

국내 코로나 확진자 데이터 분석을 통한 지역 간 감염 전파 시뮬레이션 연구*

이성호

서울과학기술대학교 ITM 학사 전공
(sean03101@seoultech.ac.kr)

정원렬

서울과학기술대학교 ITM 학사 전공
(jwr0218@seoultech.ac.kr)

황재훈

서울과학기술대학교 ITM 학사 전공
(mgs05144@seoultech.ac.kr)

심민규(교신저자)

서울과학기술대학교 산업공학과 조교수
(mksim@seoultech.ac.kr)

COVID-19과 같은 바이러스 전염병의 확산은 보균자의 지역 간 이동을 통해서 일어나곤 한다. 본 연구는 시뮬레이션 연구를 통해 2019년에 비해서 큰 폭으로 감소한 2020년의 대중교통을 이용한 지역간의 이동이 전염병의 확산에 어떤 영향을 미쳤는지 살펴본다. 예비 연구의 단계에서는 질병관리청에서 발표한 감염자-피감염자의 쌍에 대한 정보를 바탕으로 트리구조의 감염 계보를 생성하고, 한 명의 보균자가 확률적으로 몇 명의 피감염자를 발생시키는가에 대한 확률 분포가 두 개의 확률 분포의 중첩의 형태로 되어있음을 확인한다. 주 연구인 시뮬레이션 연구에서는 한국철도공사에서 제공하는 여객 승하차 데이터를 이용해 서울-부산간 대중 교통 이용량을 추정하고, 포아송 과정을 이용한 시뮬레이션 연구를 수행하였다. 이를 통해 2019년에 비해서 약 3배가 줄어든 2020년의 대중교통 이동량이 전염 수치에 대해 큰 차이를 가져오는 것을 확인할 수 있다. 이는 당국의 이동 자제 권고와 국민들의 수준높은 협조가 전염병의 확산을 저지하는데 큰 역할을 하고 있음을 의미한다.

주제어: 코로나 바이러스, 인구통계학, 극한 수치 이론, 와이블 분포, 철도 이동, 시뮬레이션

1. 서론

1.1 연구 배경

사회적 거리 두기는 약물적 처리가 불가능한 시기

에 전염성 바이러스에 대처하기 위해 인류가 오래전부터 사용해온 방법이다. 13세기의 이탈리아 도시 피스토이아에서는 흑사병을 막기 위해 보건 위생법을 제정하고 다른 도시로부터 사람과 물품의 출입을 통제하였다. 20세기 초의 스페인 독감에 대처하기 위해 미국의 도시 세인트 루이스에서는 공공기관 폐

논문투고일: 2021. 3. 30 논문수정일: 2021. 5. 11 게재확정일: 2021. 5. 12

* 이 연구는 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다.

연구목적으로 서울-부산간 발권 정보를 제공해 주신 국토교통부 철도운영 안전과에 감사드립니다.



© 한국생산관리학회. 이 저작물은 Creative Commons 저작자표시-비영리 4.0 국제 라이선스에 따라 이용할 수 있습니다. (<https://creativecommons.org/licenses/by-nc/4.0/deed.ko>)

쇄, 전차 탑승 금지등의 조치를 취하였고, 이와 대조적으로 비슷한 규모의 도시인 필라델피아에서는 약 20여만명의 시민이 참여하여 군인들을 축하하는 퍼레이드를 벌였다. 그 결과 두 도시에서의 사망주 수는 큰 차이를 보였다(Hatchett 등, 2007).

2020년에 전 세계적으로 유행한 COVID-19에 대한 대처를 위해 각 국가에서는 다양한 방식의 사회적 거리 두기를 시행하고 있다. 이들 조치는 크게 이동 봉쇄 수준의 강력한 수준의 조치와 마스크 착용 권고 등의 일상 수준의 조치로 구분할 수 있다. 봉쇄 수준의 강력한 이동제한 조치는 매우 명백한 효과를 보인다. 일례로써 2020년 미국의 Spring break 기간인 3월과 4월에는 휴양지인 플로리다주로 여행 온 대학생들로 인해 3만 7천 명의 확진자와 2천여 명의 사망자가 발생하였는데,¹⁾ 이에 대응하여 여러 주정부들에서 봉쇄 조치를 실시하며 증가 추세를 효과적으로 완화시켰다. 같은해 8월 가을학기 개강을 위한 봉쇄조치 해제로 다시 대다수의 확진자가 양산(조선일보(2020))된 것은 봉쇄 조치의 효과성을 다시 한번 입증하는 사례이다.

우리나라에서는 봉쇄 수준의 강력한 조치가 아닌 일상 생활을 유지하는 가운데에서 방역을 추구하고 있으며, 어느 정도 수준의 조치가 적절한지는 지속적인 사회적 관심사이다. 확진자의 지역간 이동은 많은 이들이 우려하는 상황이다. 2020년 9월에는 확진자 1인이 격리지침을 어기고 서울에서 대구로 열차로 이동한 사례가 적발되어 사회적으로 큰 우려를 낳았다(jtbc뉴스, 2020). 다행히 이 사건에서는 같은 칸에 탑승한 승객들의 마스크 착용과 당국의 빠른 대응으로 타지역으로의 전파를 막을 수 있었다. 우리나라에서 COVID-19은 높은 자가 방역과 함께

인구 이동량 감소라는 사회적인 현상을 낳았다. 본 연구에서 사용하는 철도를 통한 도시 간 이동에 데이터에서는 2019년에 비해서 2020년에 약 20%로 감소한 것으로 나타났다.

본 연구는 COVID-19의 감염에 대한 데이터와 대중교통을 이용한 도시 간 이동에 대한 데이터를 바탕으로 시뮬레이션 연구를 진행하여, 높은 시민의식과 감소한 인구 이동량의 현재 상황이 감염병의 확산 방지에 충분한지에 대해 조사한다.

1.2 문헌 연구

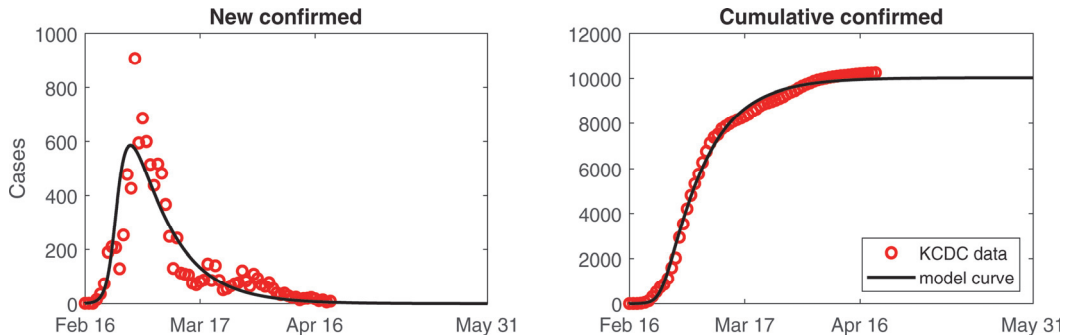
시뮬레이션 기법은 전염병 확산의 예측 혹은 방지에 관한 연구에서 널리 사용되는 방법으로, 다양한 바이러스들의 전염에 관한 연구에서 사용되어왔다. 인간의 건강을 위협하는 H1N1 바이러스 연구(Lee 등(2010))와 HIV 바이러스 연구(Ayton와 Voth(2010)) 뿐만 아니라 동물(Innocent 등(1997))과 식물(Frazer(1977))에 관한 바이러스 전파 역시 시뮬레이션 기법을 이용하여 연구 되었다.

