목적	1
제 2 절 연구의 방법 및 구성	3
제 2 장 이론적 배경	6
제 1 절 택배 서비스	6
제 1 항 택배 서비스 개요	6
제 2 항 택배 프로세스	9
제 3 항 택배 물동량 분석	12
제 2 절 수요 예측	16
제 1 항 요인분석의 개요	16
제 2 항 수요예측의 개요	17
제 3 항 예측의 정확도 측정	24
제 3 절 물동량 예측에 관한 선행연구	27
제 1 항 물동량 예측 모형	27
제 2 항 물동량 예측 영향 요인	28
제 3 항 기존 연구의 한계	30



제 3 장 연구 방법	32
제 1 절 연구모형의 설정	32
제 2 절 단계별 연구	33
제 1 항 군집화 및 데이터 정제	33
제 2 항 변수 요인분석 및 선정	35
제 3 항 예측 모형 선정 및 평가	40
제 4 장 연구 결과	42
제 1 절 핵심요인 및 상관관계 분석	42
제 1 항 핵심요인 및 신뢰도 분석	43
제 2 항 상관관계 분석	47
제 2 절 물동량 예측 모델	50
제 1 항 다중회귀분석	50
제 2 항 회귀모형별 예측 모델	59
제 3 항 시계열분석	61
제 3 절 물동량 예측 모델 평가	66
제 5 장 결 론	71
제 1 절 연구 결과 요약	71
제 2 절 한계점 및 향후 과제	72
참고문헌	74
Abstract	82



표 목차

<표 1> 수도권 내 물동 비중 현황	13
<표 2> 물량 밀도 현황	14
<표 3> 수요예측 기법	19
<표 4> 인터뷰 및 설문 조사 현황	36
<표 5> 인터뷰를 통한 예측 영향요인 선정	37
<표 6> 예측모형별 분석을 위한 활용변수 리스트	39
<표 7> 설명된 총분산	44
<표 8> 요인분석 결과	45
<표 9> 항목 총계 통계량	47
<표 10> 상관분석 결과	48
<표 11> 핵심 요인 및 입력 변수	48
<표 12> 활용변수 간 상관관계 결과	49
<표 13> 다중회귀 분석 대상 변수 조합	50
<표 14> 다중회귀 분석 대상 입력변수	51
<표 15> 입력선택법에 의한 회귀분석 결과	52
<표 16> 후진제거법에 의한 회귀분석 결과	54
<표 17> 후진제거법의 진입/제거된 변수 결과	54
<표 18> 전진선택법에 의한 회귀분석 결과	55
<표 19> 단계적 선택법에 의한 회귀분석 결과	58
<표 20> 진입/제거된 변수	59
<표 21> 입력선택법 방식의 회귀모형 결과	60
<표 22> ARIMA(1,0,1)(1,0,1) ₅ 모형 통계량	65
<표 23> ARIMA(1,0,1)(1,0,1) ₅ 모형 모수	65
<표 24> 입력선택법 방식의 회귀모형 평가	66
<표 25> 시계열모형 평가	70

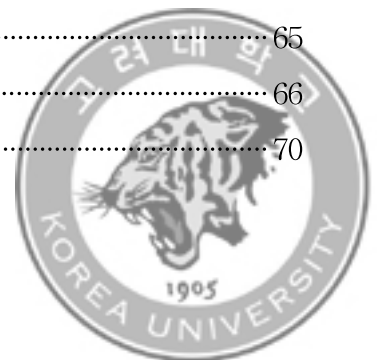
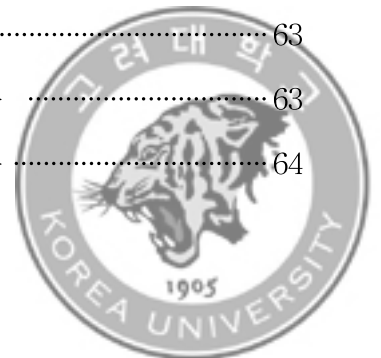


그림 목차

<그림 1> 연구 수행 절차	5
<그림 2> 택배사업 영업 구조	7
<그림 3> 택배 운영 프로세스	9
<그림 4> 택배 운영 세부 프로세스	10
<그림 5> 택배사 거점 운영 현황	11
<그림 6> 택배시장 전망	12
<그림 7> 택배 분포 현황	13
<그림 8> 서울시 구단위 집하 비중 현황	14
<그림 9> 서울시 구단위 배송 비중 현황	15
<그림 10> 연구 모형	33
<그림 11> 군집화 및 데이터 정제 방법	34
<그림 12> 요인분석 및 변수선정 방법	35
<그림 13> 기후 데이터 통계적 유의성 검증 결과	37
<그림 14> 예측 모형별 실증분석 및 평가 방법	40
<그림 15> 택배 물동량 변화 요소에 따른 분석 방법론	42
<그림 16> 물동량을 위한 예측 변수	43
<그림 17> Box-Jenkins의 ARIMA 모델 구축 순서	61
<그림 18> 수도권 택배 일별 물동량의 시계열 순서도	62
<그림 19> 수도권 택배 일별 물동량의 자기상관함수	62
<그림 20> 수도권 택배 일별 물동량의 추세 차분 적용	63
<그림 21> 수도권 택배 일별 물동량의 계절성 차분 적용	63
<그림 22> 수도권 택배 일별 물동량의 계절성 차분 적용	64



<그림 23> 회귀모델 1 추정결과	67
<그림 24> 회귀모델 2 추정결과	67
<그림 25> 회귀모델 3 추정결과	68
<그림 26> 회귀모델 4 추정결과	68
<그림 27> 회귀모델 5 추정결과	68
<그림 28> 회귀모델 6 추정결과	69
<그림 29> 회귀모델 7 추정결과	69
<그림 30> 시계열모델 추정결과	70
<그림 31> 분석 모형별 예측결과 비교	75



제 1 장 서 론

제 1 절 연구의 배경 및 목적

사회와 문화가 발전하고 경제적 성장에 따른 국민소득이 증진되면서 삶의 질은 향상되고 더 좋은 제품과 서비스를 제공받고자 하는 사회적 욕구 또한 증진시켰다. 이러한 변화는 재화의 이동을 담당하는 물류 서비스 시장에서도 일어났다. 이와 더불어 소비행태가 다양화되고 IT(정보기술)의 발달에 따른 전자상거래와 TV 홈쇼핑이나 인터넷 쇼핑몰과 같은 무점포 기반의 신유통체계가 출현하면서 물류의 수요는 지속적으로 증가하고 있다. 또한, 물류 시장의 규모가 점차 성장하고 신속함과 편리함이 강조된 생활 밀착형 서비스로 발전하게 되면서, 합리적이고 효율적인 물류시스템의 변화를 요구하고 있는 상황이다.

물류신문(2011)에 따르면 우리나라에 처음 택배 서비스가 소개된 1992년 택배 시장 규모는 연간 일인당 택배이용 건수 0.2건에 연간 총 500만개에 불과하였지만, 2015년 현재 연간 일인당 67.9건에 4.3조원에 이르는 시장으로 성장하였다(한국통합물류협회, 2015). 이와 같이 물류는 지속적인 성장을 하고 있으며 신생 업체와 타 산업 내에 물류 자회사 설립 등 공급자가 과도하게 증가하면서 물류시장에서 우위를 차지하기 위한 경쟁사 간의 치열한 경쟁은 서비스 단가하락이라는 결과를 초래하였다. 이는 기업의 수익구조를 악화시켜 다수의 기업들이 시장에서 퇴출되거나 대규모 인수합병이라는 과정을 거치게 되면서 물류 서비스 시장은 현재의 모습과 질서를 유지하게 되었다(박중권과 이명규, 2008). 이러한 흐름 속에서 물류 서비스 시장의 성장에 따른 대 고객 서비스를 위한 예측 필요성이 강조되었고, 하루 300만 건 이상의 물량이 발생되어 이에 따른 데이터양도 막대하나 데이터에 대한 분석이나 흐름의 정확성을 파악하는 데는 관심을 가지지



못하고 있는 실정이다. 기존 물류 서비스 업체도 시장점유율을 유지하고 견고히 하기 위하여 경쟁업체와 차별화되는 경쟁력을 갖추어야 하지만, 단순화된 집화-분류-배송의 프로세스로 경쟁업체와의 차별화된 품질을 제공하는 것이 어렵다(이지홍, 2017). 제조업과 물류업은 고객 수요를 사전에 파악하여 적시에 제품 또는 서비스를 제공해야 하는 공통점이 있지만, 시장 점유율의 변동성이 큰 물류업의 특성상 물류 서비스의 수요예측은 기업의 존폐를 결정지을 정도로 큰 이슈이기도 하다. 특히 택배는 발송고객(송화인)에서 수취고객(수화인)으로 배송하는 일련의 프로세스 상에서 정확한 수요 예측을 통해 운영 자원을 효율적으로 사전 배분하고 활용해야 하는 실정이다. 또한, IT의 발전으로 기존에 무시코 흘려버렸던 연간 8억건에 달하는 배달 수량에 대한 데이터들의 유의미성을 분석할 수 있는 수준에 이르렀고, 이제는 배달 데이터 기반의 물동량 예측을 통해 운영 자원을 효율적으로 배부하고 잉여 자원을 활용하여 기업의 이윤과 고객에 대한 서비스 강화의 Needs가 점차 커져가고 있는 상황이다. 이렇게 물류 서비스 및 이윤 개선을 위한 물류 물동량 예측의 필요성이 강조되고 있지만, 관련 데이터 확보의 한계성, 다양한 변수에 대한 분석, 분석의 복잡성 등 연구적 어려움으로 현재까지 택배의 물동량 예측에 관한 연구는 미흡한 실정이며 더욱이 매년 이동하는 명절 특수기, 이벤트 등 특이사항이 반영된 단기적 예측은 좀처럼 찾아보기 힘들다.

본 연구에서는 국내 최대 종합물류기업 C사의 서로 다른 특징을 가진 600여개의 택배 지점들 중 수도권을 중심으로 한 택배 물동량 수요예측을 진행하여 물동량 예측에 영향을 미치는 요인을 분석하고 여러 요인이 반영된 日 단위 수요 예측에 대한 정확도 높은 예측모형을 제시하고자 한다. 또한, 궁극적으로 빅데이터의 분석을 통해 SCM상에서 일어나는 일련의 데이터를 분석하여 물류 고유의 문제를 해결하고 고객의 트렌드 파악과 자원의 효율적인 활용에 좋은 매개체로 물동량 예측 방법을 제시하여 물류 자원 운영, 전략, 마케팅 등 다양한 기업활동의 의사결정 과정에 도움을 주어 기업의 장기 수익 개선에 기여할 수 있도록 하



고자 한다. 또한, 물류시장에서도 택배사업이 지속적으로 확대되고 있는 추세에 맞추어 질적 성장을 도모하기 위해 국내 최대 종합물류기업 C사의 수도권 내 택배지점들을 대상으로 물류 서비스 산업을 위한 택배 물동량 분석을 위한 요인분석과 수요예측 모형에 관한 연구를 수행하고자 한다.

이에 물동량 추정에 관한 본 연구의 목적은 다음과 같다.

첫째, 경험적 판단에 의존한 엑셀 기반의 수작업 물동량 예측에서 택배지점별 물동량과 시계열 및 주요 캘린더 변수가 반영된 데이터 기반의 예측으로 개선하기 위해 필요한 물동량 주요 영향요인을 파악하고자 한다.

둘째, 영향요인에 대한 택배 전문가 대상의 요인분석을 실시하여 도출된 핵심 요인과 핵심요인별 입력변수 기반의 회귀분석 모델 분석을 통해 실제 영향도가 큰 요인을 실증분석을 통해 파악하고자 한다.

셋째, 다양한 회귀분석 모형 예측치가 높은 모형과 시계열 분석모형 간 예측결과를 비교하여 실제 기업에서의 물동량 예측과 이후 자원의 효과적 운영을 위한 의사결정에 도움을 주는 예측 모형을 선정하고자 한다.

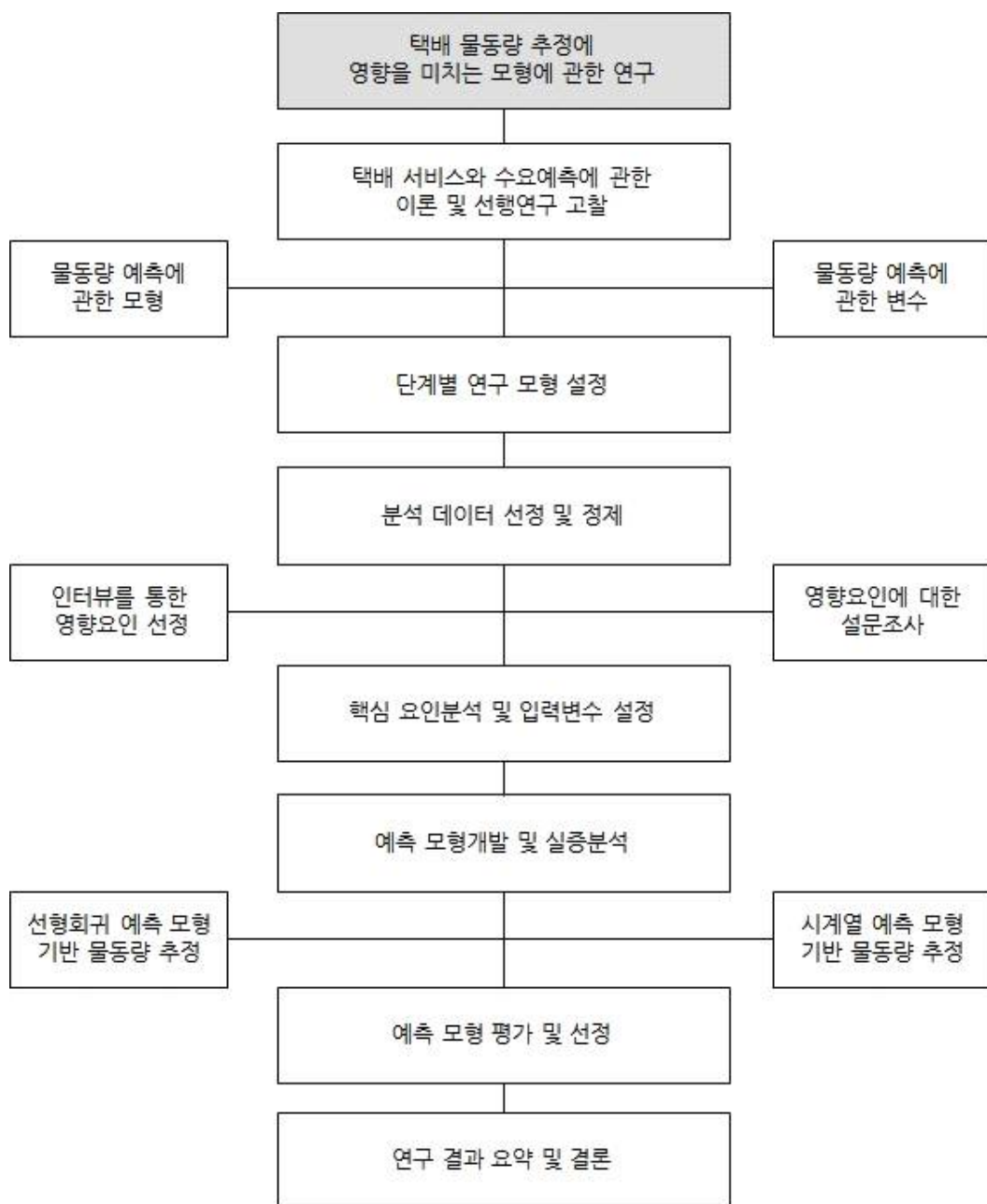
제 2 절 연구의 방법 및 구성

본 연구의 주요 목적은 과거 물동량 데이터 기반으로 트렌드 변수, 캘린더 변수를 분석하여 日 단위 택배 물동량의 예측을 위한 주요 요인 추출과 정확도 높은 예측모형을 제시하고, 나아가 기업활동에 있어 의사결정에 도움이 되는 수요예측을 위한 방법론을 제시하는 것이다. 이러한 연구목적의 달성을 위해 본 연구에서는 택배 서비스와 수요예측에 관한 이론과 선행연구에 대한 고찰을 통해 조사분석의 이론적 근거를 제시하고, 물동량 예측에 관한 모형과 물동량 예측의 두 가지 측면에서 예측 모형과 변수를 도출하고자 한다. 도출된 예측



모형과 변수를 기반으로 한 연구 수행을 위한 단계별 연구 모형을 설정하고 분석 대상 데이터에 대한 수집 및 선정과 정제 작업을 실시할 예정이다. 또한, 국내 종합물류기업 C사의 택배 경력 5년 이상의 운영 기획 Staff 대상의 인터뷰를 통해 물동량 예측 영향요인을 최종 선정하고 이들 영향요인을 토대로 핵심요인 추출을 위한 요인분석을 실시한다. 추출된 핵심요인별 입력변수에 대한 설정과 선형회귀모형 기반의 시나리오별 회귀모형을 활용하여 물동량 추정에 대한 실증분석을 수행하여, 최종 예측 결과가 높은 회귀분석모형과 시계열 예측모형 기반의 비교를 통해 실제로 활용할 수준의 예측 모형 평가 및 선정을 수행하겠다. 마지막으로 연구에 대한 결과 요약 및 시사점을 도출하고 향후 추가 수행되어야 할 과제를 제시하고자 한다.





<그림 1> 연구 수행 절차



제 2 장 이론적 배경

제 1 절 택배 서비스

제 1 항 택배 서비스 개요

택배 서비스가 활성화되기 이전에는 철도를 통한 소화물이나 정기화물 등의 물류 서비스를 이용해왔다. 하지만 이는 송화인과 수화인이 영업소에서 소화물을 직접 수탁하고 수취해야 하는 불편함과 긴 배송기간 등의 문제점을 항상 지니고 있어 편리함과 신속함이 강조되는 오늘날의 택배 서비스의 개념과 거리가 멀었다. 택배는 “소형·소량의 운송물을 고객의 주택, 사무실 또는 기타의 장소에서 수탁하여 수하인의 주택, 사무실 또는 기타의 장소까지 운송하여 인도하는 것”으로 공정거래위원회의 택배표준약관 개정안(제10026호, 2007년 12월 28일) 제2조 제1항에서 정의하고 있다. 대표적인 물류 서비스업으로 상징되는 택배 산업은 1992년 6월 택배 서비스가 국내 물류시장에 도입된 지 25년이 되었다. 또한, 물류산업의 대표적인 장치산업으로써 대형 택배업체를 중심으로 물류 인프라에 대한 구축 및 IT 시스템 영역에 대한 투자와 지속적으로 서비스 개선의 노력을 경주해왔다. 이와 더불어 홈쇼핑과 온라인 쇼핑시장이 급성장하면서 택배시장은 신규 수요를 창출하는 원동력의 역할을 하면서 2013년까지 매년 두 자릿수의 꾸준한 성장률로 시장의 규모를 확대해 왔다.

이와 같이 택배 서비스 기업들은 소화물을 신속하고 편리하게 송화인으로부터 집하하여 도시 간 및 도시 내에 수화물을 수하인에게 빠르고 안전하게 지정된 목적지까지 배송하는 서비스를 제공하고 있으며, 그에 따른 산업적 특징은 다음과 같다(물류신문, 2011).



1) 자본 집약형 장치산업

택배 서비스 산업은 기본적으로 집·배송을 위한 택배터미널, 전용차량과 수화물의 분류·반품을 처리하는 분류장치 및 분류시스템, 그리고 전국적인 물류 네트워크 등을 갖추어야 하기 때문에 막대한 자금이 투입되는 장치산업의 특성을 보인다. 이러한 장치산업의 특성으로 택배 서비스 기업들은 수도권 및 지방에 터미널 신축과 증축을 위해 대규모 시설투자를 진행하고 있으며, 최근에는 고객이 화물의 위치를 실시간으로 확인할 수 있도록 IT 시스템 구축도 필수적으로 요구되고 있다.

2) 네트워크 산업

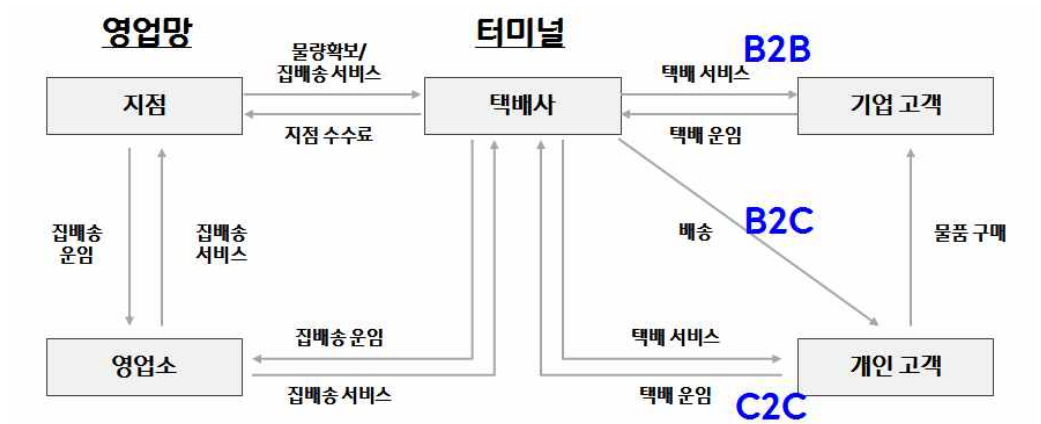
택배사업의 구조는 택배 기업이 터미널 등 거점을 확보하고 지점과 영업소를 통해 전국적인 영업망을 확보하고 있는 구조이며 서비스 대상별 B2B, B2C, C2C 유형으로 분류하여 운영된다. 즉, 집화된 수화물을 분류하고 신속·정확하게 해당 목적지로 배송할 수 있도록 각 지점과 터미널을 연결하는 물류 네트워크를 구축해야 한다. 전국적인 배송망을 구축하는 것은 물류 서비스에 즉각 반응하고, 원활한 배송을 가능하게 하여 시장수요에 대한 대응력을 높일 수 있다.

