SIMULATION OF DISEASE SPREAD MODELS WITH METAPOPULATION APPROACH

by

Fatma Elif Saygılı

B.S., Mathematical Engineering, Yıldız Technical University, 2013

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of Master of
Science

Graduate Program in Industrial Engineering
Boğaziçi University
2018

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis supervisor, Assoc. Prof. Wolfgang Hormann of the Faculty of Engineering at Boğaziçi University, for his support, patience and guidence. I also would like to thank to Prof. Refik Güllü and Assist. Prof. İsmail Başoğlu for taking part in my thesis jury.

I would like to thank to my friends, colleagues and in particular to managers for always supporting me, keeping my motivation high and giving me a change for focusing on my study.

Finally, I would like to express my gratitude to my family for their belief in me, continuous encouragement and support. Without their support and also trust in me, it would be impossible to achieve this study. I am indebted to my mother, my father, my sister, my grandmother and my uncle for always being there for me.

ABSTRACT

SIMULATION OF DISEASE SPREAD MODELS WITH METAPOPULATION APPROACH

In this thesis, a metapopulation simulation model, which is not only useful for Turkey but for any country, is built. For this purpose, a general simulation code is implemented including both infection and travel processes. Required parameters and inputs are defined and analyzed. In case travelling between cities, it will be possible to observe disease spreading behaviour and whether an outbreak will occur or not in cities. Moreover, it will be possible to predict a possible disease outbreak, the number of infected people and to take precautions such as vaccination. The model is implemented in R and so it is flexible and can be modified for different cities. In the first step, the simulation code is run for two cities with various expected number of people who travel to observe the behaviour of the disease spread and check the performance of the code. In the second step, simulation experiments are carried out with main cities in Turkey to see that how fast the epidemy will spread between cities. Before doing experiments, data is gathered for Turkey parameters in the internet. The effect of the vaccination on the disease spread behaviour in the largest three cities in Turkey is also observed.

ÖZET

HASTALIK YAYILMA MODELLERİNİN METAPOPULASYON YAKLAŞIMI İLE SİMÜLASYONU

Bu tezde, sadece Türkiye için değil aynı zamanda başka bir ülke için de kullanılmak üzere bir metapopulasyon simülasyon modeli yapılmıştır. Bu amaçla, enfeksiyon ve yolculuk süreçlerini içeren genel bir simülasyon kodu yazılmıştır. Gerekli parametreler ve girdiler tanımlanmış ve analiz edilmiştir. Şehirler arası yolculuk durumunda, hastalık yayılmasının davranışını ve salgın olup olmayacağını gözlemlemek mümkün olacaktır. Ayrıca, olası salgın, enfeksiyonlu bireylerin sayısını tahmin etmek ve aşı gibi önlemler almak mümkün olacaktır. Model R'da geliştirilmiştir, bu nedenle esnektir ve farklı şehirler için değiştirilebilecektir. İlk adımda, simulasyon kodu iki şehir ile, hastalık yayılmasının davranışını gözlemlemek, kodun performansını kontrol etmek için çeşitli sayıda beklenen yolcu sayısı kullanılarak çalıştırılmıştır. İkinci adımda, bulaşıcı hastalığın şehirler arasında ne kadar hızla yayılacağını görmek için Türkiye'deki başlıca şehirler kullanılarak simülasyon deneyleri yapılmıştır. Deneyler yapılmadan önce, Türkiye parametreleri için internetten veri toplanmıştır. Aynı zamanda, Türkiye'deki en büyük üç şehirde, aşı uygulamasının hastalık yayılması davranışı üzerindeki etkisi gözlemlenmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

In recent years, modeling disease spread has been one of the most crucial research areas. It is aimed to predict a possible disease outbreak and the number of infected people by simulation and to develop strategies to control a disease before causing serious consequences in a society. In addition, it is possible to find the most effective strategies for the use of antiviral agents and vaccine.

Classical epidemic models assume that the entire population lives in one area and is well mixed. However, the population can be living on isolated areas or in different cities and countries. In addition, demographic and disease parameters may vary spatially. This spatial heterogeneity influences the transmission of the disease. The metapopulation approach has an importance in epidemic modeling. The metapopulation model is a generalization of the model with perfect mixing in a single population. The metapopulation model assumes that perfect mixing is correct in several large populations and the infection in populations occurs due to travelling of people between these populations. There has been research about this issue in several countries but not detailed in Turkey. Using a metapopulation model, in the case of travelling between cities it is possible to observe the disease spreading behaviour and whether an outbreak will occur in all cities.

The purpose of this thesis is to build a metapopulation simulation model which is not only useful for Turkey but for any country. In order to reach that aim, a general simulation code is implemented including both infection and travel processes. Required parameters and inputs are defined and analyzed for this purpose.

R, which is a programming language and software environment for statistical computing, is used for coding the program. R is free and open source software and allows anyone to use and modify it. It is easier to interpret results and debug the code in R than in other programming languages.

Our first step is to implement and run the simulation code for a model with only two cities. For different expected number of people who travel between cities, the behaviour of the disease in both cities will be observed and also the performance of the code and the accuracy of the model will be checked. After running the simulation code with two cities, simulation experiments are conducted with main cities in Turkey. Our aim is to see how fast the epidemy will spread between cities and also how much vaccination will be needed to prevent the spreading of the disease into other cities.

This thesis is structured as follows. In Part 2, we review the literature on disease spread models such as agent based and metapopulation models. In Part 3, we explain the description and the algorithm of our metapopulation model including both infection and travel processes. We also give details about the simulation program coded in R. In Part 4, we carry out some simulation experiments with two large populations. In Part 5, we built metapopulation models for Turkey and give detailed information about the parameters and data sources for these models. In Part 6, we carry out simulation experiments for three models of Turkey and interpret the results. Moreover, we observe the effect of vaccination strategies on the disease spread behaviour. In Part 7, our conclusions are presented.

2. LITERATURE REVIEW

This section gives a brief information about SIR models, agent based models and metapopulation models.

2.1. The Deterministic SIR Model

Mathematical modeling of infectious diseases is a crucial area of mathematical biology. These models are a framework for understanding the dynamic of infections. One of the most popular among these models is the SIR model.

The model is used in order to know the number of people who need a medical care for a disease. Populations that can be humans, animals or plants are divided into different compartments in which they have same attributes according to the disease they have [1].

In the SIR model there are three compartments. These are susceptible, infectious and recovered. People in the susceptible compartment can become infected as they do not have immunity to the disease. After being infected, people move from the susceptible compartment to the infected compartment. Infected people can spread illness to susceptible ones. Some time later, people who are in the infected compartment move into the recovered compartment. Lastly, in recovered compartment, people have immunity to infection and do not have any effect on transmission of the infection to other individuals. These compartments are denoted S, I and R. The total population size is $N=S+I+R$. The sizes of each compartment at time t is denoted $S(t)$, $I(t)$ and $R(t)$ [2].

After distribution of individuals to different compartments, in order to know how the sizes of these compartments change over time, some set of equations can be formed. At time t the number of people in the infectious compartment $I(t)$ can be found according to these equations.

$$\frac{dS}{dt} = -\beta SI \tag{2.1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{2.2}$$

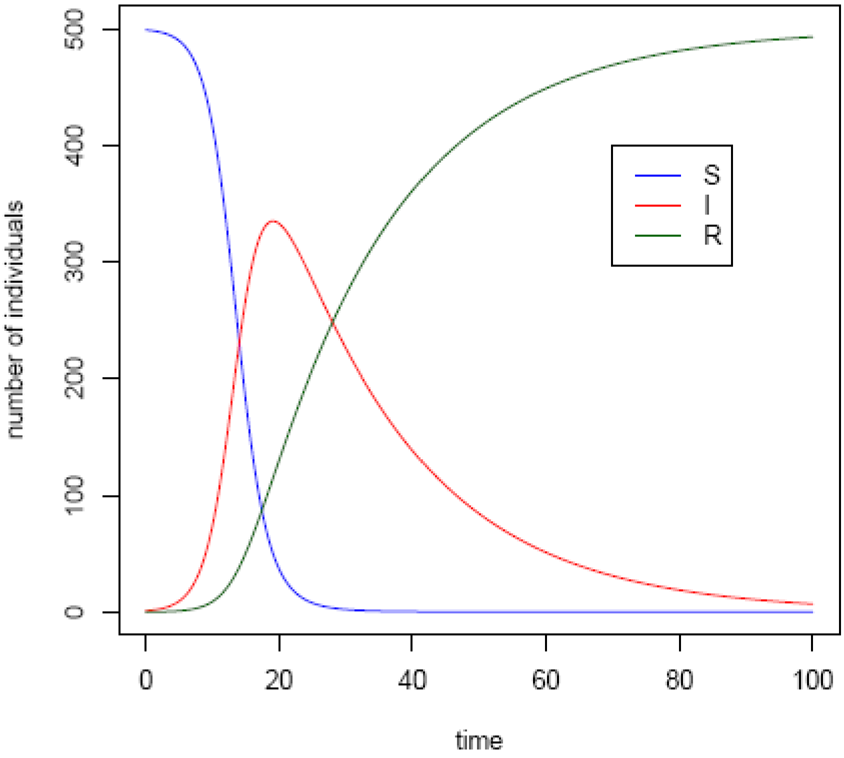

Figure 2.1. SIR epidemic model dynamics.

In these equations above, β is the transmission rate, γ is the recovery rate and 1/ γ is the mean infectious period. If all people are susceptible at time 0, (S(0)=N) an infected person introduced initially in the population can infect other individuals at the rate βN during the infectious period 1/ γ. Therefore, it is expected that this initial infected person can infect R_0 = βN/γ people [2]. The number R_0 (basic reproduction number) is the number of secondary infections produced by a single random infected person in a fully susceptible population. R_0 is an important indicator for an epidemic model. In case there is at least a single infected person and $R_0>1$, there is an outbreak in the population.

2.2. Stochastic SIR Model

Many systems such as biological processes have an uncertainty. For example the exact time of reproduction of an individual cannot be predicted. It can be possible to involve this stochasticity in a mathematical model. When an event occurs, formulating the terms as probabilities is a way of including randomness into a mathematical model.

In stochastic models, the questions we try to find the answers are for example what is the probability of an outbreak? What is the expected duration of an epidemic? These questions are based on the stochastic nature of a process.

Classic SIR models of infectious diseases assume perfect mixing between all individuals. The SIR model considers the changes in the number of individuals in the three compartments. In a stochastic SIR model, discrete numbers and process probabilities are used contrary to the continuous variables and process rates in the classical deterministic model.

In this thesis, we will use the generalization of the stochastic SIR model called metapopulation model with perfect mixing in several populations. There exist also other types of stochastic SIR models. There are three different types of stochastic model formulations which are discrete time Markov chain, continuous time Markov chain and stochastic differential equations in the scope of SIR modeling. These stochastic processes are different regarding state variables and the time. In discrete time Markov chain models, the state variables and the time are discrete. In continuous time Markov chain models, state variables are discrete, time is continuous and in stochastic differential equation model, states variables and the time are continuous [2].

In order to quantify the severity of outbreaks of infectious diseases the basic reproduction number (R_0) is used. It is defined as the expected number of secondary infections produced by a random infected person in a fully susceptible population. If R_0 is greater than one then the outbreak can lead to an epidemic, if R_0 is less than one then the outbreak will become extinct [3]. In this thesis, the general form of the formula to compute the approximate R_0 value is in the following.

$$N*\text{Expected Disease Duration}*\text{pinf} \tag{2.3}$$

In the formula, N is the population size, Expected Disease Duration is the expected number of days of disease and pinf is the transmission probability. In addition to $R_{0,}$ the overall attack rate, which is the percentage of individuals who will get sick during an outbreak in a population, can be used as a disease characteristic and estimated by using R_0 [4].

2.3. Agent-Based Stochastic SIR Model

Agent-based is the name of a class of computational models used to simulate the interactions and actions of agents and to understand their effects on the whole system. Best-known components in an agent-based model are agents, environment, process, interaction of agents and decision-making heuristics [5].

Agent-based models are used to capture the behaviour of each individual in an environment. Compartmental models can not capture the dynamics of an individual’s behaviour, especially about social networks and mobility patterns [6]. Therefore, agent-based simulations in epidemiology are powerful as they allow to observe the behaviour of a disease spreading in a granular level and also the interactions between individuals.

The agent-based SIR model is an agent-based model that uses SIR approach for epidemics to observe the behaviour of an infectious disease spreading in a closed population over time. A closed population means that there is no death, no birth and no migration to other populations. In the pass of time, the compartment of an agent can change and each agent can pass to one of three states which are susceptible, infected and recovered. Each agent interacts with other agents.

Ajelli *et al.* [7] utilizes agent-based model in order to simulate epidemic outbreaks. The agent-based model which is a stochastic, discrete time, spatially explicit model in which human individuals are represented as agents. It is thought that the infection can be spread by contacts between students in schools, household members, colleagues in workplaces. The model is based on Italian population using detailed data on a socio demographic structure. Each individual is assigned to an appropriate group due to workplace size, household size, age of individuals and so on.

Longini *et al.* [8] use a stochastic, discrete time simulation model for influenza spread for agents in a structured population in order to compare efficiency of different intervention strategies which are antiviral prophylaxis and vaccination. In this structured population, people interact in contact groups such as schools, households, day-care centers and playgroups. Also, they simulate illness attack rates based on age of agents.

Longini *et al.* [9] develop a model for the number of cases in households from a homogeneous community. Initial infection is introduced in households from the community. Infected people in houses can spread the disease to other household members. The maximum likelihood method is used for the estimation of model parameters. Symptom data of influenza and common cold are used for modeling. It is seen that influenza spreads more easily in the community than in the house.

In Egypt, Khalil *et al*. [10] develop an agent-based model for simulation of the spread of pandemic influenza. There are different types of parameters used in the model such as social agent attributes, patterns of agents' interactions and Egypt population. According to model results, transmission patterns and the conditions that an outbreak occurs can be understood.

2.4. Metapopulation Model

The metapopulation model is like the agent based model a generalization of the stochastic SIR model with perfect mixing in a single population. The metapopulation model does not use mixing group details, but assumes perfect mixing in several large populations. The infection between populations occurs due to travelling of people between these populations.

During the literature review research, We have observed that there was no study considering an epidemic model based on the metapopulation approach in Turkey. Bektaş [11] studied the simulation of agent based disease spread models in Turkey. Therefore, we used a metapopulation approach for disease spread modeling in Turkey. Our metapopulation model is not only useful for Turkey, it can be used anywhere. Details regarding the metapopulation model we built can be seen in Section 3. Due to modeling travelling between cities, it is possible to observe disease spread behaviour and whether an outbreak occurs or not in the different cities.

Metapopulation approach is one of the most popular discrete-space modeling approaches. The concept of this approach is not originated in epidemiology, but in ecology. A metapopulation is a group of populations of the same species that leave in spatially isolated areas which are called patches and connected through migration. Patches can be cities, countries, islands and so on [12].

The movement between cities can be short or long term. People visit another city for a period of time and then turn back to their home city in short term movement which has been called Lagrangian movement; on the other hand long term movement which has been called Eulerian movement is seen when people move to another location and settle there. The models behind these types of movements are different [12].

There are n cities in metapopulation epidemic models. In each city, the population is assumed to be homogeneously mixing. Also, the population has typical epidemiological classes of susceptibles, infectives and recovered. The movement of the disease results from the travelling of infected people to different cities.

Vazquez [13] considers the intra- and inter-community interactions in metapopulation modeling in order to analyze global epidemics. In the model, a disease spreads inside a community by starting from an initial infected case due to the frequent intra-community interactions and outbreaks in a community is modeled. The population is divided into two types: locals who stay in a single community and social bridges which belong to different communities. Additionaly, there are some assumptions in the model such as

(i) All communities are statistically equivalent.
(ii) The mixing between locals and bridges is homogeneous.
(iii) Social bridges belong to two populations.

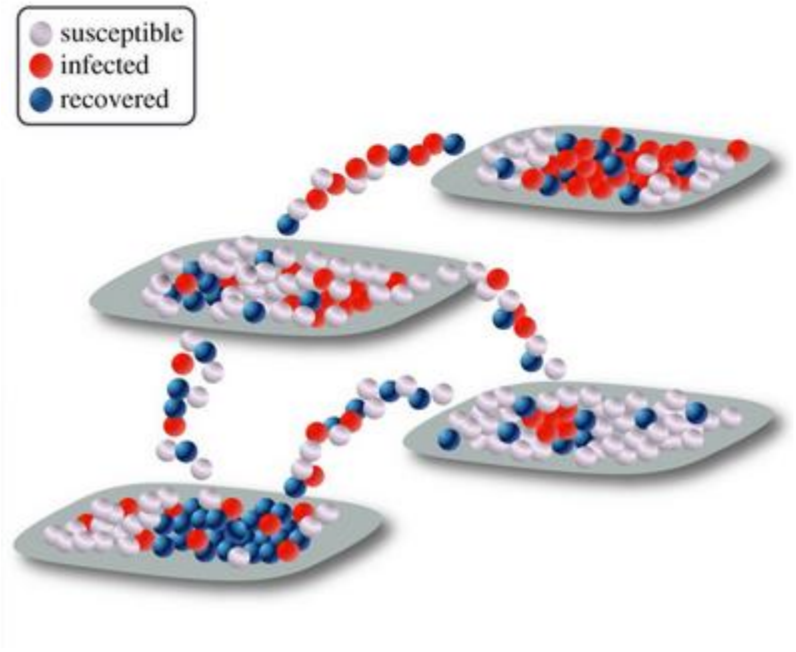

Figure 2.2. Structure of a metapopulation model.

Apolloni *et al.* [14] defines a multi-host stochastic epidemic metapopulation model according to a spatially structured population with non-homogeneous mixing and travel behaviour. In the model, different mixing patterns, population partitions and mobility structures are taken into consideration including also age partition for the early spread of the 2009 H1N1 pandemic influenza.

For their epidemic model Colizza *et al.* [15] use the airport network of major cities. The commuting patterns in inter- and intra-urban areas are characterized by networks whose nodes represent the elements of the system. In order to have a description of the system, deterministic reaction-diffusion equations are derived. Also the system is characterized by the standard epidemic threshold and a global invasion threshold providing the condition for the infection of a number of subpopulations. The first threshold defines the usual reproductive number $R_0>1$ that is just a function of the disease parameters, the second threshold defines a subpopulation reproductive number $R_*>1$ which depends on the diffusion rate of individuals among subpopulations. Lastly, the results are confirmed by Monte Carlo simulations for the infection dynamics in the metapopulation system.

Ajelli *et al.* [7] investigates the differences between the two main modeling approaches which are agent-based and metapopulation approaches. These models are developed independently and have been used in realistic scenarios and including actual data related to H1N1 pandemic. In the metapopulation approach, Global Epidemic and Mobility (GLEaM) model, which is based on high-resolution census data worldwide and integrating airline travel flow data with short-range human mobility patterns at the global scale , is used. In addition, in this model the age structure of Italy is considered. The geographical census areas centered on International Air Transport Association airports are defined by assigning each cell to the closest airport. In conclusion, the fraction of the population affected by the epidemic is larger in the metapopulation approach than in the agent-based approach due to the assumption of homogeneity and the lack of the detailed structure of contacts. Also, the detailed mobility networks in the metapopulation approach provide an information about the spreading of the disease at a global level.

Siska [16] carried out a project regarding disease dynamics in metapopulations consisting of two parts. The first part includes two different metapopulation models with well-mixed, equally-sized, coupled patches. The first model includes subpopulations with standard susceptible-infected-recovered model type. In the second model, a latent period is also included and migration exists between the patches, but migration of infected people is not allowed. In the second part, a stochastic, spatial and age-structured susceptible-exposed-infected model is developed.

2.5. Simulation Study of Real World Disease Spread

Simulation of disease spread is an important issue in order to understand the dynamics of the behaviour, outbreak process, potential effect of interventions and population properties. The most crucial part of the simulation is interpreting the results, because there can be several assumptions in the model. According to different scenarios for population and disease properties, the simulation requires multiple runs [17].

Worldwide, there have been designed various simulation models for disease spread including agent-based and metapopulation approaches. As explained in Section 2.4, during the literature review research, it is found that there was no simulation of a metapopulation model for Turkey but there has been research about this issue in several countries.

According to a study conducted by Juher *et al.* [18], Monte Carlo simulations of epidemic spread in uncorrelated networks are carried out. During the study, continuous-time equations in a susceptible-infected-susceptible model are used for the spread of infectious diseases in metapopulations. In the results, it is seen that in local populations the prevalence of the infection is not constant, but it increses with the patch connectivity.

Longini *et al.* [19] used a discrete time, stochastic influenza simulation model for typical Asian population to investigate several intervention strategies against the infection. Their aim is to represent the contact connectivity of Asian population. The model represents the number of close and casual contacts a person makes each day.

In Table 2.1, studies regarding the type of modelling approaches discussed during this section are summarized for each country. Different from other countries, Thailand is not a place in which studies are conducted but considered for modeling purposes with its population and disease properties. Also, studies carried out in Italy, Spain and USA took into account all cities in these countries and this detail is shown in the nationwide column in Table 2.1.

Table 2.1. Simulation studies in countries.

Country	Metapopulation model	Agent based model	Nationwide
Italy	[15], [7]	[7]	[15], [7]
Spain	[18]	[6]	[18]
Swedan	[16]		
USA	[13], [15]	[8], [9]	[13], [15]
Ireland		[6]	
Turkey		[11]	
Egypt		[10]	
Thailand		[19]	

According to Table 2.1, mainly for European countries, there are several studies regarding metapopulation and agent based models. In general the number of studies on agent based models is higher than for metapopulation models and there exists no study based on metapopulation models for Turkey. Therefore, we use metapopulation approach for disease spread modeling in Turkey.

In this chapter, SIR models, agent based models and metapopulation models are explained in detail. There are two types of SIR model which are deterministic and stochastic in the literature. In deterministic SIR model, there are continuous variables and process rates. In stochastic SIR model, there are discrete numbers and process probabilites. In this thesis, the generalization of the stochastic SIR model called metapopulation model with perfect mixing is used.

Agent-based model uses SIR approach to observe the behaviour of a disease spreading. The infection can be spread by contacts between students in schools, household members and colleagues in workplaces. Agent-based model is used to simulate the interactions of agents to understand their effects on whole system. Metapopulation model does not use mixing group details. The infection between populations occurs due to travelling of people between these populations.

3. DEFINITION OF OUR METAPOPULATION MODEL

In this section, we implemented a metapopulation model. The model is mainly composed of infection and travelling processes. Respectively, we explained the definition and algorithm of models including infection and travelling processes, the structure of the code we implemented and the necessary parameters used.

3.1. Model of the Infection Process

In this section, the disease spread model that is considered in this study is explained for a single population. The main idea is to simulate the disease in a large population in which all individuals have the same characteristics. We consider a dynamic model in discrete time. We use the SIR compartmental model, which is illustrated in Figure 1, According to this approach, the population is divided into different discrete compartments. In this model S denotes the number of susceptible people who have the potential of catching the disease. In the case of infection, they move into the infected compartment, denoted by I. It contains the number of symptomatically (IS) and asymptomatically (IA) infected people who can spread the disease to the susceptible people. The compartment of infected people is divided into two parts in order to simulate also the spread of asymptomatically infected individuals. This is important because asymptomatically infected people may have a big influence on intervention results. They carry the disease but do not show any symptoms. Some time later, infected people pass to the recovered compartment R in which people are assumed to be immune.

In addition to the structure of the model defined above, also some information is needed regarding the population considered in the model. First of all, the number of people in the population has to be defined. Then, in order to initialize the infection process, the number of infected people in the beginning is required. Also, the infection probability of symptomatic infected is required together with the infection probability of asymptomatically infected people.

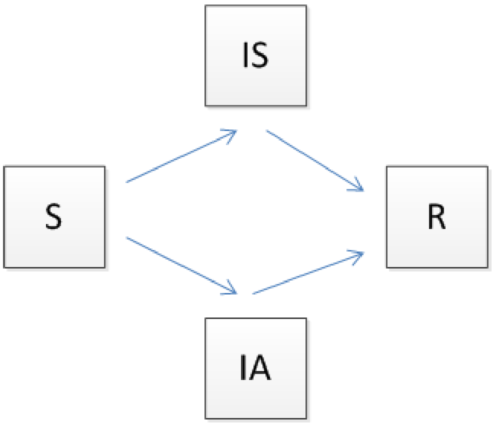

Figure 3.1. A SIR compartmental model.