COVID-19에 관해서도 감염병 예측 모델을 이용한 시뮬레이션 연구들이 수행되었다. 우리나라의 2020년 초의 1차 대유행 시기로부터의 진정 시점에 대한 예측 모형의 연구에서는 다음 <그림 1>과 같이 상당히 정확한 예측 결과를 제시하였다(Kim 등, 2020). 본 연구 역시 데이터를 활용해 수리적 모형을 구축하고 지역 간 인구 이동량을 바탕으로 바이러스 확산량에 관한 시뮬레이션 연구를 진행한다.

시계열 예측 기법을 활용한 방식과 함께, 인구의 이동과 같은 지역적인 특성을 고려한 인구의 이동으로 인한 바이러스 전파에 관한 연구도 진행되고 있다.

1) www.coronavirus.jhu.edu/data/state-timeline/new-confirmed-cases/florida

〈그림 1〉 COVID-19 확산추이 및 예측에 대한 수학적 모델과 실제 데이터 (Kim 등, 2020)



대표적으로, 중국에서의 확진자-접촉자에 관한 데이터를 이용해 고속열차 승객에 대한 전파 위험을 정량화한 연구(Hu 등(2020))가 수행되었으며, 2020년 2월 이후의 대한민국에서 서울 지하철 이용량에 대한 데이터를 바탕으로 고위험군에 속하는 노인들과 그렇지 않은 직장인들이 사회적 거리 두기에 대한 실천 방식이 달랐음을 실증하는 연구도 이루어졌다(Park, 2020).

감염자의 비율이 서로 다른 두 지역 간 인구 이동이 일어나는 경우는 많은 사람들이 우려하는 상황이다. 왜냐하면 국내 여행을 통한 지역 간 이동은 자주 발생할 뿐만 아니라, 여행자와 여행지 사람들과의 접촉으로 인해 대유행을 유발할 수 있기 때문이다(배진과 서용원, 2019). 인구 이동으로 인한 감염 확산은 두 가지 경우로 요약할 수 있다. 하나는 이동 과정에서의 전파이고, 다른 하나는 이동 후의 새로운 지역에서의 전파이다. 다행히 연구 배경에서 언급한 바와 같이 우리나라의 경우는 국민들의 높은 시민 의식으로 교통 수단의 이용 과정에서 전파가 일어난 경우는 알려진 바가 없다. 기차, 버스, 항공기에서는 COVID-19에 대한 대응으로 반드시 마스크를 착용하게 하고 있으며, 이는 열차와 같은 밀폐된 공간에서 마스크를 쓴다면 바이러스의 전파를 효과

적으로 막을 수 있다는 연구 결과에 바탕하고 있다(Chu, 2020).

그러나 이동 후의 새로운 지역에서의 전파 가능성은 여전히 지역 간의 전파 위험에 대한 우려를 낳는다. 본 연구에서는 이에 따라 대중교통 이용 중에서의 감염이 일어나지 않는다고 가정하는 한편, 대중교통을 이용해 새로운 곳으로 이동한 보균자가 지역 간 전파에 어떤 영향을 미칠 수 있는지에 대한 시뮬레이션 연구를 수행한다. 그런 뒤, 안전 인식의 증용은 안전도와 양의 상관관계를 가진다는 점을 증명하였듯(조정래와 이동원(2019)), 이동량 제한을 통한 사회적 거리 두기 또한 감염률 억제를 위한 증용이 필요한가에 대해 논의해 볼 것이다.

1.3 데이터 셋

대한민국 질병 관리청(Korea Centers for Disease Control & Prevention, KCDC)에서 발표한 확진자와 확진 경로에 관한 데이터 셋과 한국철도공사(Korea Railroad Corporation)에서 발표하는 일별/역별 승하차 데이터 셋을 활용한다. 질병 관리청에서 발표한 확진자와 확진 경로에 관한 데이터 셋은 다른 국가에서 발표한 수치와 함께 Kaggle을 통

해서 배포되고 있다.

해당 KCDC의 데이터 셋은 총 5,165개의 관측치 (2020년 6월 18일 기준)를 포함하고 있으며, <표 1>과 같은 변수들에 대한 정보를 담고 있다. 확진자의 기본적인 인구 특성에 대한 정보를 담고 있으며, 감염 사례(infection_case)는 알려진 집단 감염 사건에 대한 정보를 담고 있으며, 개인간 감염에 대해서는 Null 값을 담고 있다. 감염 경로(Infected_by)는 개인 간의 감염에서 누구로부터 감염이 되었는지에 대한 정보를 담고 있다. 확진 날짜(confirmed_

date)는 보건소 등에서 환자가 확진 판정을 받은 날짜이다.

본 연구에서는 위의 KCDC 데이터셋과 함께, 한국철도공사에서 제공하는 2019년과 2020년의 일별/역별 승하차 데이터 셋을 활용한다. 본 데이터셋은 현재 교통 데이터 거래소²⁾로부터 수집하였으며, 아래 <표 2>와 같이 특정일(RUN_DT)의 특정 역(STN_CD, STN_CD_NM)에서의 승차인원(ABRD_PRNB)과 하차인원(COFF_PRNB)을 집계한 3,504건의 자료이다.

<표 1> KCDC에서 발표한 감염 사례 데이터셋

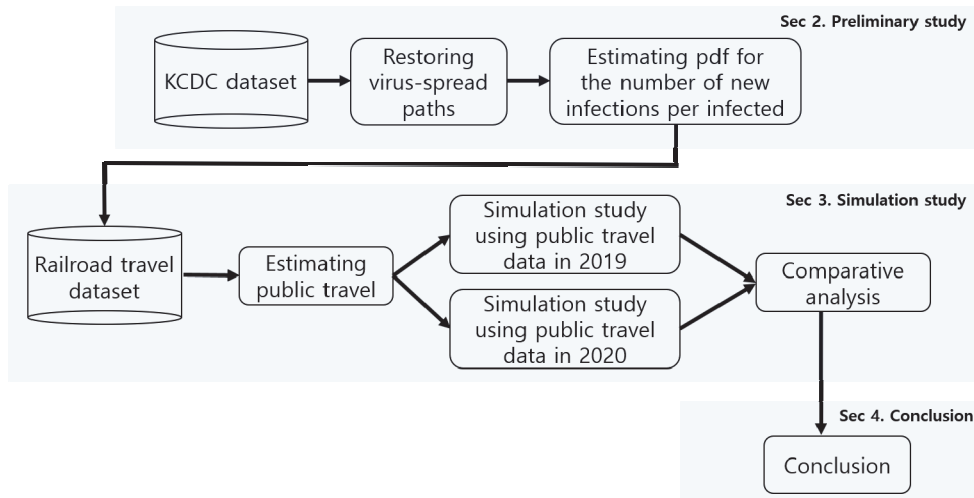
Variable	Description
patient_id	ex) 1000000001, 1000000002
sex	male or female
age	ex) 20s, 30s, 40s,...
province	ex) Seoul, Kyunggi-do, Pusan,...
city	ex) Nowon-gu, Suwon-si, Saha-gu,...
infection_case	group infection event (NaN if individual infection) ex) overseas inflow, Milal Shelter,
infected_by	patient_id of individual infection source (NaN if group infection event or infection source unknown) ex) 2002000001, 1000000002,...
confirmed_date	the date when infection is confirmed by authority (YYYY-MM-DD)