국내 택배 서비스 기업들은 Hub and Spoke 물류 네트워크, Point to Point 물류 네트워크, 그리고 앞의 물류 네트워크의 특성을 결합한 절충형 물류 네트워크 중 하나의 물류 네트워크를 선택하여 기업의 경영전략과 운영특성에 맞게 운영하고 있다(박찬석, 2006).

3) 노동집약적 산업

전국적으로 물류 네트워크를 구축한 회사라면 작은 규모의 택배업체라 하더라도 최소 1천명 이상이 근무하고 있으며, 택배 산업은 다수의 인력 투입이 요구되는 노동집약적 산업의 특성을 가진다.





<그림 2> 택배사업 영업 구조

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성.

명절 연휴기간이 다가올 때면 증폭되는 택배 물량을 처리할 인력과 차량을 추가적으로 보충·배치해야 하는데, 이는 물류 서비스 수요예측 결과와 물류자원 확보활동을 연계하여 명절 특수기 물량을 처리할 적정 인력과 차량수를 도출할 수 있다.

4) 입지형 산업

택배 서비스 산업은 입지형 산업의 특성을 가지고 있어 효율적인 집·배송을 위해 최적화된 터미널 입지를 선택할 수 있어야 한다. 시장수요가 서울도심에 주로 집중되어 있기 때문에 많은 물류 서비스 기업들이 접근성이 뛰어난 도심 인근지역에 터미널을 설립하기를 원하지만, 부지의 가격과 면적 확보의 문제에 직면하게 된다. 택배 서비스 기업들은 기업의 실정과 여러 제약조건들을 고려하여 현 상황에서 최적의 터미널 입지를 선택하기 위해 노력하고 있다.



5) 계절적 수요변화

택배 서비스 산업은 계절변동의 영향을 많이 받는 전형적인 생활물류산업 중 하나로, 명절, 연말에는 평소보다 물량이 50%이상 증가하면, 기업택배보다 개인택배의 비중이 높은 중견택배업체의 경우 지역 특산물이 수확되는 시기에는 수요가 큰 폭으로 증가함에 따라 지역별로 용차 문제가 발생하는 등 계절적 수요변화의 특성을 가지고 있다.

제 2 항 택배 프로세스

택배 상품은 권역별 대리점에서 고객(송화인)의 상품을 집하하여 Sub 터미널을 통해 Hub 터미널로 분류 및 발송 후 해당 상품을 배송주소지의 Sub 터미널로 이동하여 최종 권역 내 대리점에서 고객(수화인)에게 해당 상품을 배송하는 프로세스로 진행된다.

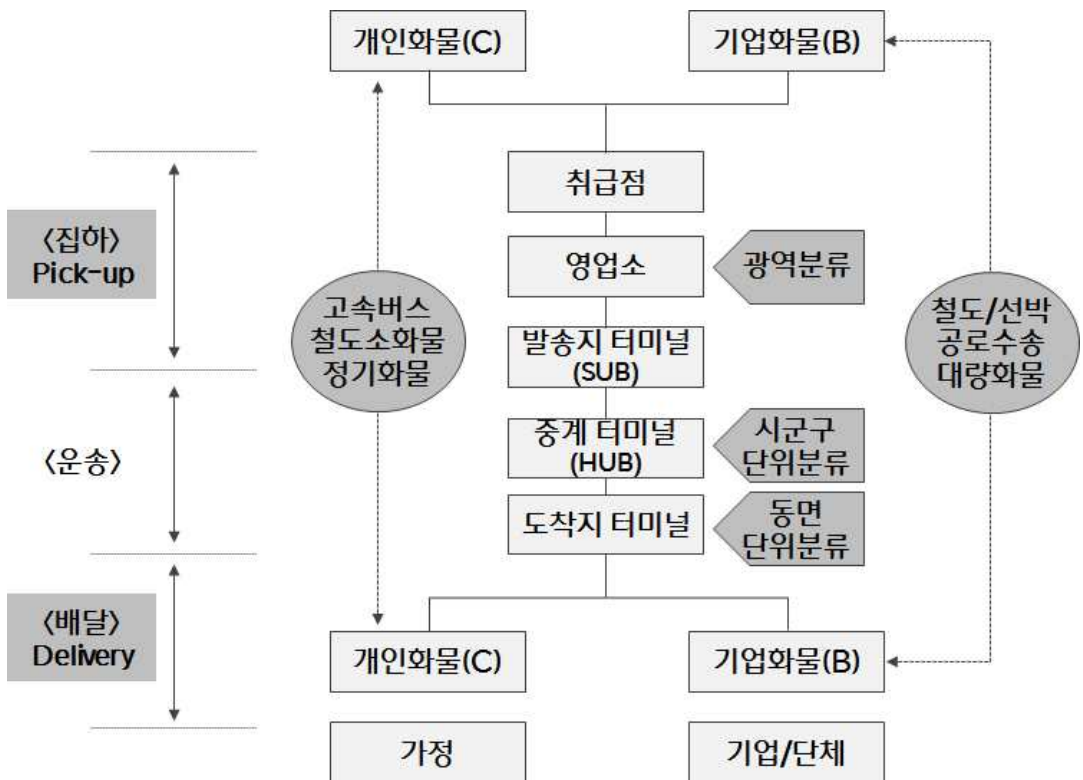


<그림 3> 택배 운영 프로세스

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성



<그림 4>에서 보듯이 고객(기업 및 개인)의 소화물을 집하하고 중간 집화지인 Sub 터미널로 이동시키기 위한 기본적 운송 체계를 갖추고, Hub 터미널은 각 권역별 Sub 터미널로부터 받은 고객 소화물을 전국적 중간 배송지인 Sub 터미널로 재분류 후 이동하여 대리점의 배달원이 고객에게 소화물을 전달하는 일련의 과정을 가지고 있다. 또한, 소량·다빈도의 불특정 다수로 구성되는 택배 서비스 이용 고객의 수요를 충족시키기 위한 운송거점 중심의 서비스 Network 구축이 필수적으로 되었고, 여러 지역에 환적 터미널을 입지시켜 터미널 간 거점운송을 하는 Sub and Spoke 방식을 통해 도시 간 익일 배송 체계를 갖출 수 있게 되었다.

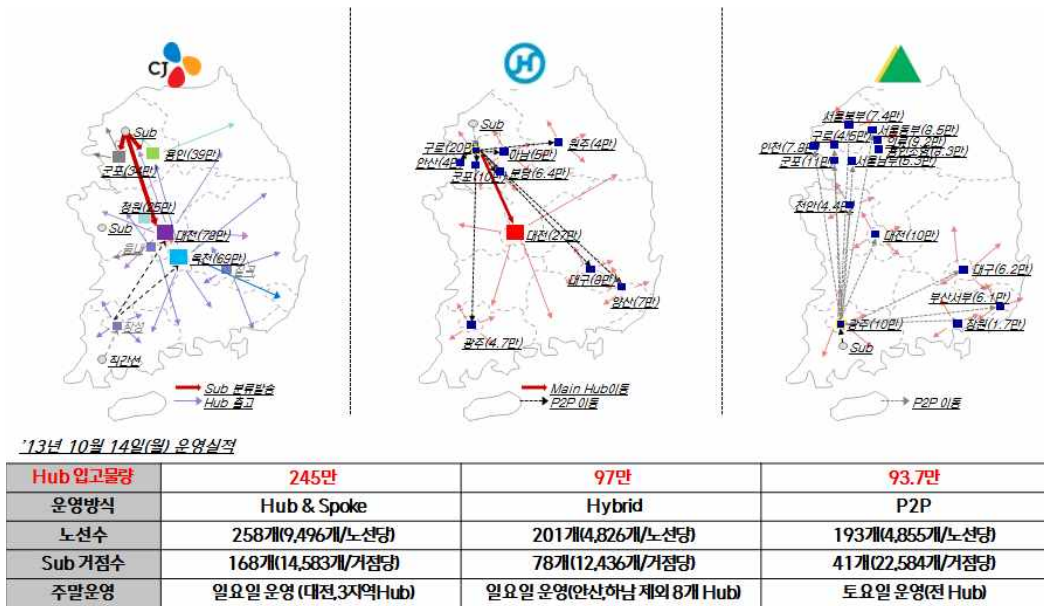


<그림 4> 택배 운영 세부 프로세스

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성



이러한 익일 배송 체계를 갖추기 위해 주요 업체별 거점에 대한 네트워크 구조는 <그림 5>과 같이 구성되어 운영되고 있으며, CJ대한통운은 168개의 거점, 한진택배는 78개의 거점, 롯데글로벌로지스(구 현대택배)는 41개의 거점을 통해 운영되고 있다(2013년 10월 기준).



<그림 7> 택배사 거점 운영 현황

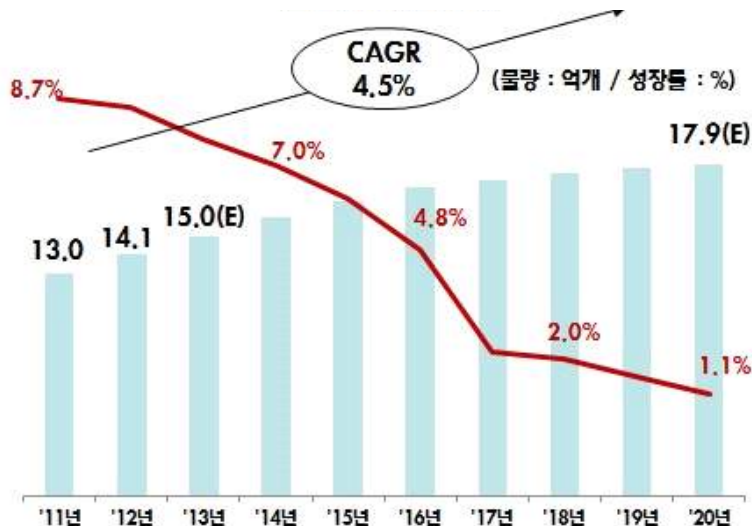
자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성

이와 같은 익일 배송 체계는 지속적으로 증가하는 택배 이용고객에 대한 서비스를 가능하게 해준다. 하지만, 이는 매일 200만 건 이상의 운송장 정보와 적재되고 있는 데이터 기반으로 분석모델에 대한 발굴과 구축을 통하여 영업 경쟁력 강화 및 자원 활용을 지속적으로 최적화 하여야 만이 가능하다. 기존 택배 서비스의 물동량 예측 및 그에 따른 차량의 배차는 수작업과 경험적 판단에 의존하여 수행되었다. 이에 각 Sub 터미널 물동량과 변수가 반영된 데이터 기반의 배차로 개선하여 자원의 효율적 활용을 하고자 한다.



제 3 항 택배 물동량 분석

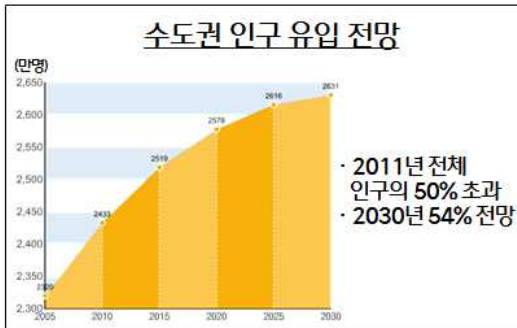
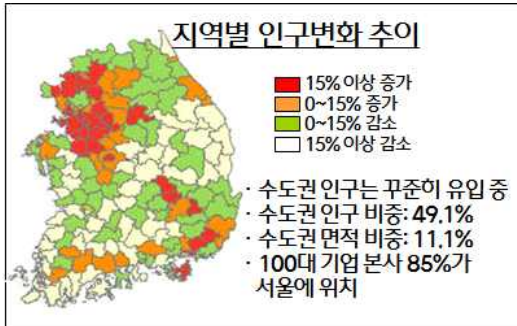
택배 사업은 최근 일 최대 물량이 500만 박스를 초과하였으며, 관련한 택배 운영 데이터도 지속적으로 증가하고 있다. 택배 물동이 발생하는 프로세스마다 물품 정보, 스캔성 정보 등의 다양한 정보가 발생되고 있는데, 이를 효율적으로 관리하고 이를 통해 프로세스 내의 문제점을 발견하고 개선하는 활동이 지속적으로 필요하다. 전자상거래 시장과 함께 고속 성장한 국내 택배시장은 성숙기에 접어들어 향후 성장률이 점차 둔화되겠지만, 온라인 쇼핑 시장의 지속적인 성장으로 2020년까지 연평균 4.5% 증가하여 약 18억개 시장을 형성할 것으로 예상된다.



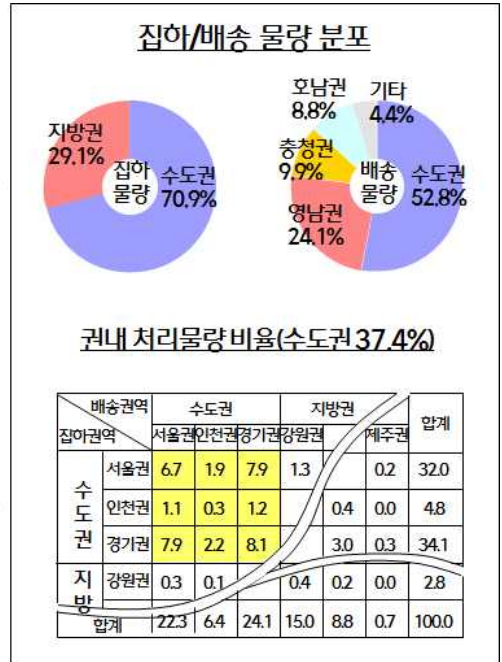
<그림 8> 택배시장 전망

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성





* 서울/인천/경기권을 수도권으로 분류, 자료: 통계청, 인구센서스 조사



<그림 9> 택배 분포 현황

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성

이중 전체 집하 물량의 70%가 수도권에서 이루어지고 있으며, 전체 물량의 40%가 수도권 집하·배송이 이루어지고 있어 수도권에서의 입지가 상당히 중요한 경쟁우위 요소로 작용하고 있다. 수도권 내 물동량의 경우 서울과 경기도로 양분되며 각 지역별 면적을 고려 시 집하·배송량 모두 서울권의 물량 밀도가 현저하게 높은 것으로 나타나고 있고, 지역별 면적과 물량 밀도를 고려 시 서울권은 순환회전 형태, 경기도는 원거리 지역 간 이동관리가 요구되고 있다.



<표 1> 수도권 내 물동 비중 현황

지역구분	물동량		전국량 대비 비중		수도권 내 비중	
	집하	배송	집하	배송	집하	배송
서울	193,640	148,491	31.4%	24.1%	44.8%	44.0%
경기	220,385	157,986	35.8%	25.7%	51.0%	46.8%
인천	18,103	31,045	2.9%	5.0%	4.2%	9.2%
수도권	432,129	337,521	70%	54.8%	100%	100%
전국	618,878	615,878	100%	100%		

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성

<표 2> 물량 밀도 현황

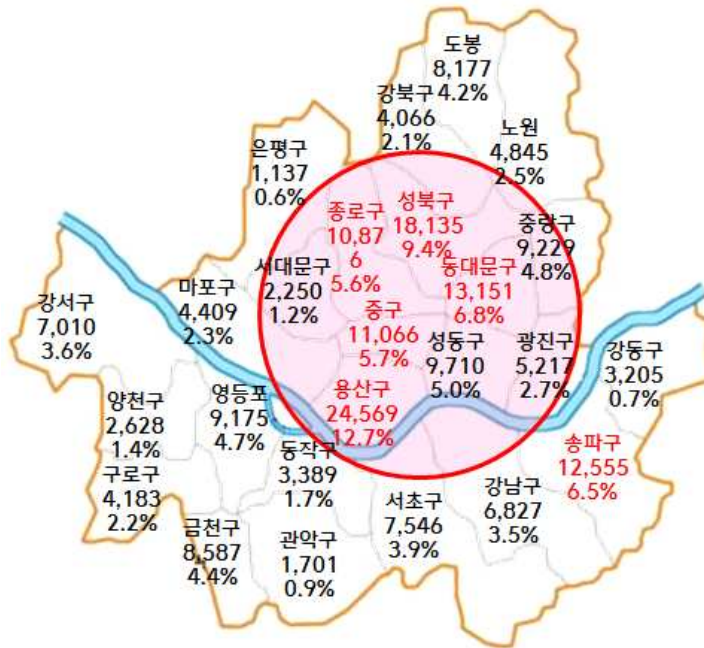
지역구분	물동량		전국량 대비 비중		면적 (km ²)
	집하	배송	집하	배송	
서울	193,640	148,491	319.9	245.3	605
경기	220,385	157,986	21.6	15.5	10,183
인천	18,103	31,045	17.9	30.7	1,010
합계	615,878	615,878			11,799

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성

또한, 물량 기반의 수도권 내 개선 검토시 인천권은 상대적으로 중요도가 낮은 것으로 나타나고 있다.



서울권을 좀 더 살펴보자면 <그림 8>과 같이 집하량은 용산을 중심으로 한 강북권 인접 5개 구에 집중되어 있으며 각 구별 집하량 편차가 매우 큰 현상을 보이고 있다. 서울권 상위 5개 구의 비중은 40.2%로 전국 대비 14.6%를 나타내고 있다.

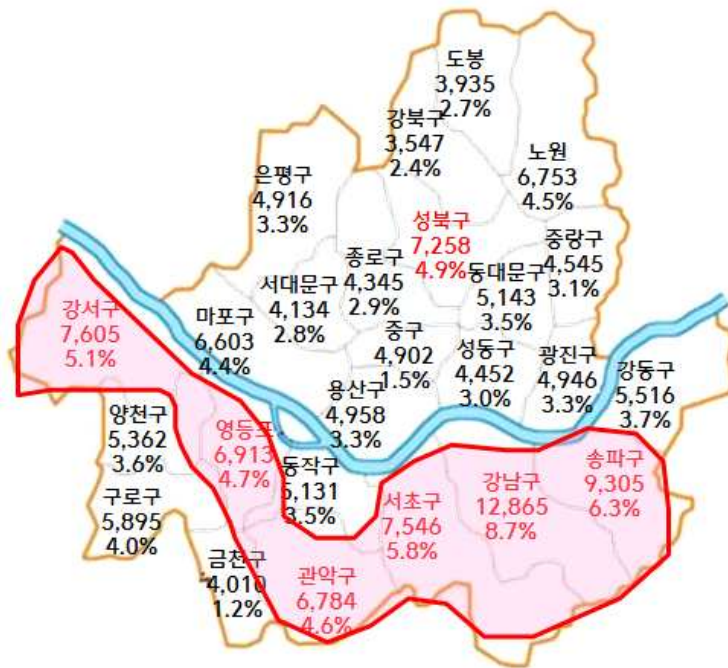


<그림 10> 서울시 구단위 집하 비중 현황

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성

서울권 배송량은 <그림 9>와 같이 강남권에 넓게 펼쳐진 형태로 집중되어 있고 각 구단위 편차는 존재하나 집하 대비 상대적으로 편차가 크지 않고 전 지역에 고른 분포를 보이고 있어 배송 네트워크 구성측면에서는 서울 전 지역을 고려한 조밀한 구성이 요구되고 있다.





<그림 11> 서울시 구단위 배송 비중 현황

자료: CJ대한통운(2015)의 택배사업의 이해 및 운영 자료를 활용하여 재구성

이와 같이 택배 시장은 수도권을 중심으로 지속적인 성장을 할 것으로 예상되며 시장의 성장은 택배 서비스 유지를 위한 자원의 투입도 함께 증가함을 의미한다고 볼 수 있다. 이를 위한 신기술의 투입과 비용 절감요소를 여러 가지 다방면으로 강구해야 할 것이며 본 논문에서는 물동량 분석을 통해 운영 자원의 효율화를 만들어 낼 수 있도록 물동량 예측 요인과 방안을 중심으로 연구하고자 한다.



제 2 절 수요 예측

제 1 항 요인분석의 개요

요인분석(Factor Analysis)이란, 변수들 간의 상호연관성 즉, 공분산과 상관관계를 분석하여 변수들 상호간 공통적으로 작용하고 있는 내재적 요인을 추출하고 전체 자료를 특성화하여 변수의 수를 줄이는 통계적 방법이다. 즉, 요인분석은 연관성 즉, 상관관계가 높은 변수(요인)들 간에 가지고 있는 공통성을 추출하여 하나의 동질적인 요인으로 묶어주는 검증방법이다. 요인분석은 변수의 연구나 모형 개발에 사용되는 변수의 수를 줄여 핵심적인 요인만으로 모형을 구성하여 설명하고, 정보와 지식을 보다 효과적으로 전달하기 위해 적용하는 통계기법으로 그 목적은 다음과 같다.

첫째, 많은 데이터의 양을 인지가 가능한 수준으로 정보를 요약해 준다.

둘째, 많은 변수들 간 존재하는 내부적 구조를 파악한다.

셋째, 요인으로 묶여지지 않는 변수 중에 중요도가 낮은 변수를 제거한다.

넷째, 동일한 개념으로 측정하려는 상호 독립적인 특성을 가진 변수들이 동일한 요인으로 묶이는지 확인한다.

다섯째, 요인분석을 통해 얻어진 요인들을 회귀분석, 판별분석 및 군집분석 등에서 설명변수로 적용할 수 있다.

주성분분석(PCA, Principle Component Analysis)은 정보 손실을 최소화하면서 기존의 요인 수 보다 적은 수의 핵심요인을 구하고자 할 때 적용하는 통계적 기법이다. 추출할 요인의 수를 결정하는 방법에는 Eigenvalue를 기준으로 하여 기준값 이상의 요인 수를 결정하는 것과 연구자가 추출하고 싶은 요인의 수를



직접 결정하는 방법이 있다. 일반적으로 Eigenvalue 1 이상을 갖는 요인을 추출하는 방법을 가장 많이 사용한다. 요인분석 항목을 기준으로 분석의 타당성을 검증하기 위해 신뢰도분석(Reliability Analysis)을 수행한다.