We assume that the infection of all susceptibles are independent events and thus independent Bernoulli trials. When there is one infected person in the city, the probability of spreading the disease to susceptible people is calculated by the formula $P=1-[(1-p_{infection})^I]$ where $p_{infection}$ denotes the transmission probability and I denotes the number of infected people and equals to 1. The number of newly infected cases per day follows the binomial distribution. Any person who is infected has random, discrete disease time.

3.2. Main Idea of the Algorithm of the Infection Process for a Single Population

In this section, we will describe how to implement the SIR model. As described in Section 3.1, we use SIR compartmental model. In order to simulate the infection process, first of all a state vector is required. This vector stores the current number of susceptible, recovered, symptomatically (IS) and asymptomatically (IA) infected people. Infected people are divided into different compartments according to the disease time length, for example, we select the maximal disease time as three. In this case, extra three compartments for infected people will be stored in the state vector. The first compartment of infected people is for those infected in the first day. The second compartment is for people who will remain infected two more days. The third one is the compartment for people who will remain infected one more day. These extra infection compartments, which are stored in the state vector for symptomatically infected people, are also considered for the asymptomatically infected people. After construction of the vector, initial values such as the number of susceptible people and the number of infected people are assigned to the related compartments. In contrast to symptomatically infected people, usually zero is assigned to the compartment of asymptomatically infected people as an initial value.

We consider discrete simulation time. In a for loop, every day newly infected people will be generated and updates in the recovered, susceptible and infected compartments of the state vector occur. The loop continues till no infected people remain.

Before generating newly infected people, people who will remain infected one more day will enter to the recovered state. People who will remain infected two more days will enter to the state in which people will remain infected one more day. People who are infected in the first day are distributed to other states in the vector randomly by using the multinomial distribution. These states are respectively the recovered state, the state in which people will remain infected two more days and the state in which people will remain infected one more day.

Right after updating the states, selecting newly infected people among the susceptible ones has to be implemented using the binomial distribution. After selection, the assignment of newly infected people to the state vector is necessary. Before assignment of the newly infected people to the state vector, first of all the total number of newly infected people is divided into two, one for symptomatic infected people, the other one is for asymptomatic infected people, by using binomial distribution. After this division, these symptomatically and asymptomatically infected people are assigned to the states in which people are infected in the first day. We have decided to implement an assignment type in such a way that all people who are newly infected are assigned into a state in which people are infected in the first day.

Finally, the number of newly infected people is subtracted from the number of susceptible people. In the end of the for loop, the vector of the daily number of infected and the final number of susceptible and recovered people are returned as main result of this simulation.

3.3. Model of the Travel Process

In this section, we will describe how we model the travel process between cities and the random structure of travelling. We describe the daily flows between cities for each day by the expected number of people who travel. We also have to know the number of cities and the population sizes in each city.

In the literature of metapopulation models, there is no clear information regarding the details of the stochastic model for travelling . However, we tried to define necessary inputs in the modeling phase. We assumed that the maximal disease length and the staying time of travellers are similar as we considered only short time visits such as tourism and business travels. This is sensible as in the case of long journeys, visitors start to behave as the local people in that city. Therefore, we assume that people return to their home after some days. Also, we define a starting city in which infection people are introduced.

For each city R_0 is the same, but pinf values which are the probability of infection are different. Each individual has the same probability to travel. Send decisions of people to other cities are independent events. Each new sending of people to other cities follows the Binomial distribution and the probability of travelling is selected such that the desired expectation, given as input parameter, is realized.

3.4. Main Idea of the Algorithm of the Travel Process

The main idea of the algorithm of the travel process between cities is explained here for the case of three cities. Each city has the same characteristics in itself, however while travelling, a heterogeneity between cities occurs. In the beginning of the process, for defining the expected number of people who travel, it is assumed that all people travel with the same probability and all of them are in their home cities. Also, the expected number of people who travel between two cities in either direction is equal.

In order to calculate the probability values of travelling to other cities and staying there, for each city, an array is constructed. The rows of this array are related to the remaining defined as the stay duration of visitors in the cities, the columns hold for the home city and all other cities. Values included in this array are calculated via multiplying staying length probabilities of visitors in other cities by the binomial travel probabilities between cities. The last column is assigned to each home city and the value in it is calculated by substracting the sum of other values in the array from one.

In order to initialize the travel process, a state array is constructed with the dimension of three. The row represents the people in different compartments in each city. Columns represent the information of home city and visitors who are coming from other cities. The third dimension is defined for the number of cities. We need to know from which city these visitors come. The number of columns belong to visitors who are coming from other cities are constructed according to the staying time length. The related formula for the number of total columns is “1+maxstay*nc” where nc is the number of cities from where visitors come, maxstay is the staying length in other cities and 1 is for the home city.

In a for loop, for each day, the travelling process is repeated until end of the simulation period. In case of generating phase, multinomial distribution is used. By the help of this distribution, it is possible to send people from one city to the other two cities at the same time. Once the number of people who travel is selected, these values are distributed to columns of visitors with different staying lengths by using again a multinomial distribution. Before assignment to the array, first of all people who are in the last day of visiting return to their home cities. People who are staying two days go to the column in which people stay one day and so on. After this update, the number of visitors with different staying lengths are assigned to the related columns of the array.

3.5. Algorithm of the general code

In the previous sections, the main ideas of the algorithm of the infection process for a single city and the travel process for three cities were described. In this section, we aim to explain the steps of the combined algoritm which includes both infection and travelling, and thus allows to observe how infection spreads between cities due to travelling. In the same for loop, infection and travelling processes are carried out day by day. We need a state array in order to know how many people are visitor in different cities and how is their disease situation, therefore we constructed a common three dimensional state array.

In the state array, each row represents susceptible, recovered and infected people respectively. Each day, for each city and each compartment of travelling matrix, newly infected cases are generated from among susceptible people following binomial distribution.While generating, infected people who are visitors are taken into consideration as well as infected people in home cities, because we assume that visitors and local people are perfectly mixing. After generation, these newly infected people are assigned to the row in which people are infected in the first day. Afterward, they are distributed to recovered state, the state in which people will remain infected two more days and the state in which people will remain infected one more day by using multinomial distribution. Before this assignment, it is required to update each row. In other words, people who are infected in the last day are assigned to recovered state. People who will remain infected two more days are assigned to the state in which people will remain infected one more day.

After infection process in which generation of newly infected people and updating of state rows are carried out, travel process begins. In the state array, columns represent the home city and visitors. Each row of the state array for each city are sent to travel process at the end of each day in order to generate the expected number of people who are travelling. While generation phase, multinomial distribution is used. By the help of this distribution, it is possible to send people from one city to all other cities at the same time. When the expected number of people who travel is attained for each staying time, these values are assigned to related columns of visitors in the state array for each city. After this assignment, infection process begins again and this continues until there is no infected people.

3.6. State array

In this section, we give details regarding the construction of the state array in the code. Because of utilizing this structure through the study in all processes carried out, it plays a crucial role.

The array is in three dimensional. One of them is based on the SIR model compartments, one is for home cities and visitors, and the last one is for the different cities. The rows in the array are defined as SIR model compartments. The first row contains the susceptible people , the second row contains the recovered people and the remaining rows contain infected people which are divided into parts according to different disease lengths. In order to make it clear, for instance, if the disease length is three then the third row includes people who are infected in the first day, the fourth row includes people who will recover after two days and the last row includes people who will recover after one day. Columns in the array include the home cities and visitors which are coming from other cities. The first column is assigned to home cities. Others, which are constructed on the basis of stay lengths, are assigned to visitors.

Day by day, updates in the array occur with respect to rows and columns. According to rows, people who will be infected for one more day in the last row are assigned to the recovered state, the number of susceptible people decreases by subtracting the newly infected people. Newly infected ones are assigned to the infected state in which people are infected in the first day. In terms of columns, in the last day of staying, people automatically turn back to their home cities at the end of each day. After that, people who are new travellers are assigned to the visitor column that has the maximum stay length. Then, all visitor columns belong to each city are updated by assigning people to the visitor columns with different stay lengths.

3.7. Inputs and Outputs of the Simulation Function

In this section, the main function written in R, which is called RepSIRdi, is described. By the help of this function, simulation of disease spread in each subpopulation can be conducted. The function returns the total number of infected people for each day and subpopulation. In this code structure, the main processes are generating infected people and travelling of people between subpopulations at the end of each day.

Furthermore, the main inputs given in Table 3.1 for the main function are defined. By using these inputs, also some variables and structures which are required to construct the code are explained. For defining the number of people who travel, it is assumed that all people travel with the same probability and in the beginning all of them are in their own populations. Also, the expected number of people who travel between two populations in either direction is equal. For example the expected number of people who travel from subpopulation 1 to subpopulation 2 is the same as from subpopulation 2 to subpopulation 1. By using this input, a rate matrix holding the probability values of travelling between each city pair is calculated by using the formula “ratem=excm/N” where excm is the expected number of people who are travelling and N is the population size. After that, the probability mass function of days infected is used for calculating the expected disease time. Then, infection probabilites in each population are calculated by the help of formula in the following where diexpected is the expected disease length.

$$pinfV=R_0/(di_{expected}*npopV*(propsympt+(1-propsympt)*facpasym)) \tag{3.1}$$

Table 3.1. Main inputs and their definitions.

Input	Meaning
N	Number of independent repetitions of simulation.
D	Last day simulated.
npopV	Number of individual in populations.
R0	The basic reproduction number is used to measure the transmission potential of a disease (see Section 2.2).
I0	Number of infected people on day zero in all cities.
Pmfdi	Probability mass function of days infected.
Pmfdexc	Probability mass function of exchange.
excV	Expected number of people who travel between cities.
Propsympt	Proportion of symptomatic people in each city.
Facpasym	The ratio of infection probability in each city for people in asymptomatic state.
Pimm	Percentage of the population that vaccination is used.

In addition to these inputs described above, also some variables used in this R function are created by doing some calculations, transformations and using structures. The calculated and newly constructed variables can be seen in Table 3.2.

Table 3.2. Calculated variables in the code.

Variable	Meaning
Nc	The number of populations.
dimax	Length of probability mass function of days infected.
dvmax	Length of probability mass function of stay numbers in populations.
pinfV	Probability of being infected in each population.
Pexc	Rate matrix of travelling between populations.

3.8. Subfunctions of the Code

In order to carry out generation of infection and travelling processes, in addition to the main function also some subfunctions have to be coded. The code structures and methodologies of each function, inputs and their brief explanations can be found in the following sections.

3.8.1. Generate Newly Infected Function

The purpose of this function is to calculate and return the newly infected cases for each subpopulation. In order to generate newly infected ones, first of all the infection probabilities in each population are calculated. After calculation of probabilities, for each population by using binomial distribution, the newly infected cases in susceptible states are generated. In order to use in this function some important input variables are defined. These inputs are state array, probability mass function of infection, the number of populations, probability of being infected and the ratio of infection probability in each city for people in asymptomatic state respectively. By using these inputs, generation process of newly infected people begins. These main inputs of the function can be seen in Table 3.3.

Table 3.3. Inputs in the generation of infection process.

Input	Meaning
Sa	The state array includes infection states, population itself and visitors for each population.
Dimax	Length of probability mass function of days infected.
Nc	The number of populations.
pinfV	Probability of being infected in each population.
facpasym	The ratio of infection probability in each city for people in asymptomatic state.

The related code of the infection process can be seen in Figure 3.2.

```
generatenewlyinfected<-
function(sa,dvmax=3,nc=3,dimax=3,pinfV=c(0.001,0.002,0.003),facpasym=1/2)

{

newinf<-array(0,dim=c(dimax,1+(nc-1)*dvmax,nc))

for(x in 1:nc){

pinfection <- 1-(1-pinfV[x])^sum(sa[3:(dimax*2+2),,x],na.rm=TRUE)*(1-pinfV[x]*facpasym)^sum(sa[(3+dimax):(2+2*dimax),,x],na.rm=TRUE)

newinf <- rbinom( (1+(nc-1)*dvmax), size=sa[1,-((x-1)*dvmax+(1:dvmax)+1),x], prob=pinfection)

            }

  return(newinf)
```

Figure 3.2. Code structure of the infection generation process.

3.8.2. Prob for Send Function

The aim of this function is to calculate probability values of staying in other subpopulations for a length of time by using pmfdexc and pexc inputs defined in Table 3.4. One of the crucial things in this function is that while constructing a matrix with the dimension of stay length and number of populations, also one column has to be included for the population itself in order to calculate probabilities of people who will stay in their homes at the end of the day. This value is found by substracting the sum of other probability values from one in each population.

Table 3.4. Inputs of probforsend function.

Input	Meaning
Nc	The number of populations.
pmfdexc	Probability mass function of exchange.
Pexc	Rate matrix of travelling between populations.

The related code structure of the function can be seen in Figure 3.3.

```
probforsend<-function(nc,pmfdexc,pexc)
 {
 dvmax<-length(pmfdexc)
 probarray<-array(0,dim=c(dvmax,nc+1,nc))
 for(p in 1:nc){
  probarray[1:(dvmax),p,p]<-NA
   }
  for(i in 1:(nc-1)){
        for(j in (i+1):nc){
```

Figure 3.3. Code structure of probforsend function.

```
probarray[,j,i]<-pmfdexc*pexc[i,j]

probarray[,i,j]<-pmfdexc*pexc[j,i]

      }

  }

 for(k in 1:nc) {

probarray[1,nc+1,k]<-1-sum(probarray[,-k,k])

 probarray[2:dvmax,nc+1,k]<-NA

    }

  return(probarray)

  }
```

Figure 3.3. Code structure of probforsend function (cont.).

3.8.3. Send Function

By the help of this function, for each day, the number of people who travel between subpopulations are attained. In order to do that, first of all “probforsend” function defined in Section 3.8.2 is called for probability values of staying in other subpopulations for a length of time. After that, “generatesend function” is called for generating the number of people who travel to other subpopulations. Once these generated values are returned, the related parts of the sa array are updated. In order to generate the number of people who are travelling between cities, some significant inputs are defined and variables are calculated. Calculated variables are mainly about the generate send function which returns the number of people who travel to other populations and stay in population itself.

Inputs are the rate matrix of travelling between populations, the number of populations, probability mass function of exchange, length of probability mass function of stay number and the state array respectively. Inputs and calculated variables of the function can be seen in Table 3.5 and Table 3.6.

Table 3.5. Inputs for send function.

Input	Meaning
dvmax	Length of probability mass function of stay numbers in populations.
Sa	The state array includes infection states, population itself and visitors for each population.
Nc	The number of populations.
pmfdexc	Probability mass function of exchange.
pexc	Rate matrix of travelling between populations.

Table 3.6. Calculated variables in the send function.

Variable	Meaning
Trv	This array includes the number of people travel to other populations which come from the “generatesend function”
updtcity	This array includes the number of people who stay in population itself
Prb	Probability values of staying in other populations for a length of time

```
send<-function(dvmax,sa,nc,pmfdexc,pexc){

prb<-probforsend(nc=nc,pmfdexc=pmfdexc,pexc=pexc)

trv<-array(0,dim=c(length(sa[,(nc*dvmax+1),1]),nc*dvmax+1,nc))

for(p in 1:nc){

trv[,(p-1)*dvmax+(1:dvmax),p]<-NA }

for(t in 1:nc){

tr<-generatesend(sa[,1,t],prb[,,t],dvmax=dvmax,nc=nc)

trv[,-((t-1)*dvmax+(1:dvmax)),t]<-tr  }

updtcity<-array(0,dim=c(length(sa[,(nc*dvmax+1),1]),1,nc))

 for(l in 1:nc){

 updtcity[,1,l]<-trv[,nc*dvmax+1,l]  }
```

Figure 3.4. Code structure of send function.

```
for(n in 1:nc){
sa[,1,n]<-updtcity[,1,n]
trv<-trv[,-(nc*dvmax+1),] }
for(i in 1:(nc-1)){
for(j in (i+1):nc){
sa[,1,i]<-sa[,1,i]+sa[,(i-1)*dvmax+2,j]
sa[,((i-1)*dvmax+2):((i-1)*dvmax+dvmax),j]<-sa[,((i-1)*dvmax+3):((i-1)*dvmax+dvmax+1),j]
sa[,(i-1)*dvmax+dvmax+1,j]<-0
sa[,(i-1)*dvmax+(1:dvmax)+1,j]<-sa[,(i-1)*dvmax+(1:dvmax)+1,j]+trv[,(j-1)*dvmax+(1:dvmax),i]
sa[,1,j]<-sa[,1,j]+sa[,(j-1)*dvmax+2,i]
sa[,((j-1)*dvmax+2):((j-1)*dvmax+dvmax),i]<-sa[,((j-1)*dvmax+3):((j-1)*dvmax+dvmax+1),i]
sa[,(j-1)*dvmax+dvmax+1,i]<-0
sa[,(j-1)*dvmax+(1:dvmax)+1,i]<-sa[,(j-1)*dvmax+(1:dvmax)+1,i]+trv[,(i-1)*dvmax+(1:dvmax),j]
  } }
  return(sa)
 }
```

Figure 3.4. Code structure of send function (cont.).

3.8.4. Generate Send Function

This function enables us to generate the number of people travelling between subpopulations by the help of the rmultinomv probability distribution function created in R package ‘FLU’. The main purpose of using this distribution for each subpopulation is sending people to the different stay lengths in other subpopulations at the same time. Inputs of the function can be seen in Table 3.7.

Table 3.7. Inputs for generatesend function.

Input	Meaning
S	The state array only includes population itself for each state.
Prb	Probability values of travelling for each population.
Dvmax	Length of probability mass function of stay numbers in populations.
Nc	The number of populations.

```
generatesend<-function(s,prb,dvmax,nc)

{ prb<-prb[!is.na(prb)]

R<-matrix(0,nrow=dvmax,ncol=length(s  R<-rmultinomv(n=length(s),size=s,prob=prb)

RN<-t(R)

return(RN)

 }
```

Figure 3.5. Code structure of generatesend function.

In this section, up to this point the definition of infection and travelling processes, the algorithm of these processes, general code structure, main function of the code and the subfunctions including input parameters are mentioned in detail. First of all model of infection process for a single population is explained in detail. Compartments of SIR model that we constructed is defined. Our assumptions regarding the model and distributions that we use in the model are mentioned. Right after defining the model, the algorithm of the process is explained for a single population in detail before implementation of the code. After that, travelling process between cities and the algorithm are defined as in the infection process. Then, the code structure of functions is shown for each processes. In addition, information about the performance of the code regarding the duration of a run and the number of repetitions that we can do will be given in Section 6.

4. BEHAVIOUR OF A METAPOPULATION MODEL CONSISTING OF TWO LARGE POPULATIONS

In this section, in order to observe the behaviour of a metapopulation model, we carried out some simulation experiments with two large populations. Before doing simulations, parameters which are used in these experiments are defined in detail. Our aim is to find the number of daily infected, the total number of disease cases and the attack rates in each city by considering the daily flows between cities for each day by the expected number of people who travel. Due to the simulation results, the behaviour of the model can be observed. For instance, we can learn at which travel rate an outbreak in the second city occurs, which intervention strategies should be applied in order to prevent a pandemic and the day on which the maximum number of infected is observed in each city.

We conducted experiments with different expected number of people who travel between cities. We began with 5000 people who travel and thought that this number is also normal for Turkey. For instance, if 10,000 people exists in a city and the percentage of infected people is 5%, the total infected people in this city is equal to 500. With this number of infected people, 1.5 R_0 value and assuming perfectly mixing, high number of people will be infected in the second city most probably. Of course, if no infected people go to the second city there will not be an outbreak. We also carried out an experiment with 1 expected number of people who travel. In case of unluckiness, infected people may go to the second city. We want to know that with which expected number of travel, the probability of an outbreak will close to 0 or 1, so that we carried out experiments.

4.1. Parameters of the Experiment

Two large populations in which all individuals have the same characteristics are taken into consideration. In other words, cities are perfectly mixing. As always, a dynamic stochastic model in discrete time is considered. As mentioned in the previous sections we use the metapopulation model based on the SIR compartmental model.

In the modeling of infection, some information is needed regarding the two populations. One of these cities has 5,000,000 and the other one has 2,000,000 inhabitants. Influenza is introduced by assigning 10 infected people to the largest city. One of our most interesting input parameters is the expected number of people who are travelling. Maximum disease time is selected as 6 days and the probability mass function of the disease time is equal to (0,0,0.2,0.4,0.3,0.1) for the days 1,2,3,4,5 and 6. The expected disease time is calculated to be 4.3. Also, maximal staying time in other cities is selected as 4 days and probability mass function of staying time is defined as (0.1,0.2,0.3,0.4). In addition, proportion of symptomatic people in each city is defined as 0.66 and an asymptomatic infection is assumed to be 50% as infectious as symptomatic infection. All these assumptions are taken from the study of Longini *et al.* [8]. Simulations are carried out with 100 repetitions for 150 days. Also, for each city using the same R_0 value we calculate the probability of infection values by using the following formula (See Section 3.7 for the definitions of inputs in the formula).

$$\text{pinfV}=R_0/(\text{di}_{\text{expected}}*\text{npopV}*(\text{propsympt}+(1-\text{propsympt})*\text{facpasym})) \qquad (4.1)$$

4.2. Simulation Experiments and Results

We asked what will happen in the second city if the number of people who are travelling from first to second city is decreasing. Is there an outbreak? If we have large R_0 value and 10 started infected people, we know that in the first city the outbreak will be fast and we expect a behaviour similar to deterministic. If the number of expected people who are travelling is large, occuring an outbreak in the second city has a high probability. If the number of expected people who are travelling is too low, most probably an outbreak will not occur in the second city. In order to know whether an outbreak occur or not in the second city, we are doing experiments. The only parameter we can change is the number of expected people who are travelling between cities.

The experiments and their results are given in the following sections. According to these experiments, attack rates, total number of daily infected people, the day of maximum number of infected and total number of infected people so far are shown respectively for each city.

4.2.1. Standard Experiment with Expected Number of Travelling is 5000

As explained in the previous section, the behaviour of the metapopulation model is analyzed with two large cities, one of them has 5,000,000 and the other one has 2,000,000 population. The expected number of people who are travelling between cities is defined as 5000. This means that approximately 0.1% people in city 1 are visiting city 2 in a day. This parameter is chosen by considering the real travelling rates between cities in Turkey. In this experiment, the number of expected travelling is not reduced.

According to simulation with 100 repetitions, the attack rates of the two cities are calculated and shown in Figure 4.1 and Figure 4.2. Attack rate is defined as the ratio of the total number of infected and recovered people at the end of the day 150 over the susceptible people in the beginning. While calculating this rate, only infected and recovered people of the home city are taken into consideration. In other words, our aim is to find the real number of recovered people in each city at the and of the simulation. Visitors who are coming from other cities are ignored in order to avoid misleading results for each city.

As seen in Figure 4.1 and Figure 4.2, in city 1 the attack rates are all close to 0.75 value for each repetition of 100. This means that approximately 75% of 5 million people are infected in city 1. We are not surprised because this result shows us that since the population size is large enough, reproductive number (R_0) is greater than 1, so chosen as 1.5, it is inevitable to observe an outbreak in city 1. Also, in city 2 attack rates are close to 0.75, which means approximately 75% of 2 million people are infected. The variation of the attack rate values is very small. The relation between R_0 and attack rate is given in the following formula where S_0 is the number of susceptible people at time zero, S_∞ is the number of susceptible people at time ∞, N is size of the population and $[1- S_\infty/N]$ is the attack rate [20].

$$\ln(S_0/S_\infty)= R_0\ [1- S_\infty/N] \tag{4.2}$$

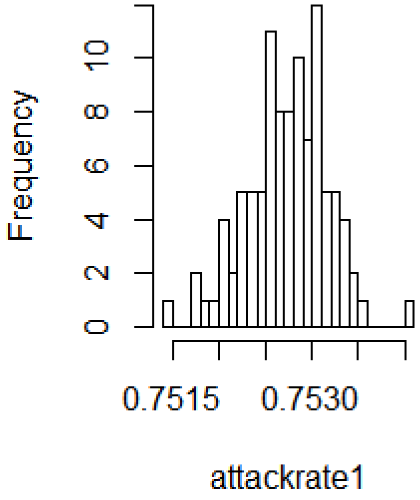

Figure 4.1. Attack rate of city 1.