<표 2> 국내 기차 탑승 데이터

Variable	Description	Example
RUN_DT	Date	20200101
STN_CD	Station code	924, 39023
STN_CD_NM	Station name	Seoul, Yongsan, Anyang
ABRD_PRNB	Number of people on board	387, 19702
COFF_PRNB	Number of people getting off the train	269, 11125

2) www.bigdata-transportation.kr

〈그림 2〉 Graphical abstract



1.4 논문의 구성

본 연구는 바이러스 보균자가 대중교통을 이용해 다른 지역으로 이동하여 전파할 경우에 관한 시뮬레이션 연구를 수행한다. 2장에서는 과거의 전파 사례를 통해 한 명의 보균자가 얼마나 많은 사람에게 바이러스를 전파할 수 있는지에 대해서 극한 수치 이론을 적용하여 통계적으로 추정한다. 3장에서는 대중교통을 통해서 이동하는 승객에 대한 데이터를 바탕으로 몇 명의 피감염자가 타 지역에 새롭게 생길 수 있는지에 대한 시뮬레이션 연구를 수행한다. 4장에서는 본 연구의 실험 결과가 도시 간 이동 봉쇄와 같은 강력한 조치의 필요성에 대해서 어떤 함의를 가졌는지 정리한다.

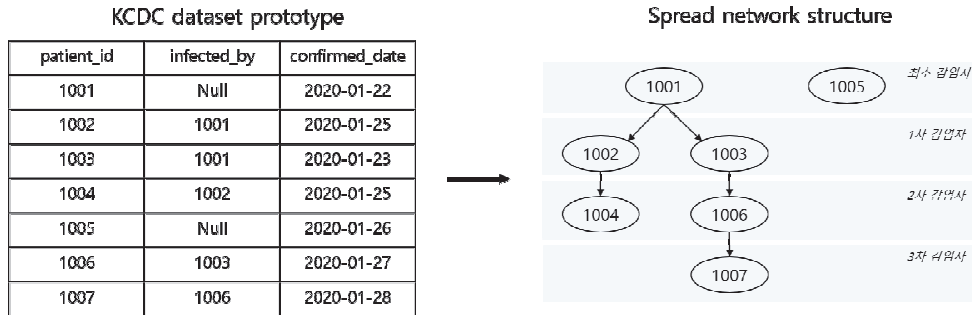
II. 예비 연구

2.1 감염경로 계보 생성

대한민국 질병 관리청에서 배포한 KCDC 데이터셋은 감염자(infected_by)-피감염자(patient_id)의 형태의 쌍대 데이터로 정리되어 있기에 본질적인 전파 메커니즘에 해당하는 네트워크 구조를 쉽게 판단할 수 없다. 본 연구에서는 우선 쌍대 데이터셋으로부터 네트워크 구조를 복원하는 데이터 전처리 작업을 수행하였다.

다음 〈그림 3〉은 네트워크 구조 복원 과정의 입력과 출력을 나타낸다. 감염경로가 불분명한, 즉, infected_by 변수의 값이 비어있는 확진자의 경우는 최초 감염자로 분류된다. 예시에서 1001번 확진자는 1002번과 1003번에게 전파하였으며, 이들은 1차 감염자로 분류된다. 이들 1차 감염자들은 각각 1명에게 전

〈그림 3〉 KCDC 데이터셋으로 부터 계보를 복구하는 과정



파하였고, 2차 감염자 중에 한 명인 1006번은 1007번이라는 3차 감염자를 낳았다. 따라서 최초확진자인 1001번 확진자는 총 5명의 확진자를 추가로 만든 것이 된다. 이들 5명의 확진자들중 가장 마지막에 확진된 1007번 확진자의 확진일인 1월 28일과 최초 확진자인 1001번 확진자의 확진일인 1월 22일은 6일의 시간 차이가 있다. 따라서 1001번 확진자는 6일간 5명에게 전파한 것이며, 1001번 확진자의 전파력은 하루를 기준으로 5/6명, 혹은 일주일을 기준으로 5.8명(=35/6)의 전파 강도에 해당한다. 1005번 확진자의 경우에는 아무에게도 감염을 시키지 않았으며, 이 경우에는 0명에게 전파한 것이다.

COVID-19의 유행 이후에 2020년 06월 18일까지 수집된 해당 데이터 셋에서 최초 확진자(감염경로가 불분명한 경우)는 총 378명이 존재했다. 즉, 전체확진자인 5,165명 중에 7.32%인 378명이 최초 확진자이다. 2.2절에서는 각 수준(최초, 1차, 2차, 3차 이상)의 확진자들에 대해 어떠한 인구 통계학적 특성을 관찰할 수 있는지 논의한다. 2.3절에서는 각각의 최초 확진자 별로 총 몇 명에게 감염을 시켰는지를 집계하여 “한명의 최초 확진자는 몇 명의 피감염자를 양산하는가”에 대한 확률 분포를 추정한다. 3장에서는 해당 확률 분포를 사용하여 시뮬레이션

연구를 수행한다.

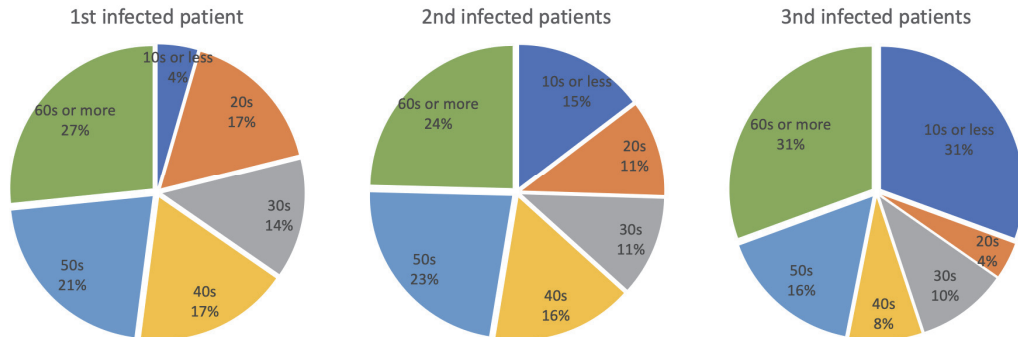
2.2 감염경로 계보 분석

연령은 사회적인 활동성과 연관성이 높기 때문에, 사회적인 활동으로 전파되는 전염병 특성상, 전파차수와 확진자의 연령의 관계는 전파 방지를 위해 중요한 분석이다. 해당 데이터 셋에서 최초 감염자 총 378명은 연령별로 치우침 없이 다양하게 분포([20대, 50대, 60대] - 50여명, [30대, 40대] - 40여명)되어 있었다. 그러나 1차-3차 감염자의 경우에는 인구통계학적 분류에 의한 특성이 존재하는 것을 관찰할 수 있었다. 해당 데이터 셋에서 수집된 1차, 2차, 3차 감염자의 수는 각각 900명, 317명, 74명이며 각 차수의 확진자 연령대는 다음 〈그림 4〉에 표현하였다.