신뢰도분석은 동일한 개념을 측정하는 여러 항목들에 대한 항목 분석과 신뢰도 계수를 계산한다. 여러 가지 질문들을 단 하나의 개념을 측정하기 위한 도구로써 사용하고자 할 때 그 질문들에 대한 반응이 동일한 척도로 측정되었기 때문에 일관성 있게 반영하여 비슷하게 나타나야 함을 전제로 하는 분석이다. 크론바흐 알파 계수(Cronbach Alpha Coefficient)값의 신뢰도 계수를 계산하여 각 변수에 대한 항목분석과 한 개념에 대한 전체 변수의 신뢰도를 측정한다.

제 2 항 수요예측의 개요

수요예측(Demand forecast)이란, 수요분석을 기초로 시장조사나 각종 예측조사의 결과를 종합하여 장래 수요를 예측하는 일이다(브리태니커 사전). 다시 말해 특정 상품의 수요동향을 결정짓는 요인들을 찾아내어 과거와 미래를 포함한 전체 수요량에 미치는 영향을 분석함으로써 미래에 대한 영향력을 사전에 짐작할 수 있는 것이다. 많은 기업들은 원하는 목표를 성취하거나 피해를 최소화하는 등 더 나은 결과를 만들어내기 위해 미래에 발생할 사건의 정확한 예측과 그 결과를 사전에 대비하고 계획함으로써 간극을 최소화하고자 노력한다. 일반적으로 예측은 관리적 활동의 필수적인 부분이다. 또한, 수요예측의 결과는 관리자들이 의사결정을 수행함에 있어 상당히 중요한 참고 자료로서의 역할을 하게 된다. 정확한 수요예측 결과를 통해 불필요한 비용을 감소시키고, 높은 수준의 고객 서비스 만족 등 조직 운영상 긍정적인 영향을 만들어낼 수 있다. 하지만, 수요예측의 결과가 불확실하다면 불필요한 비용을 초래되고 고객에게 좋은



품질의 서비스를 제공하지 못하여 신뢰를 잃는 등 전체 기업 운영에 예상치 않은 많은 악영향을 끼칠 수 있다(Miao and Xi, 2008; Hansen and Nelson, 2003; Makridakis et al, 1998). 수요예측은 효과적으로 자원을 운영할 수 있도록 도움을 주는 일정계획(Scheduling), 미래에 필요 운영자원을 사전에 파악하여 빠른 준비를 지원해 주는 자원 획득(Acquiring Resources), 관리자에게 필요한 자원에 대한 상대적으로 올바른 결정을 할 수 있도록 의사결정에 도움을 주는 자원 필요 결정(Determining Resource Requirements)의 분야에서 예측이 중요한 역할을 담당한다(Makridakis et al, 1998). 정확한 수요예측의 수행을 위해 수요에 대한 영향 요소들을 파악하는 것 또한 중요하며 이들의 변화적 흐름을 파악하고 예측치에 어떻게 반영해야 수요 예측이 보다 정확할 수 있는지는 지속적으로 기업이 풀어야할 과제일 것이다.

수요예측 방법으로는 정성적인 예측방법(Qualitative Forecasting Methods)과 정량적인 예측방법(Quantitative Forecasting Methods)으로 크게 나눌 수 있다(Makridakis et al, 1998). 정성적 예측방법은 미래 수요를 예측하고자 할 때 과거 분석 대상 데이터의 수집이 쉽지 않거나 분석할 만큼의 충분한 데이터를 수집하지 못한 경우 주로 사용되고 있으며, 전문가가 시장에 대한 전문적이고 해박한 지식을 가지고 있을 때 가장 적합한 방법이고 미래 새로운 산업이나 시장에 대한 장기적 예측을 수행할 경우 유용하게 활용될 수 있는 예측하는 방법이다. 정성적 예측방법에는 시장 상황에 대한 자료를 수집하기 위해서 소비자 패널을 사용하거나 설문지, 서베이(survey) 등을 사용하는 시장조사법(Market Research), 소비자, 영업사원, 경영자들을 모아서 패널을 구성하고 이들의 의견을 모아서 예측치로 활용하는 패널동의법(Panel Consensus), 전문가들을 대상으로 하여 우편을 통한 질문과 응답에 대한 통계, 피드백 과정을 의견의 일치를 볼 때까지 반복적으로 사용하는 델파이법(Delphi Method) 이외 수명주기 유추법(Life Cycle Analogy), 판매원의견합성법(Sales force composite), 중역의견법(Executive Opinions) 등이 있다.



<표 3> 수요예측 기법

예측방법 분류			
정성적 분석법	단순예측법	집행부 의견수렴법, 판매력 평가법	
	기술적 예측법	탐사적 예측법	델파이 예측법, 형태학적 분석법, 시나리오 설정 법, 횡단면 영향분석, S곡선 분석, 전문가 판단에 의한 모형
		규범적 예측법	관련나무분석, 의사결정분석
정량적 분석법	시계열 분석법	단순 추세분석법, 이동평균법, 분해법, 지수평활법 단변량 박스-젠킨스 방법 : 일반 ARIMA, 전이함수모형, X-11-ARIMA	
	인과관계 분석법	척도법, 마켓분석법, 클로슨 방법, 다변량 회귀수요분석 다변량 박스-젠킨스 방법 : 다변량 ARIMA, 개입모형, 상태공간모형 공간적 상호작용모형: 중력모형, 개재기회모형 성장 시나리오 모형, 결합예측법	
	시스템모형	시스템 다이내믹스, 투입-산출분석	

자료: Archer(19765, 1987), Sheldon & Var(1982), Uysal & Crompton(1985) 및 Choy(1984)의 자료를 이용하여 최영문·김사현(1998)이 재구성한 것을 인용.

이용 가능한 과거 데이터가 존재하고, 계량적으로 표현이 가능하며, 과거에 존재하는 추세나 패턴이 미래에도 지속적으로 연계되는 현상을 보일 것이라는 가정이 충족된 경우 정량적 예측방법의 적용 가능하다. 다시 말해, 정량적 예측 방법은 과거의 수요를 기반으로 하는 분석 데이터를 통해 미래의 수요를 예측하는 방법으로, 크게 시계열 모형(Time-Series Model)과 인과적 모형(Casual Model)으로 구분할 수 있다.

시계열 모형은 과거 분석대상 데이터에서 발견한 일정의 변동성 패턴이 미래에도 지속되는 현상을 보일 것이라는 가정을 두고 시계열 자료를 이용하여 미래의 수요를 예측하는 방법으로, 수요예측 방법 중에서 가장 많이 사용되고 있는 방법 중 하나이다(Chambers et al 1971; 김연형, 2001). 시계열 모형에는 분해법(Decomposition), 이동평균법(Moving Average), 지수평활법(Exponential



Soothing), Box and Jenkins의 ARIMA(Auto-Regressive Integrated Moving Average)모형이 있다. 1920년대 Persons의 연구는 시계열과 계절조정에 관한 연구에 많은 영향을 주었으며, 시계열이 추세(Trend), 계절변동(Seasonal Fluctuation), 순환변동(Cyclical Movement) 그리고 불규칙변동으로 구성되어 있다고 보았다(Zhang and Qi, 2005; Bell and Hillmer, 2002). 시계열 자료를 구성하는 변동을 살펴보면 다음과 같다.

(1) 추세 (Trend)

분석대상 데이터가 장기적으로 변화해 가는 큰 흐름을 나타내는 것을 추세라 한다. 데이터의 흐름이 장기적 관점에서 변화가 커지는지, 작아지는지 또는 변화가 없는지를 나타내 주는 요소를 추세라 보면 된다. 장기적으로 커지거나 작아지는 현상일 때 그것이 선형인지, 지수적으로(Exponentially) 증가 또는 감소하는 것인지를 함수관계로 나타내 줄 수도 있을 것이다.

(2) 순환변동 (Cyclical Movement)

순환변동은 경제적인 전반 또는 특정 산업에 대한 부침(Ups and Downs)를 표현해 주는 것을 말하며 분석 대상 데이터가 상하의 움직임을 반복하는 변동 중 계절변동에 따라 설명되는 부분을 제외한 것을 말한다. GNP(국민총생산), 주택수요, 산업생산지수 등과 같은 예시들이 시계열에 대한 순환변동을 갖는다고 할 수 있다. 또는 주식가격, 이자율, 자동차 판매대수, 통화공급량 등도 순환변동을 내포하고 있다고 간주된다.

(3) 계절변동 (Seasonal Fluctuation)

계절변동은 1년보다 짧은 관측단위기간(4분기 또는 월별)의 시계열(時系列) 데이터에서 볼 수 있는데, 일정한 주기를 갖으면서 반복되면서 같은 패턴으로



변화하는 현상을 말한다. 온도(1일 주기), 전력소비량(1일 주기), 보너스 지급기(1년 주기), 강우량(1년 주기) 등의 예시와 같이 시계열 자료이 일정한 주기로 같은 패턴을 반복하며 변화하는 현상을 찾을 수 있다. 계절변동의 패턴은 대체적으로 일정하다고 보일 수 있으나 장기적 관점으로 보면 제도적 변경이나 생산, 소비 측면의 기술적 진보로 계절변동의 진폭이 점점 좁아지는 추세로 변화하고 있다. 순환변동과 계절변동의 차이는 계절변동이 일정한 주기(1년, 1달, 1주 등)를 갖는 반면 순환변동은 정해지지 않은 기간 동안 사이클(Cycle)적인 현상을 보여주는 것이라 할 수 있다.

기술한 바와 같이 시계열자료는 불규칙하게 변화하지만 세 가지의 변동을 내포하는 것이므로 다음과 같이 시계열자료(Y_t)를 표현할 수 있다.

$$Y_t = f(\text{추세, 순환, 계절}) + \text{오차}$$

여기서 오차(Error)라는 항이 보이는데 이는 불규칙 변동으로 설명되는 항이다. 실제 시계열자료의 값과 추세, 순환변동, 계절변동의 조합에 따른 패턴인 $f(\text{추세, 순환, 계절})$ 가 갖게 되는 차이라고 보면 된다. 따라서 시계열자료는 실제로 위에서 설명한 네 가지 요소가 섞여 있는 것으로 간주하여 표현된다.

$$Y_t = f(T_t, C_t, S_t, I_t)$$

여기서, Y_t 는 t 시점의 시계열 값,

T_t 는 t 시점의 추세 요소,

C_t 는 순환변동 요소,

S_t 는 계절변동 요소,

I_t 는 오차로서의 불규칙 요소를 나타낸다.



인과형 예측기법(Causal Method)은 미래에 영향을 주는 결정요인들을 파악하고 이들 간의 인과관계를 바탕으로 미래를 예측한다. 인과형 예측기법은 과거 데이터를 바탕으로 예측 대상 값을 종속변수로 하고 이에 영향을 미치는 요인들을 독립변수로 하여 회귀모형을 추정하고, 각각의 독립변수들을 시계열 기법을 통해 예측한 값들을 회귀모형에 대입해 예측한다.

회귀분석이란, 독립변수(Independent Variable)와 종속변수(Dependent Variable) 사이의 관계를 검증을 통해 독립변수가 종속변수에 미치는 영향력을 알아보기나, 변화하는 독립변수에 따른 종속변수의 변화를 시계열적으로 예측하기 위하여 사용되는 통계적인 분석방법이다. 이에 종속변수에 영향을 미치는 독립변수의 개수에 따라 크게 둘로 종류를 나눌 수 있는데, 한 개의 독립변수를 가지고 있는 경우는 단순회귀분석(Simple Regression Analysis), 두 개 이상의 독립변수를 취하는 경우를 다중회귀분석(Multiple Regression Analysis)이라 한다(김윤식, 2010). 이러한 회귀분석이 사용되는 이유는 종속변수 즉, 결과에 대한 여러 원인들인 독립변수들을 한 번에 분석할 수 있기 때문이다. 또한, 회귀분석은 종속변수에 대한 개별적 독립변수들의 순수 영향을 각각 분석할 수 있는 장점 때문에 다른 변수를 통제하였을 때 다른 각각의 독립변수가 종속변수의 변화에 얼마나 기여하고 있는지에 대한 정도를 쉽게 판단할 수 있다(김태근, 2006). 하지만 각 독립변수 간의 상호관련성을 무시하고, 단 방향의 상호 관계만을 다루는 특성을 갖고 있으며 측정 오차를 인정하지 않는 특징이 있다. 따라서 다중공선성 또는 독립변수 내 관련성에 대한 문제를 해결할 수 있는 비교적 단순한 인과모형을 대상으로 하는 경우, 종속변수에 대한 독립변수 간 상호영향력의 우위를 가늠할 수 있는 아주 유용한 통계기법 중 하나로 볼 수 있다(성혁제 외, 2002).



다중회귀분석에서 회귀모델의 구성 시 종속변수에 대해 영향을 미치는 독립변수들을 회귀모델에 포함해야 하는데, 이들 독립변수를 투입하는 방식은 아래와 같이 크게 네 가지 방법으로 구분된다.

(1)입력선택방식(All Possible Selection)

한 번에 모든 독립변수들을 포함시켜 분석하는 방식이 입력선택방식이다. 이 방법을 통해 다른 독립변수들이 분리된 상태에서 특정 독립변수의 영향력을 파악할 수 있으며, 연구자가 고려하고자 하는 종속변수에 대한 모든 독립변수들의 영향력 수준을 알 수 있다.

(2)전진선택법(Forward Selection)

가장 단순하게 절편만 있는 상수항의 모형을 기초 모형으로 지정하고 독립변수들 중 가장 중요할 것으로 생각되는 순서의 독립변수부터 차례로 모형에 추가해 나가는 방법이다. 변수를 추가해 나가는 과정을 거치면서 더 이상의 유의한 추가적 설명변수가 없다는 판단이 되면, 최종으로 선택된 설명변수들만을 사용하여 최적의 모형을 만들어 나가는 방법이다. 이 때문에 한 번 선택되어 사용된 설명변수가 추후에 다른 설명변수가 추가되어 더 이상 유의하지 않은 상태가 되더라도 제거될 수 없다는 단점을 가지고 있다.

(3)후진제거법(Backward Elimination)

불필요한 독립변수를 제거시켜 나가는 절차를 가지고 있는 방법이 후진제거법이다. 초기모형은 완전한 모형으로 시작하는데 첫 단계는 모든 후보가 되는 독립변수들을 포함하는 회귀모형에서 시작하고 기여도가 가장 적은 변수부터 하나씩 차례대로 제거해 나가는 것 방법이다. 해당 방법은 한번 제거된 독립변수가 추후에 다른 독립변수의 제거로 인해 이전 보다 유의하게 되어 필요한 상황



이 발생되더라도 모형에 포함시킬 수 없는 단점을 지니고 있다. 하지만 후진제거법은 중요한 설명변수들을 모형에서 제외할 가능성이 적은 방법이기 때문에 비교적 안전한 방법으로 볼 수 있다.

(4)단계적선택법(Stepwise Selection)

전진선택법과 후진제거법이 가진 단점을 보완하기 위하여 고려된 방법이 단계적선택 방식이다. 해당 방식은 모형에 독립변수를 개별로 하나씩 추가하면서 모형 내 이미 영향을 주고 있는 독립변수들 중에서 추가되는 독립변수로 인해 더 이상 유의하지 않게 되는 변수를 제거시키는 과정을 반복하며, 추가되거나 제거되는 변수가 더 이상 없을 때까지 추가와 제거의 과정을 지속적으로 반복수행한다. 단계적선택법은 결국 중요한 독립변수를 추가하면서 하나씩 선택을 하되, 모형에 이미 적용된 독립변수들이 제거될 수 있는지도 단계별로 검토하는 방식이다. 다시 말해, 전진선택 방식과 후진제거 방식을 교대로 수행하면서 가장 최적의 모형을 선택해 내는 방법이라고 할 수 있다.

제 3 항 예측의 정확도 측정

미래에 대한 예측의 정확도가 높을수록 이를 기초로 하여 물류 운영 계획을 합리적으로 마련하고 불필요한 운영 Cost의 낭비를 최소화할 수 있다. 새로운 화주 고객에 대한 서비스를 제공하려하거나 현 시점의 화주의 물류를 운영하는 모든 물류 운영 이해 관계자들에게 물동량 예측의 정확도(Accuracy)는 의사결정에 중대한 영향을 미친다. 예측치의 정확도는 예측오차의 정도이며 이는 관측치와 예측치 간의 차이로 볼 수 있다. 예측방법의 유효성을 평가하기 위하여 예측오차를 기준으로 활용한다.

정확도는 예측오차(Forecast Error)에 의해 결정되고 정확도는 오차가 작으면



작을수록 좋아진다. 정확도를 측정하는 방법으로는 오차제곱평균(MSE: Mean Square Error), 평균오차제곱근(RMSE: Root Mean Square Error), 오차제곱합(SSE: Sum of Squared Error), 절대평균편차(MAD: Mean Absolute Deviation), 절대평균오차비율(MAPE: Mean Absolute Percentage Error), 테일 계수(Theil-U Coefficient) 등이 있다. 이 중에서 MAPE와 테일 계수는 계량모형 간에 예측오차 비교가 용이하고, 신뢰도가 높아 자주 이용되고 있다(이충기, 2003: 27 재인용). 일반적으로 SSE, MSE, RMSE방법은 수치만 제시할 뿐, 예측모형이 어느 정도 정확한지를 설명해 주지 못한다. 그러나 특히 MAPE 방법은 예측모형이 어느 정도 정확한지를 설명해 주며, 그 결과 관광수요 예측을 위해 사용해도 좋은 모형인지 여부를 명확히 밝혀준다는 점에서 가장 유용한 정확도 평가방법이라고 할 수 있다(이충기, 2003: 28). 이러한 주장은 선행연구를 통해서도 증명되었다.

Li와 Song 그리고 Witt(2005)은 1990년부터 2004년까지 주요 저널에서 발표된 총 84편(1980년대 이후 3편 포함)의 논문을 검토하였다. 이 중 4편을 제외한 대부분의 논문은 MAPE값을 기준으로 오차크기의 정확도를 측정하였다. 이와 같이 MAPE가 가장 많이 이용되는 이유는 오차의 정도를 백분율로 똑같이 가중함으로써, 예측기간이나 관측단위가 다른 예측모형 간 비교가 용이할 뿐만 아니라 해석이 용이하기 때문이다(송근석 · 이충기, 2007: 105 재인용).

MAPE값은 낮을수록, Theil-U지수는 $U < 1$, 즉 0에 가까울수록 더 우수한 예측 모형이다. MAPE와 Theil-U의 산출공식과 해석 방법은 다음과 같다.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right| \times 100$$

주: X는 관측치(실제치), F는 예측치, 그리고 n은 관측치의 기간 수

계산된 MAPE값은 다음과 같이 해석하여 판단할 수 있다(Lewis, 1982).



0% ≤ MAPE값 < 10%: 매우 정확한 예측으로 판단됨

10% ≤ MAPE값 < 20%: 비교적 정확한 예측으로 판단됨

20% ≤ MAPE값 < 50%: 비교적 합리적 예측으로 판단됨

MAPE값 ≥ 50%: 부정확한 예측으로 판단됨

(이충기, 2003).

$$Theil - U = \sqrt{\frac{\sum (F_t - X_t)^2}{\sum (X_t)^2}}$$

계산된 Theil-U 값은 다음과 같이 해석될 수 있다(이종원, 2006)

Theil-U < 1 : 사용된 모형이 나이브모형보다 정확하며 U값이 가까울수록 더 정확한 모형으로 판단할 수 있음

Theil-U > 1 : 사용된 모형이 나비브모형보다 정확치 않으므로 이용할 가치가 없는 것으로 간주할 수 있음(우주리, 2007)

실제수요와 예측수요의 비교를 통해 예측오차를 도출하여 예측모형의 정확도를 측정하고 평가할 수 있다.

본 연구에서는 평균절대비율오차(MAPE) 외 평균절대편차(MAD)와 평균제곱오차(MSE)를 활용하여 제시한 모형들의 예측력 평가하고자 한다. 평균절대편차(MAD)는 절대오차를 모두 더한 후 이를 기간수로 나눈 값으로서, 오차의 절대치를 의미한다. 평균제곱오차(MSE)는 절대오차 제곱의 합을 기간 수로 나눈 값으로서, 각 기간의 오차에 제곱을 하기 때문에 모형이 일관성 있는 예측력을 보여주지 못하고 오차가 클수록 평균제곱오차 값이 커지게 된다. 예를 들어 평균절대편차(MAD)의 값이 같은 모형이라도 일정한 오차를 유지하는 모형보다 일정한 오차를 유지하지 못하고 예측기간의 어느 한순간 오차의 폭이 큰 모형의 경우 평균제곱오차(MSE) 값이 훨씬 커지게 된다. 예측모형의 정확도를 평



가하는데 사용한 예측오차 평가척도의 수식은 아래와 같다.(김연형, 2001; 이원우, 2011)

$$MAD = \frac{1}{n} \sum_{t=1}^n |X_t - F_t|$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (X_t - F_t)^2$$

정확한 예측이란 실제수요와 예측수요의 예측오차를 최소화하는 것으로, 실제수요에 근접한 예측치를 얻기 위해서는 예측형태, 예측기간, 예측비용 등의 다양한 요인을 고려하여 적절한 예측방법을 선택해야 할 것이다(Chambers et al, 1971; 김연형, 2001).