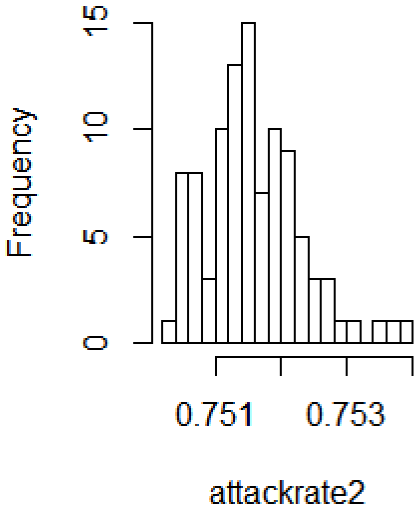

Figure 4.2. Attack rate of city 2.

In Figure 4.3, for each city the number of daily infected people is shown including 10 different paths and the average number of infected people for each day that is shown with a red line. While calculating the number of daily infected cases, both home cities' infected people and infected visitors in the city are taken into account. We expect an outbreak in the first city because we use a high R_0 value which is too effective on producing newly infected cases. In the second city, starting of the infection will be late comparing to first city. According to results, an outbreak is observed in both cities which is not surprising us; but it does not begin at the same time. In the second city, it begins later.

Approximately after day 50, 500,000 people are infected at most in city1; however in city2 after day 50 approximately 200,000 people are infected at most. The relationship between day of maximum number of infected of two cities can be shown also in Figure 4.4. In general, outbreak occurs in city 1 previously as we expected, but as seen in the plot some repetitions show us that the time of beginning an outbreak is almost the same in both cities.

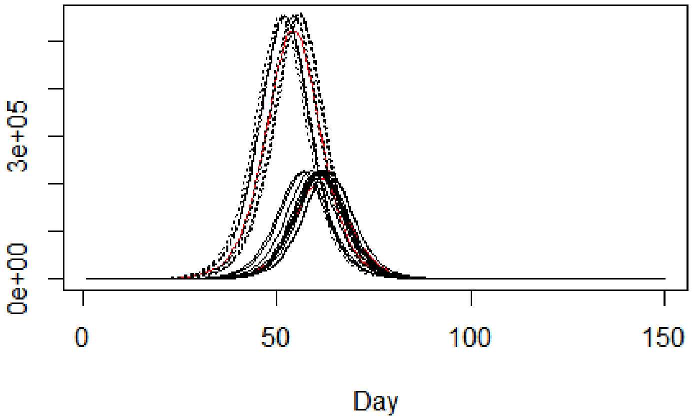

Figure 4.3. Number of daily infected cases in cities.

In Figure 4.5, the total number of infected people in both cities are shown. Approximately 4,000,000 people in city 1, 1,500,000 people in city 2 are being infected. As explained in attack rate calculation, while calculating the total number of infected cases only home cities' infected and recovered people are taken into consideration. Visitors coming from the other city are ignored. After peak time in all cities, the number of infected cases are starting to be constant as there are no new infections.

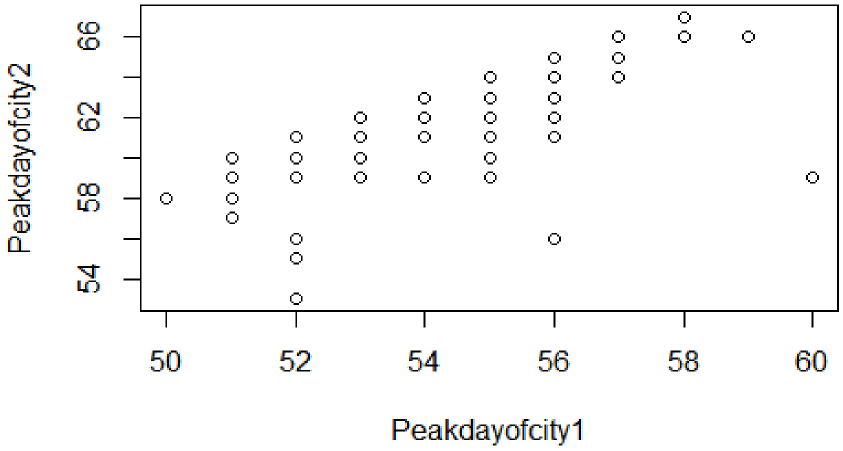

Figure 4.4. Peak time of infection in cities.

In Figure 4.5, the total number of infected people in both cities are shown. Approximately 4,000,000 people in city 1, 1,500,000 people in city 2 are being infected. As explained in attack rate calculation, while calculating the total number of infected cases only home cities' infected and recovered people are taken into consideration.

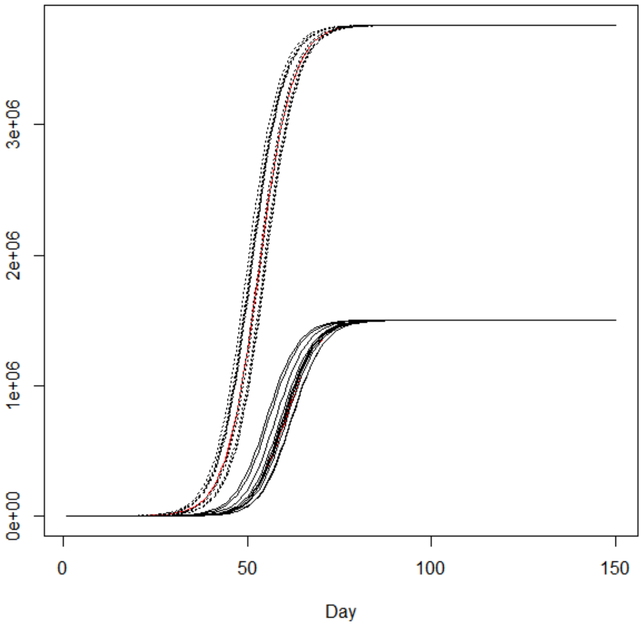

Figure 4.5. Total number of infected people in cities.

4.2.2. Experiment 2 with Expected Number of Travelling is 50

In this experiment, the only difference from the previous experiment is the expected number of people who are travelling. All other input parameters remain the same. Our aim is to observe the infection behaviour in the case of low travelling rates between cities, therefore the expected number of travelling people is decreased from 5000 to 50. This means that approximately 0.001% people in city 1 are visiting city 2 in a day.

Attack rates of all cities are calculated to see the disease spread behaviour. Attack rates are similar in both cities. As seen in Figure 4.6 and Figure 4.7, as we expected, in city 1 the attack rates are again close to 0.75 for each repetition. This means that approximately 75% of 5 million people are infected in city 1 and outbreak is still observed. In city 2, attack rates are also close to 0.75. Since the expected number of people who are travelling between cities are chosen as 50 and existence of an outbreak in city 1, a considerable number of infected people are going from city 1 to city 2, therefore an outbreak is observed also in city 2.

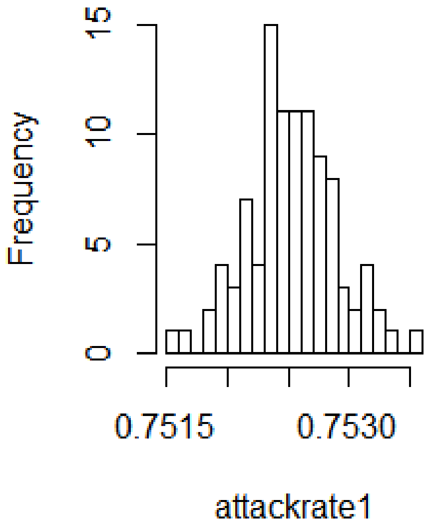

Figure 4.6. Attack rate of city 1.

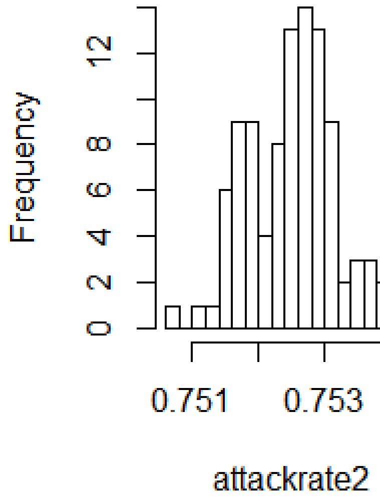

Figure 4.7. Attack rate of city 2.

According to results in Figure 4.8, an outbreak is observed in both cities; but it does not begin at the same time. In the second city, it begins later. Contrary to the previous experiment, outbreak in the second city is delayed longer, on average the delay is about 25 days as seen in Figure 4.9.

Approximately after day 50, 500,000 people are infected at most in city1, however in city2 after day 90 approximately 200,000 people are infected at most. The relationship between day of maximum number of infected of two cities can be shown also in Figure 4.9.

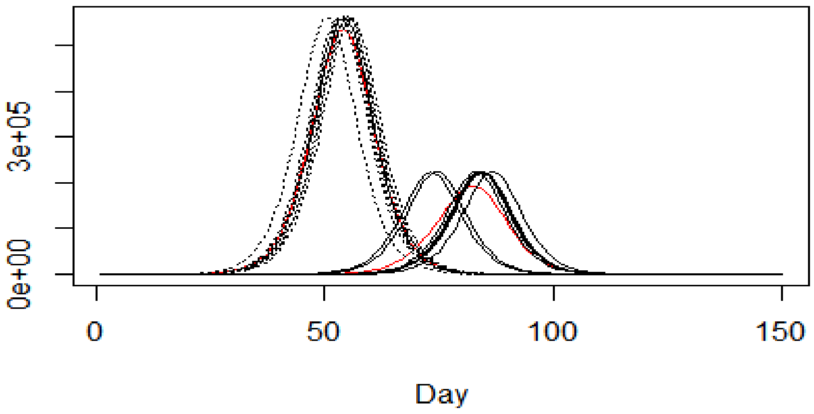

Figure 4.8. Number of daily infected cases in cities.

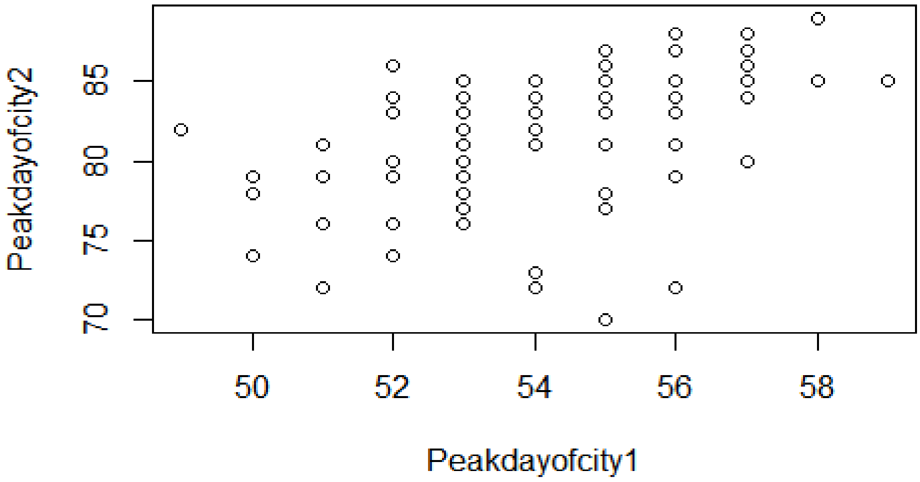

Figure 4.9. Peak time of infection in cities.

In Figure 4.10, the total number of infected people in both cities are shown. Approximately 4,000,000 people in city 1, 1,500,000 people in city 2 are being infected. As explained in the previous plot, due to decreased expected number of people who are travelling, in the second city peak times in each repetition are greater than that in the standard experiment.

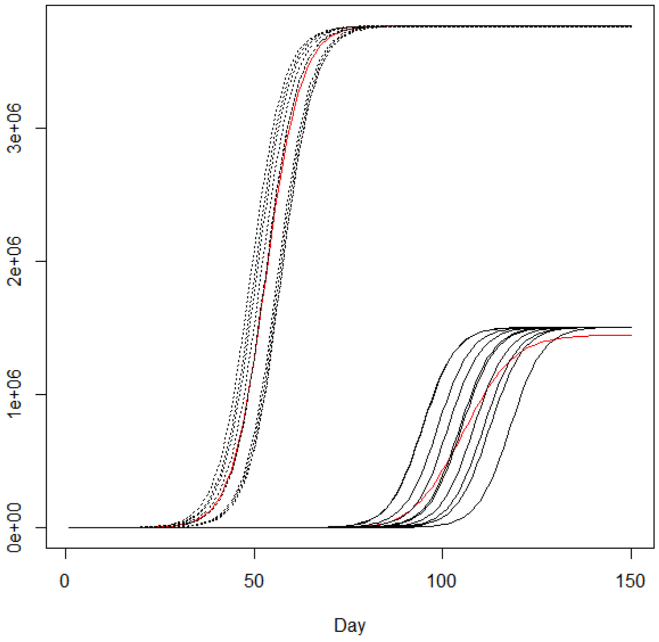

Figure 4.10. Total number of infected people in cities.

4.2.3. Experiment 3 with Expected Number of Travelling is 1

In this experiment, the expected number of travelling people is decreased from 50 to 1. This means that approximately 0.00002% people in city 1 are visiting city 2 in a day. Apart from that everything regarding input parameters remains the same. It is expected that due to high decrease in the expected number of people who are travelling, in the second city the probability of an outbreak will be low. In city 1 outbreak will still occurs.

As in the previous experiments attack rates are calculated in both cities. According to results in Figure 4.11, there is no change in attack rates of city 1 as expected. They are still distributed around 0.75 value for each repetition of 100. This means that approximately 75% of 5 million people are infected in city 1 and outbreak is still observed.

When we compare city1 with city2, Due to the decrease of the expected number of people who are travelling, approximately in 5% of repetitions no outbreak occurs in city2. In the remaining repetitions, we still observe an outbreak in city 2. In more than 80% of repetitions we see an outbreak in city2.

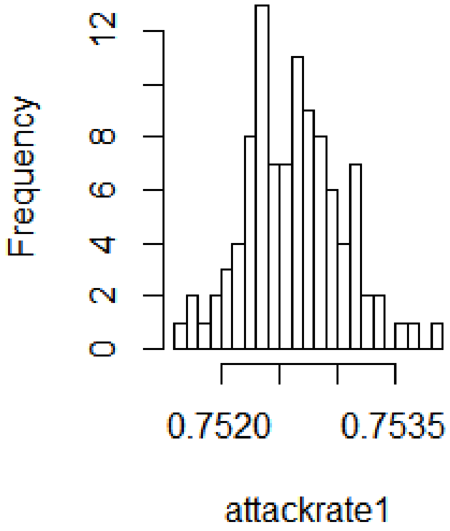

Figure 4.11. Attack rate of city 1.

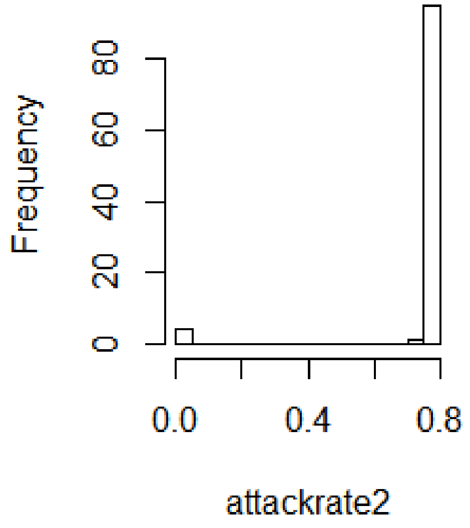

Figure 4.12. Attack rate of city 2.

According to results in Figure 4.13, an outbreak is in most cases observed in both cities; but it does not begin at the same time. In the second city, it begins later. Contrary to previous experiment, the outbreak in the second city begins later. Also, the histogram of the attack rates of city 2 shows that there is no outbreak in some repetitions. Approximately after day 50, 500,000 people are infected at most in city1; however in city2 after day 100 approximately 200,000 people are infected at most. The relationship between peak times of two cities can be shown also in Figure 4.14.

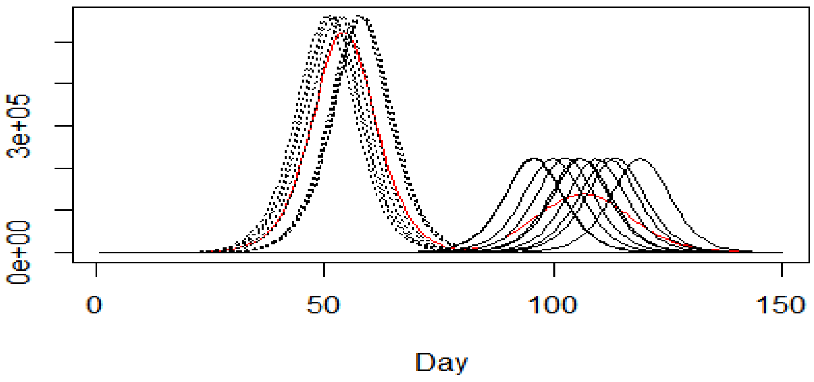

Figure 4.13. Number of daily infected cases in cities.

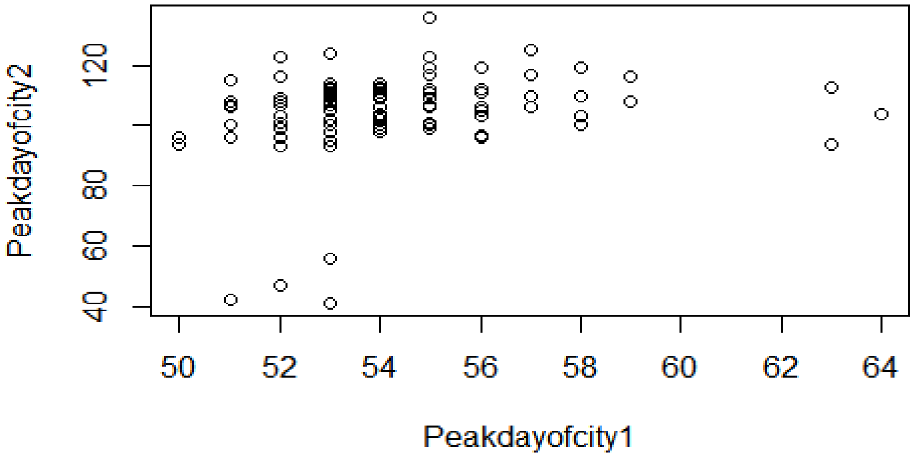

Figure 4.14. Peak time of infection in cities.

In Figure 4.15, the total number of infected people in both cities are shown. As explained in the previous plot, due to decreased expected number of people who are travelling, in the second city peak time is greater than in the previous experiments, but in some repetitions, peak times of the first and second city are almost the same or peak time of the first city is greater than the second city.

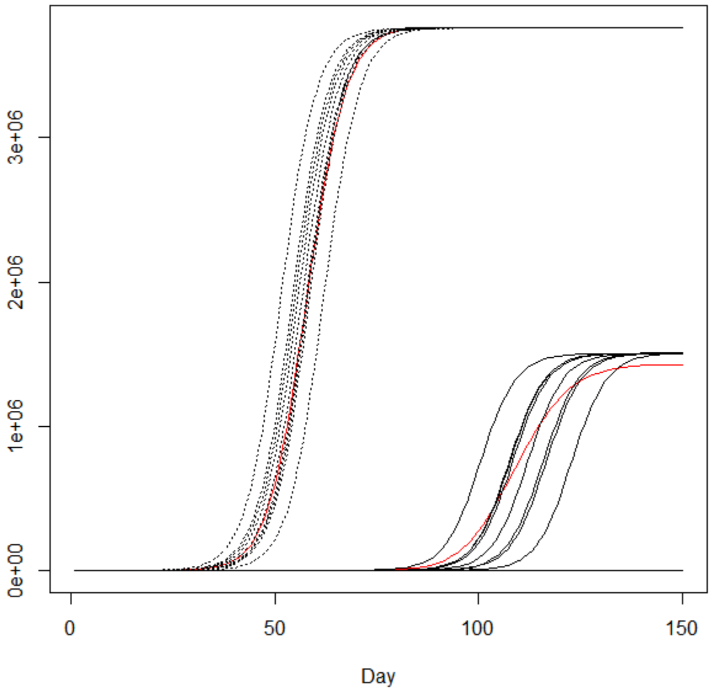

Figure 4.15. Total number of infected people in cities.

4.2.4. Experiment 4 with Expected Number of Travelling is 0.1

In this experiment, the expected number of travelling people is decreased from 1 to 0.1. This means that approximately 0.000002% people in city 1 are visiting city 2 in a day. Apart from that everything regarding input parameters remains the same. We expected that in city 1 there will be no change regarding outbreak occurrence.

In Figure 4.16, there is no change in attack rates of city 1. They are still distributed around 0.75 value for each repetition of 100. This means that approximately 75% of 5 million people are infected in city 1 and outbreak is still observed. However in city 2, due to high decrease in the expected number of people (0.1 which is less than 1) who are travelling, in the great majority of repetitions (approximately 80%) an outbreak is not observed. But in some repetitions (approximately 20%), there is an outbreak with attack rates close to 0.75.

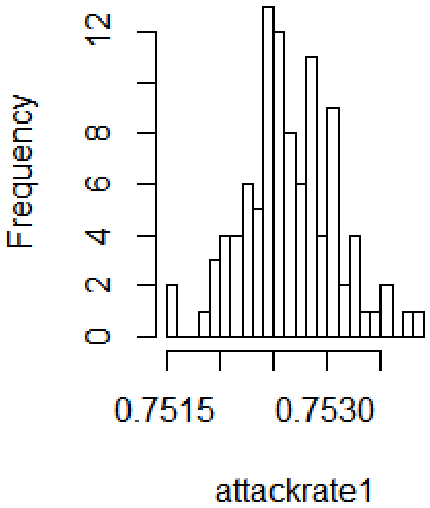

Figure 4.16. Attack rate of city 1.

According to results in Figure 4.18, an outbreak is observed in city 1, but in the second city, an outbreak is not observed in the great majority of repetitions. Due to these cases in which an outbreak is not observed, the number of infected people in average are too low to compare single paths.

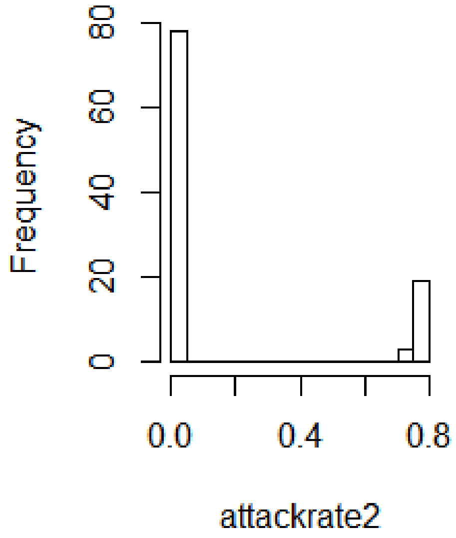

Figure 4.17. Attack rate of city 2.

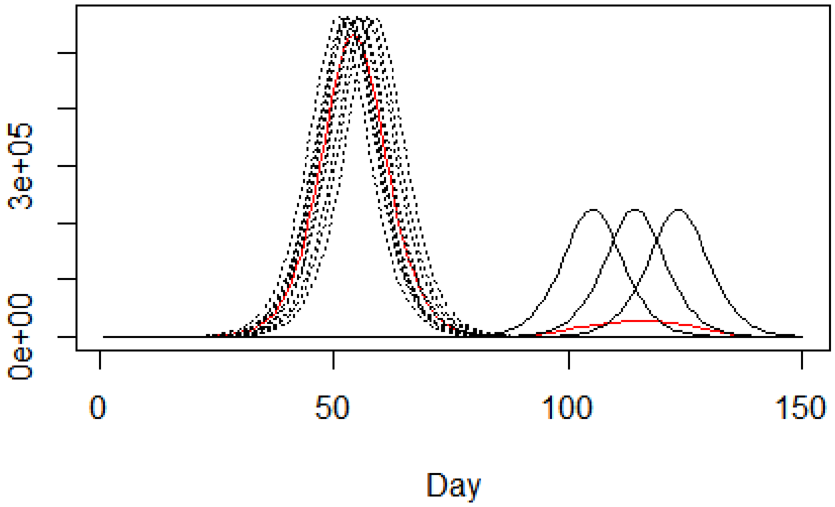

Figure 4.18. Number of daily infected cases in cities.