상대적으로 최초와 1차 감염자의 경우에는 고연령분포를 보였지만, 10대 이하 감염자의 비율은 1차에서 4.5%, 2차에서 14.7%, 3차에서는 30.6%로 증가하는 예외적인 모습을 보였다. 즉, 10대 이하의 감염자 비율은 1차에서 매우 낮고, 차수가 올라감에 따라서 높아진다.

10대 이하는 바이러스의 창궐 이후에 학교의 휴업

〈그림 4〉 n차 감염 환자들의 연령 분포



조치등으로 집 밖을 돌아다니는 일이 적었기에 외부 활동으로부터 감염되는 1차 감염의 비율이 매우 낮았다. COVID-19 이후의 서울시 지하철 승하차 인원 에 대한 연구(Park(2020))에 따르면, 상대적으로 10대들이 많이 하차하는 지하철 역인 홍대입구역, 신촌역, 건대입구역의 하차 승객수는 크게 감소하였다. 대조적으로 가산디지털단지역과 같이 제조업 중심이라 재택 근무가 어려운 사업장이 밀집되어 있는 지하철 역의 승객의 하차인원은 감소폭이 작았다. 10대 이하의 감염자 비율이 차수가 올라감에 따라서 높아지는 현상은 이들 미성년자들이 가정에서 부모들로부터 감염되는 경우가 많다는 것을 의미한다. 이는 사회활동의 제한이 상대적으로 용이한 10대와 는 달리 20-50대의 인구의 사회활동을 제한하는 것은 힘든 일이기에 이들에 대한 적극적인 방역 지원 조치가 바이러스의 확산을 막는 데에 결정적이라는 것을 의미한다.

연령에 대한 인구 통계학적인 분석 외에 본 연구의 주된 관심에 해당하는 지역 간 감염경로에 관한 통계적 특성도 조사하였다. 감염경로가 밝혀진 1,325 건 중에서 광역시도 수준에서 같은 지역으로 감염시킨 경우는 총 1,141건으로 약 86%에 해당한다. 나

머지 184건(14%)의 경우에는 타 광역시도로 전파시킨 경우이다. 타 광역시도로의 전파의 절대적인 수치는 같은 지역간의 감염에 비해서 높지 않지만 새로운 지역에 바이러스를 전파시킬 수 있다는 가능성 때문에 높은 우려 사항이다.

이들 184건의 타지역으로의 감염 중에서는 총 140건의 경우가 서울, 인천, 경기도의 수도권 지역 내에 일어난 감염이었다. 수도권은 인구가 밀집된 구역일 뿐만 아니라 수도권에 거주하는 인구들은 생활 반경이 넓고, 일상적으로 이동하기 때문에 이와 같은 결과가 일어난 것으로 보인다. 수도권은 인구 밀집도가 매우 높을 뿐만 아니라 출퇴근 등 일상적인 이동이 많기에 수도권 내 지역과 지역 간 인구 이동률 또한 매우 높다. 또한 과거 프랑스 대도시 주민들의 집단 감염 사례를 보면 알 수 있듯이 프랑스의 주요 도시들이 위치하는 동부에서 서부로 확산되어 프랑스 모든 지역에서 COVID-19 확진자들이 쏟아져 나왔다(Cauchemez 등, 2020). 그 이후, 프랑스 정부는 극단적인 봉쇄 조치를 통해 가장 심각한 피해를 입은 두 지역의 확진자 수를 완화하는 데 성공했을 뿐만 아니라 다른 지역 사람들의 감염을 예방하는 데에도 성공한 것으로 밝혀졌다. 대한민국 수도

권과 프랑스의 매트르폴리탄의 유사성을 고려한다면, 수도권에서의 방역 실패는 전국적 대유행을 야기시킬 수 있다는 것을 의미한다. 따라서 국가적 차원의 방역 노력과 사회적 거리 두기와 같은 정책적 조치를 상대적으로 수도권에 집중하여 수도권 내 타 지역으로의 전파를 최대한 억제하는 정책적 조치가 중요한 것을 시사한다.

2.3 확률 분포 함수 추정

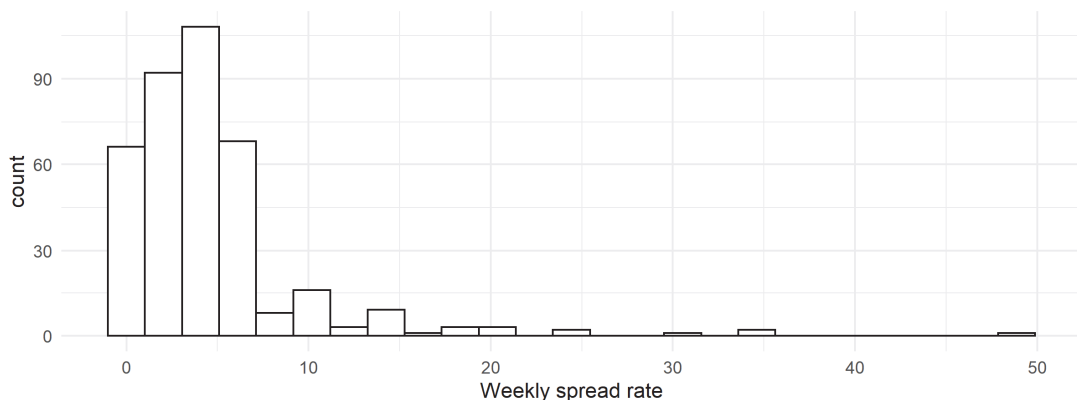
본 절에서는 최초확진자가 몇 명의 확진자를 양산할 것인지에 대한 분포함수를 추정한다. 위의 2.1절에서 총 378명의 최초확진자를 집계하였으며, 이들의 일주일 간의 전파율에 대한 분포함수를 추정한다. 본 연구에서 전파율을 일주일로 보정하는 이유는 최초확진자 이래로 하나의 계보에서 전파를 일으키는 기간이 매우 길거나 짧아서 표준화가 필요하며, 현재 주어진 데이터 셋은 전파가 완료된 것이 아니라 미래에도 계속적으로 최초확진자 밑의 계보가 확산될 수 있다는 점이다. 따라서 일정 기간 내에서의 전파강도를 측정하여 이를 바탕으로 분포함수를 추정

하고 시뮬레이션 연구를 진행한다.

아래의 <그림 5>는 최초 감염자가 일주일에 몇 명에게 확산을 시키는지에 대한 히스토그램이며, 오른쪽으로 꼬리가 긴 분포를 보인다. 한 명의 최초 감염자는 평균적으로 4.7명의 확진자를 일주일간 양산하며, 중간값은 3.5명이다. 상위 25%의 수치는 7명이며, 최대값은 49명이다. 오른쪽으로 극단적으로 꼬리가 긴 분포이며, 전염병 계보의 특성상 어떤 차수에서 다수의 피감염자를 양산한다면 그 효과가 기하급수적으로 늘어나 이와 같이 매우 큰 극단값을 보이기 때문에 자연스러운 현상이라고 할 수 있다. 이를 더 자세히 살펴보면 X축인 주간 감염강도가 낮은 경우에는 분포가 종형의 모습을 보이지만, 일정 주간 감염강도 이상에서는 분포가 완만히 감소하는 모습을 보인다.