제 3 절 물동량 예측에 관한 선행연구

제 1 항 물동량 예측 모형

국내 물류서비스 중 하나인 우편 물류서비스는 전면적으로 차량을 통한 배송이 이루어지며, 국내에서 우편물량 예측에 관한 연구가 활발하게 진행되어 왔다. 김태영과 황승준(2010)은 다중회귀분석과 분해법을 활용하여 국내 우편 물류 수요를 예측할 수 있는 시스템을 개발하였다. 김의창과 최헌탁(1999)은 집중국별 최적의 운송차량 할당 계획을 도출하기 위해 우편물량을 일반통상, 특수통상, 소포로 구분하여 예측하였으며, 이덕주 등(2005)은 단기예측에 주로 사용되는 ARIMA모형을 이용해 우편시장 환경 변화에 영향을 주는 경제 성장률 및 이메일 대체 효과를 고려하여 우편물량을 예측하였다. 또한 이성준 등 (2003)은 국내총생산(GDP) 자료를 가지고 우편 물량을 예측하는 회귀식과 그 결과를 도출하였고, 나아가 성장모형 로지스틱 곡선을 사용하여 우편물량이 지속적으로 증가하지만 그 증가율이 점차 둔화될 것이란 결과를 보여주었다. 우편물처럼 단순히 하나의 품목만을 다루는 예측뿐만 아니라 화물 품목 별 물동량 예측 연구도 진행되었다. 이금숙과 백호중(1990)은 단면적 분석을 통해 권역별·품목별 화물 수송 수요를 예측하는 모형을 구축하고, 정확한 화물 수송 수요 산출을 위해 직선 회귀식과 지수 회귀식을 결합시킨 혼합 회귀모형을 사용하여 교통 권역별·품목별 물동량 예측결과를 제시하였으며, 이강대(2001)는 경제기획원의 화물품목대분류에 따라 화물품목별 지역간 화물 유입·유출 모형을 제시하여 이들 모형을 통해 어떤 사회-경제적 변수가 화물발생량에 영향을 미치는지 보여주었다. 김권일(2016)은 항공수출 화물 운임 예측에 관한 연구에서 예측 기법을 시계열 기법과 회귀분석에 기반한 인과형 기법을 모두 적용하여 예측오차가 가장 적은 기법을 채택하여 미래 항공운임을 예측하였다. 또한, 항공운임에 영



향을 미칠 것으로 기대되는 요인들을 선정하고 요인들간 상관분석 및 회귀분석을 실시하였으며 모형 적합도 검증과정을 거쳐 최종으로 사용될 결정요인 및 회귀모형을 선정하였다. 다음으로는 회귀모형에 사용된 요인들에 대해 지수평활법, Box-Jenkins(ARIMA 모형) 기법 등 시계열 분석을 통해 독립적으로 예측을 실시하고, 이러한 예측치들을 회귀모형에 대입해 장래 항고운임을 예측하였다.

해외에서는 실제 LTL 운송업체의 데이터를 바탕으로 실증연구가 진행되었다. Garrido와 Mahmassani(1998)은 화물수요의 공간-시계열적 특징을 포함한 STARMA(Space-Time Auto Regressive Moving Average)모형을 제시하고 실제 자동차운송회사의 데이터를 이용하여 STARMA모형과 ARIMA모형의 수요 예측 결과를 비교함으로써 제시한 모형의 우수성을 증명하였다. Ling(2006)등은 독일 운송회사의 데이터를 가지고 신경망 분석을 통해 단기간 LTL 운송에 대한 수요예측을 실시하여 그 결과를 ARIMA모형과 지수평활법의 예측결과와 비교해 그 우수성을 증명하였다. Zhang과 Zhu(2009)는 선형의 ARIMA모형과 비선형의 ANN모형을 결합한 예측모형을 제시하여 한 회사의 운송량 데이터를 이용해 수요예측을 진행한 결과, 결합 예측모형이 단일 모형들보다 정확성면에서 더 우수한 것을 증명함으로써 단일 예측모형의 결점을 보완하였다.

제 2 항 물동량 예측 영향 요인

송승훈 외(2011)는 택배산업의 성장 요인에 관한 실증분석에서 택배 성장의 주요 요인을 택배매출액, 전자상거래액, GDP성장률, 판매신용액, 창고업종사자수에서 발견하였고, 기존에 학계나 산업계에서 일반적으로 인식하고 있는 택배산



업의 성장이 전자상거래의 성장에 의하여 영향을 받았다고 하는 점을 실질적으로 처음 실증분석을 시도하였다. 그러나, 택배산업에 한 자료를 5개 업체로 제한하고 택배산업 전체에 대한 자료를 사용하지 않고, 자료 획득의 어려움으로 인하여 변수의 설정에 제한이 있어 통계 분석상의 편의가 발생할 수 있는 한계 점을 보이고 있다. 노경호 외(2000)는 전자상거래 및 택배 물동량 추정에 관한 연구에서 전자 상거래 물동량, TV/홈쇼핑 물동량, 카탈로그통관 물동량, 기존 택배 물동량의 요인을 통해 전자상거래 발전에 따른 택배물동량을 시계열 분석 방법을 사용하여 추정하였다. 기업간 전자상거래의 경우 일부 물량만 택배 처리하는 것을 제외한 나머지 요인별 택배 물량 처리는 100%로 가정하여 추정하였다. 고상필 외(2008)는 국내 물류산업 시장규모와 성장요인에 관한 시계열 인과관계 분석에서 물류산업 연간총생산액, 농수산업 및 관업 연간총생산액, 제조업 연간총생산액, 경공업 연간 총생산액, 중화학공업 연간총생산액, 연간 수출액, 연간 수입액, 운수장비설비투자, 전국도로 총연장길이, 전국 항만 연간 총하역능력, 물류산업 1인당 부가가치와 같은 성장요인과 관련한 시계열 변수들에 관한 연구로서, 수익측면에서 물류산업시장규모의 성장에 요인을 미치는 변수들과의 관계에 대한 고찰을 목적으로 연구하였다. 또한, 도로의 소통, 물가와 임금, 에너지 등의 변수들 보다는 물류산업 성장 자체와 인과관계가 있을 것으로 예상되는 물동량 변수와 투자 관련 변수를 성장요인관련 변수로 설정하였다. 전찬영 외(2007)는 수출입 컨테이너물동량의 변동요인에 대한 민감도 분석에서 국내투자, 최종소비, 수출입비율, 2차산업 생산비율, 제조업 해외직접 투자금액, TEU당 컨테이너 수출금액, 인구의 변동 요인을 연구하였으나, 현실적으로 다변량에 대한 민감도 분석이 보다 의미가 있음에도 불구하고 변수의 조합에 따른 경우의 수가 너무 많은 관계로 미 시행하였다. 각 변량마다 표준화된 단위의 개념 또한 달라 민감도에 대한 정확한 해석이나 비교, 평가가 용이하지 않음을 알 수 있었다. 서선애 외(2012)는 산업구조와 수출입 상품 특성이 항만



물동량에 미치는 영향에서 국내 산업구조(1,2,3차 산업 비중 변화 포함), 수출입 상품 구조, 항만물동량간 변화추이의 요소를 통해 항만 물동량에 미치는 연구를 하였다. 여기서는 항만 물동량이 경쟁력이나 항만의 우위성을 파악하는데 연구를 단순화하고 명확하게 할 수 있다는 가정하에 종속변수를 항만 물동량으로 한정하고, 국내 여러 항만의 물동량을 하나의 단위로 하는 것은 독립변수의 영향요인을 분산시키고, 물동량 특성을 저해할 수 있으므로 대표항인 부산항의 물동량으로 한정하여 연구를 하였다. 이지홍(2017)은 빅데이터 분석을 통한 물동량 분석 예측 모델에 관한 연구에서는 기간별 물동량, 월/요일 단위 효과 반영 가능한 캘린더 변수, 휴일 등 이벤트 특성이 반영된 캘린더 변수와 전국 대상의 물동량 크기에 따른 군집으로 재구성하여 접근하였으나, 예측 모델에 대한 실증적 분석은 진행하지 않았다.

제 3 항 기존 연구의 한계

앞에서 설명된 기존 연구들은 잘 만들어진 계산식으로 물동량 예측에 대한 도움을 주고 있다. 하지만 물동량 예측에는 지난 물동량과 달력 효과(Calendar Effect)가 변동폭에 큰 영향을 주고 있는 것을 알 수 있었고, 기존 예측 모델에는 달력 효과나 실제 물류 실행단에서 발생할 수 있는 변수에 대한 추가적 요인을 반영을 할 수 없어 실제 현장에서는 사용하기가 어려웠다. 또한, 각 예측 모델이 가지고 있는 아래의 한계점으로 인해 전체 Sub 터미널 기준의 물동량을 분석하기에는 기존의 예측 모델을 각각의 특징에 맞는 적절한 변수와 그 변수의 사용에 대한 가중치를 두어야 하였다.

첫째, 최소제곱추정치(회귀분석)는 표본의 크기(Sample Size)가 작거나 독립 변수사이의 강한 상관관계(Correlation)로 인한 다중공선성(Multicollinearity)이 존재하는 경우 회귀계수의 분산이 커져 추정 회귀식의 예측력(Prediction Accuracy)이 떨어지는



문제점이 있다. 독립 변수의 개수가 증가하면 변수에 대한 해석력이 떨어지므로 이를 보완하기 위해 부분집합선택법과(Subset Selection)과 계수축소법(Coefficient Shrinkage)이 있다. 따라서 위와 같은 문제가 있을 시 OLS의 방법 보다는 Best Subset이나 Shrink-age 방법을 이용하는 방법이 훨씬 더 회귀계수를 측정하는데 도움이 될 것이다(Dunis, 2002).

둘째, Subset Selection의 경우 영향력 있는 변수를 선택함으로써 해석이 가능한 모형을 제공하며 적절한 모형의 복잡도(변수의 개수) 유지로 완전모형(Full Model)에 비해 낮은 예측 오류(Prediction Error)를 가진다. 하지만 변수선택 기준(AIC, BIC 등)에 따라 최종모형이 달라질 수 있고, 데이터의 작은 변화에 민감하기 때문에 예측력이 떨어지며, 독립변수의 개수가 증가하면 계산량이 많아지는 문제점이 있다(김은령, 2011).

셋째, 변동폭이 큰 계절요인이나 추세만으로는 설명하기 힘든 불규칙요인을 포함하고 있는 시계열자료의 경우에는 추세에 대한 정확한 파악이 어렵다. 또한 이는 시간적 분석에 의한 값을 가져오는데 변동을 주고 요일적 요인도 음력을 사용하는 한국 사회에서 명절의 예측력은 떨어지고 오차를 발생시키게 된다(김연형, 2001).

선행 연구에서는 물동량 예측에 영향을 미치는 요인에 대한 도출 및 상호 연관성을 파악하기 위한 요인분석이 미흡하였으며, 도출된 모든 요인을 독립변수로 한 번에 포함하여 회귀분석을 수행하였다. 이에 본 연구에서는 물동량 예측을 위한 주요 영향요인에 대한 분석과 핵심요인을 설명할 수 있는 독립변수를 설정하여 한정된 변수를 기반으로 회귀 예측 모형 선정에 관하여 심도 있는 연구를 수행하고자 한다. 또한, 회귀 분석을 통한 예측 모형과 시계열 분석을 통한 예측 모형 간 실증분석을 진행하고자 한다.



제 3 장 연구 방법

제 1 절 연구모형의 설정

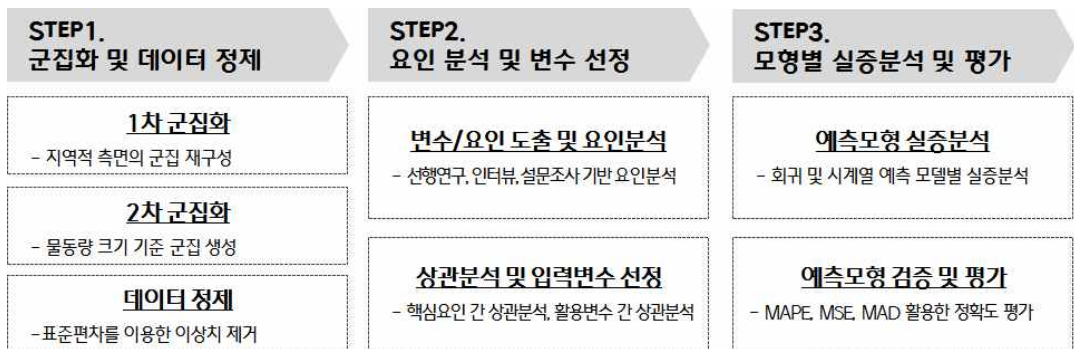
지금까지 물류 사업과 연관된 성장요인, 물동량 추정에 관한 연구는 주로 전자상거래의 성장과 국내소득의 증가에 영향을 받는 현상에 대한 실증분석이 주를 이루고 있고, 기타 항만 물동량과 같은 물류 전반의 거래량에 대한 연구 또한 산업 전반의 구조와 산업 총생산액, 수출입액 등 거시경제지표를 기반으로 연구가 추진되었다. 이로 인해 산업을 대상으로 한 성장요인과 물동량 추정에 관한 대부분의 연구는 年 단위 추정에 근거한 실증분석에 한계성을 보이고 있었다. 또한, 선행 연구를 통해 검토한 바와 같이 물류 산업에 있어 물동량 예측은 모두 年 단위 기반의 추정을 토대로 수행이 되어 거시경제지표를 요인으로 한 실증분석이 수행되었다. 따라서, 본 연구는 지금까지 제한적으로 접근하였던 日 단위 물동량 추정을 위해 활용될 요인을 도출하고 그 요인들이 실제로 물동량 추정에 어느 정도 설명력을 가지고 있는지 택배 물류 서비스를 제공하고 있는 국내 최대 종합물류기업 C사 택배부문의 실제 물동량 데이터를 근거로 실증분석을 하고자 한다.

C사의 경우 예측 모델에 대한 반영을 모든 Sub 터미널로 한다면 240여 개가 넘는 터미널의 모든 특성을 고려해야하기 때문에 각각의 터미널 단위로 반영하지 않고 지역적 위치와 日 단위 Sub 터미널의 집합을 묶는 군집화(Clustering)를 실시하였다. 즉, 1차로는 전국을 3개 단위로 나누어 지역적 측면에 의한 군집화를 진행하고 2차로는 예측 정확도의 중요도를 고려하여 日 단위 물량에 의한 군집의 재구성을 하였다. 다음으로 과거 물동 데이터 기반의 분석 마트를 구성하고 마트 구성 시 데이터 정제를 위해 표준 편차를 이용한 이상치를 사전



에 제거하여 시계열적 흐름에 대한 기초적인 물동 데이터 마트를 구성하였다. 또한, 선행연구 및 인터뷰를 통해 도출된 기준별 변수를 요인 분석을 통해 주요 요인을 선정하고 회귀와 시계열 예측 모형을 통해 시뮬레이션 하여 결과값을 토대로 가장 정확도가 높은 예측 모형을 선정하는 데 중점을 두었다.

먼저 전국 대상의 전체 물동량이 아닌 집하의 시작점인 sub 터미널별 물동량 분석에 중점을 두고 Sub 터미널을 중심으로 예측 모델의 검증을 위한 데이터의 군집화, 데이터 정제, 요인 분석 및 설명변수의 설정 그리고 모형별 실증 분석 및 검증과 평가를 수행하는 순서로 구성되어 있다.



<그림 12> 연구 모형



제 2 절 단계별 연구

제 1 항 군집화 및 데이터 정제

본 연구를 위해 국내 종합물류기업 C사의 2014년 11월 21일부터 2017년 10월 31일에 이르는 3년 가까운 기간의 Sub 터미널별 日 단위 실제 물동량 정보를 수집하였다. 물동량 예측 결과의 정확도를 높이기 위해 2가지의 기준으로 Sub 터미널에 대한 군집화를 2단계에 걸쳐 수행하였다.

STEP1. 군집화 및 데이터 정제

1차 군집화

- 지역적 측면의 군집 재구성
- 수도권 중심 분석(전체 3/4 차지)

수도권	1군집 (전체물량의 73.2%)
경상권	2군집 (전체물량의 13.3%)
기타 (충청, 호남 등)	3군집 (전체물량의 13.5%)

2차 군집화

- 예측 정확성의 중요도를 고려한 물동량 크기 기준 군집 생성

1군집	일평균 물동량 $\geq 1,000$
2군집	일평균 물동량 $< 1,000$

데이터 정제

- 데이터 탐색 후 표준편차를 이용한 이상치 제거(단가성 이벤트 물량 위주)



<그림 13> 군집화 및 데이터 정제 방법

첫 번째 기준으로는 예측의 정확성 측면의 향상을 위해 물동량 기준으로 군집화를 수행하였다. 먼저 일평균 물동량이 30,000건 이상의 물량을 가지고 있는 Sub 터미널을 1군집으로 구성하고, 일평균 물동량 1,700건 이하의 물량을 가지고 있는 Sub 터미널을 5군집으로 분류 후 2~4군집을 Two-Step에 의한 군집분석을 실시하였다. 이는 Sub 터미널별 물동량을 크기순으로 정렬하여 그래프로 분석 후 급격히 하락하는 구간을 기준으로 차수별 군집화를 수행하였다. 다음으로 2차 군집화는 지역적 물동량의 분포를 고려하여 총 3가지의 부분을 하였



고, 전체 물량의 73.2%를 차지하고 있는 수도권과 13.3%를 차지하고 있는 경상권을 주요 2군집으로 설정 후 나머지 충청권, 호남권 등을 3군집으로 나누어 물량 분석의 군집을 선 진행하였다.

제 2 항 변수 요인분석 및 선정

택배서비스는 개인간 거래(C2C), 업체 대 개인 거래(B2C), 그리고 업체간 거래(B2B) 등 거래 조건에 따라 물동량의 대한 차이가 발생할 수 있는 가능성이 크며, 택배의 운영 프로세스 측면에서 접근해 볼 때 日 단위 물동량 예측을 위한 분석 관점을 고객별 예측, Sub 터미널별 예측과 고객별XSub 터미널별 예측으로 구분할 수 있다. 여기에서 고객별 예측을 위한 요인 분석은 기업고객별 영업 전략 수립에 많은 영향도를 보이고 있고, 고객별XSub 터미널별 예측은 일별 물량이 작아 예측 정확도가 매우 낮은 것으로 판단되어, 단위 Sub 터미널별 예측 요인에 대한 분석으로 수행하고자 한다. 또한, 동 분야에 대한 선행연구를 보면 年 단위의 물동량에 영향을 미치는 요인들이 단순히 특정한 시설이나 현상에 머물지 않고 다소 복잡적이고 다양한 요인들이 될 수 있음을 보이고 있고 연구자 마다 활용하는 요인과 변수에 차이가 있다.

따라서 본 연구에서는 인터뷰를 통해 주요 요인들을 도출하고 설문을 통한 탐색적 요인분석을 통해 요인을 축소 후 요인별 특성을 가진 대표 변수를 선정하여 예측을 수행하고자 한다.

본 연구를 위한 인터뷰 및 설문 모집단은 종합물류기업 C사 택배사업부문으로 하며, 1개 기업의 택배사업부를 대상으로 설문을 수행하는 이유는 현재 C사 국내 택배 점유율 40% 이상을 차지하는 1위 택배 기업이며 전국 240여개의 Sub 터미널에서 매일 일별 물동량을 수작업으로 예측을 하고 있기 때문에 택



배와 관련된 물동 움직임의 예측을 위한 분석 대상으로 종합적인 판단이 가능하다는 당위성이 있기 때문이다.

STEP2. 요인 분석 및 변수 선정



<그림 14> 요인분석 및 변수선정 방법

인터뷰는 1차로 택배 운영 기획부서 내 물동량 계획 담당자와 인터뷰를 실시하였고, 이후 5년 이상의 택배 경력을 가진 각 지역 지점장을 대상으로 실시하였다.

<표 4> 인터뷰 및 설문 조사 현황

구 분	택배 기획부서원	택배 지점장
인 원	2	5
일 시	'17.10.16	'17.10.17~19
내 용	日 단위 택배 물동량 예측을 위한 영향요인 도출	

日 단위 택배 물동량 예측에 영향을 미치는 요인을 도출하기 위해 선행연구를 통해 도출된 영향요인을 기반으로 전문가를 통한 인터뷰를 통하여 물동량에 영



향을 미치는 독립변수를 선정하였다. 이상의 인터뷰를 통하여 물동량 예측의 영향요인은 주단위/월단위/년단위의 동요일 물동량과 공휴일 효과, 명절 효과, 요일별 특성치 효과, 연말연초 효과 등 총 10개를 선정하였다.

<표 5> 인터뷰를 통한 예측 영향요인 선정

분 류	코드	영향요인
수량적 요소	VAR1	주단위 동요일 평균 물동량
	VAR2	월단위 동요일 평균 물동량
	VAR3	분기단위 동요일 평균 물동량
	VAR4	년단위 동요일 평균 물동량
명목적 요소	VAR5	명절 효과
	VAR6	공휴일 효과
	VAR7	연속 휴일 효과
	VAR8	연말연초 효과
	VAR9	간절기 효과
	VAR10	요일별 특성치 효과

日 단위의 물동량을 예측하기 위하여 고려하여야 할 첫 번째 영향요인은 과거의 동요일 물동량이며 일정 기간단위의 동요일 물동량이 평균이 중요한 요인이 됨을 인터뷰를 통해 확인하였으며, 특히 공휴일 여부에 따른 물동량의 변동량이 큰 것으로 조사되었다. 선행연구에서는 최근 3년 동안(2013년 6월 ~ 2015년5월)의 3.4mm 이상의 일평균 강수량을 기준으로 강수 여부와 물동량 변화와의 통계적인 검증을 실시하였고, 검증의 결과 통계적으로 하루 전일의 강수 여부는 당일 물동량에 변화에 많은 영향을 주지 않는 것으로 판단하고 있다. 결과적으로 기후적 데이터 기반의 변수 사용은 제외하였다(이지홍, 2017).