In Figure 4.19, the total number of infected people in both cities are shown. Approximately 4,000,000 people in city 1 are being infected. Due to high decrease in the expected number of people who are travelling, in some repetitions there are no infected cases are observed.

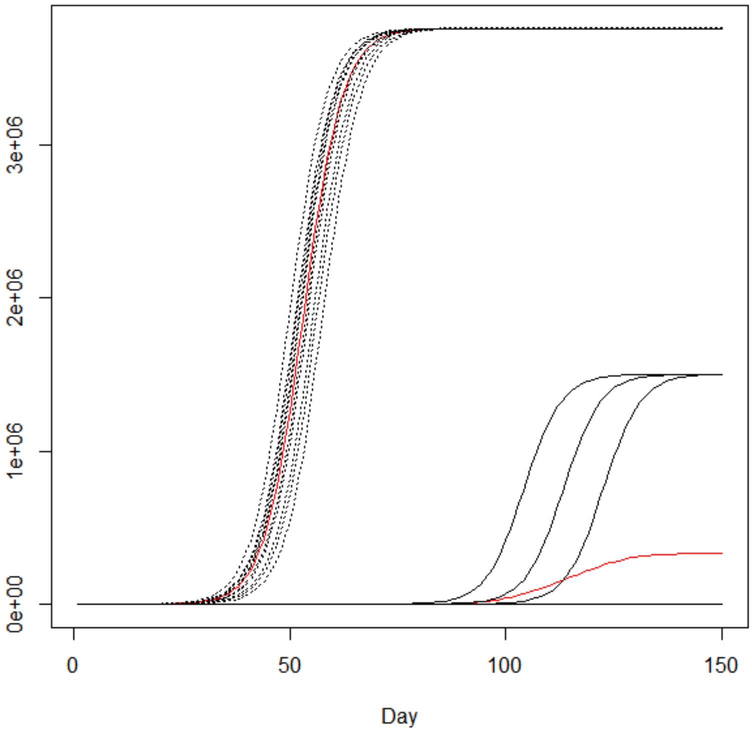

Figure 4.19. Total number of infected people in cities.

According to results of experiments which are conducted above, with the decreasing travel rate, the probability of an outbreak in the second city also decreases. Also delays will be longer in the second city for each decreased expected number of people who are travelling between cities.

In this section, we mainly conducted four artificial simulation experiments according to different expected number of people who are travelling. We tried to show the results by using different plots such as attack rate plots, total number of infected people and number of daily infected people. We observed the disease behaviour by reducing the number of travelling between two large cities in each experiment. Although the number of travelling is decreased considerably, it is difficult to prevent an outbreak occurrence in cities. We saw that if an outbreak occurs in a city, the disease will most probably spread to other one unless the travelling rate is extremely low.

5. BUILDING METAPOPULATION MODELS FOR TURKEY

In this section, we built three different metapopulation models which are including three cities, six cities and ten cities for Turkey. We considered large cities in order to observe the disease spread behaviour between the major metropolitan areas. We want to learn that if a pandemic arises in the world, how it will affect Turkey. İstanbul has one of the largest airports of the world and very likely a pandemic enters into Turkey from this city. We collected information about cities only from the internet. We investigated several transportation services in order to gather information about travelling between cities.

5.1. Parameters and Data Sources for Models

While doing simulations, we need some parameters regarding Turkey. İstanbul has the biggest airport in Turkey, so we assumed that during the first 50 days on average every day 1 infected person comes to İstanbul from abroad. In this way influenza is introduced into İstanbul. The basic reproductive number (R_0) is selected as 1.5. Maximum disease time is selected as 6 days and pmf of disease time is defined as (0,0,0.2,0.4,0.3,0.1).

People will remain infected three days with probability of 0.2, four days with probability of 0.4, five days with probability of 0.3 and six days with probability of 0.1. The maximal staying time in other cities is selected as 4 days and pmf of staying time is defined as (0.1,0.2,0.3,0.4). This means that people will stay in other cities 1 day with probability of 0.1, 2 days with probability of 0.2, 3 days with probability of 0.3 and 4 days with probability of 0.4.

In order to estimate the expected number of people who travel between these cities, as a transportation type daily bus and airplane services were investigated. Two airlines [21,22] and three bus companies [23-25] which are the most preferred and have a considerable extent of passenger capacity in Turkey were taken into consideration. For identifying the number of flights and bus services, the related firm's reservation systems were examined in detail for each city pairwise. For each transportation type, daily travelling flow information between all cities were gathered. In order to prevent misleading results, summer season travels, which are mostly preferred by people for short-term holiday purposes, are ignored. The results according to both transportation types can be seen in Figure 5.1 and Figure 5.2.

Cities	İstanbul	İzmir	Muğla	Antalya	Ankara	Adana	Gaziantep	Van	Trabzon	Samsun	Kayseri
İstanbul	-	20	10	15	30	15	8	7	9	6	7
İzmir	-	-	6	3	10	0	0	0	0	0	0
Muğla	-	-	-	0	0	0	0	0	0	0	0
Antalya	-	-	-	-	3	0	0	0	0	0	0
Ankara	-	-	-	-	-	4	4	4	4	2	1
Adana	-	-	-	-	-	-	0	0	0	0	0
Gaziantep	-	-	-	-	-	-	-	0	0	0	0
Van	-	-	-	-	-	-	-	-	0	0	0
Trabzon	-	-	-	-	-	-	-	-	-	0	0
Samsun	-	-	-	-	-	-	-	-	-	-	0
Kayseri	-	-	-	-	-	-	-	-	-	-	-

Figure 5.1. Number of rounds (airplane) between cities.

Cities	İstanbul	İzmir	Muğla	Antalya	Ankara	Adana	Gaziantep	Van	Trabzon	Samsun	Kayseri
İstanbul	-	20	6	10	20	10	4	2	3	8	6
İzmir	-	-	6	5	6	1	1	1	3	2	4
Muğla	-	-	-	3	7	1	0	0	0	0	0
Antalya	-	-	-	-	7	6	2	0	6	7	7
Ankara	-	-	-	-	-	10	2	8	7	15	16
Adana	-	-	-	-	-	-	6	4	6	6	4
Gaziantep	-	-	-	-	-	-	-	2	3	8	6
Van	-	-	-	-	-	-	-	-	2	2	9
Trabzon	-	-	-	-	-	-	-	-	-	20	5
Samsun	-	-	-	-	-	-	-	-	-	-	8
Kayseri	-	-	-	-	-	-	-	-	-	-	-

Figure 5.2. Number of rounds (bus) between cities.

In Figure 5.1, “0” means that there is no direct flight between city pairs, so in these cases the expected number of people who are travelling between cities can be calculated only considering buses. However, in Figure 5.2 “0” means that there is no bus round between cities such as between Muğla and Gaziantep, Van, Trabzon, Samsun and Kayseri.

While estimating the total expected number of people who are travelling between these cities, some assumptions are taken into account.

(i) Only direct flights between cities are considered.
(ii) Seat occupancy rate for airplane transportation type is assumed to be 80 %, for bus 40 %.
(iii) The number of available seats on any given connection is assumed to be 150 for an airplane and 45 for a bus.
(iv) The expected number of people who travel between two cities in either direction is equal.

There are not only two bus companies and three airlines in Turkey. Therefore, in order to take other bus firms and airlines into account for the number of people who are travelling, the estimated expected number of people who are travelling by bus is multiplied by 2 and the estimated expected number of people who are travelling by airplane is multiplied by 1.1. Also, it was not possible to reach the informaton about the number of people who are travelling by car, so in order to consider people who are travelling by car between cities, final results were multiplied by 2 for cities which are close to (below 200 km) each other and by 1.5 for distant cities. In Table 5.1, the calculation method for the largest three cities which are İstanbul, İzmir and Ankara is shown. It is seen in Table 5.1 that the number of people who are travelling between İstanbul and Ankara is higher than those between İstanbul and İzmir and Ankara and İzmir. Also, final expected number of people who are travelling between cities are shown in Table 5.2. It is seen in Table 5.2 that the number of people who are travelling between large cities is the highest. In addition, a travelling flow is not observed between some cities such as Gaziantep-Muğla.

Table 5.1. Travel flow between cities.

Cities	Transportations						
	Airplane	Bus	Airplane	Bus	Total	Total (considering car)	Total (final)
İstanbul-İzmir	20*150*0.80*2*1.1	20*45*0.40*4*2	5280	2880	8160	8160*1.5	12240
İstanbul-Ankara	30*150*0.80*2*1.1	20*45*0.40*4*2	7920	2880	10800	10800*2	21600
Ankara-İzmir	10*150*0.80*2*1.1	6*45*0.40*4*2	2640	864	3504	3504*1.5	5256

Table 5.2. Travel flow between cities.

Cities	İstanbul	İzmir	Muğla	Antalya	Ankara	Adana	Gaziantep	Van	Trabzon	Samsun	Kayseri
İstanbul	-	12240	5256	8100	21600	8100	4032	4932	6588	4968	5148
İzmir	12240	-	4896	3024	5256	216	216	216	648	432	864
Muğla	5256	4896	-	864	1512	216	0	0	0	0	0
Antalya	8100	3024	864	-	2700	1296	432	0	1296	1512	1512
Ankara	21600	5256	1512	2700	-	3744	2016	3312	3096	5376	3852
Adana	8100	216	216	1296	3744	-	1728	864	1296	1296	1152
Gaziantep	4032	216	0	432	2016	1728	-	576	648	1728	1728
Van	4932	216	0	0	3312	864	576	-	432	432	1944
Trabzon	6588	648	0	1296	3096	1296	648	432	-	4320	1080
Samsun	4968	432	0	1512	5376	1296	1728	432	4320	-	1728
Kayseri	5148	864	0	1512	3852	1152	1728	1944	1080	1728	-

5.2. Building a Metapopulation Model for İstanbul, Ankara and İzmir

In this section, in order to observe the behaviour of metapopulation approach in Turkey, we designed a model with three cities which are İstanbul, Ankara and İzmir. These cities have the largest populations and importance in terms of their centricity, preferred by many tourists to visit, and people to work or study. Therefore, these cities are good choice for observing disease spread behaviour in Turkey and also selecting the correct parameters for the code that we implemented.

İstanbul is the most populous and largest city with a population of 14,992,171 in Turkey and the country's economic, cultural, and historic center. Approximately 12.56 million foreign visitors arrived in İstanbul in 2015, five years after it was named a European Capital of Culture, making the city the world's fifth most popular tourist destination. Therefore, İstanbul has an importance for Turkey to investigate the disease spread behaviour.

Ankara is the capital and second largest city in Turkey with a population of 5,630,998 after İstanbul. It is the administrative center of Turkey and a huge university town, so it has a large population of government workers and university students. İzmir is a metropolitan city in the western extremity of Anatolia and the third most populous city in Turkey after İstanbul and Ankara with a population of 5,496,889. Especially in summer terms, it hosts a huge number of local and foreign tourists every year. Cities which are close to İstanbul, Ankara and İzmir are also taken into account. İzmir is combined with Manisa and Aydın, İstanbul is combined with Bursa and Çorlu, Ankara is combined with Çankırı and Kırıkkale. Due to high the number of people who are travelling each day between these cities, it is logical to take them into consideration as a single city. Therefore, it is assumed that disease dynamics are the same and a perfect mixing occurs in each combinaton. After this combination, the final population sizes for the three cities are shown in Table 5.3.

Table 5.3. Population sizes of the three cities.

Cities	Population
İstanbul (with Bursa and Çorlu)	18,123,801
Ankara (with Çankırı and Kırıkkale)	6,082,669
İzmir (with Manisa and Aydın)	7,943,430

5.3. Building a More Detailed Metapopulation Model for Turkey

In this section, we designed a model with more than three cities which are İstanbul, İzmir, Muğla, Antalya, Ankara, Adana, Gaziantep, Van, Trabzon, Samsun and Kayseri in order to observe the behaviour of metapopulation approach in Turkey. These cities include the largest airports in Turkey. Different from the model described in Section 5.2, we wanted to cover all regions in Turkey, therefore some cities that have no airport are assigned to the closest airport as long as the distance between cities and the airport is less than 200 km. Choosing these regions, approximately 59% of Turkey can be investigated, so that we have a chance to observe the dynamics of the disease in a large part of the country. Airport centers and cities in each center can be seen in Table 5.4.

5.4. Six Cities Model

In this section, we designed a model with six cities which are İstanbul, Ankara, İzmir, Muğla, Antalya and Adana in order to observe the behaviour of a metapopulation model for Turkey. These cities include the six airports in Turkey which have the largest passenger transportation capacity. Some cities that have no airport are assigned to the closest airport as long as the distance between these cities and the airport is less than 200 km. We observed that every day a great number of people are travelling between these cities and it is logical to combine them instead of evaluating them seperately. For example people who are living in İzmit and Adapazarı usually use the airport in İstanbul to travel other cites, so they are taken into consideration as one population. We observed that the number of people who are travelling in cities which are close to each other and large is higher than those which are far away from each other and less central. Airport centers, cities in each center and distance between cities can be seen in Table 5.5 and Table 5.6 respectively.

Table 5.4. Population sizes of airport centers.

Airport centers	Cities	Total population size
İstanbul	İstanbul+İzmit+Adapazarı+Bursa	18,123,801
İzmir	İzmir+Manisa	6,880,706
Muğla	Muğla+Aydın	1,986,370
Antalya	Antalya+Burdur	2,481,224
Ankara	Ankara+Bolu+Çankırı+Kırıkkale	6,111,924
Adana	Adana+Mersin+Osmaniye+Tarsus	4,801,036
Gaziantep	Gaziantep+Kahramanmaraş	3,077,395
Van	Van+Ağrı	1,656,623
Trabzon	Trabzon+Rize	1,102,472
Samsun	Samsun+Sinop	1,486,127
Kayseri	Kayseri+Malatya	2,133,621

Table 5.5. Population sizes of airport centers.

Airport centers	Cities	Total population size
İstanbul	İstanbul+İzmit+Adapazarı	15,607,324
İzmir	İzmir+Manisa	6,880,706
Muğla	Muğla+Aydın	1,986,370
Antalya	Antalya+Burdur	2,481,224
Ankara	Ankara+Bolu+Çankırı+Kırıkkale	6,111,924
Adana	Adana+Mersin+Osmaniye	4,471,036

Table 5.6. Distance between cities.

Cities	Distance<200 km
İstanbul-İzmit	117 km
İstanbul-Adapazarı	162 km
Ankara-Bolu	191 km
Ankara-Çankırı	131 km
Ankara-Kırıkkale	83 km
İzmir-Manisa	38 km
Muğla-Aydın	99 km
Antalya-Burdur	122 km
Adana-Mersin	83 km
Adana-Osmaniye	86 km

5.5. Ten Cities Model

In this section, we designed a model with ten regions in Turkey which are Southeastern Anatolia, Eastern Anatolia, Eastern Black Sea, Central Black Sea, Marmara, Aegean, Mediterranean, East Mediterranean, Central Anatolia and East Central. In each region, the two largest cities are taken into consideration in order to observe the behaviour of the metapopulation approach. Choosing these regions, approximately 59% of the Turkish population is included in the simulation. The list of regions and cities can be seen in Table 5.7.

Table 5.7. Population sizes of each region.

Region	City	Total population
Southeastern Anatolia	Gaziantep,Kahramanmaraş	3,077,395
Eastern Anatolia	Van,Ağrı	1,656,623
Eastern Black Sea	Trabzon,Rize	1,102,472
Central Black Sea	Sinop,Samsun	1,486,127
Marmara	İstanbul,Bursa	17,870,250
Aegean	İzmir,Aydın	6,559,613
Mediterranean	Antalya,Burdur	2,481,224
East Mediterranean	Mersin,Adana	3,952,393
Central Anatolia	Ankara,Kırıkkale	5,900,593
East Central Anatolia	Kayseri,Malatya	2,133,621

6. SIMULATION EXPERIMENTS FOR TURKEY

In this section, we conducted simulation experiments for the three models of Turkey explained in Section 5. We learned from Section 4 that for frequent travel between cities and R_0 equals to 1.5, a pandemic spreads to all cities. Differently from the experiments in Section 4, we now consider also vaccination to see how it will affect the behaviour of the disease spread. We also try to find out what is the appropriate percentage of vaccination that we should use.

6.1. Experiment for Three Cities Baseline Model

In this section we did experiments for three cities with not using any vaccination. In all cities R_0 is assumed to be 1.5. The number of people who are travelling between cities is high. Thefore, there is high probability for an outbreak in all cities. In order to observe the behaviour of the model, we analyzed attack rates, number of daily infected, total number of infected and the disease length for all cities. In addition to these analyses, we want to observe the maximal number of infected days (peak days) of infection in each city. 1000 repetitions are conducted. All other inputs and assumptions used in the experiment are defined in Section 5.2.

Average attack rates in all cities are close to 0.75 for each repetition. As seen in Table 6.1 below, the model behaves similar to a deterministic model. Approximately 75% of people are infected in all cities. We are not surprised because the travelling flows between cities are high , R_0 is greater than 1 and the cities are large. Therefore, an outbreak occurs always in all cities.

In Figure 6.1, the daily and total number of infected people in two different repetitions are shown. An outbreak is observed in all cities. Outbreak begins in İstanbul firstly and continues with Ankara and İzmir in all examples. Only slight changes are observed in the time of the outbreak. Also, it is seen that the peak days of infection do not change considerably in the different repetitions.

Table 6.1. Attack rates of cities.

Average attack rate	İstanbul	Ankara	İzmir
Maximum	0.7536312	0.7536213	0.7533000
Average	0.7530125	0.7515816	0.7516806
Minimum	0.7523348	0.7502591	0.7505933

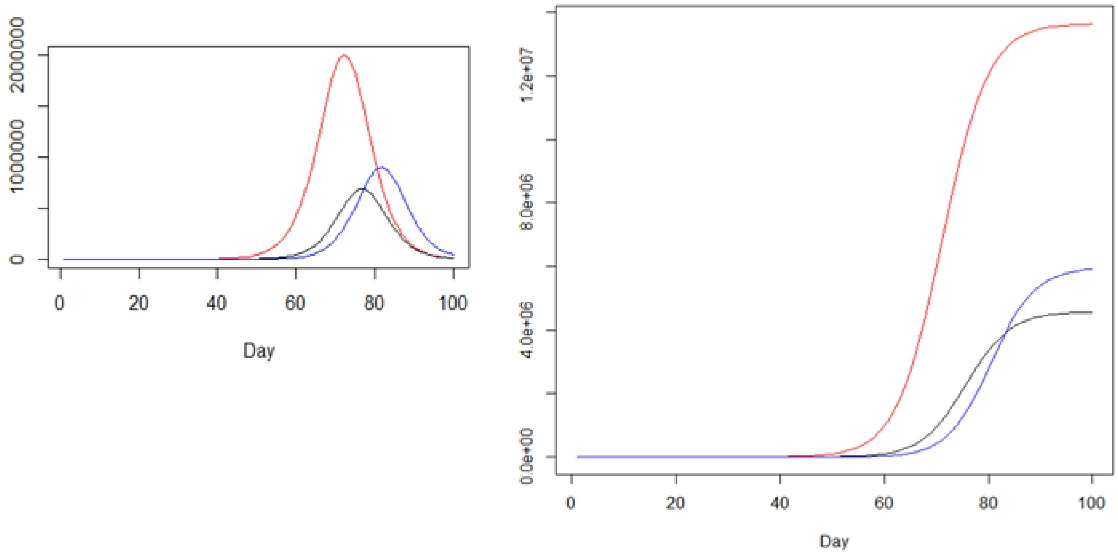

Figure 6.1. Two examples of the simulated number of daily and total infected cases in three cities.

In addition to two examples of the simulated number of daily and total infected in İstanbul, Ankara and İzmir, in Figure 6.2, the average of number of daily infected people for all cities is shown. As seen in the two different examples, an outbreak is observed in all cities which is not surprising us, but it does not begin at the same time. Firstly it begins in İstanbul as we assumed that the infection starts from a visitor arriving in İstanbul, then in İzmir and Ankara. Also, peak times of infection in all cities are diffrent as seen in the plot. The peak times in Ankara and İzmir are lower than İstanbul. When we look at the plot in detail, approximately after day 40, 2,000,000 people are infected at most in İstanbul. In Ankara and İzmir, at most approximately between 750,000 and 1,000,000 people are infected.

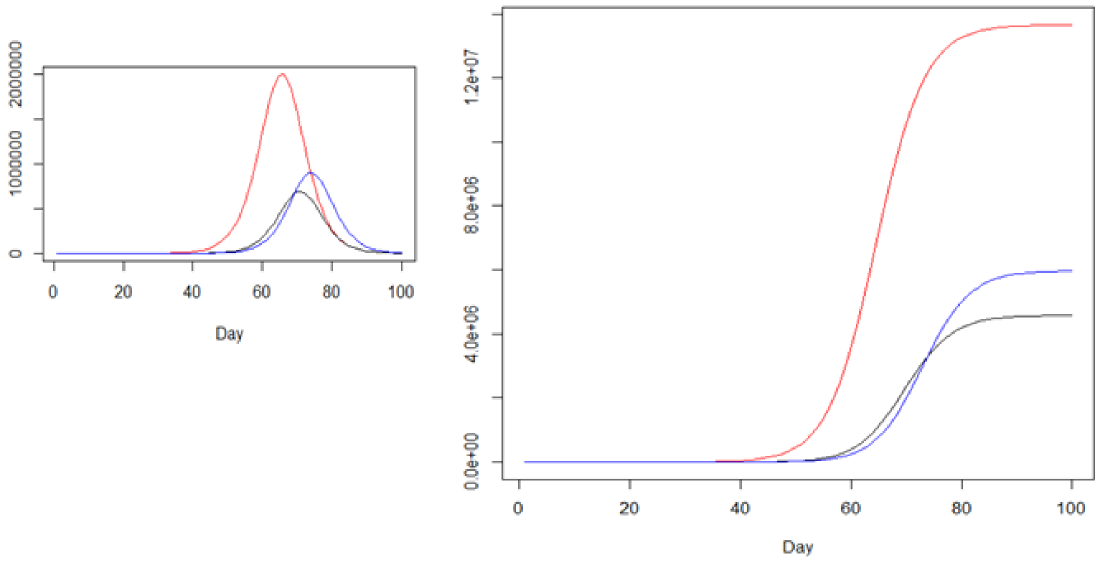

Figure 6.1. Two examples of the simulated number of daily and total infected cases in three cities (cont.).

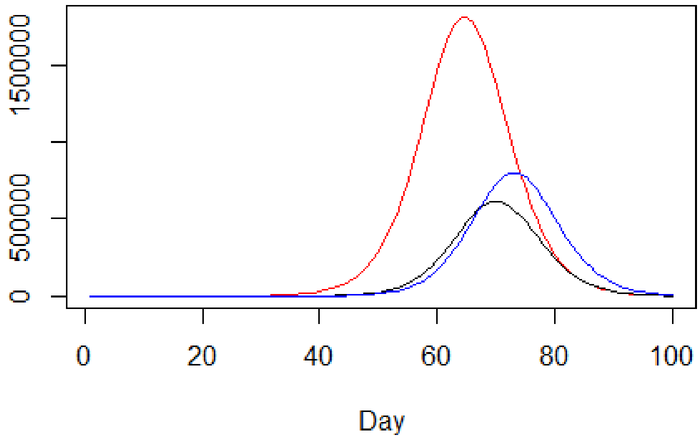

Figure 6.2. Average number of daily infected people.