이처럼 주간감염강도의 분포가 마치 두 개의 분포가 합쳐진 형태를 가지고 있기에, 본 연구는 주간감염강도 7명을 기준으로 하여 2개의 독립적인 분포함수를 추정하고 이를 중첩하는 형태로 모형화 한다. 이와 같이 하나의 현상을 2개 이상의 분포함수로 모형화하는 기법은 극한 수치 이론(Extreme Value

<그림 5> 최초 감염자가 1주일간 감염시킨 사람들의 수



Theory, EVT)의 근간이 되는 기법이다. 극한 수치 이론은 자연재해(Jindrova와 Pacakova(2016))나 금융 시장(Gencay와 Selcuk(2004)) 등의 사회 현상에 대해, 일반적인 경우에 해당하는 경우와 특정한 이벤트에 따라서 극단적인 경우에 해당하는 경우로 나누어 각기 모델링 한 후, 일반적인 경우가 일어날 확률과 극단적인 경우가 일어날 확률에 대한 추정을 바탕으로 현상을 통합적으로 모형화 하는 기법이다. 본 연구는 이와같은 극한 수치 이론의 일반적인 방법을 사용하여 주간감염강도를 모형화한다.

구체적으로는 본 연구에서는 1명의 최초확진자가 일주일간 7명 이하와 7명 초과인 경우로 나누어 모형화 한다. 그 이유는 <그림 5>에서 1) 관찰할 수 있듯이 7명을 기준으로 두 개의 다른 패턴이 관찰되며, 2) 7명은 평균적으로 매일 1명에 해당하므로 직관적으로 다루기 쉬우며, 3) 결과적으로 모델링된 분포가 위의 실험 관찰값을 잘 설명하기 때문이다. 최초 확진자 중에서 1주일간 7명 이하의 전파강도를 가진 경우에는 최초감염자를 “normal case”라고 부르고, 7명을 초과하는 전파강도를 가진 경우에는 최초 감염자를 “extreme case”라고 부른다. 최초감염자가 normal case가 되는 비율은 87.21%에 해당하며, extreme case가 되는 비율은 12.79%에 해당한다. 최초 감염자가 extreme case가 되는 경우는 최초 감염자 혹은 그 하위차수의 감염자들이 왕성한 전파를 일으켜 감염 계보의 폭과 깊이를 크게 만든 경우에 해당한다.

최초 감염자의 87.21%에 해당하는 normal case의 감염자들은 평균 3.24명, 중간값 3.50명, 표준편차 2.13명의 전파강도를 보였다. Q1값과 Q3값은 각각 1.40명과 4.67명이다. 이 분포는 대체로 좌우로 대칭이며 종형의 모습을 보이기에 정규분포로 근사하였다. 최초 감염자의 12.79%에 해당하는

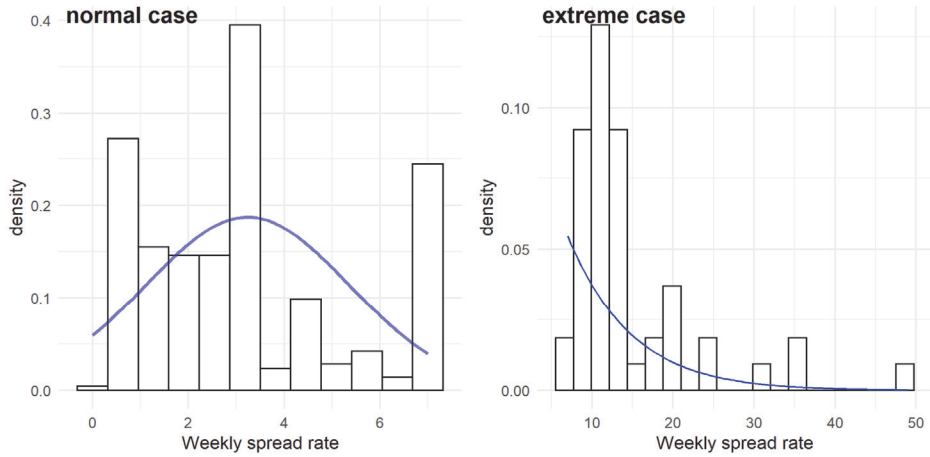
extreme case의 피감염자들은 평균 14.76명, 중간값 11.26명, 표준편차 8.32명의 전파강도를 보였다. Q1값과 Q3값은 각각 10.0명과 15.75명이며, 최대값은 49명이다. 이 분포는 극단적으로 오른쪽으로 꼬리가 두꺼운 모습을 보이며, 평균값과 Q3값이 비슷한 점으로도 유추가 가능하다. 이 분포는 shape와 scale의 두 개의 파라미터를 가지고 있으며, 극한 수치이론에서 자주 사용되는 와이블 분포를 사용하여 근사하였다. 와이블 분포의 두 개의 파라미터를 추정된 결과, 형상 매개변수(shape parameter)의 수치는 1.046이며 척도 매개변수(scale parameter)는 7.915이다. 다음 <그림 6>은 각각의 case에 대해서 추정된 분포와 실제값의 적합도를 보여준다.

다음 <그림 7>의 왼쪽 그림은 실제 378명의 최초 확진자에 대한 히스토그램이고, 오른쪽 그림은 본 연구에서 추정된 분포를 바탕으로 378개의 난수를 발생시켜서 그린 히스토그램이다. Normal case와 extreme case로 나누어서 추정된 분포에 대한 모형이 실제 수치와 잘 적합되는 것을 확인할 수 있다.

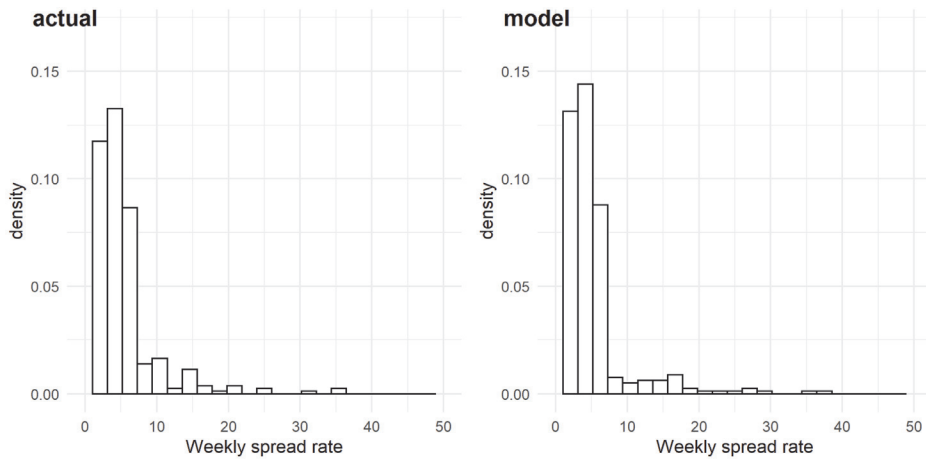
다음의 <그림 7>의 유사한 히스토그램이 실제로 동일한 분포에서 나왔다는 것을 확인할 수 있는 동질성에 대한 피어슨 카이제곱 검정을 수행하였다. 각각의 378개 값을 이용해 다음 <표 3>와 같은 빈도표를 준비하였고, 이를 기준으로 통계 검정을 수행하였다. 통계 검정의 결과로서 자유도 16, 유의확률 0.220이 확인되었으며, 이는 두 분포가 다르다는 통계적 근거가 일반적인 신뢰수준에서 유의하지 않다는 것을 의미한다.