✓ 최근 3년 동안의 일 평균 강수량(3.4mm)을 기준으로
강수여부와 물동량 변화와의 통계적인 검증을 실시

구분	강수여부-D1	빈도	평균	평균차이	신뢰수준
구간1	O	64	1,518,673	78,656	47.0%
	X	301	1,440,017		
구간2	O	55	1,953,409	147,502	61.9%
	X	310	1,805,907		
전체기간	O	119	1,719,602	93,945	62.7%
	X	611	1,625,657		

통계적
유의성
결과

- 통계적 결과로는 전일 강수여부가 당일 물동량 변화에 영향을 주지 않는 것으로 판단됨
- 전체기간의 경우 비에 의해 물동량 평균이 더 높게 나타나는 경우는 62.7%로 비가 온 날 다음날에 물동량이 많다는 결과를 신뢰할 수 없다

- 구간 1 : 2013년06월~2014년05월
- 구간 2 : 2014년06월~2015년05월
- 전체기간 : 2013년06월~2015년05월

<그림 15> 기후 데이터 통계적 유의성 검증 결과

자료: 이지홍(2017) 자료를 활용하여 기후 데이터 유의성 검증 부분 재구성

최종적으로 선정된 영향요인은 기간별 물동량 및 공휴일, 명절 등의 효과를 반영할 수 있는 달력 변수와 요일별 특성을 반영할 수 있는 요일 변수를 사용하여 물동의 변동 사항에 대한 초기 요인을 10개로 책정하여 진행하고자 하며 택배에서 실제 물동량 예측을 수행하고 있는 각 Sub 터미널별 담당자에 대한 설문 조사를 기반으로 요인 분석을 통해 통계적 유의성의 신뢰도를 높이고 예측의 정확도를 높여 나아갈 수 있도록 진행하였다.

설문지 배포 및 회수기간은 2017년 11월 2일~11월 9일까지 C사 택배사업부문의 기획부서의 도움을 통해 수도권 대상 총 116개 Sub 터미널 중 군집화를 통해 선정된 97개 Sub 터미널을 대상으로 실시되었다. 응답률은 76명으로 78%이며 이 중 유효한 설문은 총 72개로 최종 74%의 설문을 대상으로 요인분석을 실시하고자 한다. 요인분석을 통해 인터뷰를 통해 도출된 변수들 사이의 공분산과 상관관계 등을 사용하여 변수들 간의 상호관계를 분석하고, 그 결과를 토대로 설문 문항 내 변수들 간의 상관성과 구조를 파악하여 여러 변수들이 지닌 정보를 보다 적은 수의 요인으로 묶어서 회귀분석 시 변수를 간소화 하고자 한다. <표 6>는 선행연구 및 인터뷰를 통해 파악된 변수이며 최종적으로 요인분



석을 통해 최종적으로 간추려진 요인을 기반으로 대표 변수를 선정하여 회귀모형을 통한 분석을 실시할 예정이다.

<표 6> 예측모형별 분석을 위한 활용변수 리스트

활용변수		
물동량 기준	달력 기준	기후 요소
최근 3일 물동량 평균	휴일 전일 여부	당일 강수 여부
최근 7일 물동량 평균	휴일 익일 여부	1일전 강수 여부
1주일전 동요일 물동량	2연휴 익일 여부	2일전 강수 여부
2주일전 동요일 물동량	3연휴 익일 여부	3일전 강수 여부
3주일전 동요일 물동량	4연휴 익일 여부	당일평균기온
최근 3주 동요일 물동량 평균	휴일 징검다리 여부	1일전평균기온
최근 3개월 동요일 평균 물동량	연말 효과 여부	2일전평균기온
최근 6개월 동요일 평균 물동량	월 여부 변수	3일전평균기온
2년 월별 동요일 평균 물동량	요일 여부 변수	
최근 1년 월별 동요일 평균 물동량	명절 2-10일 전 여부	
전월 동요일 평균 물동량	명절 1-3일전 여부	
최근 1년 물동량	월 초/중/말 여부	
최근 2년 물동량	최근 1년 여부	
물동량 월 물동량 집중도		
물동량 요일 물동량 집중도		



제 3 항 예측 모형 선정 및 평가

군집화와 데이터 정제를 통해 분석마트를 설계하였고, 예측의 주요 영향요인의 도출 및 최종 선정을 통해 예측 모형을 개발할 수 있는 기초를 마련하였다. 일반적으로 기업의 수요예측이나 물동량 예측기법은 크게 정성적 혹은 정량적 기법으로 구분하고 이 중 정량적 기법은 세부적으로 인과형 기법 및 시계열 기법으로 구분한다.

STEP3. 예측 모형별 실증분석 및 평가



<그림 16> 예측 모형별 실증분석 및 평가 방법

인과형 예측기법(Causal Method)은 물동량 예측에 영향을 주는 주요요인들을 파악하고 이들 간의 인과관계를 고려하여 미래 물동량을 예측한다. 인과형 예측기법은 과거 물동 데이터를 바탕으로 日 단위 물량을 종속변수로 하고 이에 영향을 미치는 요인들을 독립변수로 하여 회귀모형을 추정 후 개별 파생변수들을 시계열 기법을 통해 예측한 값들로 회귀모형에 대입하여 물동량을 예측한다.



시계열 예측기법(Time Series Method)은 과거 물동 데이터를 분석하여 시간에 따른 물동 변화의 패턴을 파악하고 이를 바탕으로 미래를 예측하는 방법이다. 시계열 기법은 과거의 물동량에 대한 패턴이 미래의 물동량에도 적용된다는 가정에 기초로 하고 있으며 주로 중·단기 예측에 적합하고 사용되는 기법으로는 이동 평균법(Moving Average), 지수평활법(Exponential Smoothing), ARIMA(Autoregressive Integrated Moving Average), 최소자승법 (Least Square Method) 모형 등이 있다.

日 단위 물동량 예측 모델 개발을 위한 방법으로는 최종으로 선정된 영향요인별 파생 변수를 통해 선형 회귀 예측모델과 물량의 변동 폭의 측정과 시간의 흐름을 이해하기에 좋은 시계열 예측모델을 선정하여 진행하고자 한다.

본 연구는 택배화물을 대상으로 수도권 내 각 Sub 터미널별로 택배 집하에 대한 2014년 11월부터 2016년 12월까지 24개월 간 C사의 택배 물동 데이터를 바탕으로 2017년 1월부터 2017년 10월까지 10개월간 수도권의 택배 물동량 예측을 목적으로 한다. 예측기법으로는 회귀분석에 기반한 인과형 기법과 시계열 기법을 모두 적용하여 예측오차가 가장 적은 기법을 채택하여 택배 물동량을 예측한다.

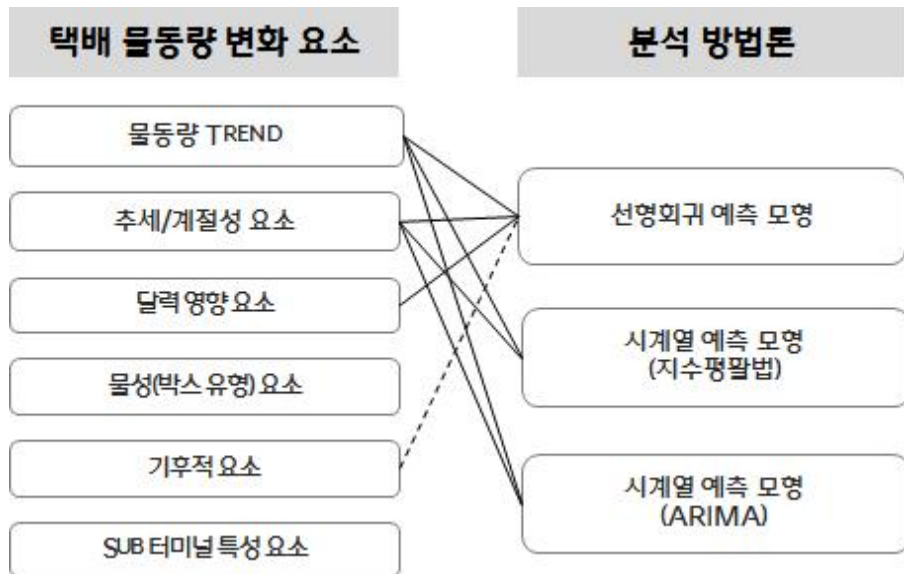
물동량에 영향을 미칠 것으로 기대되는 주요 요인들을 선정하고 요인들 간 신뢰도분석 및 상관분석을 실시하고 회귀분석(regression analysis)을 수행하며 모형 적합도 검증과정을 거쳐 본 연구 물동량 예측에 활용될 결정요인 및 회귀모형을 확정한다. 다음으로 지수평활법, Box-Jenkins의 ARIMA 모형 등 시계열 모형을 통해서도 예측하며 예측오차를 회귀모형의 결과와 비교하여 예측오차가 더 작은 모형을 바탕으로 한 예측치를 최종 선택한다.



제 4 장 연구 결과

제 1 절 핵심요인 및 상관관계 분석

택배 日 단위 물동량 예측모델을 개발하기 위해 기존 연구에서 적용된 변수와 C사 택배 사업부문 운영자료 분석을 통해 도출된 변수는 지난 물동량에 대한 트렌드 변수, 달력 영향 여부 변수, 기후적 요소가 반영된 변수 등 총 40여 개에 달한다.



<그림 17> 택배 물동량 변화 요소에 따른 분석 방법론

선정된 많은 변수 간 핵심 내재 요인을 간추려 예측 요인에 대한 이해도를 높이고, 변수들 사이에 다중공선성이 발생하지 않으면서 최소의 변수를 활용하여 최대의 효과를 산출할 수 있는 예측 모델을 만들고자 한다. 하여, 전문가 집단

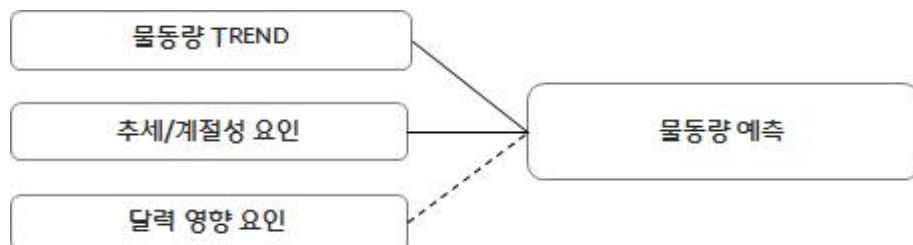


의 인터뷰를 통하여 최종 10개의 요인으로 압축하여 선정하였고 이 요인들을 토대로 요인분석 및 상관관계분석을 실시하여 핵심 내재 요인에 따른 대표 변수로 예측 모델에 대한 연구를 수행한다.

제 1 항 핵심요인 및 신뢰도 분석

요인분석은 R-type 요인분석과 Q-type 요인분석으로 나뉜다. R-type 요인분석은 평가항목인 변인들을 기준으로 요인들을 구분하는 것이고, Q-type 요인분석은 개별 설문 응답자들에 대하여 case별로 상이한 특성을 가지는 개인들을 상호 동질적인 몇 개의 집단으로 구분하는 것이다. 이번 연구에서는 R-type 요인분석을 통해 도출된 변인들을 요인별로 추출하고 최종 변수로 설정할 예정이다.

<그림 16>와 같은 가설적 예측 연구모델을 진행하기 위해 설문조사를 실시하였다. 물동량 예측에 영향을 끼칠 수 있는 변인을 총 10개 문항으로 조사를 실시하였으며 그 결과를 토대로 요인분석을 실시하였다.



<그림 18> 물동량을 위한 예측 변수



해당 변인들을 가지고 요인 분석을 시행하게 되면 사용된 변수의 개수만큼 요인이 추출되고 각각의 요인에 고유값이 정해진다. 다음의 <표 8>은 설문에 따른 분석결과 중 요인들이 가진 변인의 분산 설명도를 보여주고 있다. 총 성분이 10가지로 확인되고 있으나, 고유값이 1 이상인 요인을 추출하는 것으로 설정하여 설정대로 추출된 3개 요인으로 회전 전/후의 고유값을 판단하면 된다. 추출된 요인들 3개가 전체 입력변수의 68.853%를 설명하고, 추출된 적재값은 60.280%를 설명하고 있다.

추출된 요인 적재값(추출 제곱합 적재값 중 % 누적 열의 마지막 행의 값)은 일반적으로 사회과학 연구에서는 60% 이상, 자연과학에서는 95% 이상으로 확인한다. 해당 연구는 사회화학에 초점이 맞추어져 있으므로 추출된 요인들은 충분한 설명력을 지닌 것으로 판단할 수 있다.

<표 7> 설명된 총분산

요인	초기 고유값			추출 제곱합 적재값			회전 제곱합 적재값 a
	합계	% 분산	% 누적	합계	% 분산	% 누적	
1	4.117	41.169	41.169	2.374	23.737	23.737	2.089
2	1.699	16.986	58.155	.925	9.250	32.988	2.095
3	1.070	10.698	68.853	2.729	27.292	60.280	3.092
4	.907	9.075	77.927				
5	.829	8.294	86.222				
6	.457	4.567	90.789				
7	.331	3.306	94.095				
8	.251	2.507	96.602				
9	.197	1.966	98.568				
10							

추출 방법: 최대우도.

a.요인이상관된경우전체분산을구할때제곱합적재값이추가될수없습니다.

요인분석을 통해 설문 문항인 변인들을 서로 연관성이 있는 것끼리 묶어 최종 3개 핵심요인으로 간결하게 표현을 할 수 있음을 확인하였다. 하지만 이 분



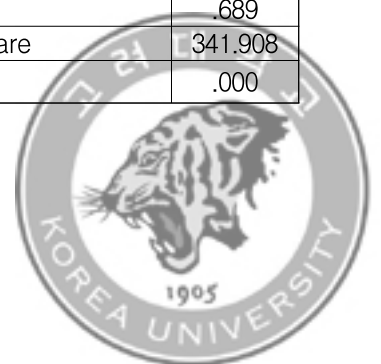
석 결과만으로는 연구모형에 대한 방향 제시를 할 수가 없으며 요인분석으로 도출된 핵심요인 간 신뢰성이 있는지에 대해서 판단해야 한다. 따라서 연구대상에 대한 반복적 측정을 가정하였을 때, 동일한 값을 얻어낼 수 있다는 가능성을 확인하는 신뢰도 분석을 진행하였다. 신뢰도를 판단할 때는 크론바흐 알파(Cronbach α) 값을 사용한다. 크론바흐 알파계수를 구하는 식은 아래와 같다.

$$(\text{크론바흐 } \alpha) = \frac{N\rho_{12}}{\{1 + \rho_{12}(N-1)\}} = \frac{(\text{문항수}) \times (\text{상관항목들의 평균값})}{1 + (\text{상관항목들의 평균값}) \times \{(\text{문항수}) - 1\}}$$

신뢰도 분석의 주 목적은 측정한 결과가 얼마나 동일하게 유지한지에 대해 알아보는 분석이기 때문에 이는 곧 각 요인들이 얼마나 정확하게 나뉘었는지를 확인할 수 있다. 따라서, 요인분석 결과에 따른 3개의 요인으로 나누어 각각의 요인에 대한 신뢰도분석을 진행하였다.

<표 8> 요인분석 결과

요인	성분			Cronbach Alpha
	요인1 (요일별특성치)	요인2 (최근물동트렌드)	요인3 (캘린더적용치)	
요일별 특성치 효과	1.032	-.064	-.091	.720
명절 효과	.547	.078	.075	
월단위 동요일 물동량	.204	.966	-.183	.605
주단위 동요일 물동량	-.052	.493	.257	
연말연초 효과	-.121	.175	.798	.844
년단위 동요일 물동량	.148	.063	.792	
분기단위 동요일 물동량	-.081	.022	.738	
연속 휴일 효과	.114	.370	.499	
간절기 효과	.350	.235	.480	
공휴일 효과	.059	-.193	.448	
KMO (Kaiser-Meyer-Olkin 측도)				.689
Bartlett의구형성 검정 (Bartlett Test of Sphericity)		Chi-Square		341.908
		df(p)		.000



요인분석 및 신뢰도분석의 결과는 <표 9>와 같다. 먼저 변인들간의 편상관을 확인하기 위해 변인들의 숫자에 대한 적절성을 나타내는 표본 적합도를 의미하는 KMO(Kaiser-Meyer-Olkin) 측도를 살펴본다. KMO 값은 높을수록 좋으며 요인분석을 실시하는 것이 적절하다고 판단하는 기준은 일반적으로 0.5보다 큰 경우 그렇게 판단할 수 있다. 간혹 그 값이 0.6 이상 또는 0.8 이상이어야 한다는 주장도 있지만 연구문제에 따라 값은 달라질 수 있기 때문에 보통 사회과학에서는 0.5 이상의 수준을 요구하고 있다. 또한, 요인분석을 실시할 때 사용되는 상관계수의 행렬이 대각행렬이면 요인분석을 하는 것이 부적절하기 때문에, Bartlett 값을 확인하여 $p < 0.05$ 이면 대각행렬이 아님을 의미하므로 요인분석을 하는 것이 적절함을 확인하여야 한다. 이에 요인분석의 결과 KMO 값은 0.689로 > 0.5 , Bartlett의 p 값은 0.000으로 < 0.05 를 만족하므로 현재 진행 중인 요인분석은 적절한 것으로 판단할 수 있다.

3개의 핵심요인에 대한 신뢰도는 크론바흐 알파계수를 기준으로 평가할 수 있다. 보통 크론바흐 알파의 절대값이 0.7 이상이면 항목들 간 어느 정도는 일관성이 있다는 판단을 하고 0.8 이상이면 일관성이 좋다고 판단할 수 있으며, 0.6 이하이면 일관성이 없다는 판단을 하는 것이 일반적이다. 또한, 예를 들어 다중공선성(multicollinearity) 문제와 같이 불필요한(Redundant) 항목들이 있는 것으로 판단하는 기준은 0.9 또는 0.95 이상으로 값이 매우 큰 경우에 고려할 수 있다. 크론바흐 알파가 0.8 정도 되는 것이 이상적(Ideal) 값이라고 하는 선행연구도 있었지만, 이는 상당히 주관적 판단일 수 있으며 가설검정(Hypothesis Test)의 수행 시 유의수준(Significance Level)이 어느 정도(몇 %)이어야 하는 지에 대한 정답이 없듯이 크론바흐 알파가 얼마 이상의 값을 가져야 하는 지도 마찬가지로 볼 수 있다. <표 9>에서 확인할 수 있듯이 요인분석을 통해 도출된 3개 핵심요인의 크론바흐 알파 계수가 각각 0.720, 0.605, 0.844이므로 2번째 요인을 제외하고는 모두 충분한 신뢰성을 가지고 있다고 판단할



수 있다.

추가적으로 3번째 핵심요인에 대한 항목 총계 통계량을 <표 10>과 같이 확인할 수 있는데, 여기서 공휴일의 효과를 제거할 경우 기존의 크론바흐 알파계수 값인 0.844 보다 큰 값을 갖게 되는 것을 확인할 수 있다. 이는 일정 수준의 신뢰도를 확보하기 위해 삭제해야할 문항을 선택할 수 있는 기준을 마련해 주는 것이고 신뢰도가 낮게 나온 경우에는 항목 총계 통계량을 확인하면서 항목을 삭제하면서 신뢰도를 높이면 된다.

하지만 신뢰도를 높이는데만 집중하여 항목을 너무 많이 제거해서는 안되며, 최소 2개 이상의 항목은 유지를 해야 한다. 이 경우는 크론바흐 알파계수 값이 0.8 이상으로 이미 높은 신뢰성을 가지고 있기 때문에 굳이 해당 값을 제외하지 않아도 무방할 것으로 판단한다.

<표 9> 항목 총계 통계량

3번째 핵심요인	항목이 삭제된 경우 척도 평균	항목이 삭제된 경우 척도 분산	수정된 항목-전체 상관관계	항목이 삭제된 경우 Cronbach 알파
요일별 특성치 효과	23.82	18.037	.681	.809
명절 효과	25.31	16.497	.678	.807
월단위 동요일 물동량	23.60	16.554	.727	.797
주단위 동요일 물동량	22.89	19.959	.373	.862
연말연초 효과	24.18	17.108	.692	.804
년단위 동요일 물동량	26.04	17.393	.610	.821



제 2 항 상관관계 분석

다중회귀분석 모형에 사용될 변인들 간 상호관계를 알아보기 위해 요인분석을 통해 도출된 핵심요인에 대한 상관관계 분석을 수행하였다. 상관관계 분석은 연관성분석의 대표적인 방법 중 하나로 연관성분석은 각각의 변수들이 서로 독립적(연관성이 없음)인지, 아니면 어떠한 연관이 있어서 영향을 서로 주고 받는지($0 < \text{연관성} \leq 1$)를 알기 위한 분석 방법이다. 연관성분석은 크게 4가지 분석 방법으로 나뉘는데, 어떤 척도를 사용하는가에 따라 적용 방법이 달라진다. 분석 방법에는 Spearman 서열 상관분석, Pearson 상관분석, 편상관분석, 교차분석 등이 있다. 본 연구에서는 가장 보편적으로 활용되는 pearson 상관분석을 실시하였다.

<표 10> 상관분석 결과

변수	평균	표준편차	상관관계		
			1	2	3
1. 요일별 특성치	6.451	0.998995	1		
2. 최근 물동 트렌드	5.507	0.997883	.264*	1	
3. 캘린더 적용치	4.861	0.94196	.253*	.264*	1

*.상관계수는0.05수준(양쪽)에서유의합니다.

3가지 핵심요인에 대한 상관분석을 실시한 결과는 <표 11>과 같으며 모든 변수 간 상관관계가 유의한 것으로 확인되었다. 해당 결과를 토대로 각 핵심요인 별 특성에 해당되는 변수를 선정하여 다중회귀분석을 실시하고자 하며, 분석을 위한 변수 선정은 변수간 유의한 상관관계를 형성하고 있는 조합을 <표 12>를 참고하여 선정 후 다중회귀분석을 실시한다.