In Figure 6.3, the average total number of infected people in all cities are shown. Approximately more than 12,000,000 people in İstanbul, 4,000,000 in Ankara and 6,000,000 in İzmir are being infected. Also, in this plot it is seen that an outbreak begins in

İstanbul earlier than other cities. After peak time in all cities, number of infected cases are starting to be constant as there are no new infections.

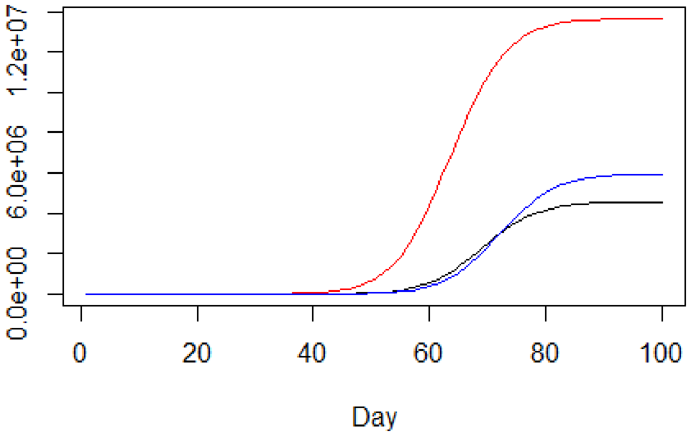

Figure 6.3. Average total number of infected people so far.

In Figure 6.4, we show the peak days of infection in each city for each repetition. It is seen that outbreak begins firstly in İstanbul then it continues with Ankara and İzmir. Also in Table 6.2, we can see that the average peak days of İstanbul are lower than those of the other cities. Peak days in other cities are almost the same.

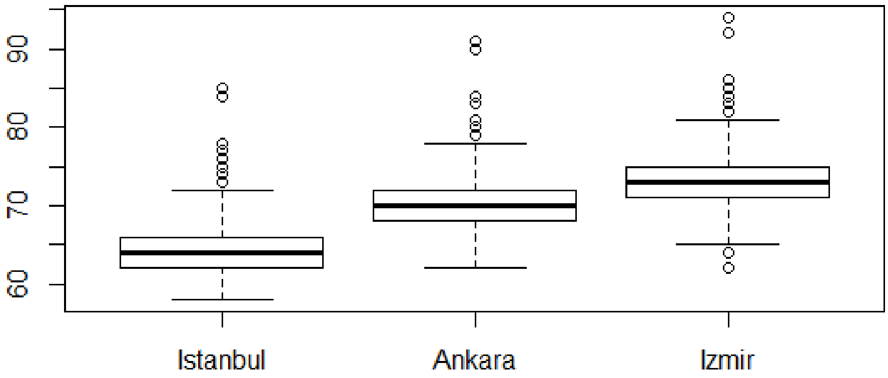

Figure 6.4. Peak days of infection.

We want to learn the difference of peak days between İstanbul and other cities in order to observe the time that an outbreak begins compare to İstanbul. In the following figure peak delays of all cities are presented.

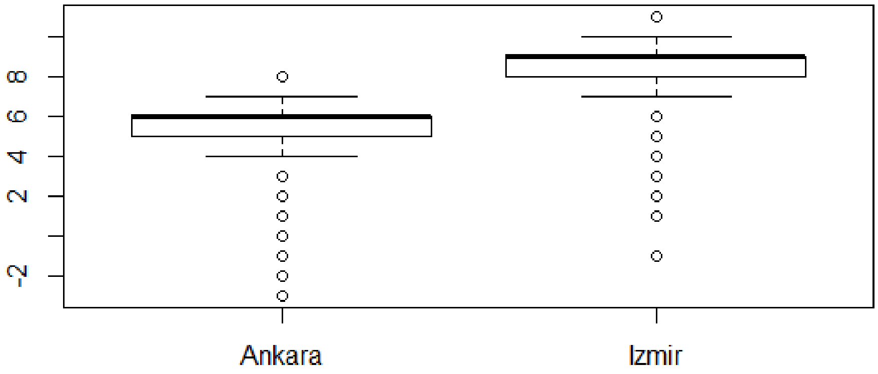

Figure 6.5. Delay of peak days between İstanbul and other cities.

Table 6.2. Average peak days of cities.

İstanbul	Ankara	İzmir
64	70	73

Table 6.3. Average delay of peak days.

Ankara	İzmir
5.414	8.386

Table 6.4. Travelling flows between İstanbul and other cities.

	Ankara	İzmir
Flow	21600	12240

It is seen in Figure 6.5 that the differences of peak days between İstanbul and Ankara are lower than those between İstanbul and İzmir. Therefore, an outbreak begins in Ankara after İstanbul and continues with İzmir. Also, the travelling flows between İstanbul and other cities justifies this result. The number of people who are going to Ankara from İstanbul is larger than the number of people who are going to İzmir from İstanbul, so that the possibility of spreading the disease to Ankara is higher than İzmir. In very few repetitions, outbreak begins in Ankara and İzmir before İstanbul. The quantities of these cases can be seen as negative values in the following figures.

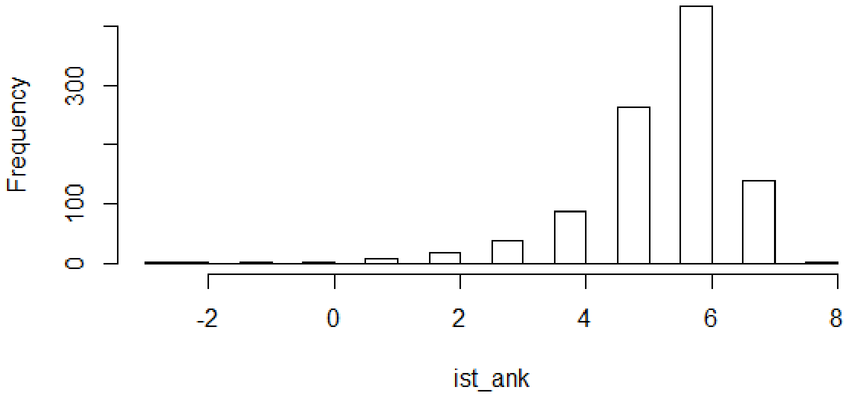

Figure 6.6. Delay of peak days between İstanbul and Ankara.

Also, according to peak day differences between İstanbul and other cities, it is seen in Figure 6.6 and Figure 6.7 that relative difference of peak delays to İstanbul are different. We carried out a t-test to see whether there is a significant difference between the peak delay means of Ankara and İzmir or not. Null hypothesis is that the means of these two groups are the same. Alternative hypothesis is that the means of these two groups are not the same.

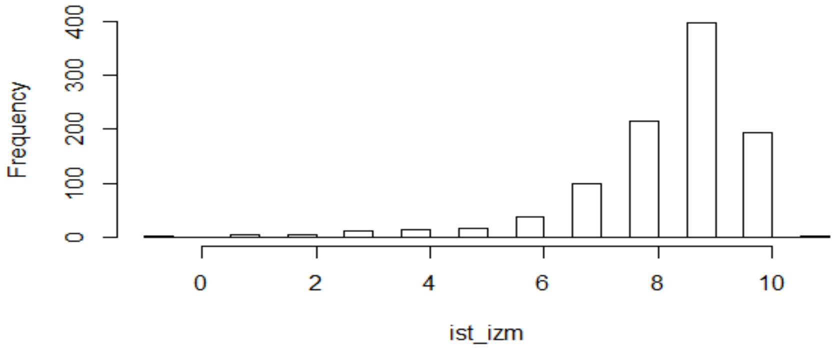

Figure 6.7. Delay of peak days between İstanbul and İzmir.

It is seen in Figure 6.9 that with 95% confidence and 0.05 significance level, the p-value is very close to 0 and less than the 0.05 significance level, so the means of two samples are significantly different. Also, t value calculated in the result is outside of the lower and upper bounds of the 95% confidence interval, so that we reject the null hypothesis.

According to confidence interval of difference of the means, with 95% confidence the difference between mean of Ankara and mean of İzmir is between 3.099731 and -2.844269 values indicating that we can expect that peak day delay of İzmir is approximately 3 days larger than mean of Ankara.

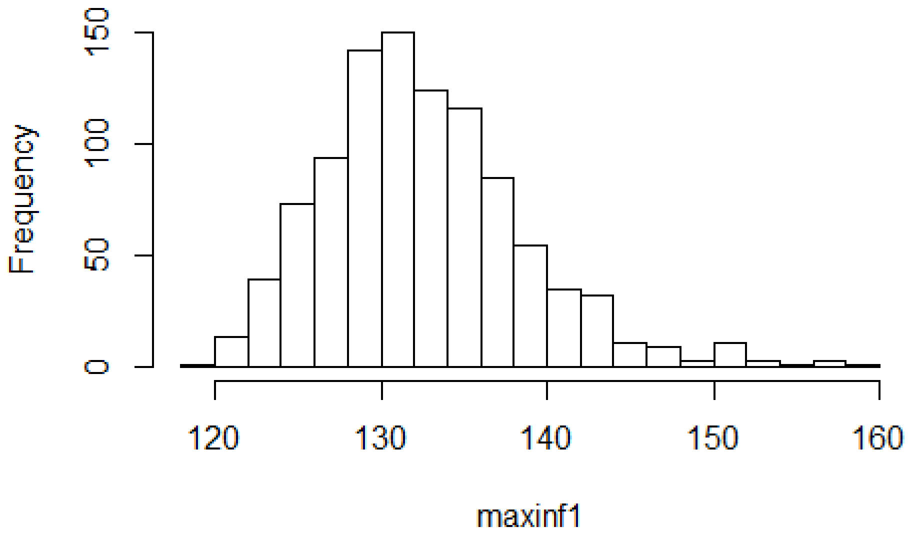

Figure 6.8. Maximum infected days of İstanbul.

In Figure 6.8, Figure 6.10 and Figure 6.11, the days of infection ends for each repetiton and city are shown. It is seen in the histograms that maximum infected days of İstanbul, Ankara and İzmir are similar. Although maximum infected days are similar, the average maximum infected day in Ankara is larger than those in İstanbul and İzmir. Values are distributed around between 120 and 150 days. There are also some cases distributed around 150 and 160 days in each city.

```
Welch Two Sample t-test

data:  ist_ank and ist_izm

t = -45.632, df = 1933.1, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95% confidence interval:

-3.099731 -2.844269

sample estimates:

mean of x mean of y

5.414     8.386
```

Figure 6.9. Two sample t-test results.

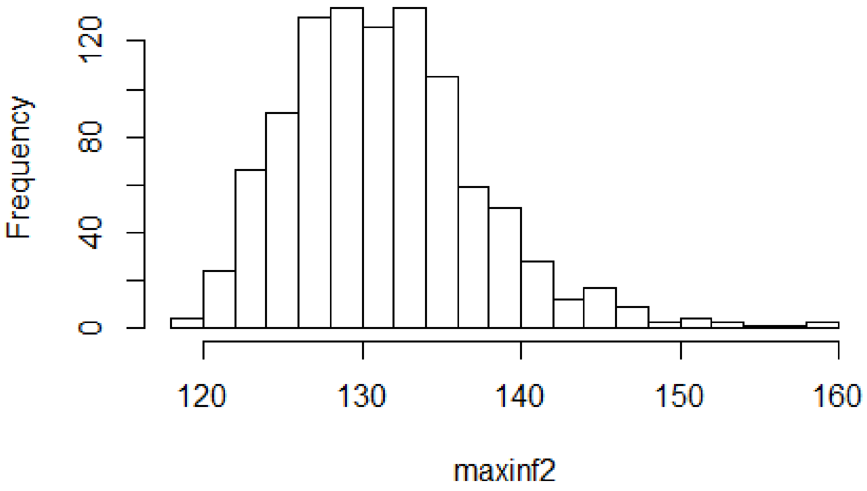

Figure 6.10. Maximum infected day of Ankara.

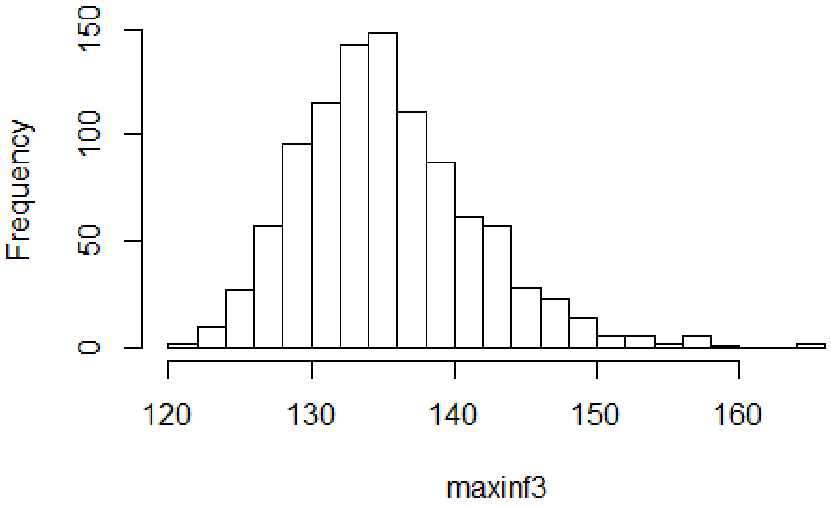

Figure 6.11. Maximum infected day of İzmir.

Table 6.5. Average of maximum infected days.

İstanbul	Ankara	İzmir
166	167	165

6.2. Experiment for Three Cities with Vaccination Rates 50% in İstanbul, 50% in Ankara and 25% in İzmir

In this section, we carried out experiments with three cities by using vaccination in order to observe the effect of vaccination on the behaviour of the disease . We want to observe especially the influence of the vaccination rates in the different cities on the outbreak probabilities. The second main question is which vaccination rate will be appropriate to prevent the spreading of the disease. In İstanbul and Ankara vaccination rate is selected as 50%, in İzmir it is selected as 25%. Due to these vaccination rates R_0 value is reduced from 1.5 to 0.75 in İstanbul and Ankara and to 1.1 in İzmir. Increasing vaccination rate reduces the value of R_0 for that city. In other words, the probability of an outbreak will decrease. The new R_0 values are calculated using the following formula. As R_{0old} we use 1.5 and “vacc” denotes the vaccination rate.

$$R_{0new}=R_{0old} * (1\text{-vacc}) \qquad (6.1)$$

According to new R_0 values, in İzmir, the R_0 value is still greater than 1 but occurrence of an outbreak is not certain. In İstanbul and Ankara, the R_0 value is less than 1 and without travelling it is clear that an outbreak is impossible as herd immunity is observed. Herd immunity means that when a certain percentage of people in a community is immunized an outbreak is impossible. Thus other people, who are not vaccinated are also protected. However, in case of travelling if an outbreak occurs in İzmir, this situation may affect İstanbul and Ankara. Therefore in order to know the real behavior, we have to carry out simulations. Herd immunity for the 3 cities is only guaranteed, if R_0 is below 1 in all cities.

According to attack rate histograms in Figure 6.12, Figure 6.13 and Figure 6.14, in Ankara and İstanbul, outbreak does not occur in 50% of the 1000 repetitions. In 50% of the repetitions, although the attack rate is low, due to the large population of Ankara and İstanbul there are still a large number of infected people. It shows us that in case of travelling and due to unsufficient rate of vaccination in İzmir, infected people come to İstanbul and Ankara from İzmir and causes new infections. In İzmir, no outbreak is observed in 50% of 1000 repetitions, in the remaining repetitions we observe a high number of infected people. Using high vaccination rates in İstanbul also affects İzmir.

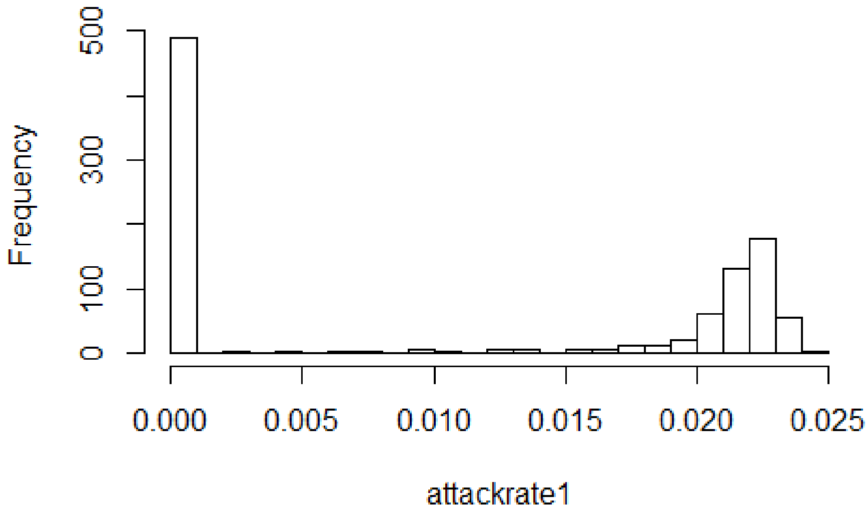

Figure 6.12. Attack rate of İstanbul.

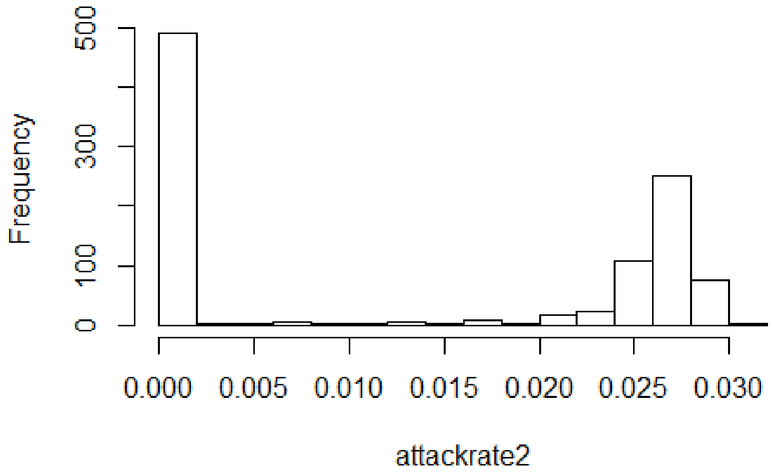

Figure 6.13. Attack rate of Ankara.

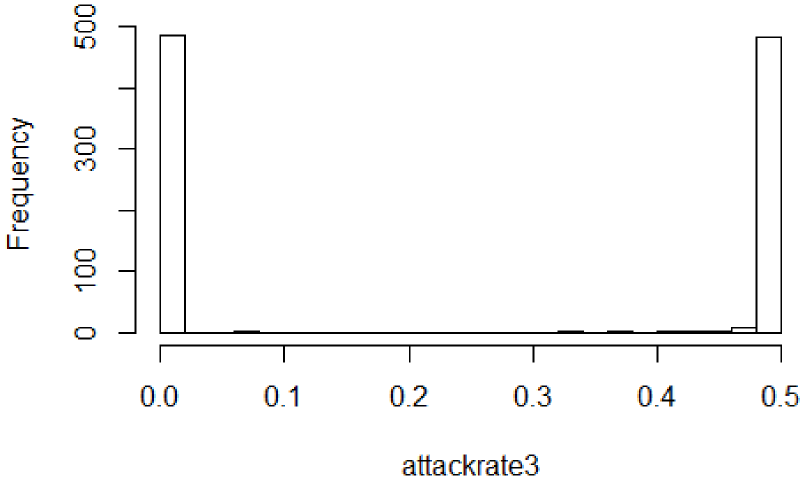

Figure 6.14. Attack rate of İzmir.

Table 6.6. Attack rates of cities.

Average attack rate	İstanbul	Ankara	İzmir
Maximum	0.02444376	0.03017074	0.49676260
Average	0.01609929	0.01480188	0.39606070
Minimum	3.752581e-06	4.273104e-06	3.357511e-07

We also investigated disease behaviour in İzmir without travelling between cities with 100 repetitions. We observed that with 1.1 R_0 there is an outbreak in the city in all repetitions. However, with using vaccination and in case travelling between cities, it is seen in Figure 6.14 that there is no outbreak in the 50% of repetitions in İzmir. The attack rate in İzmir without travelling and vaccination can be seen in Figure 6.15.

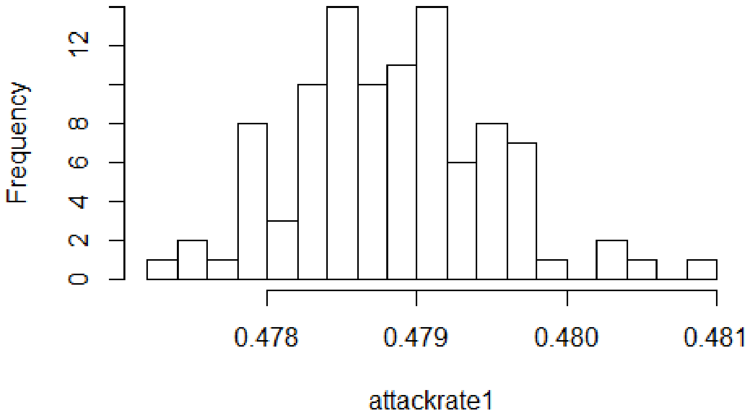

Figure 6.15. Attack rate of İzmir.

We wanted to see the attack rates by changing the number of people seed at the beginning. Therefore, in addition to analyses above, we also analyzed new attack rates in all cities after changing the number of infected people seed at the beginning. When the number of seeding is increasing from 1 to 2, the number of people in high attack rates in each city is also increasing. It shows us that attack rate can change according to number of infection seeding at the beginning of the process.

When we change the number of infected seeding at the beginning, the probability of an outbreak can change considerably. In İstanbul, Ankara and İzmir, approximately in 25% of repetitions, there seems no outbreak. The probability of an outbreak in İzmir is higher than the probability of an outbreak in İstanbul and Ankara. The results can be seen in Figure 6.16, Figure 6.17 and Figure 6.18.

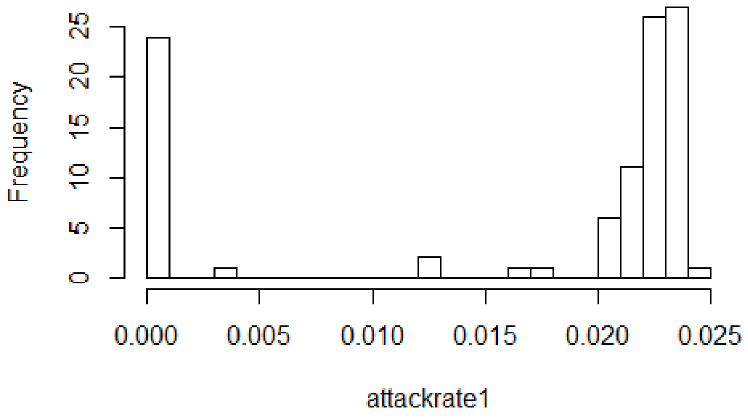

Figure 6.16. Attack rate of İstanbul.

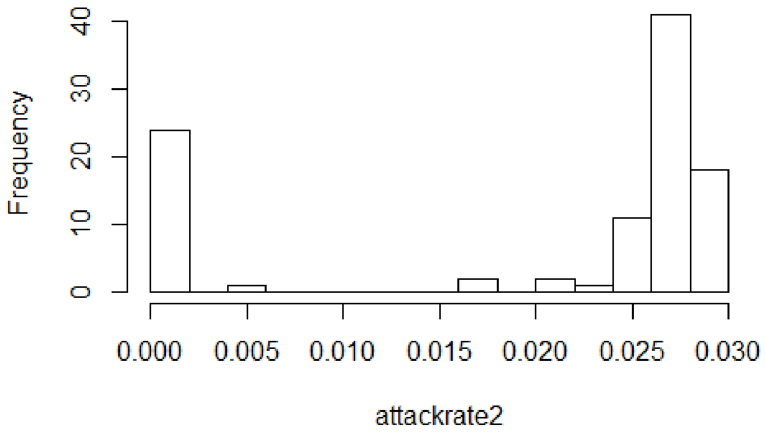

Figure 6.17. Attack rate of Ankara.

In Figure 6.19, the behaviour of the infection is shown in two different repetitions with daily and total number of infected people. In the first three plots an outbreak is observed in all cities. First of all it begins in İzmir and then continues with İstanbul and Ankara. As it is explained before, in some repetitions, an outbreak is not observed and the number of people who are infected is low.