다음 <표 4>는 378개의 실제 관찰값과 추정된 분포로 생성된 378개의 난수의 주요 통계량이며, 모형이 실제 분포를 대체로 잘 추정하는 것을 다시 한번 확인할 수 있다. 이어지는 3장에서는 추정된 분포함수와 절도 이용 승객의 데이터를 활용하여 시뮬

〈그림 6〉 Normal case와 extreme case의 주간 전파율 분포



〈그림 7〉 실제 데이터와 모델의 히스토그램 분포



〈표 3〉 두 분포의 동질성 검정을 위한 빈도표

범위	[0, 1]	(1, 3)	(3, 5)	(5, 7)	(7, ∞]
Actual	63	94	107	70	49
Model	41	110	118	56	58

〈표 4〉 실제 데이터와 모델 결과값의 수치적 요약

	Minimum	1 st quartile	Median	Mean	3 rd quartile	Maximum
Actual	0.00	1.75	3.50	4.71	7.00	49.00
Model	0.00	2.04	3.66	4.40	5.30	37.29

레이션 연구를 수행한다.

III. 시뮬레이션 연구

3.1 이동 인원 추정

본 연구에서는 우리나라에서 높은 인구 밀집도로 바이러스 전파가 가장 많이 일어나는 서울에서 탑승한 확진자가 두 번째로 큰 도시인 부산으로 이동하는 경우를 상정한다. 한국철도공사에서 제공한 데이터 셋을 이용해 2019년과 2020년의 3월 셋째주 일주일에 걸쳐 서울(서울역, 용산역, 영등포역, 청량리역, 수서역)에서 출발해 부산(부산역, 부전역, 구포역, 신해운대역)으로 도착한 KTX, 무궁화호, 새마을호, 그리고 SRT³⁾ 승객의 총 합을 일별로 집계

하였다.

철도가 아닌 다른 수단으로 이동한 인원을 추정하기 위해 한국교통연구원의 교통수단별 수송분담률 조사 결과 활용하였다. 국토교통부가 의뢰해 2010년 한국교통연구원이 조사한 교통수단별 수송분담률 조사에 따르면 300km 이상의 이동의 교통수단별 수송분담률은 승용차 33.6%, 버스 16.0%, 철도 31.9%, 항공 18.1%, 해운 0.4%이다. 만약에 해당 수송분담률이 COVID-19의 상황에서도 유지되었다면 전체 인구 이동량을 추산할 수 있겠지만, COVID-19의 확산으로 얼마나 많은 여행수요가 대중교통에서 승용차로 대체되었는지는 정확하게 알려진 바가 없다. 본 연구는 따라서 대중교통을 이용한 이동만을 고려한다. 이에 따라 추정된 대중교통을 이용한 이동 인원의 수는 아래 <Table 5>와 같다.

아래의 <표 5>는 우리나라 국민들이 COVID-19의 상황에서 대중교통 이용을 상당히 자제한 모습을

<표 5> 서울발 부산행 이동 승객 수

2019년 (총 이동 인원 167,743명)							
	3/17 (일)	3/18 (월)	3/19 (화)	3/20 (수)	3/21 (목)	3/22 (금)	3/23 (토)
철도	12,537	9,041	8,347	9,620	10,414	17,541	15,904
버스	6,288	4,535	4,187	4,825	5,223	8,798	7,977
항공	7,113	5,130	4,736	5,458	5,909	9,953	9,024
합계	25,938	18,705	17,269	19,902	21,546	36,291	32,905
2020년 (총 이동 인원 53,255명)							
	3/15 (일)	3/16 (월)	3/17 (화)	3/18 (수)	3/19 (목)	3/20 (금)	3/21 (토)
철도	3,315	3,413	3,136	3,242	3,786	5,480	3,368
버스	1,663	1,712	1,573	1,626	1,899	2,749	1,689
항공	1,881	1,937	1,779	1,839	2,148	3,109	1,911
합계	6,858	7,062	6,488	6,707	7,834	11,338	6,968

3) 경부선 SRT의 여객수는 정확한 자료를 구하지 못하여 경부선 KTX 여객수와 경부선 SRT 승객수의 비례식을 이용하여 추산하였다. (2019년 3월 KTX와 SRT의 경부선 이용 승객수는 각각 290.6만과 140.3만이며, 2020년 3월은 각각 80.2만과 47.6만이다.)

보여준다. 전체적으로 주중과 주말에 걸쳐 철도 이동량이 급격하게 줄었으며, 특히 이동 수요가 높은 금요일과 토요일의 이동량이 급격히 감소하였다.

3.2 시뮬레이션 가정

서울시의 인구는 약 970만명(2020년 기준)이며, 2020년 3월 15일 서울시에는 247명의 확진자가 존재한다. 이 비율은 약 0.0027%(=247/970만)에 해당하는 매우 낮은 비율이다. 본 연구에서는 1) 서울시 인구의 0.0027%가 실제 감염자인 경우와 2) 서울시 인구의 0.01%가 실제 감염자인 경우의 두 가지 경우에 대해서 고려한다. 해당 두 가지 경우를 2019년과 2020년의 이동량 시나리오와 교차하여 시뮬레이션 연구를 수행한다.

예를 들어 서울시 인구의 0.0027%가 감염자인 경우, 하루에 서울에서 부산으로 이동하는 인구가 6,858명이라면 여기에 답승하는 감염자의 수는 모수가 0.185($6,858 \times 0.0027\%$)인 포아송 분포를 따른다. 본 연구에서는 감염의 시발점이 되는 최초 확진자만이 감염을 일으킬 수 있다고 가정하며, 이들의 비율은 2.1절에서 살펴본 바와 같이 전체 확진자 중에서 7.32%에 해당한다. 따라서 6,858명의 일일 이동 시나리오에서, 답승하는 최초 확진자의 수는

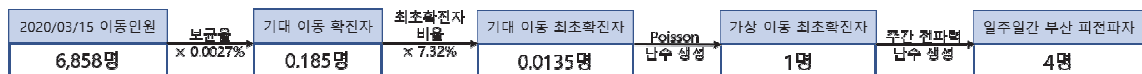
모수가 0.0135명($0.185 \times 7.32\%$)인 포아송 분포를 따른다. 이들 최초 감염자들은 서울에서 부산으로 이동하여 각각 앞의 <그림 7>과 같은 분포로 부산에서 피감염자를 생성한다. 이 과정을 아래 <그림 8>은 시뮬레이션 실험의 과정의 예시로서 나타내었다. 해당 몬테카를로 시뮬레이션 실험은 10,000번의 반복에 걸쳐 시행된다.

3.3 시뮬레이션 결과

2020년 3월 15일의 감염자 비율인 0.0027%를 2020년의 서울-부산 이동량에 적용하였을 때의 시나리오를 10,000번 시뮬레이션 한 결과, 일주일간 부산에는 평균 0.48명의 피감염자가 발생한다. 같은 감염자 비율을 2019년의 서울-부산 이동량에 적용하였을 때에는 시뮬레이션 결과 일주일간 부산에는 평균 1.70명의 피감염자가 발생한다. 이들 두 시나리오의 통계량은 아래 <표 6>와 같다.