<표 11> 핵심 요인 및 입력 변수

핵심 요인	변인	입력 변수
요일별 특성치	요일별 특성치 효과	월요일 여부
	명절 효과	명절 15일 전부터 3일 후
최근 물동 트렌드	월단위 동요일 물동량	1개월 전 동요일 평균 물량
	주단위 동요일 물동량	최근 2주 동요일 평균 물량
캘린더 적용치	연말연초 효과	연말효과 여부
	년단위 동요일 물동량	최근 2년 동월 동요일 평균물량
	분기단위 동요일 물동량	3개월 전 동요일 평균 물량
	연속 휴일 효과	2일에서 4일 연속휴일 여부
	간절기 효과	간절기
	공휴일 효과	휴일 전일&익일 여부

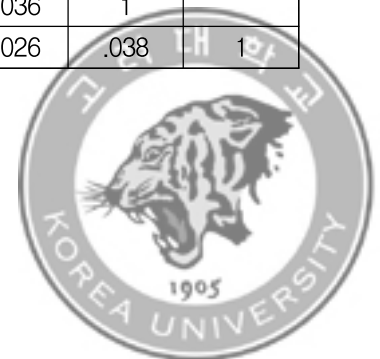


<표 12> 활용변수 간 상관관계 결과

구 분		요인 1		요인 2		요인 3					
		월요일 여부	명절 15일 전부터 3일 후	1개월 전 동요일 평균 물량	최근 2주 동요일 평균 물량	연말 효과 여부	최근 2년 동월 동요일 평균물 량	3개월 전 동요일 평균 물량	2일에서 4일 연속휴 일 여부	간절기	휴일 전일& 익일 여부
요인 1	월요일 여부	1									
	명절 15일 전부터 3일 후	-.008	1								
요인 2	1개월 전 동요일 평균 물량	.724**	-.067	1							
	최근 2주 동요일 평균 물량	.731**	-.074*	.966**	1						
요인 3	연말효과 여부	.005	-.187**	-.110**	-.044	1					
	최근 2년 동월 동요일 평균물량	.894**	.029	.760**	.757**	-.160**	1				
	3개월 전 동요일 평균 물량	.713**	.010	.944**	.950**	-.107**	.759**	1			
	2일에서 4일 연속휴일 여부	.070	.190**	.035	.035	-.049	.093*	.046	1		
	간절기	.001	.101**	.027	-.044	-.137**	-.032	.050	-.036	1	
	휴일 전일 & 익일 여부	.014	.388**	.000	-.003	.005	.018	.015	-.026	.038	1

** . 상관계수는 0.01 수준(양쪽)에서 유의합니다.

* . 상관계수는 0.05 수준(양쪽)에서 유의합니다.



제 2 절 물동량 예측 모델

제 1 항 다중회귀분석

핵심요인 분석과 상관관계 분석 기준으로 다중회귀분석을 시행할 입력변수를 정리하면 <표 14>와 같다. 선정 기준은 요인에 해당하는 입력 변수를 선정 후 해당 입력 변수와 상관관계가 가장 높은 다른 요인의 입력 변수를 선정하였다. 즉, 1번 요인인 요일별 특성치와 관련된 입력 변수 중 월요일 여부를 먼저 선정하고 해당 변수와 상관관계가 높은 다른 요인의 입력 변수인 최근 2주 동요일 평균 물량과 최근 2년 동월동요일 평균 물량을 선정한다.

<표 13> 다중회귀 분석 대상 변수 조합

요인	입력 변수	1번 요인 입력 변수	2번 요인 입력 변수	3번 요인 입력 변수
1.요일별 특성치	월요일 여부		최근 2주 동요일 평균 물량	최근 2년 동월 동요일 평균물량
	명절 15일 전부터 3일 후		최근 2주 동요일 평균 물량	휴일 전일 익일 여부
2.최근물동 트렌드	1개월 전 동요일 평균 물량	월요일 여부		3개월 전 동요일 평균 물량
	최근 2주 동요일 평균 물량	월요일 여부		3개월 전 동요일 평균 물량
3.캘린더 적용치	연말효과 여부	명절 15일 전부터 3일 후	1개월 전 동요일 평균 물량	
	최근 2년 동월 동요일 평균물량	월요일 여부	1개월 전 동요일 평균 물량	
	3개월 전 동요일 평균 물량	월요일 여부	1개월 전 동요일 평균 물량	
	2일에서 4일 연속휴일 여부	명절 15일 전부터 3일 후	-	
	간절기	명절 15일 전부터 3일 후	-	
	휴일 전일&익일 여부	명절 15일 전부터 3일 후	-	



이렇게 각 요인에 해당하는 입력 변수별 조합을 수행하게 되면 최종 7개의 변수 조합을 얻어 낼 수 있다. 도출될 수 있는 총 10개의 변수 조합 중 3번 요인의 2일에서 4일 연속휴일 여부, 간절기, 휴일 전일&익일 여부 등 3개의 입력 변수 기준의 조합은 물동량 기준의 입력 변수가 선정되지 못한 관계로 제외하였다.

<표 14> 다중회귀 분석 대상 입력변수

구분	입력 변수					
회귀모형 1	X_1	월요일 여부	X_2	최근 2주 동요일 평균물량	X_3	최근 2년 동월동요일 평균물량
회귀모형 2	X_1	명절 15일전부터 3일후	X_2	최근 2주 동요일 평균물량	X_3	휴일 전일익일 여부
회귀모형 3	X_1	1개월 전 동요일 평균물량	X_2	월요일 여부	X_3	3개월전 동요일 평균 물량
회귀모형 4	X_1	최근 2주 동요일 평균물량	X_2	월요일 여부	X_3	3개월전 동요일 평균 물량
회귀모형 5	X_1	연말효과 여부	X_2	명절 15일전부터 3일후	X_3	1개월전 동요일 평균 물량
회귀모형 6	X_1	최근 2년 동월동요일 평균물량	X_2	월요일 여부	X_3	1개월전 동요일 평균 물량
회귀모형 7	X_1	3개월 전 동요일 평균물량	X_2	월요일 여부	X_3	1개월전 동요일 평균 물량

이론적 배경에서 살펴본 다중회귀모델의 구축방법으로는 입력 변수의 선택방법에 따라 입력선택법(All Possible Selection), 단계적선택법(Stepwise Selection), 전진선택법(Forward Selection), 후진제거법(Backward Elimination)이 있다.



(1) 입력선택법

전체 독립변수를 한 번에 포함시켜 회귀모델을 추정하는 방식이다. 해당 방식을 통해 다른 통제된 독립변수의 관여 없이 특정 독립변수의 영향력을 알 수 있고, 연구자가 고려하고자 하는 모든 독립변수들이 동시에 종속변수를 설명하는 정도를 알 수 있다. 입력선택법을 통해 추정한 결과는 <표 15>와 같다.

<표 15> 입력선택법에 의한 회귀분석 결과

종속변수	독립변수	비표준화 계수		표준화 계수	t	유의확률	VIF	Durbin-Watson	R 제곱
		B	표준오차	베타					
택배물량	(상수)	-38532.218	95482.118		-4.036	.000		1.228	.807
	월요일 여부	-22726.4185	53904.225	-.158	-4.216	.000	5.166		
	최근2주 동요일 평균물량	.882	.028	.803	31.284	.000	2.423		
	최근2년 동월 동요일 평균물량	.324	.047	.269	6.879	.000	5.620		

먼저 월요일 여부를 살펴보면, B값이 -38532.218로 나타났다. 검정통계량 t값을 보면 -4.036이고 유의확률이 0.000으므로 통계적 유의한 음의 영향을 주고 있다. 표준화된 베타의 값이 -0.158이므로 월요일의 경우에는 물동량이 -0.158(15.8%)로 감소함을 알 수 있다. 다음으로 최근 2주 동요일 평균 물량 변수를 보면 B값이 0.882로 나타났다. 검정통계량 t값을 보면 31.284이고 유의확률이 0.000으므로 통계적 유의한 양의 영향을 주고 있다. 표준화된 베타의 값이 0.803이므로 최근 2주 동요일 평균 물량이 1 단위씩 증가하는 경우에는 물



동량이 0.803(80.3%)로 증가하게 된다. 마지막으로 최근 2년 동월 동요일 평균 물량 변수를 보면 B값이 0.324로 나타났다. 검정통계량 t값을 보면 6.879이고 유의확률이 0.000으므로 통계적 유의한 양의 영향을 주고 있다. 표준화된 베타의 값이 0.269이므로 최근 2년 동월 동요일 평균 물량이 1 단위씩 증가하는 경우에는 물동량이 0.269(26.9%)로 증가한다.

입력선택법에 의한 분석결과 택배 물동량(Y) 예측 회귀모델은 다음과 같다.

$$Y = -385322.218 - 227264.185 * X1 + 0.882 * X2 + 0.324 * X3$$

Y : 택배 물동량(수도권)

X1 : 월요일 여부

X2 : 최근 2주 동요일 평균 물량

X3 : 최근 2년 동월 동요일 평균물량

(2) 후진제거법

모든 변수가 포함된 분석모형으로부터 필요하지 않은 독립변수들을 개별적으로 하나씩 제거하면서 반복적 과정을 통해 모형을 단순화해 나가는 방법으로 남은 독립변수를 활용하여 회귀모델을 추정하는 방식이다. 기본적으로 독립변수의 유의확률이 0.05 보다 높은 변수 중 하나씩 후진제거법으로 자동 제거한 결과는 <표 16>과 같다. 먼저 월요일 여부를 살펴보면, B값이 -385322.218로 나타났다. 검정통계량 t값을 보면 -4.036이고 유의확률이 0.000으므로 통계적 유의한 음의 영향을 주고 있다. 표준화된 베타의 값이 -0.158이므로 월요일의 경우에는 물동량이 -0.158(15.8%)로 감소함을 알 수 있다. 다음으로 최근 2주 동요일 평균 물량 변수를 보면 B값이 0.882로 나타났다. 검정통계량 t값을 보면 31.284이고 유의확률이 0.000으므로 통계적 유의한 양의 영향을 주고 있다. 표준화된 베타의 값이 0.803이므로 최근 2주 동요일 평균 물량이 1 단위씩 증가하는 경우에는 물동량이 0.803(80.3%)로 증가하게 된다. 마지막으로 최근 2년



동월 동요일 평균물량 변수를 보면 B값이 0.324로 나타났다. 검정통계량 t값을 보면 6.879이고 유의확률이 0.000으므로 통계적 유의한 양의 영향을 주고 있다. 표준화된 베타의 값이 0.269이므로 최근 2년 동월 동요일 평균 물량이 1 단위 씩 증가하는 경우에는 물동량이 0.269(26.9%)로 증가한다.

<표 16> 후진제거법에 의한 회귀분석 결과

종속 변수	독립 변수	비표준화 계수		표준 화 계수	t	유의 확률	VIF	Durbin -Wats on	R 제 곱
		B	표준오 차	베타					
택배 물량	(상수)	-38532 2.218	95482. 118		-4.036	.000		1.228	.807
	월요일 여부	-22726 4.185	53904. 225	-.158	-4.216	.000	5.166		
	최근2주 동요일 평균 물량	.882	.028	.803	31.284	.000	2.423		
	최근2년 동월 동요일 평균물량	.324	.047	.269	6.879	.000	5.620		

후진제거법에 의한 분석결과 택배 물동량(Y) 예측 회귀모델은 다음과 같다.

$$Y = -385322.218 - 227264.185 * X1 + 0.882 * X2 + 0.324 * X3$$

Y : 택배 물동량(수도권)

X1 : 월요일 여부

X2 : 최근 2주 동요일 평균 물량

X3 : 최근 2년 동월 동요일 평균물량

후진제거법에 의한 분석 결과를 자세히 보면 후진제거법으로 분석을 했음에도 입력선택법과 같은 결과를 보이고 있으며, 회귀모형의 결과 중 진입/제거된 변수 부분을 보면 방법이 입력선택법을 선택하여 진행된 것을 볼 수 있다.



<표 17> 후진제거법의 진입/제거된 변수 결과

모형	진입된 변수	제거된 변수	방법
1	최근2년 동월 동요일 평균물량, 최근2주 동요일 평균 물량, 월요일 여부	.	입력

이는 후진제거법의 방식이 처음에는 모든 변수를 다 투입하는 입력선택법의 방법을 자동으로 선택하여 분석하다가 회귀식에 유의하지 않은 변수가 있으면 하나씩 삭제하며 분석하는 방법을 사용하기 때문에 유의하지 않은 변수가 없는 경우에는 후진제거법의 방법이 나타나지 않기 때문에 생기는 현상으로 볼 수 있다. 즉 모든 독립변수의 유의확률이 의미가 있는 것으로 예측되어 생기는 현상이다.

(3) 전진선택법

전진선택 방식은 단순히 절편만 존재하는 상수항 모형을 기초 모형으로 지정하고 진입 변수들 중에서 가장 유효하다고 판단되는 변수부터 차례대로 모형에 추가하며 회귀모형을 추정하는 방식이다. 전진선택법을 통해 추정한 결과는 <표 18>과 같다. 먼저 최근 2주 동요일 평균 물량 변수를 보면 B값이 0.882로 나타났다. 검정통계량 t값을 보면 31.284이고 유의확률이 0.000이므로 통계적 유의한 양의 영향을 주고 있다. 표준화된 베타의 값이 0.803이므로 최근 2주 동요일 평균 물량이 1 단위씩 증가하는 경우에는 물동량이 0.803(80.3%)로 증가하게 된다.

다음으로 최근 2년 동월 동요일 평균물량 변수를 보면 B값이 0.324로 나타났다. 검정통계량 t값을 보면 6.879이고 유의확률이 0.000이므로 통계적 유의한 양의 영향을 주고 있다.



<표 18> 전진선택법에 의한 회귀분석 결과

종속 변수	독립 변수	비표준화 계수		표준 화 계수	t	유의 확률	VIF	Durbin -Wats on	R 제 곱
		B	표준오 차	베타					
1	(상수)	119319 .453	39544. 595		3.017	.003		1.228	.793
	최근2주 동요일 평균 물량	.979	.019	.891	52.301	.000	1.000		
2	(상수)	-34784 .661	47497. 202		-.732	.464			.802
	최근2주 동요일 평균 물량	.860	.028	.783	30.684	.000	2.338		
	최근2년 동월 동요일 평균 물량	.172	.031	.143	5.604	.000	2.338		
3	(상수)	-38532 2.218	95482. 118		-4.036	.000			.807
	최근2주 동요일 평균 물량	.882	.028	.803	31.284	.000	2.423		
	최근2년 동월 동요일 평균 물량	.324	.047	.269	6.879	.000	5.620		
	월요일 여부	-22726 4.185	53904. 225	-.158	-4.216	.000	5.166		

표준화된 베타의 값이 0.269이므로 최근 2년 동월 동요일 평균 물량이 1 단위씩 증가하는 경우에는 물동량이 0.269(26.9%)로 증가한다. 마지막으로 월요일 여부를 살펴보면, B값이 -227264.185로 나타났다. 검정통계량 t값을 보면 -4.216이고 유의확률이 0.000으므로 통계적 유의한 음의 영향을 주고 있다. 표준화된 베타의 값이 -0.158이므로 월요일의 경우에는 물동량이 -0.158(15.8%)로 감소함을 알 수 있다.



전진선택법에 의한 분석결과 택배 물동량(Y) 예측 회귀모델은 다음과 같다.

$$Y = -385322.218 - 227264.185 * X1 + 0.882 * X2 + 0.324 * X3$$

Y : 택배 물동량(수도권)

X1 : 월요일 여부

X2 : 최근 2주 동요일 평균 물량

X3 : 최근 2년 동월 동요일 평균물량

(4) 단계적선택법

전진선택법과 후진제거법의 단점을 보완하기 위해 고려된 방법으로 모형에 설명변수를 하나씩 추가하면서 그 변수로 인하여 유의하지 않게 되는 것을 제거시키는 과정을 반복하여 회귀모델을 추정하는 방식이다. 더 이상 제거하거나 추가할 변수가 나타나지 않을 때까지 추가와 제거의 과정을 반복적으로 시행한다. 단계적선택법을 통해 추정한 결과는 <표 19>와 같다. 먼저 최근 2주 동요일 평균 물량 변수를 보면 B값이 0.882로 나타났다. 검정통계량 t값을 보면 31.284이고 유의확률이 0.000으므로 통계적 유의한 양의 영향을 주고 있다. 표준화된 베타의 값이 0.803이므로 최근 2주 동요일 평균 물량이 1 단위씩 증가하는 경우에는 물동량이 0.803(80.3%)로 증가하게 된다. 다음으로 최근 2년 동월 동요일 평균물량 변수를 보면 B값이 0.324로 나타났다. 검정통계량 t값을 보면 6.879이고 유의확률이 0.000으므로 통계적 유의한 양의 영향을 주고 있다. 표준화된 베타의 값이 0.269이므로 최근 2년 동월 동요일 평균 물량이 1 단위씩 증가하는 경우에는 물동량이 0.269(26.9%)로 증가한다. 마지막으로 월요일 여부를 살펴보면, B값이 -227264.185로 나타났다. 검정통계량 t값을 보면 -4.216이고 유의확률이 0.000으므로 통계적 유의한 음의 영향을 주고 있다. 표준화된 베타의 값이 -0.158이므로 월요일의 경우에는 물동량이 -0.158(15.8%)로 감소함을 알 수 있다.



<표 19> 단계적 선택법에 의한 회귀분석 결과

종속 변수	독립 변수	비표준화 계수		표준화 계수	t	유의 확률	VIF	Durbin -Wats on	R 제곱
		B	표준오 차	베타					
1	(상수)	119319 .453	39544. 595		3.017	.003		1.228	.793
	최근2주 동요일 평균 물량	.979	.019	.891	52.301	.000	1.000		
2	(상수)	-34784 .661	47497. 202		-.732	.464			.802
	최근2주 동요일 평균 물량	.860	.028	.783	30.684	.000	2.338		
	최근2년 동월 동요일 평균물량	.172	.031	.143	5.604	.000	2.338		
3	(상수)	-38532 2.218	95482. 118		-4.03 6	.000			.807
	최근2주 동요일 평균 물량	.882	.028	.803	31.284	.000	2.423		
	최근2년 동월 동요일 평균물량	.324	.047	.269	6.879	.000	5.620		
	월요일 여부	-22726 4.185	53904. 225	-.158	-4.21 6	.000	5.166		

단계적선택법에 의한 분석결과 택배 물동량(Y) 예측 회귀모델은 다음과 같다.

$$Y = -385322.218 - 227264.185 * X1 + 0.882 * X2 + 0.324 * X3$$

Y : 택배 물동량(수도권)

X1 : 월요일 여부

X2 : 최근 2주 동요일 평균 물량

X3 : 최근 2년 동월 동요일 평균물량

단계적선택법에 의한 분석 결과를 자세히 보면 단계적선택법으로 분석을 했음



에도 전진선택법과 같은 결과를 보이고 있으며, 회귀모형의 결과 중 진입/제거된 변수 부분을 보면 방법이 입력선택법에서 진행된 진입 변수의 순서와 같은 현상을 발견할 수 있다.

<표 20> 진입/제거된 변수

모형	진입된 변수	단계적 선택법	전진선택법
1	최근2주 동요일 평균 물량	단계선택 (기준: 입력할 F의 확률 $\leq .050$, 제거할 F의 확률 $\geq .100$).	전진 (기준: 입력할 F 확률 $\leq .050$)
2	최근2년 동월 동요일 평균물량	단계선택 (기준: 입력할 F의 확률 $\leq .050$, 제거할 F의 확률 $\geq .100$).	전진 (기준: 입력할 F 확률 $\leq .050$)
3	월요일 여부	단계선택 (기준: 입력할 F의 확률 $\leq .050$, 제거할 F의 확률 $\geq .100$).	전진 (기준: 입력할 F 확률 $\leq .050$)

이는 단계적선택법이 전진선택법에서 정하고 있는 입력할 F 확률의 기준이 ≤ 0.050 로 같기 때문에 발생하는 현상으로 파악된다.

4가지의 변수 투입방법별 회귀모델의 분석 결과 입력 변수의 수가 3개로 한정적이기 때문에 분석 결과가 모두 같은 것으로 도출되었다. 이후 진행될 7가지의 회귀모형에 대한 회귀모델 산정 방식은 모든 입력변수를 한 번에 투입하여 회귀모형을 추정하는 방법인 입력선택법을 대표로하여 진행하고자 한다. 이 방법은 강제로 모든 변수를 투입하게 되므로 유의미한 변수와 그렇지 않은 변수에 대한 모든 정보를 산출할 수 있기에 가장 적절한 방법으로 판단된다.



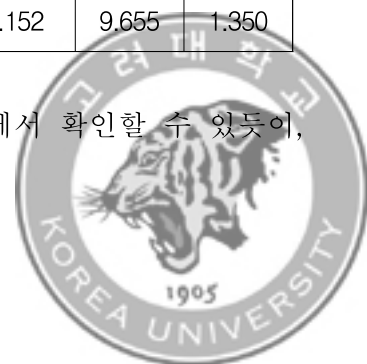
제 2 항 회귀모형별 예측 모델

이미 <표 15>에서 총 7가지의 회귀모형을 구분하였다. 해당 구분에 따른 회귀모형별 분석을 실시하여 통계적으로 가장 유의한 회귀 예측 모델을 구하고자 한다. 회귀모형의 결과를 관찰 시 중요한 요소로 회귀선에 대한 만족감을 표현하는 R제곱값과 독립변수의 수와 표본의 크기를 고려하여 수정된 값인 수정된 R제곱값을 살펴보고 Durbin-Watson 수치를 기준으로 독립적 여부를 판단한다. 또한, 각 변수들 간 상관계수와 유의확률 및 공차와 VIF를 수치를 통해 다중공선성을 확인한다. 공차는 공차한계(tolerance)를 의미하고, VIF는 분산팽창계수(variance inflation factor)를 의미하는데 VIF와 공차는 역수의 관계이므로 VIF 수치를 통해서 공선성을 판단할 예정이다. VIF는 1 ~ ∞ 값을 가지는데 10 미만이면 다중공선성의 문제가 없다고 판단한다.