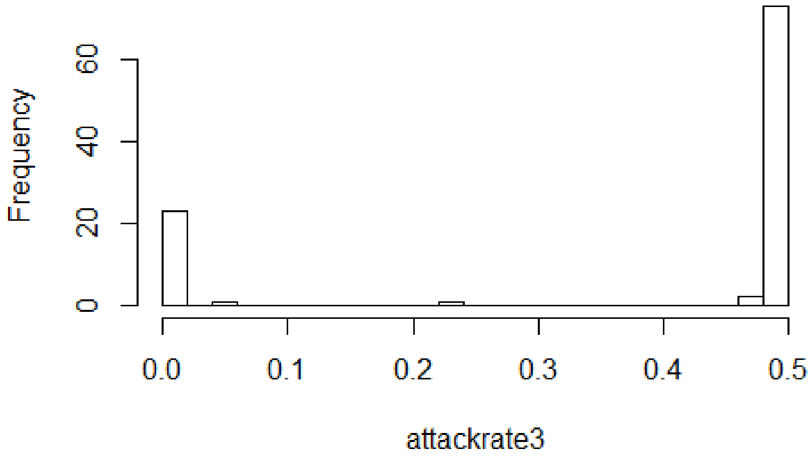

Figure 6.18. Attack rate of İzmir.

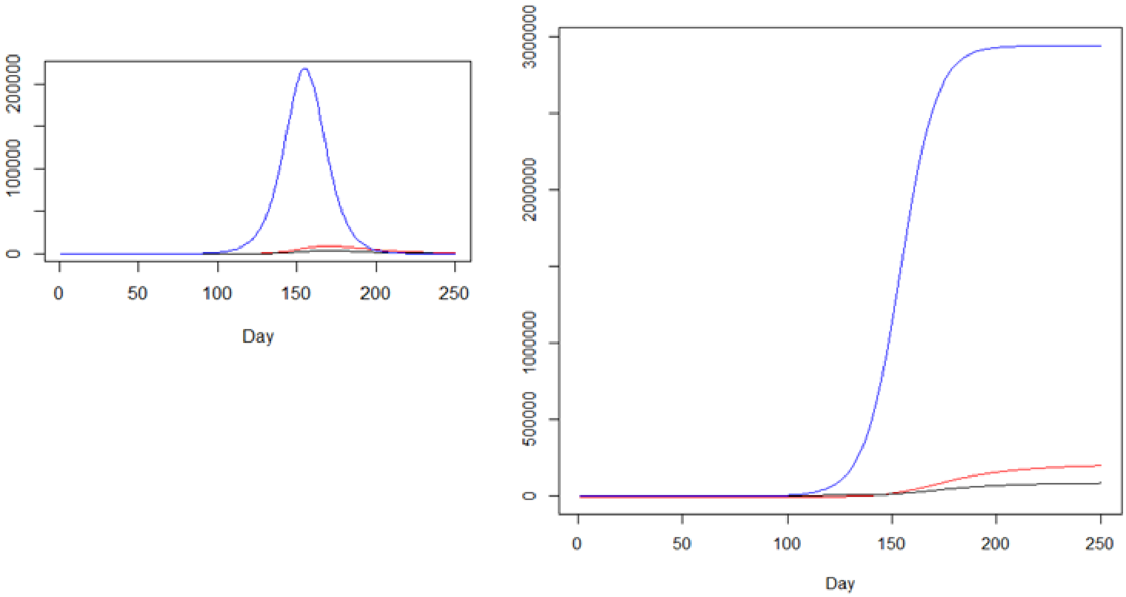

Figure 6.19. Two examples of the simulated number of daily and total infected cases in three cities.

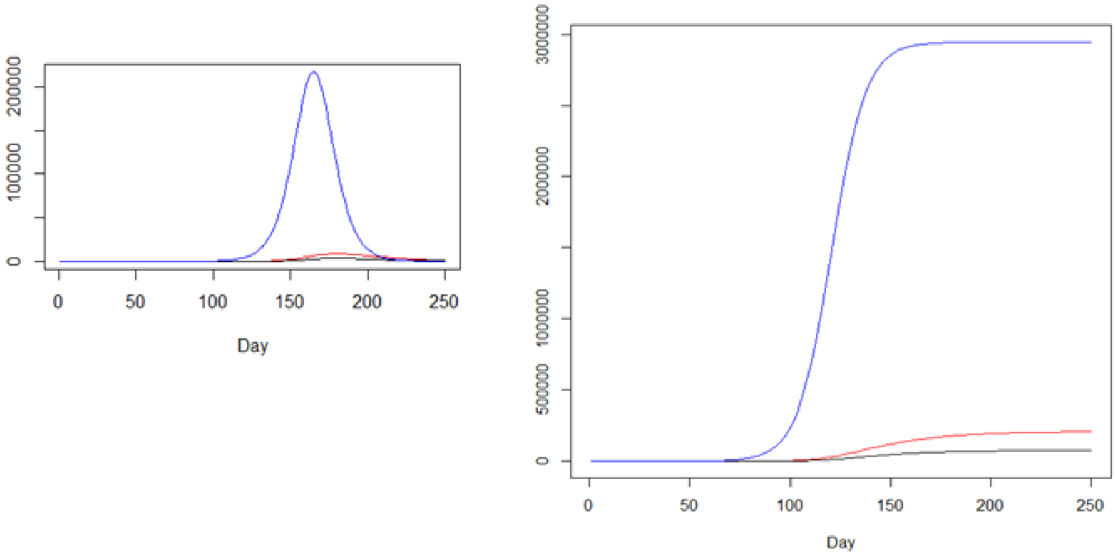

Figure 6.19. Two examples of the simulated number of daily and total infected cases in three cities (cont.).

In Figure 6.20, the average number of daily infected people for all cities is shown. As seen in Figure 6.19, an outbreak is observed in all cities but it does not have the same severities. Firstly it begins in İzmir, then in İstanbul and Ankara. Approximately after day 50, 50,000 people are infected at most in İzmir.

In İstanbul and Ankara, less than 10,000 people are infected. In Figure 6.21, the average total number of infected people in İstanbul, Ankara and İzmir are shown. It is seen in the plot that approximately more than 1,500,000 people in İzmir, less than 100,000 people in Ankara and İstanbul are being infected.

Also, in this plot it is seen that an outbreak begins in İzmir earlier than İstanbul and Ankara. In addition, peak time of infecton in İstanbul and Ankara is lower than peak time of infection in İzmir. After peak time in all cities, number of infected cases are starting to be constant as there are no new infections.

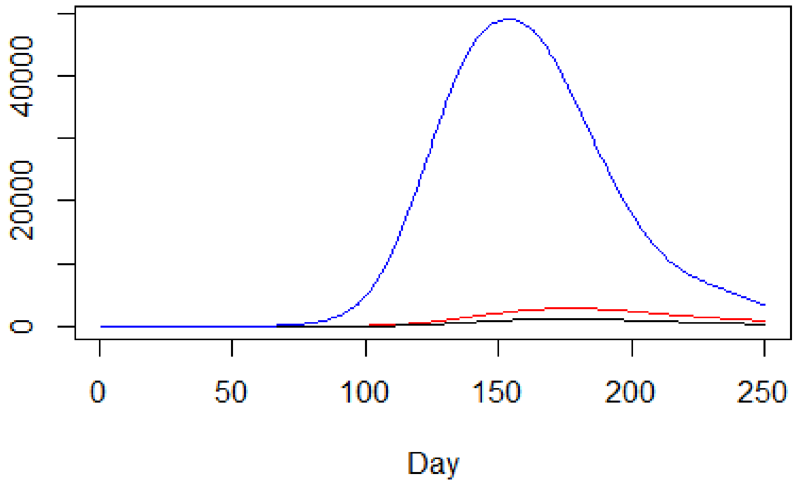

Figure 6.20. Average number of daily infected people.

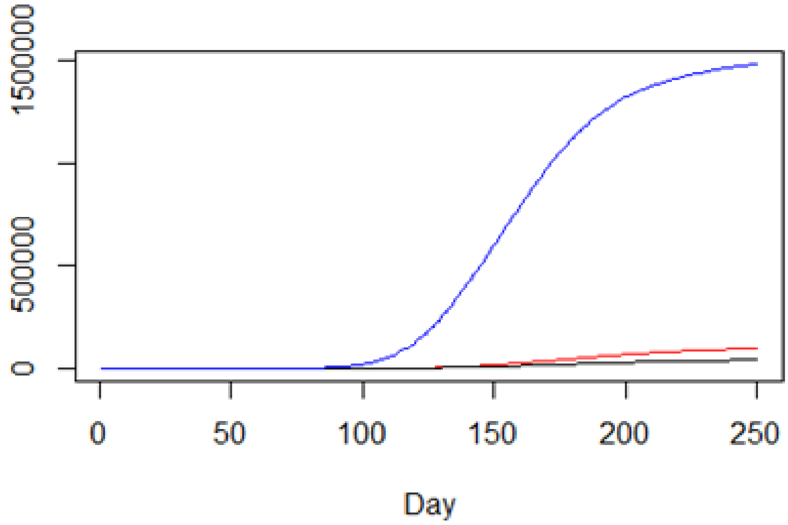

Figure 6.21. Average total number of infected people so far.

In this experiment, our aim is to see the effect of vaccination on disease spread. Therefore, by using some number of vaccination rates in all cities we try to observe disease spread dynamics and find the number of people who are infected in all cities.

6.3. Experiment for Three Cities with Vaccination Rates 50% in İstanbul, 25% in Ankara and 25% in İzmir

In this section, we carried out experiments with three cities by using vaccination. In İzmir and Ankara vaccination rate is selected as 25%, in İstanbul it is selected as 50%. With these vaccination rates R_0 is reduced from 1.5 to 1.1 in İzmir and Ankara, to 0.75 in İstanbul. In İzmir and Ankara, R_0 is still greater than 1. In İstanbul, R_0 is less than 1 and without travelling it is clear that an outbreak will not occur, but in case travelling if an outbreak occurs in İzmir and Ankara, this situation may affect İstanbul.

Also, İstanbul is larger than other cities and infection seeding is conducted in İstanbul at the beginning. One infected person is coming from abroad to İstanbul every day during 50 days. R_0 is 0.75, and the expected disease length is 4.3 in İstanbul. Approximately 200 new infected people are generated with 0.75 R_0. Due to the expected disease length which is 4.3, approximately 800 number of infected days are generated in total. Therefore, infected people travel from İstanbul to İzmir and Ankara and cause new infection generations in these cities. In order to observe the real behavior, we have to carry out simulations.

According to attack rate histograms in Figure 6.22, Figure 6.23 and Figure 6.24, in İstanbul, outbreak does not occur in 20% of 1000 repetitions. In 80% of the repetitions, although the rate of being an outbreak is low, due to large population number of İstanbul there are still a number of infected people. It shows us that in case of travelling and due to unsufficient rate of vaccination in Ankara and İzmir, infected people come to İstanbul from İzmir and Ankara and cause new infections. In İzmir, outbreak is not observed in 20% of 1000 repetitions, in remaining repetitions we observe a high number of infected people. Decreasing of the vaccination rate in Ankara affects the number of outbreak cases in a negative way.

It is observed that the number of outbreaks increases by changing the vaccination rate in Ankara comparing to the previous experiment. For example, due to unsufficient vaccination in Ankara, the number of infected people are increasing also in İstanbul. In other words, the vaccination rate that is used in a city also affects other cities. It is not enough to use vaccination only in a city to prevent an outbreak. Also, the rate of vaccination is significant to prevent an outbreak and spreading of the disease to other cities.

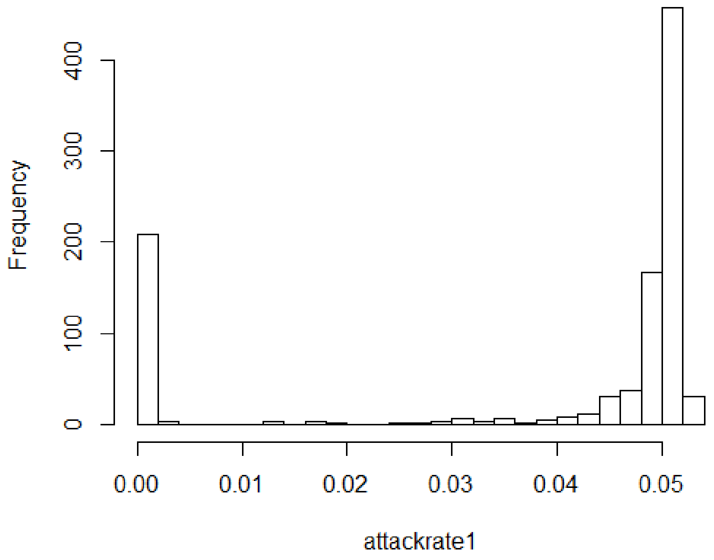

Figure 6.22. Attack rate of İstanbul.

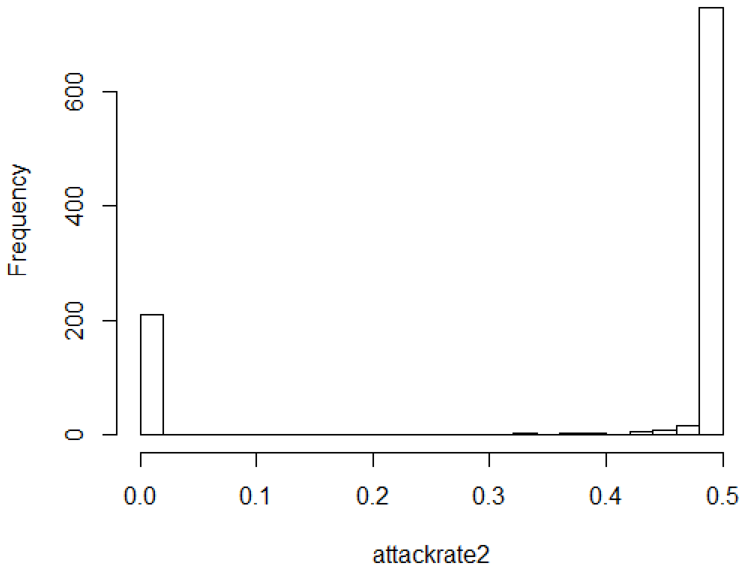

Figure 6.23. Attack rate of Ankara.

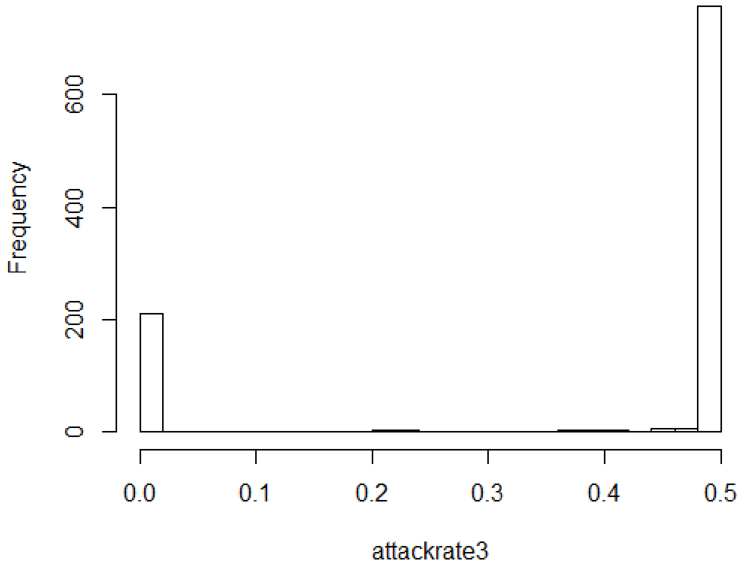

Figure 6.24. Attack rate of İzmir.

Table 6.7. Attack rates of cities.

Average attack rate	İstanbul	Ankara	İzmir
Maximum	0.05327778	0.49049686	0.49937468
Average	0.04713265	0.46623991	0.47415375
Minimum	3.410735e-05	2.039088e-05	4.533112e-06

In Figure 6.25, the daily and total number of infected people in two different repetitions are shown. An outbreak is observed in İstanbul, Ankara and İzmir, but when we compare with the previous experiment, it begins at a later day. It seems that outbreak begins firstly in Ankara and then continues with İstanbul and İzmir. In İzmir the number of daily infected people is more than those in Ankara, because İzmir is a larger city than Ankara, so a large number of people are exposed to the infection which is coming from İstanbul.

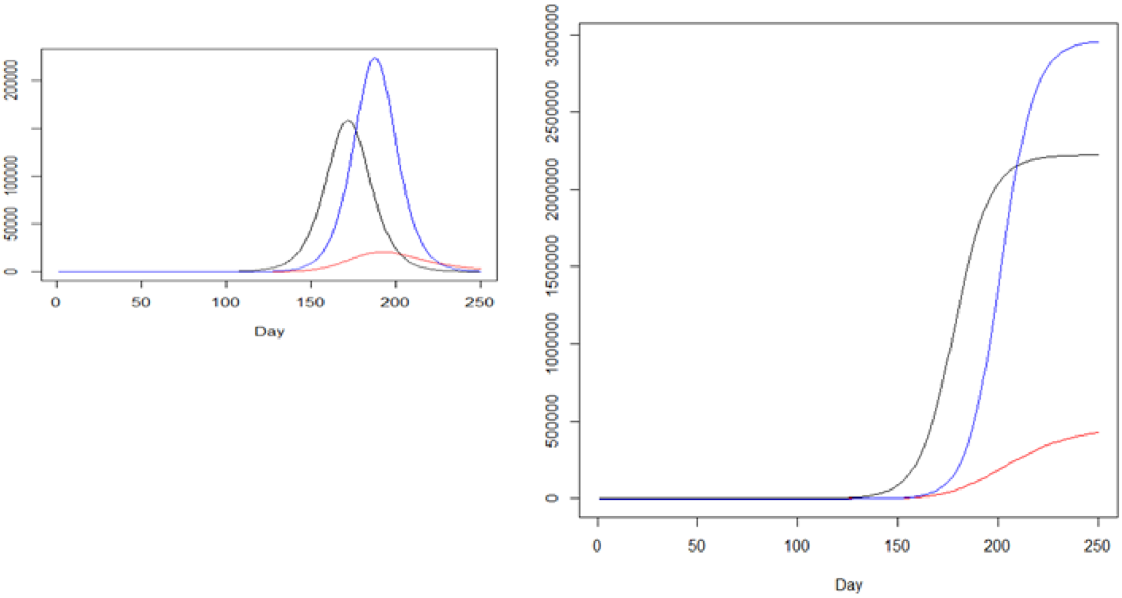

Figure 6.25. Two examples of the simulated number of daily and total infected cases in three cities.

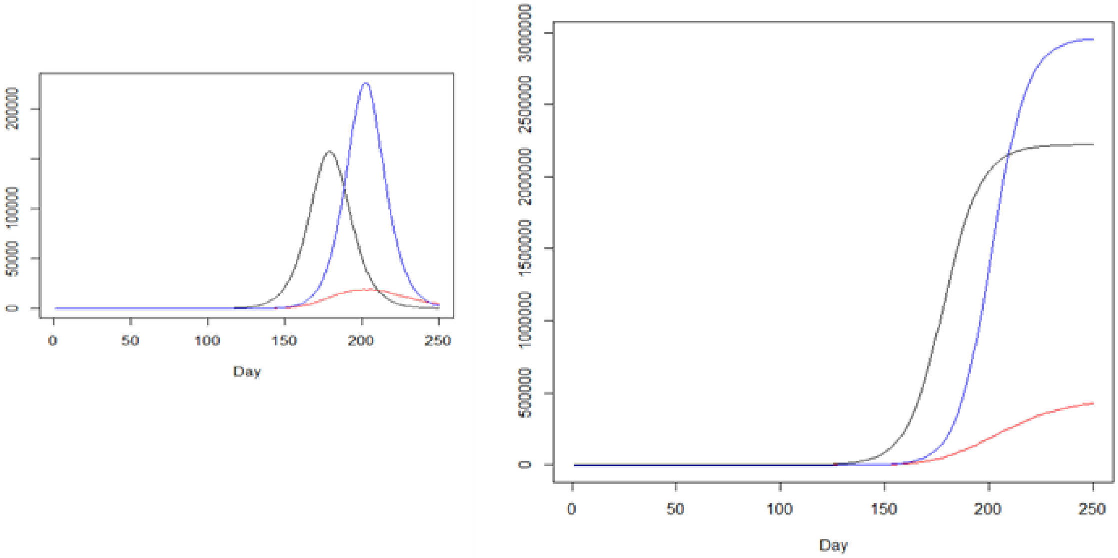

Figure 6.25. Two examples of the simulated number of daily and total infected cases in three cities (cont.).

In Figure 6.26, the average number of daily infected people for all cities is shown. As in Figure 6.25, an outbreak is observed in all cities but the severity is different. In other words, for example the number of infected cases in İzmir is more than in the other cities. In average, time of beginning an outbreak is a bit later and almost the same. Approximately after day 100, 10,000 people are infected in İstanbul. In Ankara and İzmir, at most approximately between 60,000 and 90,000 people are infected.

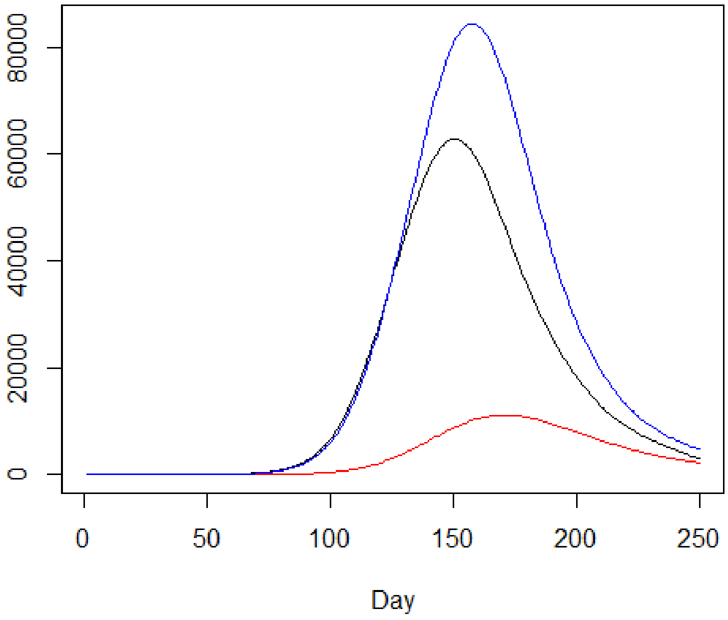

Figure 6.26. Average number of daily infected people.

In Figure 6.27, the average total number of infected people so far in İstanbul, Ankara and İzmir are shown. As seen in the plot that approximately 3,000,000 people in İzmir, 2,250,000 in Ankara and 500,000 in İstanbul are being infected. Also, in this plot it is seen that an outbreak begins in Ankara earlier than in the other cities. Peak times of İzmir and İstanbul are lower than peak time in Ankara. After peak time in all cities, the number of infected cases are starting to be constant as there are no new infections.

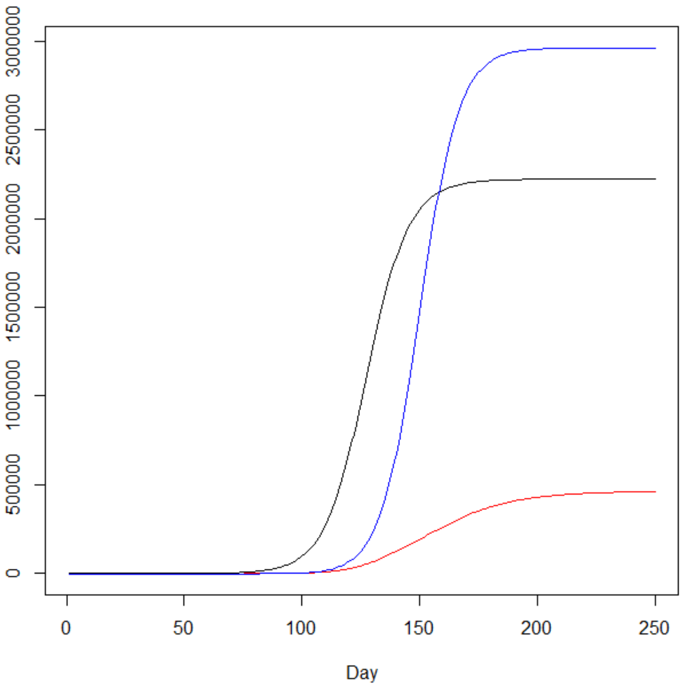

Figure 6.27. Average total number of infected people so far.