2019년에 비해서 2020년에 대폭으로 감소한 도시간의 이동은 가상 시나리오에서 부산의 피감염자수를 낮추는 효과가 있음을 확인할 수 있다. 10,000번의 반복 실험에서 2020년의 경우에는 9,104번의 경우에 새로운 감염이 발생하지 않았고, 2019년의 경우에는 7,237번의 경우에 새로운 감염이 발생하

<그림 8> 시뮬레이션 진행 과정



<표 6> 10,000번의 시뮬레이션 수행 결과(감염률이 0.0027%인 경우)

	Min.	1 st Qu.	Med.	Mean	3 rd Qu.	Max.
Under travel pattern in 2020	0.00	0.00	0.00	0.48	0.00	43.00
Under travel pattern in 2019	0.00	0.00	0.00	1.70	2.00	50.00

지 않았다. 불운한 25%의 경우에 해당하는 경우인 제3사분위수인 경우에는 2020년에는 0명의 감염이 발생하였고, 2019년에는 2명의 피감염자가 발생하였다. 두 경우에 모두 매우 극단적으로 불운한(10,000번 반복중에서 가장 최악의 시나리오) 경우에는 최대 43명이나 50명의 새로운 감염을 피감염자를 발생시키는 것을 확인할 수 있었다.

2020년 3월 15일의 감염자 비율인 0.0027%를 적용한 시나리오에서는 대부분의 경우에 기하급수적인 피감염자를 낳지는 않았다. 이번에는 감염자 비율이 만 명중에 1명에 해당하는 0.01%를 적용하여 위의 실험을 다시 시행하였다. 이는 서울시 인구 970만 명 중에서 970명의 비율로 감염이 되어있는 사례이며, 바이러스의 확진 상황에 따라서 얼마든지 일어날 수 있는 사례이다. 질병의 전파와 관련된 시나리오 분석에서는 이처럼 불운한 상황을 가정하여 시뮬레이션 하는 것이 적합하다.

감염자 비율 0.01%를 2020년 서울-부산 이동량에 적용하였을 때의 시나리오를 10,000번 반복 시뮬레이션 한 결과, 일주일간 부산에는 평균 1.83명의 피감염자가 발생하였다. 같은 감염자 비율을 2019년의 서울-부산 이동량에 적용하였을 때에는 일주일간 부산에 평균 5.94명의 피감염자가 발생하였다. 이들 두 시나리오의 통계량은 아래 <표 7>와 같다.

감염자 비율 0.01%의 경우에도 인구 이동이 줄어든 2020년의 경우에 피감염자의 수가 2019년의 이동량 시나리오에 비해서 훨씬 적은 것을 확인할 수 있다. 10,000번의 반복실험에서 2020년 이동 패턴

의 경우에 6,967번의 경우에 새로운 감염이 생기지 않았고, 2019년 이동 패턴의 경우에는 3,176번의 경우에 새로운 감염이 생기지 않았다.

<표 7>의 첫 번째 행인 0.01% 감염률 하에서 2020년 이동 패턴의 통계량과 <표 6>의 두 번째 행인 0.0027% 감염률 하에서의 2019년 이동 패턴의 통계량이 10%의 차이를 보이며 대체로 유사하다. 이는 2019년에 비해서 크게 줄어든 2020년의 도시 간 이동량의 효과가 10,000명당 1명이 감염이라는 높은 감염률을 2020-03-15의 감염률인 0.0027%로 낮추는 만큼의 효과를 발휘할 수 있음을 의미한다. 즉, 2020년의 대폭 줄어든 도시 간 이동량은, 2019년의 두 번째 행에 해당하는 상대적으로 우려할만한 대규모 도시 간 이동으로 인한 감염을 방지하는 효과가 있음을 의미한다.

<표 6>와 <표 7>을 종합하여 고려해보면, 평균적으로 생성되는 피감염자의 수는 이동하는 인원의 수와 서울시의 감염률에 대체로 비례하는 것을 알 수 있다. 감염률이 0.0027%인 시나리오에 대해서 2019년과 2020년 이동량의 패턴이 만들어낸 1.70명과 0.48명의 표본 평균은 두 개의 독립 모집단에 대한 모평균의 차이에 관한 t-검정을 실시 하였을 때에 유의확률 $2.22e-16$ 이하의 매우 강한 통계적인 근거를 보였으며, 감염률이 0.01%인 시나리오에 대해서 5.94명과 1.83명의 경우에도 같은 t-검정에서 유의확률 $2.22e-16$ 이하의 강한 통계적 근거를 나타냈다. 즉, 각 감염률에 대해서 이동량의 감소가 통계적으로 매우 유의미한 차이를 만들어 낸 것이라고 해

<표 7> 10,000번의 시뮬레이션 수행 결과(감염률이 0.01%인 경우)

	Min.	1 st Qu.	Med.	Mean	3 rd Qu.	Max.
Under travel pattern in 2020	0.00	0.00	0.00	1.83	2.00	57.00
Under travel pattern in 2019	0.00	0.00	4.00	5.94	8.00	70.00

석할 수 있다. 하지만 2.3절에서 논의한 바와 같이 감염력의 분포가 매우 긴 꼬리를 가진 형태이기에 최악의 경우를 나타내는 최대값의 경우에는 철도 이동량이 적은 경우라고 하더라도 수십명의 감염이 발생할 수 있는 것을 확인할 수 있었다. 이는 본질적으로 바이러스 전파에 대해서는 어떤 경우에서도 방심할 수 없음을 의미한다.

IV. 결론

질병과 전쟁과 같은 특수 위험 상황에 관한 시뮬레이션 연구에서는 비판적인 가정하게 연구를 수행하는 것이 바람직하다. 이에 따라서 본 연구에서는 현재의 통계량뿐 아니라 최악의 상황에 해당하는 1) 2019년의 이동량으로 인구 이동이 일어날 경우, 2) 서울시의 감염률이 만분의 일에 해당할 경우에 대해서도 고려하여 시뮬레이션 연구를 진행하였다.

시나리오의 비교는 임의의 특별한 이벤트가 없는 3월 일주일을 고려하여 시뮬레이션을 진행하였다. 일주일 동안 확진자가 몇 명 나올지에 대한 시뮬레이션 연구이므로 1) 일주일 이상의 오랜 기간을 고려하거나, 2) 상대적으로 낮은 확진율로 인해서 서울 외 지역에서 경각심이 낮을 수 있음을 고려하거나, 3) 부산뿐만 아니라 대구, 대전 등 여러 대도시에서 동시다발적으로 확진자가 이동한다면 본 시뮬레이션 결과보다 더 많은 피감염자가 발생할 수 있다. 즉, 같은 사건이 전국적으로 오랜 기간 걸쳐 반복된다면 지역 간 이동으로 인해 수많은 확진자가 나올 수 있는 결과를 초래할 것이다.