<표 21> 입력선택법 방식의 회귀모형 결과

모형	회귀모델	수정된 R제곱	유의 확률	VIF	Durbin-Watson
1	$Y_1 = -385322.218 - 227264.185 * X_1 + 0.882 * X_2 + 0.324 * X_3$.806	.000	5.620	1.228
2	$Y_2 = 21683.738 + 234235.928 * X_1 + 0.993 * X_2 + 131620.550 * X_3$.834	.003	1.185	1.561
3	$Y_3 = 90814.281 + 0.451 * X_1 + 50096.159 * X_2 + 0.529 * X_3$.803	.152	9.655	1.350
4	$Y_4 = 90787.559 + 0.547 * X_1 + 32899.612 * X_2 + 0.445 * X_3$.808	.343	10.834	1.317
5	$Y_5 = -25279.291 + 82980.124 * X_1 + 264499.989 * X_2 + 0.986 * X_3$.817	.001	1.053	1.641
6	$Y_6 = -304075.723 + 0.292 * X_1 - 160070.247 * X_2 + 0.851 * X_3$.790	.004	5.750	1.265
7	$Y_7 = 90814.281 + 0.529 * X_1 + 50096.159 * X_2 + 0.451 * X_3$.803	.152	9.655	1.350

<표 21>의 입력선택법 방식을 통한 회귀모형 분석결과에서 확인할 수 있듯이,



회귀모형 2번의 수정된 R제곱값, 즉 수정결정계수의 값이 0.834로 회귀선의 만족감에 대해 83.4%를 설명하고 있으며 Durbin-Watson 수치는 1.561로 2에 가까우므로 독립적이라 볼 수 있다. 또한, 유의확률 및 다중공선성 측면에서도 각각 p값이 0.003과 VIF 값이 1.185로 타당한 것으로 판단된다. 회귀모형 2번이 전반적으로 다른 모형들 보다 통계적 유의성이 크다고 판단되며 택배 물동량 (Y) 예측 회귀모델은 다음과 같다.

$$Y = 21683.738 + 234235.928 * X1 + 0.993 * X2 + 131620.550 * X3$$

Y : 택배 물동량(수도권)

X1 : 월요일 여부

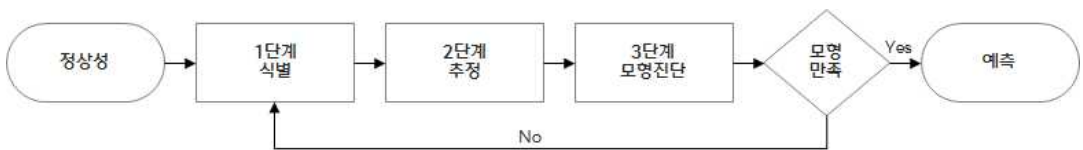
X2 : 최근 2주 동요일 평균 물량

X3 : 최근 2년 동월 동요일 평균물량



제 3 항 시계열분석

Box-Jenkins는 정확도 높은 ARIMA 모형을 구축을 위해 일반적으로 3단계의 과정을 제안하였다. 이 3단계의 ARIMA 모형 구축과정을 <그림 17>과 같이 도표로 제시하였다. 또, ARIMA 모형으로 분석을 할 경우에는 최소 50개 이상의 표본(관측치)이 필요하며, 표본의 수가 많으면 많을수록 예측이 원활하게 잘 이루어지고, 더 정확한 값을 지니게 되는 것으로 설명하고 있다. 즉, 월별 데이터 기준으로 최소 4년 이상, 분기별 데이터 기준으로 최소 12년 이상의 시계열 자료를 구축할 것을 제안하고 있다. 본 연구의 경우 일별 데이터 기준으로 2년 이상의 데이터를 확보하기 있기 때문에 충분한 표본수를 확보했다고 판단하며 <그림 17>의 방식으로 연구를 진행하였다.

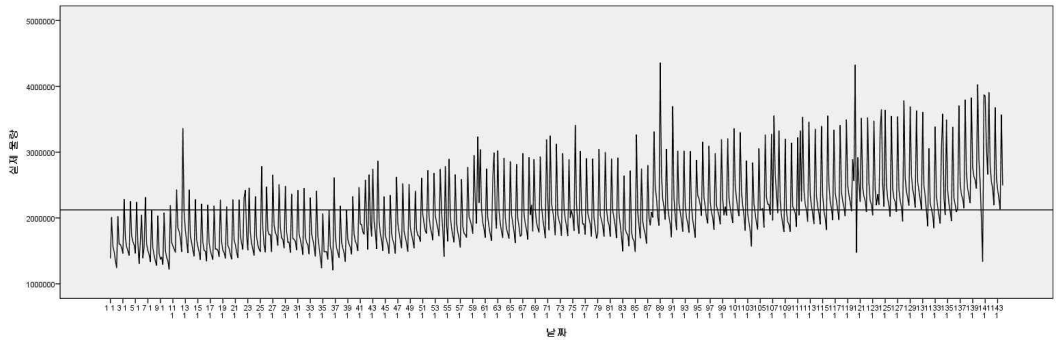


<그림 19> Box-Jenkins의 ARIMA 모델 구축 순서

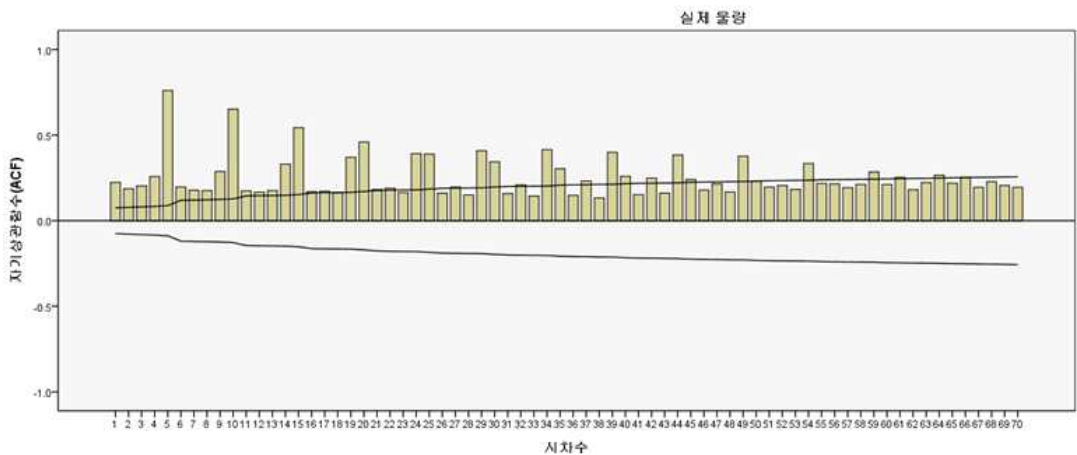
(1) 평균과 분산의 정상성 확인

본 연구에서는 2014년 11월부터 2017년 10월까지의 C사 택배 사업부문의 수도권 지역에 대한 물동량 시계열 자료를 순서도로 표시한 결과 <그림 18>과 같이 나타났다. 원시계열 도표는 시간이 흐름에 따라 추세가 나타나고 있으며, <그림 19>를 통한 원자료의 자기상관의 분석 시 5일 단위의 일정한 주기의 증감이 발견되므로 계절성이 있음을 알 수 있다.





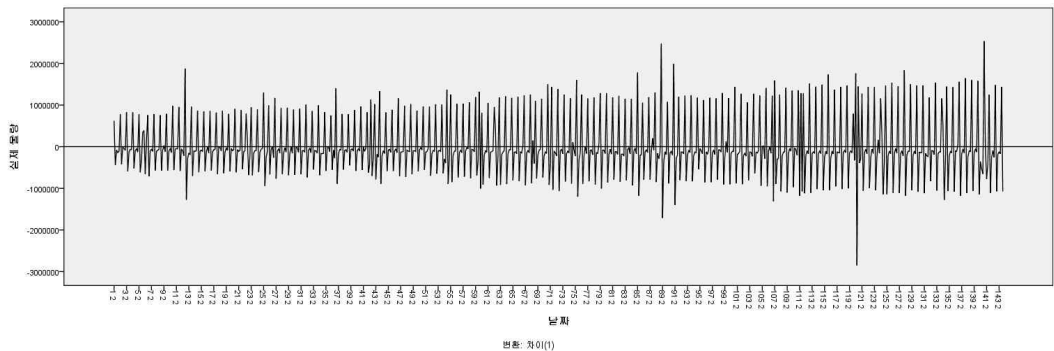
<그림 20> 수도권 택배 일별 물동량의 시계열 순서도



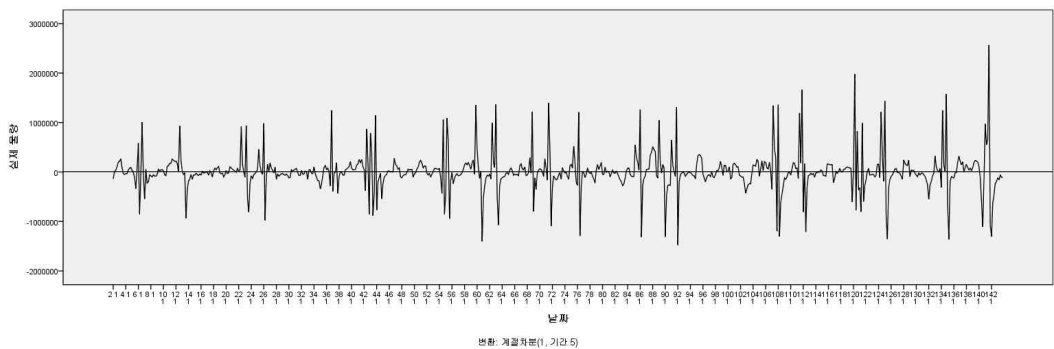
<그림 21> 수도권 택배 일별 물동량의 자기상관함수

또한, 증가적 추세 및 계절성에 의한 비정상성이 나타나므로 차분과 계절 차분을 각각 적용할 필요성이 있다고 판단하고, 차분 및 계절차분을 수행 한 결과가 각각 <그림 20>, <그림 21>과 같이 나타났다.





<그림 22> 수도권 택배 일별 물동량의 추세 차분 적용

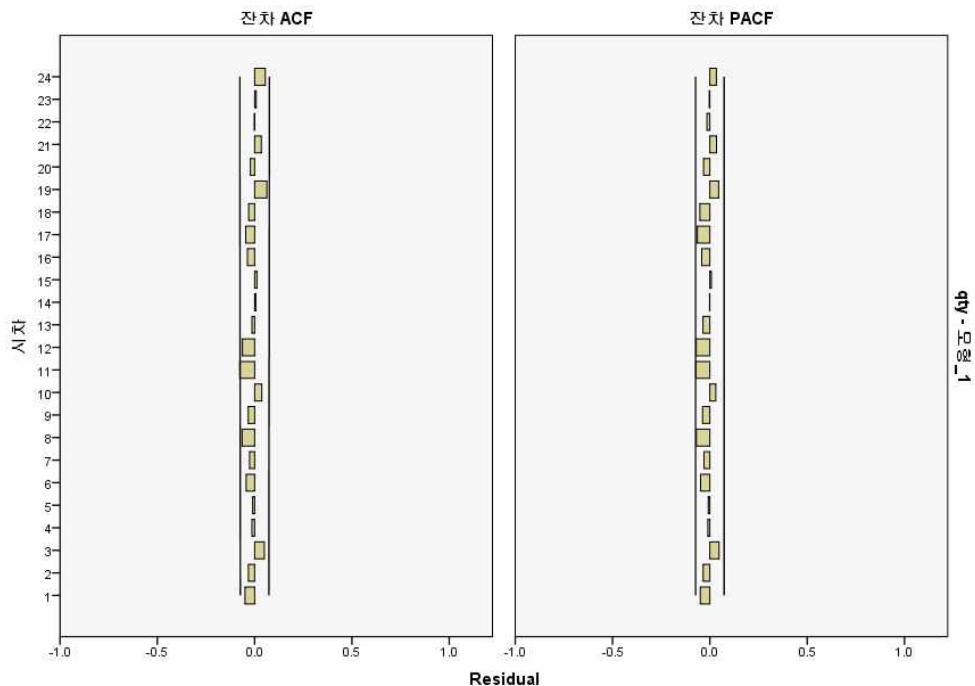


<그림 23> 수도권 택배 일별 물동량의 계절성 차분 적용

(2) 모형 추정 및 검증

최적의 모수를 추정하기 위하여 SPSS의 자동모형생성기를 통해 ARIMA 모형만으로 계절모형을 고려하여 모수를 추정하였으며, 도출된 예측모형으로 ARIMA(1,0,1)(1,0,1)5 모형을 생성한 결과 <그림 22>와 같이 나타났으며, 신뢰한계선 범위 안에 모두 존재하므로 ARIMA(1,0,1)(1,0,1)5 모형을 수도권 택배 물동량 예측모형으로 선정하였다.





<그림 24> 수도권 택배 일별 물동량의 계절성 차분 적용

정상성 및 모형식별, 모수추정의 단계를 거쳐서 최종적으로 선정된 계절 ARIMA 모형은 $ARIMA(1,0,1)(1,0,1)_5$ 모형으로 채택되어졌으며, 해당 모형에 대한 적합성을 분석하기 위해서 <표 22>, <표 23>와 같이 모수 추정치를 비교하였다

우선, 해당 모형은 61.1%의 설명력을 보이고 있으며, 비교적 낮은 BIC와 Q-통계량으로 추측되어지고, 유의확률에서 두 모형 모두 0.05보다 큰 것으로 나왔기 때문에 독립적으로 백색잡음이 존재하므로, 예측모형으로서 적합한 것으로 판단 할 수 있다. 따라서, 채택된 모형은 적합한 것으로 검증되었다.



<표 22> ARIMA(1,0,1)(1,0,1)5 모형 통계량

모형	정상 R제곱	정규화된 BIC	Ljung-Box			이상값 수
			통계량	자유도	유의확률	
(1,0,1)(1,0,1)5	.611	25.630	21.639	14	.086	0

<표 23> ARIMA(1,0,1)(1,0,1)5 모형 모수

모형			추정값	SE	t	유의 확률
(1,0,1)(1,0,1)5	상수항		2190753.268	681207.059	3.216	.001
	AR	시차 1	.999	.003	330.654	.000
	MA	시차 1	.986	.010	101.087	.000
	AR, 계절	시차 1	.787	.035	22.689	.000
	MA, 계절	시차 1	.194	.053	3.655	.000



제 3 절 물동량 예측 모델 평가

지금까지 2014년 11월부터 35개월 간의 정제된 日 단위 물동 데이터를 토대로 日 단위 물동 예측을 위한 회귀모형별 모델을 설정하였다. 이제 요인별 특성이 반영된 회귀모형별 모델에 대한 예측력을 확인하고자 한다. 모델의 예측력을 평가하기 위해서 평균절대편차(MAD; Mean Absolute Deviation), 평균제곱오차(MSE; Mean Squared Error), 평균절대비율오차(MAPE; Mean Absolute Percent Error)를 활용하여 제시한 모형들의 예측력을 평가하고자 한다.

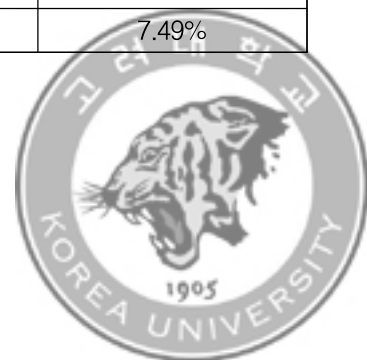
2016년 11월부터 17년 10월까지 1년간의 日 물동량 예측결과를 실제 물동량과 비교하여 평균절대편차(MAD), 평균제곱오차(MSE), 평균절대비율오차(MAPE) 값으로 제시하였다.

(1) 회귀모형 평가

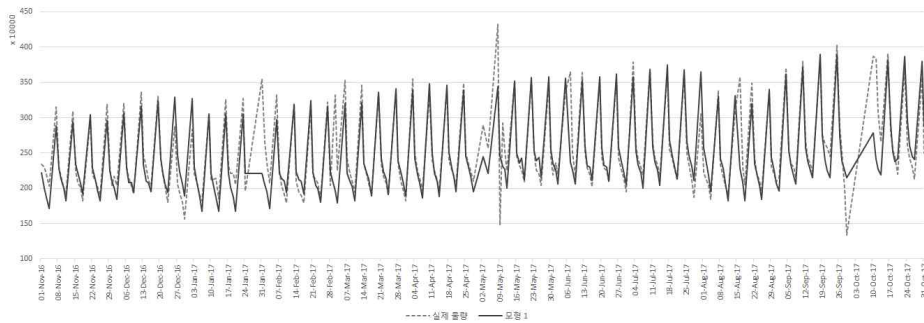
수도권 전체 Sub 터미널에 대한 예측 정확도는 2016년 11월부터 1년간을 기준으로 7개 모든 회귀모델을 보았을 때 92.4% ~ 93.5% 사이에 있다.

<표 24> 입력선택법 방식의 회귀모형 평가

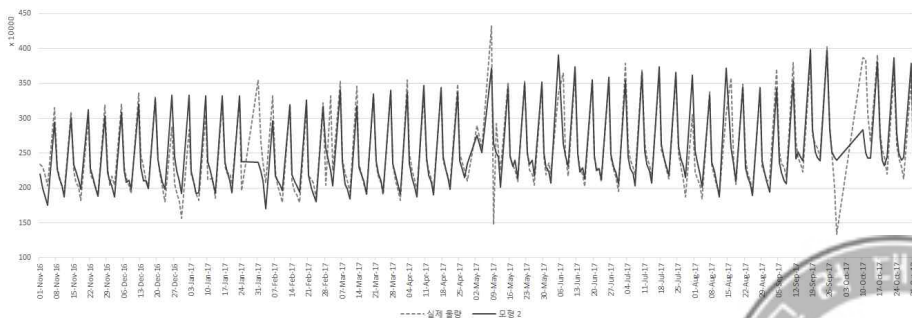
예측 모델	MAD	MSE	MAPE
회귀모델 1	177567	85414331797	7.09%
회귀모델 2	160315	70968125455	6.53%
회귀모델 3	188259	85662147912	7.49%
회귀모델 4	173731	84339906349	6.94%
회귀모델 5	191090	75755985377	7.63%
회귀모델 6	190855	84117823702	7.50%
회귀모델 7	188259	85662147912	7.49%



<그림 23> ~ <그림 28>에서 보듯이 요일별 특성은 제대로 반영이 되지만 명절 및 어린이날 특성이 제대로 반영되지 않음을 확인할 수 있다. 해당 회귀모델에 반영된 입력변수가 요일별 특성을 반영하는 월요일 여부의 영향이 크게 반영은 되었지만 캘린더 적용치 요인인 휴일 관련 변수와 장기간에 걸친 동요일 평균 물량의 변수가 어느 정도 영향을 미친 것으로 판단한다. 전체 예측치는 90%는 상회하여 단 3개의 입력변수를 사용한 모델로서의 유의적 측면에서는 비교적 양호하나 실제 Sub 터미널별 물량 예측이 운영 자원과 직접적인 영향도가 높기 때문에 좀 더 신중한 접근이 필요하고 회귀모형을 통한 예측 모델은 많은 변수를 넣어서 모델링하여 적중률을 높이는 것도 중요한 접근으로 예상된다.