6.4. Experiment for Three Cities with Vaccination Rates 50% in İstanbul, 50% in Ankara and 0% in İzmir

In this experiment, in İstanbul and Ankara vaccination rate is selected as 50%, in İzmir vaccination is not applied. With these vaccination rates R_0 is reduced from 1.5 to 0.75 in İstanbul and Ankara. Due to not using a vaccination rate in İzmir, R_0 is still 1.5.

According to new R_0 values, the probability of an outbreak in İstanbul and Ankara is low. Due to the no vaccination rate in İzmir, this situation may effect these cities, because in case of travelling if an outbreak takes place in İzmir, a large number of infected people may come to İstanbul and Ankara and cause an outbreak. In order to observe the behaviour of infection, we carried out simulations. According to attack rate histograms in Figure 6.28, Figure 6.29 and Figure 6.30, in İzmir an outbreak is observed in more than 60% of 1000 repetitions. In the remaining repetitions, due to a high vaccination rate that is used in İstanbul and Ankara, there is no outbreak. When we look at İstanbul and Ankara, in a large number of repetitions an outbreak is not observed. There are some infected people in these cities, but it does not have an huge effect. These infected people most probably come from İzmir due to travelling. Also, it can be inferred that although a high vaccination rate is used in İstanbul and Ankara, this does not prevent an outbreak in İzmir as the R_0 value is 1.5.

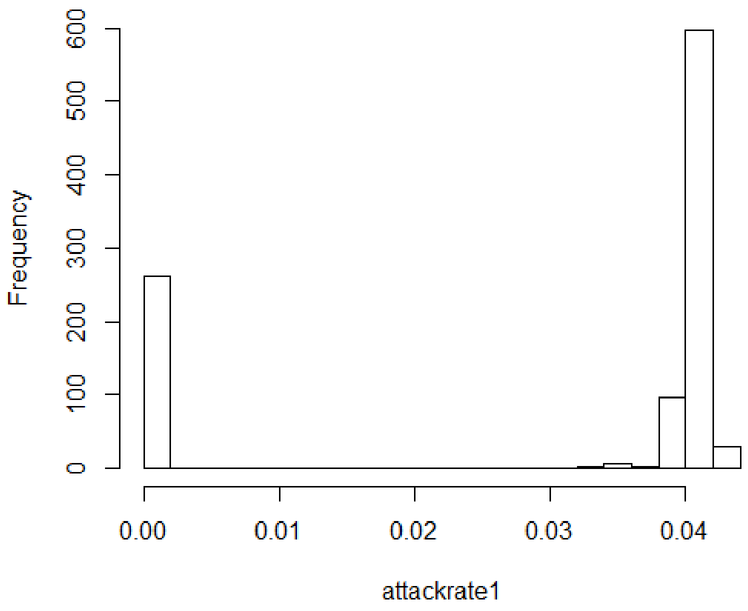

Figure 6.28. Attack rate of İstanbul.

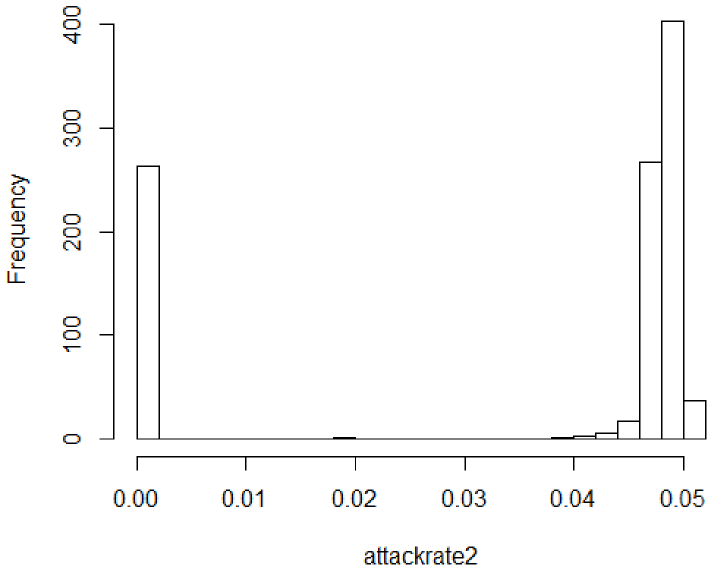

Figure 6.29. Attack rate of Ankara.

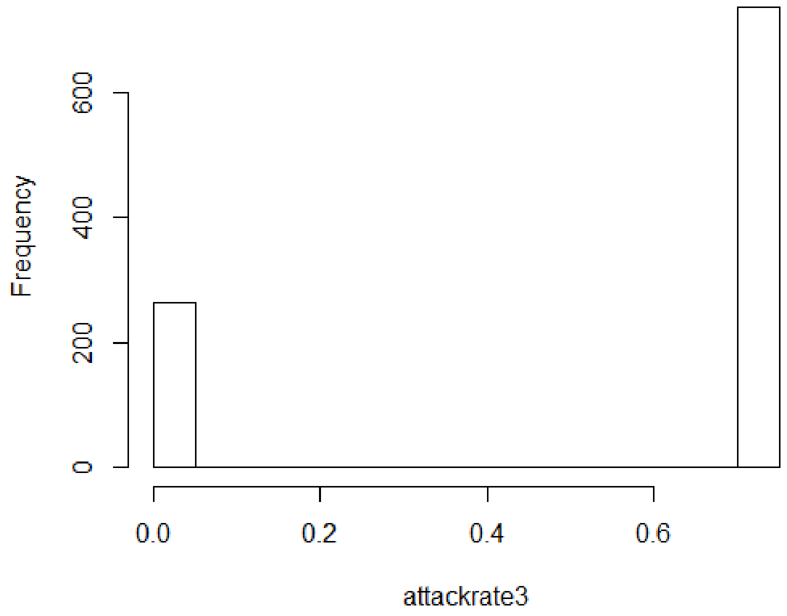

Figure 6.30. Attack rate of İzmir.

Table 6.8. Attack rates of cities.

Average attack rate	İstanbul	Ankara	İzmir
Maximum	0.04294927	0.05162766	0.74737777
Average	0.03448602	0.04060580	0.74501789
Minimum	3.310589e-06	3.616557e-06	1.825408e-05

In Figure 6.31, the behaviour of the infection with daily and total infected is shown in two different repetitions. In some repetitions an outbreak occurs in İzmir, Ankara and İstanbul, in some of them it does not. It shows us the stochasticity of the disease spread dynamics. Also the time of occurring an outbreak is later in all cities when we compare to previous experiment. In some repetitions, we do not see an outbreak, but when we look at the results in average, an outbreak occurs in all cities as seen in Figure 6.32.

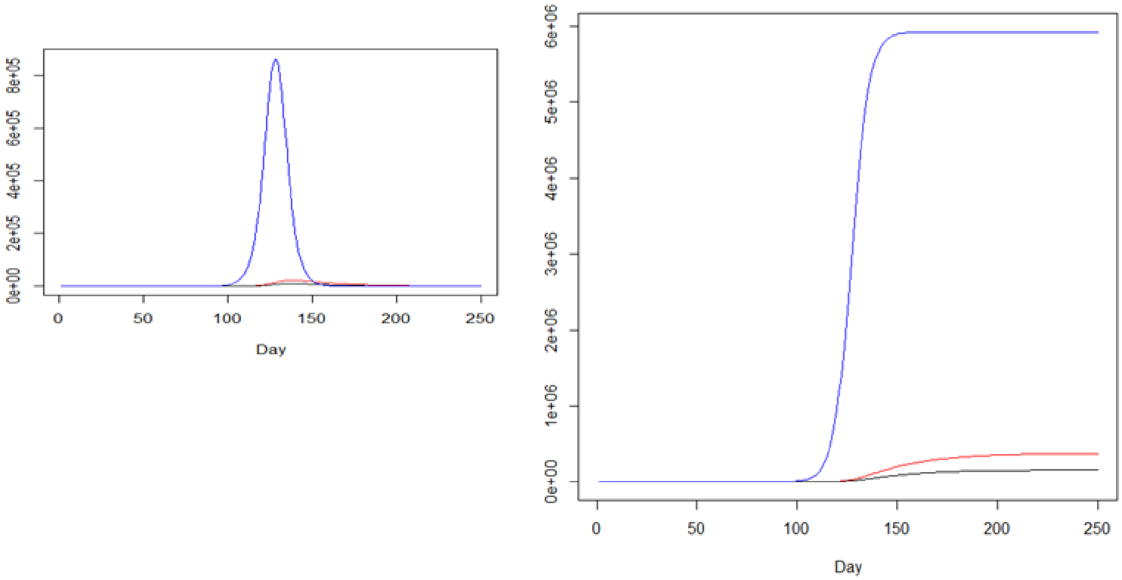

Figure 6.31. Two examples of the simulated number of daily and total infected cases in three cities.

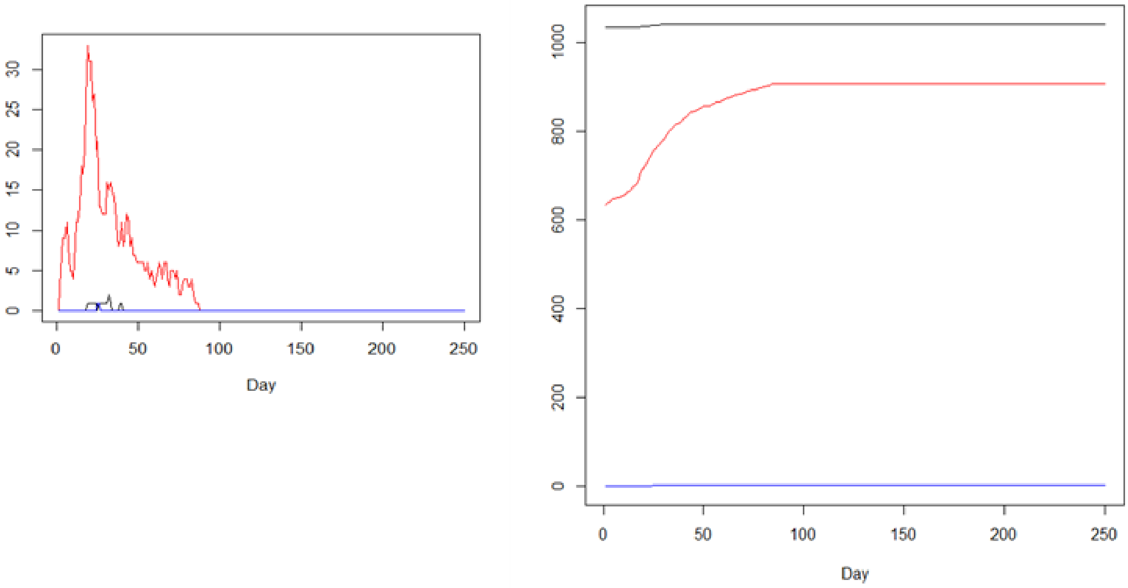

Figure 6.31. Two examples of the simulated number of daily and total infected cases in three cities (cont.).

In Figure 6.32, the average number of daily infected people for İstanbul, Ankara and İzmir is shown. As in Figure 6.31, in average an outbreak is observed in İstanbul, Ankara and İzmir but the severity is different. In other words, for example the number of infected cases in İzmir is more than those in other cities. Also, the time of beginning an outbreak in İzmir is earlier than İstanbul and Ankara. Approximately after day 50, 150,000 people are infected in İzmir. In Ankara and İstanbul less than 20,000 people are infected.

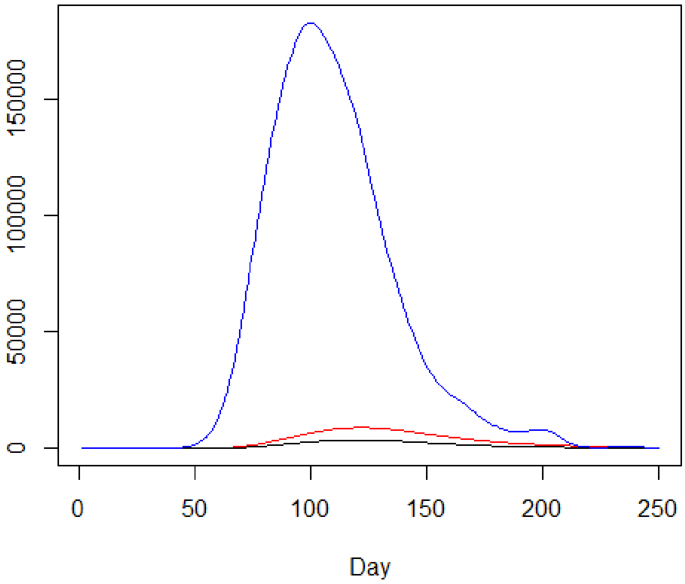

Figure 6.32. Average number of infected people.

In Figure 6.33, the average total number of infected people so far in İstanbul, Ankara and İzmir are shown. It is seen in the plot that approximately 4,000,000 people in İzmir, less than 500,000 people in Ankara and İstanbul are being infected. Also, in this plot it is seen that an outbreak begins in İzmir earlier than in the İstanbul and Ankara.

When we look at Figure 6.33, the time of beginning an outbreak in İstanbul is the same with Ankara. However in İzmir, the time of beginning an outbreak is earlier than İstanbul and Ankara. Peak times of infection in Ankara and İstanbul are lower than peak time of infection in İzmir. After peak time in all cities, the number of infected cases are starting to be constant as there are no new infections.

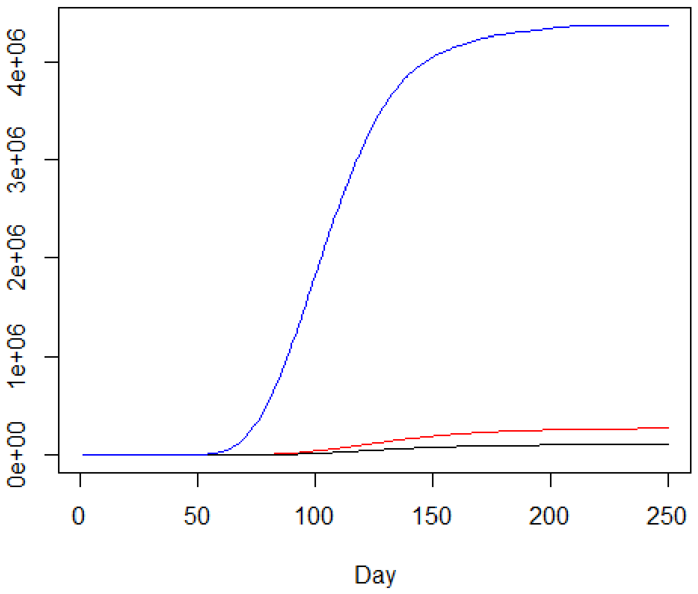

Figure 6.33. Average total number of infected people so far.

6.5. Six Cities Model Experiment

In this section, a simulation experiment is carried out for six cities in Turkey. 100 repetitions during 250 days are conducted. R_0 is assumed to be 1.5. All other inputs, assumptions used in the experiment and information regarding cities are defined in Section 5 in detail. Since the cities are large enough, the expected number of people who are travelling between cities is high and R_0 is chosen as 1.5 so greater than 1, there is high probability for an outbreak in all cities.

Average, minimum and maximum attack rates are calculated in order to see disease spread dynamics in all cities. Average attack rates in all cities are close to 0.75 in each repetition. As seen in Table 6.9, there is not much difference between minimum and maximum attack rates. Approximately 75% of people are infected in all cities. We are not surprised because the travelling flows between İstanbul, İzmir, Muğla, Antalya, Ankara and Adana are high enough, R_0 is greater than 1 and cities are large. Therefore, an outbreak occurs in all cities.

Table 6.9. Attack rates of cities.

Average attack rate	İstanbul	İzmir	Muğla	Antalya	Ankara	Adana
Maximum	0.7538564	0.7525407	0.7546978	0.7543853	0.7531777	0.7529620
Average	0.7533484	0.7515865	0.7516843	0.7517962	0.7517116	0.7516484
Minimum	0.7526322	0.7506263	0.7499977	0.7505868	0.7507847	0.7508407

In Figure 6.34, the behaviour of the infection is shown in four different repetitions and almost the same for each cities. In order to see the disease spread dynamics, we decided to carry out this experiment with more than one repetition. An outbreak is observed in all cities which is not surprising us, but it does not begin at the same time. Firstly it begins in İstanbul, then in İzmir, Muğla, Antalya, Ankara and Adana. Also, peak time of infection in İstanbul is higher than peak times of infection in İzmir, Muğla, Antalya, Ankara and Adana. Approximately after day 50, more than 1,500,000 people are infected at most in İstanbul. In other cities, at most approximately between 250,000 and 750,000 people are infected.

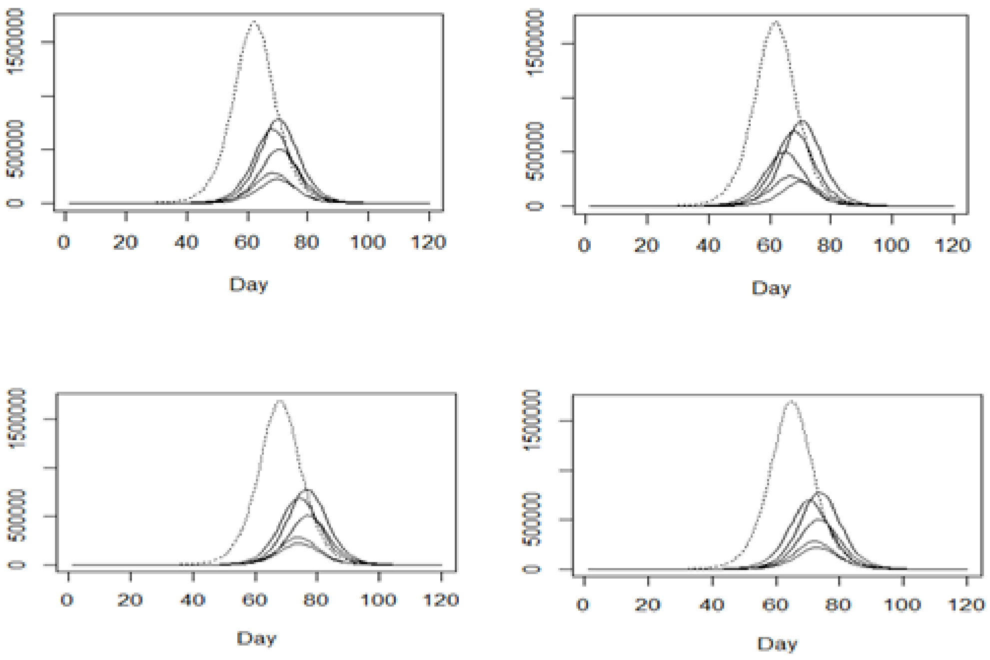

Figure 6.34. Number of daily infected people in cities.

In Figure 6.35, we want to see the peak days of infection in each city for each repetition. It is seen that outbreak begins firstly in İstanbul. Also in Table 6.10, the average peak days of İstanbul is lower than other cities. Peak days in other cities are almost the same.

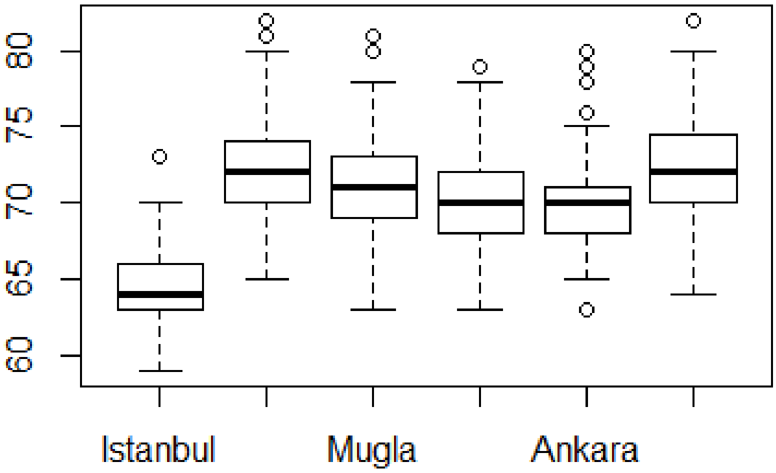

Figure 6.35. Peak days of infection.

Table 6.10. Average peak days of cities.

İstanbul	İzmir	Muğla	Antalya	Ankara	Adana
64	72	71	70	70	72

We want to learn the difference of peak days between İstanbul and İzmir, Muğla, Antalya, Ankara and Adana in order to observe the time that an outbreak begins compare to İstanbul, so that in Figure 6.35 peak delays of each cities are presented.

According to results seen in Figure 6.36 and Table 6.11, an outbreak begins in Ankara after İstanbul and continues with Antalya, Muğla, İzmir and Adana respectively.

Due to high travelling flow between İstanbul and Ankara, it seems logical that the outbreak begins in Ankara after İstanbul.

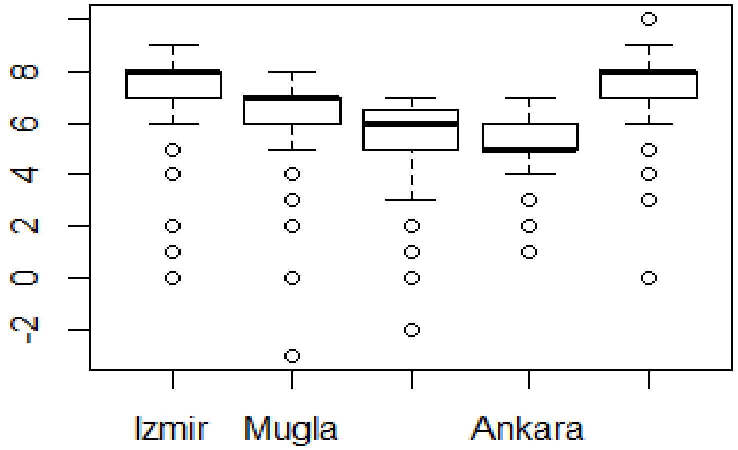

Figure 6.36. Delay of peak days between İstanbul and other cities.

Table 6.11. Average delay of peak days.

İzmir	Muğla	Antalya	Ankara	Adana
7.46	6.55	5.59	5.26	7.57

Table 6.12. Travelling flows between İstanbul and other cities.

	İzmir	Muğla	Antalya	Ankara	Adana
Flow	12240	5256	8100	21600	8100

In order to see whether the difference of peak days between İstanbul and other cities are significantly different, we carried out one way ANOVA test. Our H_0 hypothesis is that the mean of peak delays are the same. Alternative hypothesis is that the mean of peak delays are not the same. According to the results of the test, in 95% confidence level, with 0.05 significance level, we see in the results that p value Pr(>F)= 2.2e-16 value is smaller than 0.05, so we reject the null hypothesis and it can be said that the means of two sample are significantly different.