본 연구의 비교실험은 다른 지역으로의 전파에 있어서 2019년에서 2020년으로 줄어든 인구의 이동

이 소스가 되는 서울의 감염률을 낮추는 것과 같은 효과가 있음을 입증한다. <표 5>에 의하면 2019년에 비해서 2020년에는 이동량이 31%로 감소하였으며, 이는 <표 6>에 의하면 평균적으로 28%로 감염자가 감소하는 효과를 보인다. 지역 간 이동률과 감염률이 양의 상관관계를 가진다는 또 하나의 사례로서 일본 정부의 2020년 7월 고 투 트래블(Go To Travel) 정책을 예로 들 수 있다. 일본 정부는 여행 경비의 최대 50%를 지원하는 해당 정책에 10월 부터는 도쿄 지역까지 포함을 시켰고, 그 결과로서 일일 확진자가 2,000명을 넘어서는 사태까지 벌어졌다(Anzai와 Nishiura, 2021).

반면 2019년과 2020년의 철도 이동량의 차이에서 관찰할 수 있듯이 과거와 달리(강가경 등(2018)) 우리나라 국민들은 생업에 덜 필수적인 주말의 이동을 대폭 줄였으며, 생업에 필수적이라고 생각할 수 있는 주중의 이동도 상당히 줄이면서 보건당국의 지침을 충실히 따르고 있다. 따라서 본 연구는 이러한 당국의 정책과 국민들의 거리 두기 협조가 전염병의 확산을 저지하는데 크게 기여하고 있음을 보인다고 할 수 있다.

참고문헌

- 강가경, 이우균, 이수열 (2018), “기후변화가 비즈니스에 미치는 위험, 기회 인식 차이가 탄소경영활동과 성과에 미치는 영향: 탄소정보공개(CDP)를 이용한 탐색적 연구,” 한국생산관리학회지, 29(4), 463-494.
- 국토해양부 교통정책실 종합교통정책과 (2010, 9), “교통수단별 수송분담률 발표,” <https://eiec.kdi.re.kr/policy/materialView.do?num=109270>

- 배진, 서용원 (2019), “여행상품 인술자 스케줄링 문제에 대한 휴리스틱 알고리즘 연구,” *한국생산관리학회지*, 30(3), 251-266.
- 이옥진, “개강파티가 뭐기에...美대학 곳곳서 코로나 집단 감염,” *조선일보*, 2019년 8월 19일.
- 조정래, 이동원 (2019). “안전관리시스템 요인이 안전성과에 미치는 영향: 현장 종사자를 중심으로,” *한국생산관리학회지*, 30(3), 207-225.
- 최승훈 “60대 확진자, 서울서 KTX...동대구역까지 가서야 잡아,” *JTBC 뉴스*, 2019년 9월 16일.
- Anzai, A. and H. Nishiura (2021), “Go To Travel Campaign and Travel-Associated Coronavirus Disease 2019 Cases: A Descriptive Analysis July-August 2020,” *Journal of Clinical Medicine*, Vol.10, No.3, 398.
- Ayton, G. S. and G. A. Voth (2010), “Multiscale computer simulation of the immature HIV-1 virion,” *Biophysical Journal*, Vol.99, No.9, 2757-2765.
- Cauchemez, S., C. T. Kiem, J. Paireau, P. Rolland and A. Fontanet (2020), “Lockdown impact on COVID-19 epidemics in regions across metropolitan France,” *The Lancet*, Vol.396, No.10257, 1068-1069.
- Chu, D. K., E. A. Akl, S. Duda, K. Solo, S. Yaacoub, H. J. Schünemann,... and A. Hajizadeh (2020), “Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis,” *The Lancet*, Vol.395, No.10242, 1973-1987.
- Gencay, R. and F. Selcuk (2004), “Extreme value theory and Value-at-Risk: Relative performance in emerging markets,” *International Journal of Forecasting*, Vol.20, No.2, 287-303.
- Frazer, B. D. (1977), “Plant virus epidemiology and computer simulation of aphid populations. In *Aphids as Virus Vectors*,” *Academic Press*, 413-431.
- Hatchett, R. J., C. E. Mecher and M. Lipsitch (2007), “Public health interventions and epidemic intensity during the 1918 influenza pandemic,” *Proceedings of the National Academy of Sciences*, Vol.104, No.18, 7582-7587.
- Hu, M., H. Lin, J. Wang, C. Xu, A. J. Tatem, B. Meng,... and H. Xie (2020), “The risk of COVID-19 transmission in train passengers: an epidemiological and modelling study,” *Clinical Infectious Diseases*.
- “Impact of opening and closing decisions by state,” (2020 Aug), Retrieved from <https://gizmodo.com/heck-yes-the-first-free-wireless-plan-is-finally-here>
- Innocent, G., I. Morrison, J. Brownlie and G. Gettinby (1997), “A computer simulation of the transmission dynamics and the effects of duration of immunity and survival of persistently infected animals on the spread of bovine viral diarrhoea virus in dairy cattle,” *Epidemiology & Infection*, Vol.119, No.1, 91-100.
- Jindrova, P. and V. Pacakova (2016), “Modelling of extreme losses in natural disasters,” *International Journal of Mathematical Models and Methods in Applied Sciences*, Vol.10, Issue.2016.
- Kim, S., Y. Ko, Y.-J. Kim and E. Jung (2020), “The impact of social distancing and public behavior changes on COVID-19 transmission dynamics in the Republic of Korea,” *PLoS ONE*, Vol.15, No.9. <https://doi.org/10.1371/>

- journal. pone.0238684
- Lee, B. Y., S. T. Brown, G. W. Korch, P. C. Cooley, R. K. Zimmerman, W. D. Wheaton,... and D. S. Burke (2010), "A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 H1N1 influenza pandemic," *Vaccine*, Vol.28, No.31, 4875-4879.
- Park, J. (2020), "Changes in subway ridership in response to COVID-19 in Seoul, South Korea: Implications for social distancing," *Cureus*, Vol.12, No.4.

A Simulation Study on the COVID-19 Infection Spread in Korea Using the Infection Database*

Sungho Lee** · Wonryul Jung*** · Jaehun Hwang**** · Min Kyu Sim*****

Abstract

The spread of viral epidemics such as COVID-19 occurs through the movement of carriers between regions. This simulation study examines how the inter-regional movement using public transportation in 2020, which has significantly decreased compared to 2019, has affected the spread of infectious disease. In the preliminary research, a tree-structured infection genealogy is generated based on the information on the carrier-infected person pair published by the Korea Centers for Disease Control and Prevention, and the probability distribution of how many infected people are probabilistic by one carrier. It is confirmed that the probability distribution is a convolution of two different distributions. In the main simulation study, the amount of public transportation used between Seoul and Busan was estimated using the passenger boarding and disembarkation data provided by the Korea Railroad Corporation, and a simulation study was performed using the Poisson process. It was confirmed that the amount of public transport travel in 2020, which is about three times less than in 2019, makes a big

* This study was supported by the Research Program funded by the SeoulTech(Seoul National University of Science and Technology)

** Undergrad student, ITM department, Seoul National University of Science and Technology

*** Undergrad student, ITM department, Seoul National University of Science and Technology

**** Undergrad student, ITM department, Seoul National University of Science and Technology

***** Assistant Professor, Department of Industrial Engineering, Seoul National University of Science and Technology (Corresponding Author)



© Korean Production and Operations Management Society. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
(<https://creativecommons.org/licenses/by-nc/4.0/deed.ko>)

difference in the number of contagions. This substantiates that the authorities' recommendation to refrain from movement and the high level of citizen cooperation play big roles in preventing the spread of infectious disease.

Keywords: COVID-19, Demographics, Extreme value theory, Weibull distribution, Railroad travel, Simulation