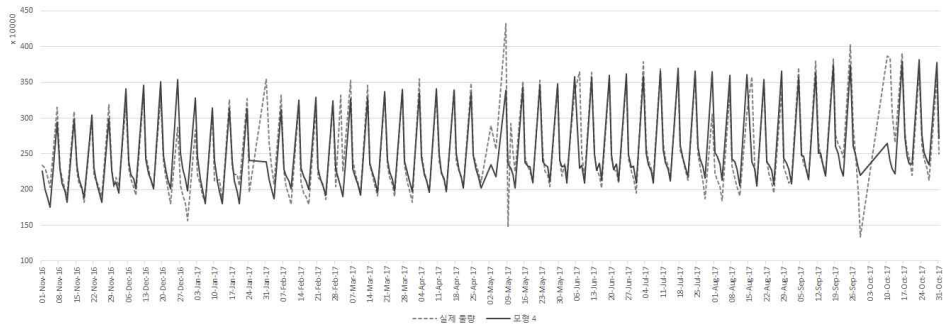


<그림 25> 회귀모델 1 추정결과

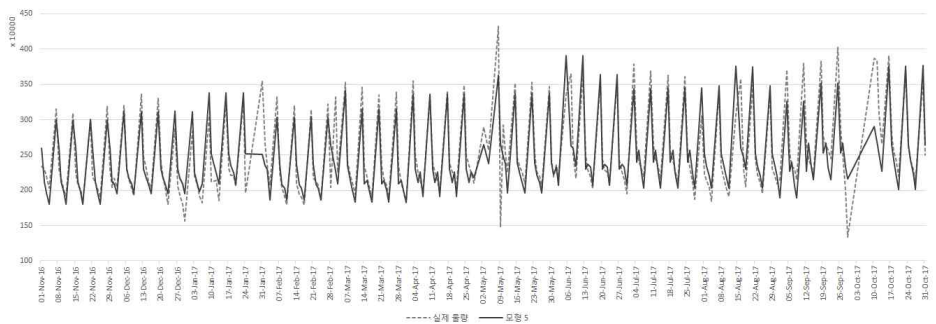


<그림 26> 회귀모델 2 추정결과

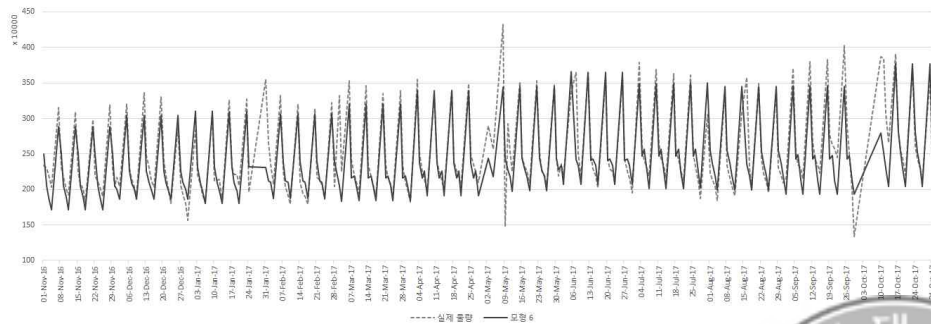




<그림 27> 회귀모델 3 추정결과

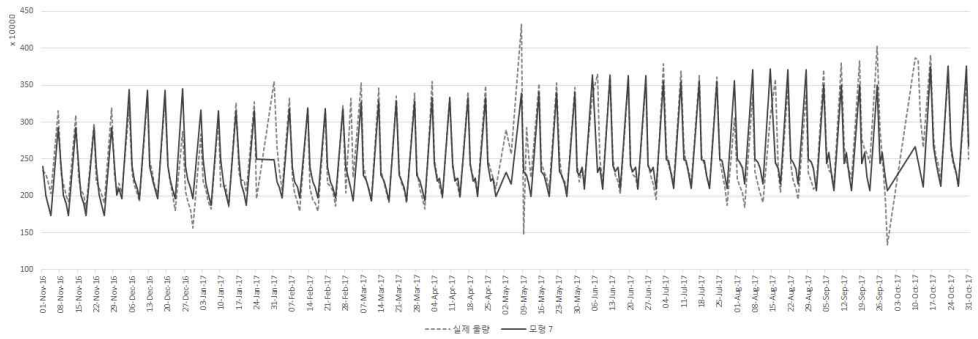


<그림 28> 회귀모델 4 추정결과

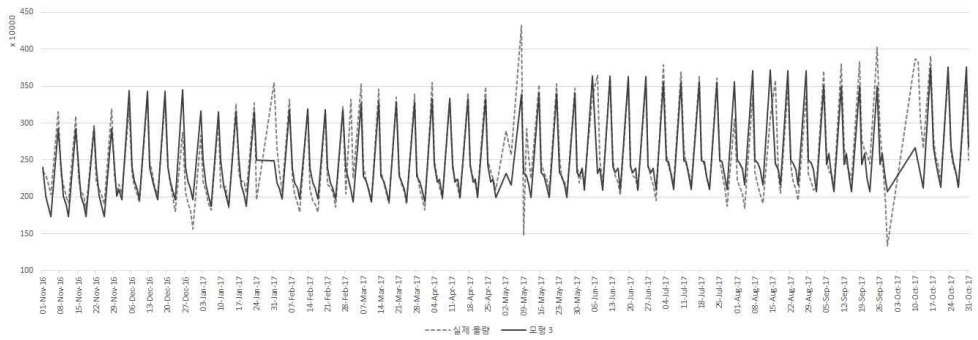


<그림 29> 회귀모델 5 추정결과





<그림 30> 회귀모델 6 추정결과



<그림 31> 회귀모델 7 추정결과

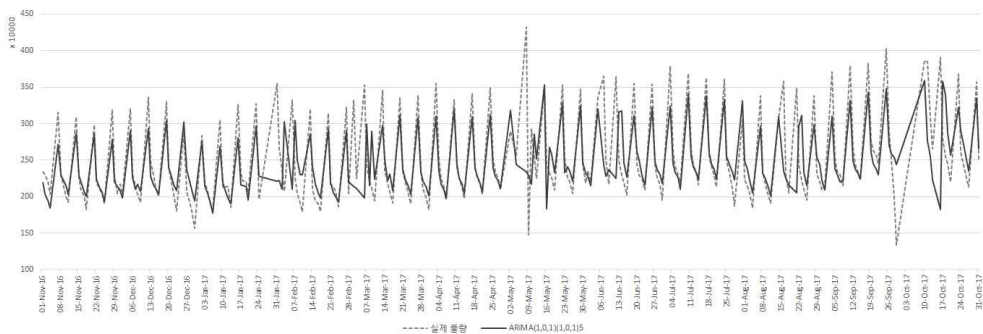


(2) 시계열모형 평가

수도권 전체 Sub 터미널에 대한 예측 정확도는 2016년 11월부터 1년간을 기준으로 시계열모형을 보았을 때 89.7%로 회귀모델 보다는 다소 낮은 것으로 평가되었다.

<표 25> 시계열모형 평가

예측 모델	MAD	MSE	MAPE
ARIMA(1,0,1)(1,0,1)5	268753	178026168543	10.28%



<그림 32> 시계열모형 추정결과



제 5 장 결 론

제 1 절 연구 결과 요약

물류 서비스는 Lastmile 배송과 같은 서비스 품질, 가격, 박스 형태의 취급 단위와 포장 등에서 차별화가 어렵기 때문에 물류를 본업으로 하는 기업들은 시장에서 우위를 확보하기 위한 경쟁력을 만들어 가는데 어려움을 가지고 있다. 이런 환경 하에서 물류 서비스가 가지고 있는 물동적 특성을 파악하고 이를 반영한 물동량 예측방법을 통해 좀 더 정확하고 효율적인 물동 예측 결과를 도출할 수 있다면, 경영상 낭비되고 있는 불필요한 비용지출을 원천적으로 제거시켜 수익구조 면에서 기업 경쟁력을 극대화 할 수 있을 것이고 이를 통해 차별화 포인트를 만들어 낼 수 있을 것이다. 이에 미래에 대한 물류 물동 예측의 필요성이 강조되고 있지만 원천 데이터의 확보, 분석 복잡도에 따른 어려움으로 이에 대한 연구는 미흡하게 진행되고 있는 실정이다.

본 연구는 택배 시장이 확대되고 있는 추세에 맞추어 택배 물동량의 효과적인 분석을 통해 미래를 예측하고 운영 자원의 선제적인 배치를 통해 질적 성장을 도모하고자 본 연구를 수행하고자 하였다. 먼저 선행연구를 검토를 기반으로 연구모형의 구조적 타당성을 검증하기 위해 관련 업계 전문가 인터뷰 및 설문 을 통한 실증적 요인분석을 실시하였고, 年 단위의 장기적인 물동량 분석에서 한 걸음 나아가 日 단위 물동량 분석을 효과적으로 수행할 수 있는 모델을 발굴하여 물동량 분석의 최적화 및 이를 기반으로 한 운영 효율화를 이루고자 하였다. 결과는 다음과 같다.

첫째, 물동량에 영향을 미치는 조건변수들에 대한 체계를 정하고자 전문가 인터뷰 및 선행연구를 통해 주요 변인 10개를 추출하였으며 설문에 따른 요인분



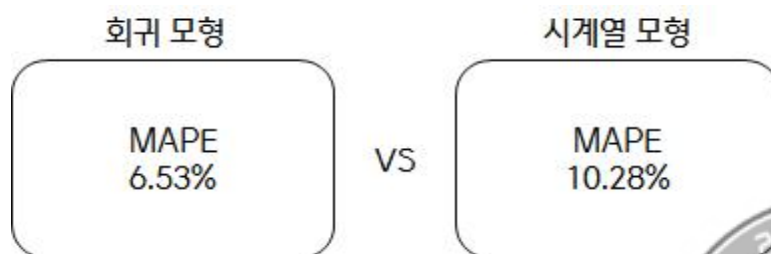
식을 통해 주요 영향요인을 도출하였다. 최종적으로 日 단위의 단기 택배 물동량 예측을 위해서는 요일별 특성치와 최근 물동 트렌드, 캘린더 적용치의 3가지 주요 영향요인이 가장 중요한 요인으로 판별되었다.

둘째, 주요 영향요인 기반의 입력 변수를 설정하고 변수 간 상관관계를 고려하여 총 7가지 회귀분석모형을 예측 모델을 도출하였다. 7가지의 회귀분석모형의 분석 결과 ‘X1: 월요일 여부’, ‘X2: 최근 2주 동요일 평균 물량’, ‘X3: 최근 2년 동월 동요일 평균물량’의 3가지 독립변수를 가진 회귀모형이 가장 평가치가 좋은 것으로 분석되었다. 최종 선택된 회귀모형은 아래 식과 같다.

$$Y = 21683.738 + 234235.928 * X1 + 0.993 * X2 + 131620.550 * X3$$

셋째, 물동량 분석을 통하여 물동량 예측모델의 효과를 실증적으로 규명하였다. 2014년 11월부터 35개월 간 데이터 기준으로 예측 모델을 평가한 결과 일평균 92% 이상의 적중률로 다소 미흡하지만 어느 정도는 신뢰할 수 있는 결과를 취득하였다. 이는 단 3개의 입력변수의 조합으로 만들어낸 결과이기에 좀 더 많은 물동량 영향 요인과 관련 변수를 도출한다면 좀 더 높은 적중률의 결과를 만들어 낼 수 있을 것이다.

넷째, 회귀분석모형을 통한 예측 모델이외에 ARIMA를 통한 시계열 분석을 실시하였다. 최종적인 ARIMA 모형은 ARIMA(1,0,1)(1,0,1)5 모형으로 선정되었으며 해당 모형과 회귀분석모형 간 예측 결과 <그림 30>과 같이 회귀분석모형을 통한 예측 결과가 좀 더 우세한 것으로 분석되었다.



<그림 33> 분석 모형별 예측결과 비교



다섯째, 보통 요일별 특성치와 명절에 관련된 캘린더 특성치가 물동량 예측에 가장 많은 영향을 끼치는 것이 정설이다. 이에 요일별 특성치, 특히 월요일에 대한 특성치가 가장 많은 영향도를 보이고 있다는 것을 확인하였다.

여섯째, 최근 물동 트렌드 변수와 휴일 특성치의 반영만으로도 충분한 예측이 가능하고 최근 1개월 이전의 물동 트렌드 변수가 과거 1년 이상 지난 물동 트렌드 보다 적합도가 큰 것으로 분석되었다.

제 2 절 한계점 및 향후 과제

본 연구는 택배 日 단위 물동량 분석을 이해하는데 이론적·실무적 시사점을 제공하였으나 향후 추가로 필요한 연구 과제와 관련하여 다음과 같이 한계점과 개선사항을 제안하고자 한다.

첫째, 본 연구는 C사의 데이터를 기초로 연구되었는데, 토요일, 일요일에 해당되는 주말 공휴일에 대한 데이터를 고려되지 않았으나 일부 실제 운영상에는 해당 일에도 물동량이 발생한 것으로 나왔다. 이는 Sub 터미널별 특성치가 반영되지 않은 문제가 발생한 것인데 이는 수도권 전체를 군집하여 진행하였기 때문에 해당 데이터가 무시된 상태에서 연구가 진행되었다. 추후에는 주말 휴일의 특성치가 반영된 연구가 진행되어야 할 것이다.

둘째, 예측 모델에 대한 추가 정의가 필요하다. 본 연구는 회귀분석을 통해 예측을 위한 모델링이 진행되었는데, 이는 한가지의 모델이 전체를 대변할 수 있다는 가정으로 수행된 것이다. 실질적인 물동량 예측의 영향도를 보았을 때 주 단위의 특성이 강한 특징상 요일별 예측 모델, 계절별 예측 모델 등 다양한 특성치 기반의 모델링을 수행할 수 있을 것으로 보인다. 특히 수도권 전체 Sub



터미널에 대한 군집화를 기준으로 수행된 모델링이기 때문에 실제 운영을 위해서는 각 Sub 터미널 단위로 예측이 되어야 한다. 지역별 Sub 터미널이 가지고 있는 고객과 지역적 특성치를 고려하여 추가적인 연구가 필요하다.

셋째, 본 연구는 수도권 물량이 전체의 3/4을 차지하고 있어 전체를 대변할 수 있다는 가정하에 수도권만을 기준으로 진행되었다. 특히나 수도권도 서울과 경기도 등의 좀 더 상세한 구분을 하고 다른 지역을 추가하여 지역적 물동량의 특성을 파악할 수 있을 것으로 보인다. 지역적으로 물동량에 많은 영향도를 미치는 요인에 대한 분석과 실증 연구가 필요할 것으로 판단된다.

본 연구는 물동량의 예측을 日 단위 예측이 활성화되기 전에 연구를 진행하였다는데 큰 의의가 있다. 앞으로 물동량의 단기 예측을 수행하는데 이론적으로나 실무적으로 유의미한 시사점을 제공했기를 기대해 본다. 또한, 본 연구를 통해 물동량 분석에 관련된 연구들이 활성화되기를 기대하며 이를 위한 가이드의 역할을 할 수 있기를 기대해 본다. 물동량 예측을 단순 운영 자원 최적화에만 집중하지 말고 센터간의 간선 차량에 대한 체적과 연결시킨다면 W&D 내 시너지 부분은 급격히 상승할 것으로 예상된다. 앞으로는 이러한 많은 연구를 통해 점진적인 Low Cost 운영 구조를 확립하고 Operation Excellence를 구현하여 궁극적으로는 Cost Leadership을 만들어 낼 수 있는 기반을 마련하였으면 한다.



참고문헌

- [1] 김권일. (2017). “우리나라 항공수출 화물 운임 결정요인 분석 및 예측에 관한 연구”, 인하대학교 석사학위논문, pp.25-33
- [2] 김영민(2007). “우체국택배의 경쟁력 강화방안에 관한 연구”, 물류학회지, Vol.17, pp.183-203.
- [3] 김윤식(2010). “사업초기단계의 공사비 예측을 위한 CBR기반 MRA 보정모델”, 서울시립대 석사학위논문, pp.18-20
- [4] 김은령(2010). “변수선택과 모형집합성을 위한 능형회귀와 LASSO기법에 관한 연구”, 동국대학교
- [5] 김의창, 최현탁(1999). “집중국별 우편물량 예측에 따른 차량 할당”, 한국 정보시스템학회, 춘계학술대회 발표논문집, pp.252-259
- [6] 김상원(2011). “윈터스 지수평활법을 이용한 제주도 관광객 수요 예측”, 한국지역경제연구(제19집)
- [7] 김 신(2017). “다중선형 회귀분석을 이용한 현혈횟수 예측에 관한 연구”, 성균관대학교 석사학위논문
- [8] 김재호. (2014). “다중회귀분석 기반 지하도상가 시설물의 유지관리비 예측”, 서울시립대학교 석사학위논문
- [9] 김태영, 황승준(2010). “중장기 우편물류를 위한 수요예측 시스템 개발”, 한국 SCM 학회지, 제 10권 제2호, pp.73-80
- [10] 민경창, 전영인, 하현구(2013). “계절성 ARIMA 모형을 이용한 항공화물 수요예측”, 대한교통학회지, 제 31권 제3호, pp.3-18
- [11] 박귀환(2006). “택배시장의 문제점과 현황”, 한국유통과학회 학술대회, Vol.2006, pp.805-819
- [12] 박득희, 이계희, 이민정(2015). “방한 중국 관광객 수요 예측모델의 비교



- 연구”, 관광연구저널, Vol.29, pp.29-40
- [13] 박중권, 이명규(2008). “우체국택배 사업의 수익 다변화 전략”, 우정정보, 2008 권 4호, 정보통신정책연구원, pp.23-49
- [14] 박찬석(2006). “택배 시장의 동향과 시사점: 우체국택배에 미치는 영향”, 우정정보, 제 66권 제3호, pp.1-22
- [15] 방지영(2015). “음력효과를 반영한 물류 서비스 산업의 택배 물동량 수요 예측”, 한양대학교 대학원 석사학위논문
- [16] 서선애, 오가영(2012). “산업구조와 수출입 상품 특성이 항만물동량에 미치는 영향”
- [17] 성혁제 외(2002). “데이터 분석을 위한 통계기법 및 SPSS활용”, 동일 출판사
- [18] 손은호, 서진우, 정명보(2005). “ARIMA 모형을 이용한 호텔식 음료 매출액의 예측: 경주지역 특급호텔을 중심으로”, 관광·레저연구, pp.117-132
- [19] 신현규, 김홍렬, 이준재(2014). “대전지역 호텔객실에 대한 시계열 예측방법에 관한 연구”, 관광연구저널, Vol.28, pp.49-58
- [20] 우주리(2007). “시계열 모형을 이용한 방한 일본인 관광객 수요예측에 관한 연구”, 경희대학교 석사학위논문
- [21] 이강대(2001). “화물품목별 지역간 발생예측모형 개발에 관한 연구”, 로지스틱스연구, 제 9권 제1호, pp.5-26
- [22] 이금숙, 백호종(1990). “화물 수송 수요 예측을 위한 혼합 회귀 모형”, 응용지리, 제 13권, pp.35-63
- [23] 이덕주, 김기홍, 안재경, 김종화, 이성준(2005). “예측 및 시계열 분석: 시장환경 변화를 고려한 국내 다량우편물 물량예측”, 대한산업공학회, 춘계학술발표논문집, pp.784-793
- [24] 이성준, 장태우, 남윤석, 오형식(2003). “우편물량에 대한 예측: 로지스틱



- 곡선을 중심으로”, 대한산업공학회, 추계학술대회 논문집, pp.308-312
- [25] 이지홍(2017). “빅데이터 분석을 통한 물동량 분석 예측 모델에 관한 연구”, 중앙대학교 석사학위논문, p.2
- [26] 이충기 (2003). “관광응용경제학”, 일신사, 서울, pp.27-28/p.52/p.105
- [27] 이충기, 송학준(2007). “최적 시계열 수요예측 모델선택에 관한 연구”, 관광학연구, pp. 289-311
- [28] 정성환, 강경우(2015). “계량경제모형간 국내 총화물물동량 예측정확도 비교 연구”, 大韓交通學會誌, Vol.33, pp.61-69
- [29] Archer. Brian(1987). “Demand Forecasting and Estimation in Travel”, Tourism and Hospitality Research. John Wiley & Sons.
- [30] Bell, W. R., Hillmer, S. C.(2002), “Issues Involved with the Seasonal Adjustment of Economic Time Series”, Journal of Business & Economic Statistics, Vol. 20(1), pp.98-127
- [31] Chambers, J. C., Mullick, S. K., Smith, D. D.(1971). “How to Choose Right Forecasting Technique”, Harvard business review, Vol. 49(4), pp.45-74
- [32] Garrido, R. A., Mahmassani, H. S.(1998). “Forecasting Short-Term Freight Transportation Demand: Poisson STARMA Model”, Transportation Research Record: Journal of the Transportation Research Board, Vol. 1645(1), pp.8-16
- [33] Hansen, J. V., Nelson, R. D.(2003), “Forecasting and Recombining Time-Series Components by using Neural Networks”, Journal of the Operational Research Society, Vol. 54(3), pp.307-317
- [34] Ling, Z., Heimann, B., Clausen, U(2006). “Forecasting a Logistic Service Demand Based on Neural Network”, In Service Systems and Service



- Management, 2006 International Conference on, Vol. 1, pp.530-534
- [35] Makridakis, S., Wheelwright, S. C., Hyndman, R. J.(1998), “Forecasting: Methods and Applications”, New York: John Wiley & Sons Inc
- [37] Miao, X., Xi, B.(2008), “Agile Forecasting of Dynamic Logistics Demand”, Transport, Vol. 23(1), pp.26-30
- [38] Zhang, G. P., Qi, M.(2005). “Neural Network Forecasting for Seasonal and Trend Time Series”, European journal of operational research, Vol. 160(2), pp.501-514
- [38] Zhang, J., Zhu, J.F.(2009). “Logistics Amount Forecasting Based on Combined ARIMA and ANN Model”, In Grey Systems and Intelligent Services, GSIS 2009. IEEE International Conference on, pp.594-597



Abstract

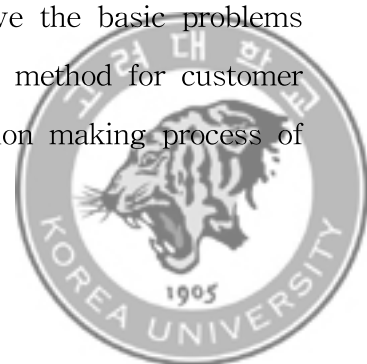
A Study on Factors Influencing Forecasting of Short-term Trade Volume in Parcel Service Industry

Nahm, Kee Chan

Department of Management of Technology
Graduate School of Management of Technology
Korea University

In today's, the necessity of forecasting for big customer service is more emphasized due to the growth of logistics service market. The existing logistics service providers should possess the differentiated competitiveness in order to maintain and solid the market share; however, it is difficult to provide the differentiated quality from competitors through a simplified process of consolidation, sort and deliver.

In this research, the high-accuracy forecasting model based on the daily demand forecast is proposed through the analysis of the factors, influencing the forecast of freight volume by examining in details of demand forecast around the metropolitan area among the key delivery points of the largest logistics company C in Korea. Ultimately, it is to solve the basic problems in logistics and propose the freight volume forecasting method for customer trends and resource planning; moreover, helps in decision making process of



various corporate activities such as logistics resource management, strategy, and marketing in order to contribute to long-term profit improvement.

In the case of Company C, clustering is performed by grouping the local locations and daily sub-terminal without reflecting the forecast model to each terminal unit. The final result is mainly to select the high-accuracy forecasting model based on simulation results of regression forecasting model by deciding the main criterion and analyzing variables with the precedent study and interviews.

This research has highly significant insights in the theoretical and practical point of view, and in the fact that the research has been carried out before the daily freight volume forecast is activated.

key word : Forecasting, Factor Analysis, Time Series Analysis, Multiple Regression Analysis, ARIMA Model, Mean Absolute Deviation, Mean Absolute Percentage Error, Mean Square Error



남 기 찬 의 技術經營學碩士學位論文

審査를 完了함.

2017年 12月 16日

委員長 김 영 준



委員 윤 지 환



委員 조 광 희