Table 6.13. Result of one way ANOVA test.

```
Analysis of Variance Table


Response: dat
          Df  Sum Sq Mean Sq F value    Pr(>F)
groups      4  481.27  120.32  52.537 < 2.2e-16 ***
Residuals 495 1133.62    2.29
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In addition, in Figure 6.36 it is seen that an outbreak begins in Muğla and Antalya earlier than İstanbul in one of the repetitions. In İzmir, Muğla, Antalya and Adana outbreak begins almost in the same time with İstanbul in one of the repetitions. Also, in İzmir, Muğla and Adana median of the peak delays are equal to upper quantile. In Ankara, median of the peak delays are equal to lower quantile. In Figures 6.37-6.42, the days of an infection ends for each repetiton and city are shown. According to results, the infection duration in Ankara and İzmir are longer than those in other cities.

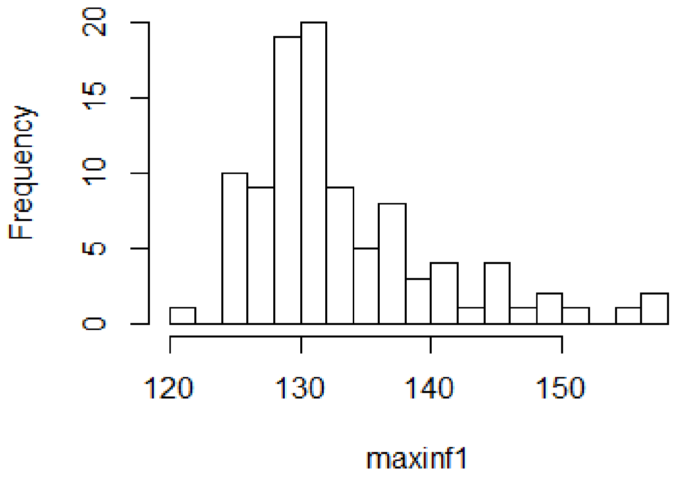

Figure 6.37. Maximum infected days of İstanbul.

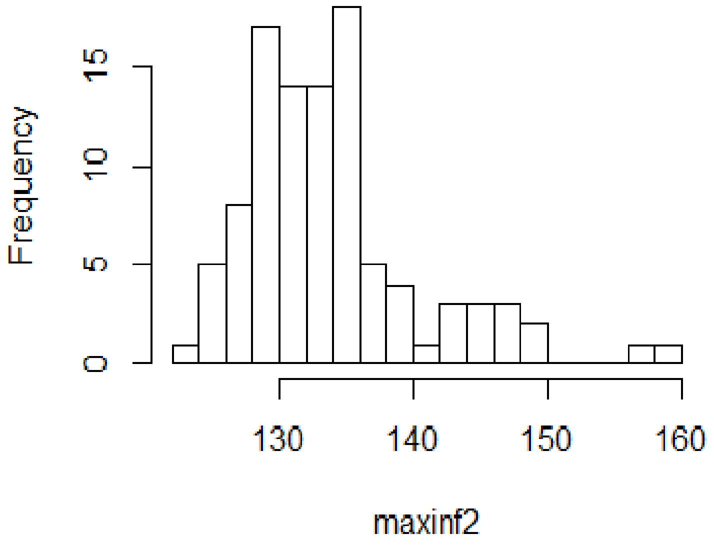

Figure 6.38. Maximum infected days of İzmir.

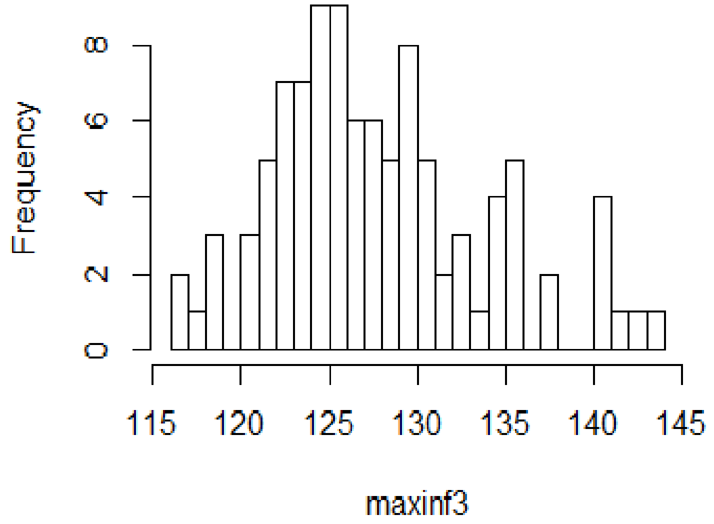

Figure 6.39. Maximum infected days of Muğla.

In İzmir, maximum infected days are distributed around 120 and 140 days. When we compare the results with Muğla, we see much higher maximum infected days in İzmir than Muğla.

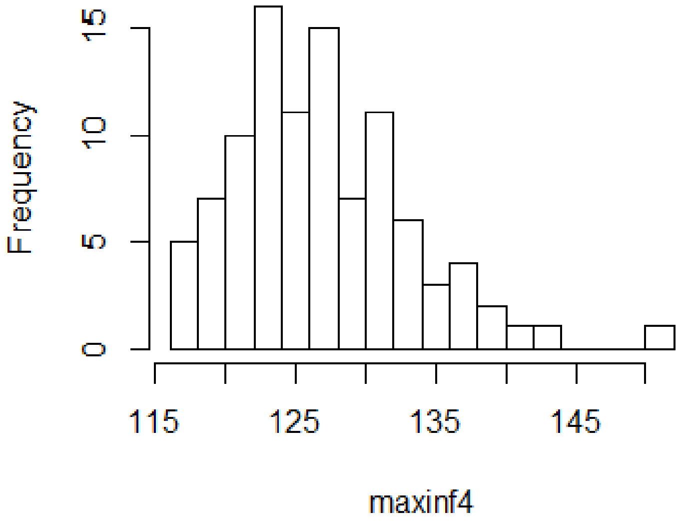

Figure 6.40. Maximum infected days of Antalya.

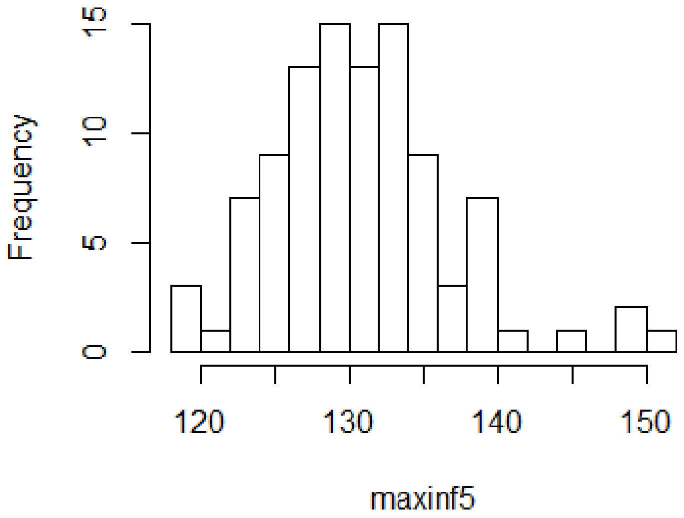

Figure 6.41. Maximum infected days of Ankara.

In Antalya, maximum infected days are distributed around 120 and 135 days. When we compare the results with Ankara, we see much higher maximum infected days in Ankara than Antalya.

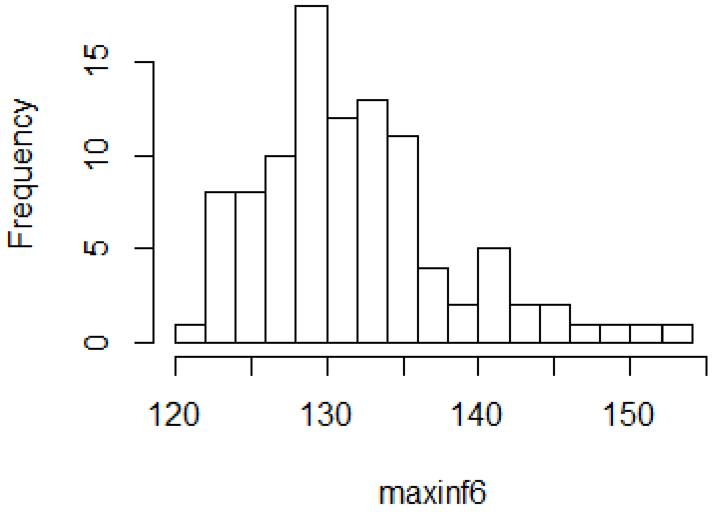

Figure 6.42. Maximum infected days of Adana.

Table 6.14. Average of maximum infected days.

İstanbul	İzmir	Muğla	Antalya	Ankara	Adana
149	159	147	150	154	149

In Figure 6.43, the total number of infected people in İstanbul, İzmir, Muğla, Antalya, Ankara and Antalya is shown. As seen in the plot that approximately 12,000,000 people in İstanbul, between 1,500,000 and 5,000,000 people in other cities are being infected. Also, in this plot it is seen that an outbreak begins in İstanbul earlier than other cities. It continues with Ankara, Antalya, Muğla, İzmir and Adana respectively.

Also, we can see that due to high travelling flow between İstanbul and Ankara, disease spreads from İstanbul to Ankara fast. Peak times of infection in İzmir, Muğla, Antalya, Ankara and Adana are lower than peak time of infection in İstanbul. After peak time in all cities, number of infected cases are starting to be constant as there are no new infections.

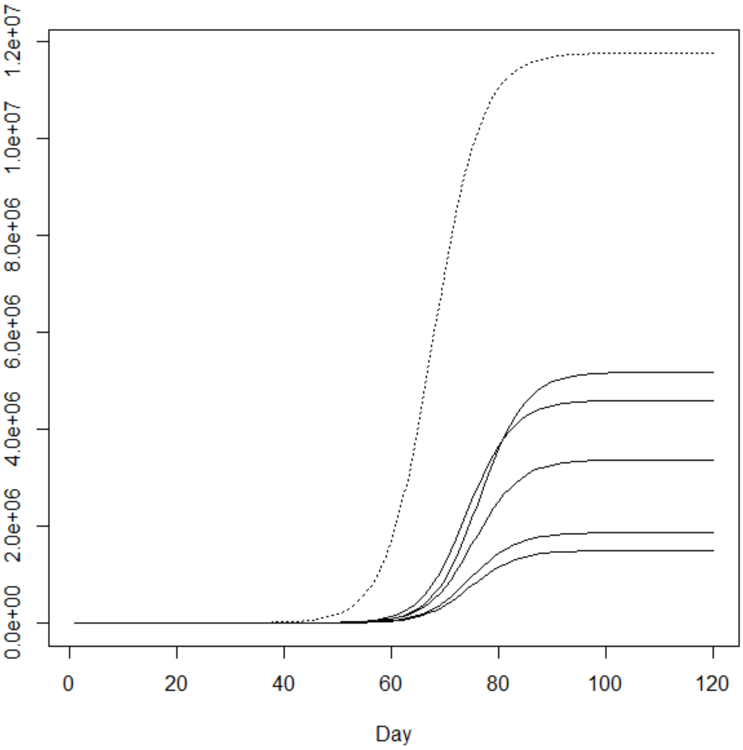

Figure 6.43. Total number of infected people so far.

6.6. Ten Cities Model Experiment

In this section, a simulation experiment is carried out for ten cities in Turkey. 100 repetitions during 250 days are conducted. R_0 is assumed to be 1.5. All other inputs and assumptions used in the experiment are defined in Section 5 in detail.

Since the cities are large enough, the expected number of people who are travelling between cities is high and R_0 is chosen as 1.5 so greater than 1, there is high probability for an outbreak in all cities. Outbreak begins in İstanbul in which the infection starts with a single initial infected person.

Average attack rates in all cities are close to 0.75 in each repetition. As seen in Table 6.15, there is no much difference between minimum and maximum attack rates. Approximately 75% of people are infected in all cities. We are not surprised because the travelling flows between cities are high enough, R_0 is greater than 1 and cities are large. Therefore, outbreak occurs in all cities.

Table 6.15. Attack rates of cities.

Average attack rate	Gaziantep	Van	Trabzon	Samsun	İstanbul	İzmir	Antalya	Adana	Ankara	Kayseri
Maximum	0.7528826	0.7538927	0.7573190	0.7557483	0.7540048	0.7533092	0.7538650	0.7537320	0.7534605	0.7543378
Average	0.7513591	0.7516547	0.7530257	0.7521947	0.7534084	0.7514548	0.7517770	0.7514633	0.7517765	0.7515248
Minimum	0.7498680	0.7488831	0.7510848	0.7501243	0.7526124	0.7504023	0.7498396	0.7503596	0.7505956	0.7497733

In Figure 6.44, the behaviour of the infection is shown in four different repetitions and almost the same for each cities. An outbreak is observed in both cities which is not surprising us, but it does not begin at the same time. Firstly it begins in İstanbul, then in other cities. Approximately after day 50, more than 2,000,000 people are infected at most in İstanbul. In other cities, at most approximately between 100,000 and 750,000 people are infected. In Figure 6.45, we want to see the peak days of infection in each city for each repetition. It is seen that outbreak begins firstly in İstanbul. Also in Table 6.16, the average peak days of İstanbul is lower than those in other cities. Peak days in other cities are almost the same.

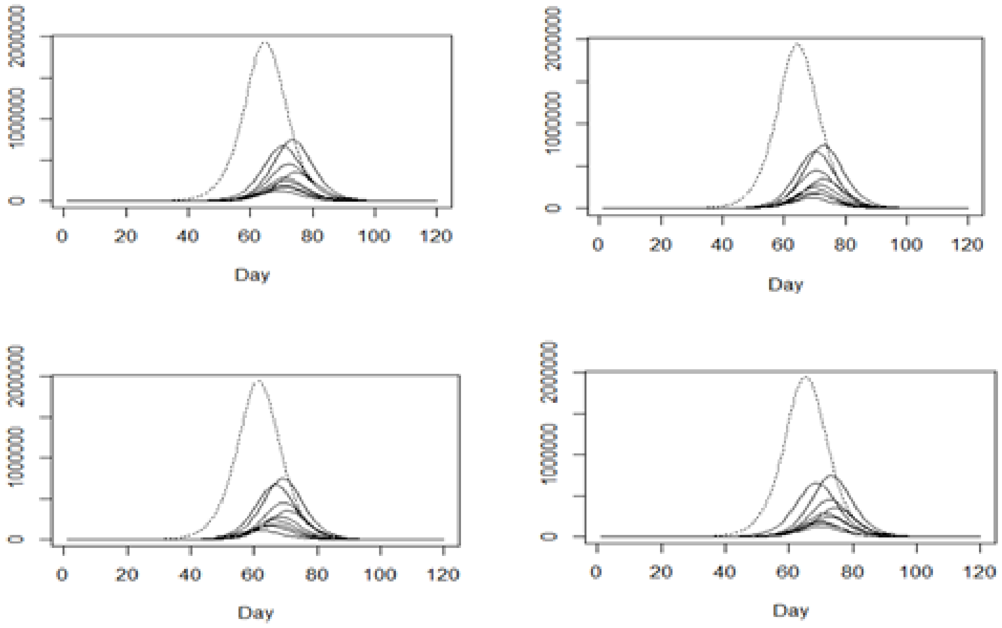

Figure 6.44. Number of daily infected people in cities.

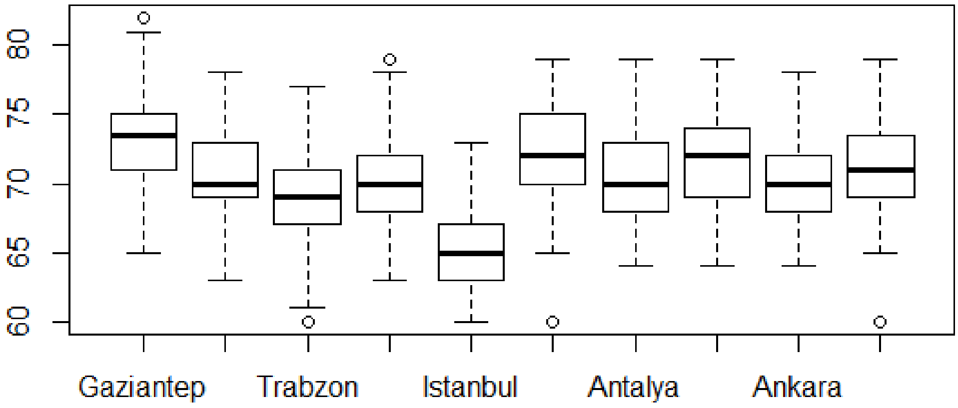

Figure 6.45. Peak days of cities.

Table 6.16. Average peak days of cities.

Gaziantep	Van	Trabzon	Samsun	İstanbul	İzmir	Antalya	Adana	Ankara	Kayseri
73	71	69	70	65	72	70	72	70	71

We want to learn the difference of peak days between İstanbul and other cities in order to observe the time that an outbreak begins compare to İstanbul.

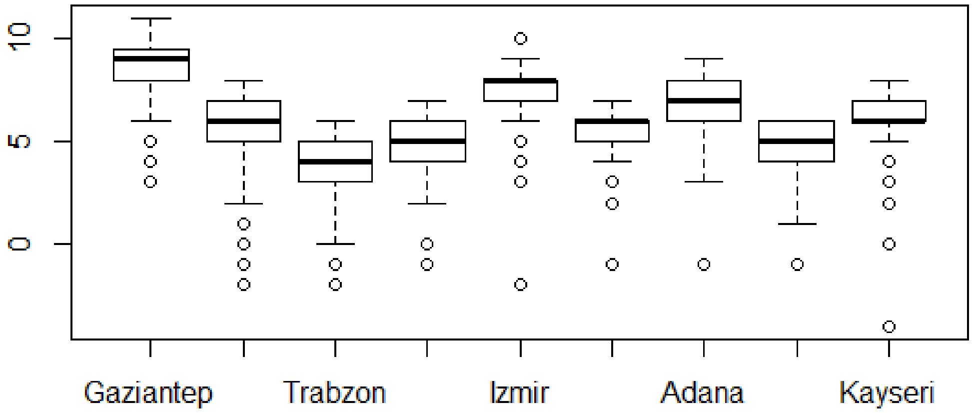

Figure 6.46. Delay of peak days between İstanbul and other cities.

Table 6.17. Average delay of peak days.

Gaziantep	Van	Trabzon	Samsun	İzmir	Antalya	Adana	Ankara	Kayseri
8.45	5.66	3.74	5.06	7.26	5.42	6.75	5.01	6.02

Table 6.18. Travelling flows and population.

	Gaziantep	Van	Trabzon	Samsun	İzmir	Antalya	Adana	Ankara	Kayseri
Flow	4032	4932	6588	4968	12240	8100	8100	21600	5148

According to results seen in the Figure 6.46 and Table 6.17, an outbreak begins in Trabzon after İstanbul and it continues with Ankara, Samsun, Antalya, Van, Kayseri, Adana, İzmir and Gaziantep respectively. In addition, it is seen that an outbreak begins earlier in cities except Gaziantep than İstanbul in one of the repetitions, in Van in two repetitions. In Van, Trabzon, Samsun and Kayseri, outbreak begins almost in the same time with İstanbul in one of the repetitions. Also, in İzmir and Antalya median of the peak delays are equal to upper quantile. In Kayseri, median of the peak delays are equal to lower quantile. In Figures 6.47-6.56, the days of an infection ends for each repetiton and city are shown. According to results, the infection duration in İstanbul is longer than other cities.

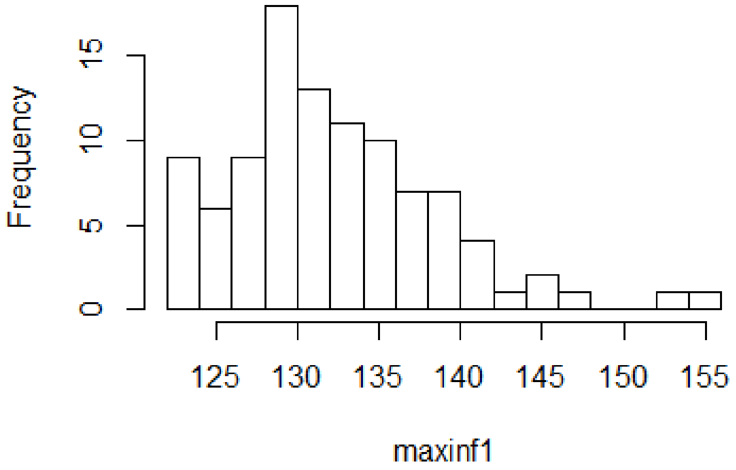

Figure 6.47. Maximum infected days of Gaziantep.

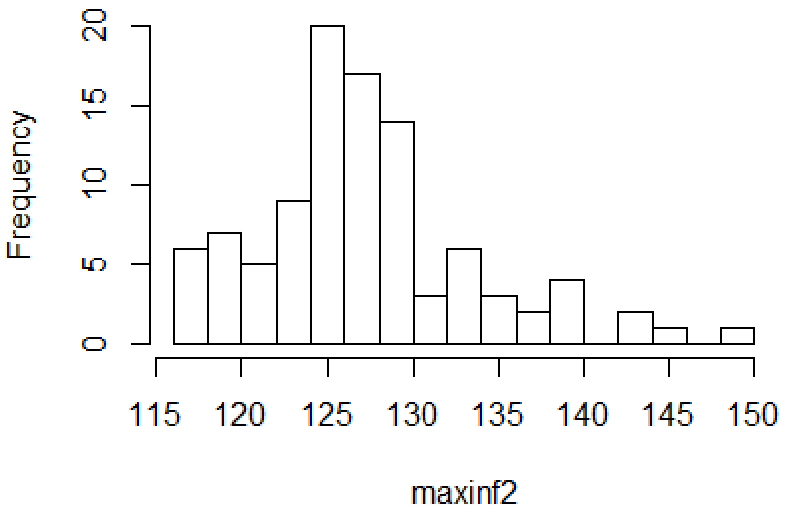

Figure 6.48. Maximum infected days of Van.

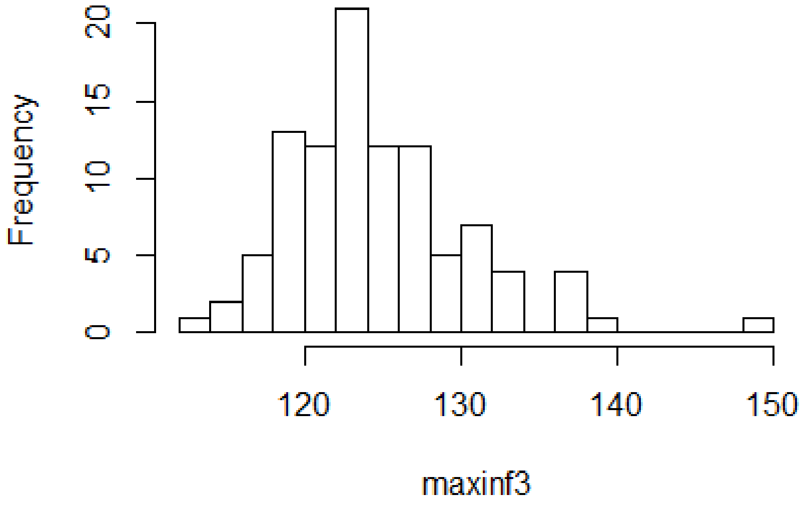

Figure 6.49. Maximum infected days of Trabzon.

In Van, maximum infected days are distributed around 120 and 135 days. When we compare the results with Trabzon, we see much higher maximum infected days in Van than Trabzon.

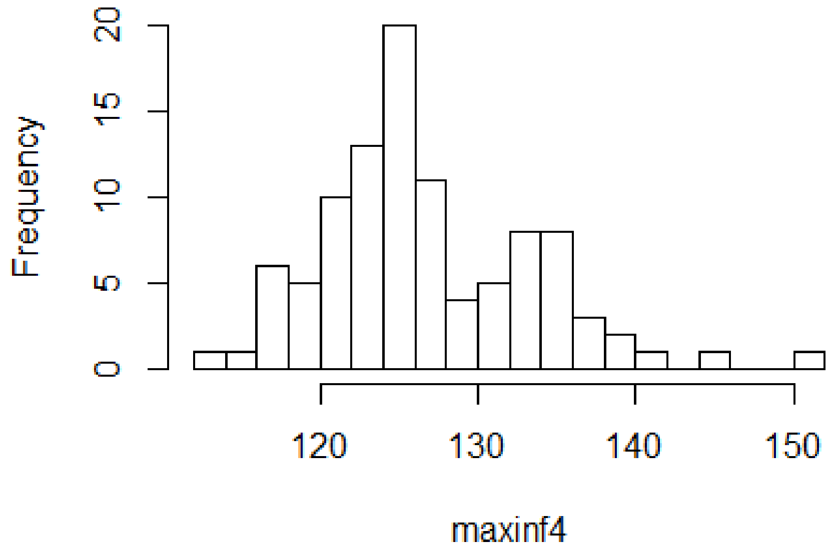

Figure 6.50. Maximum infected days of Samsun.

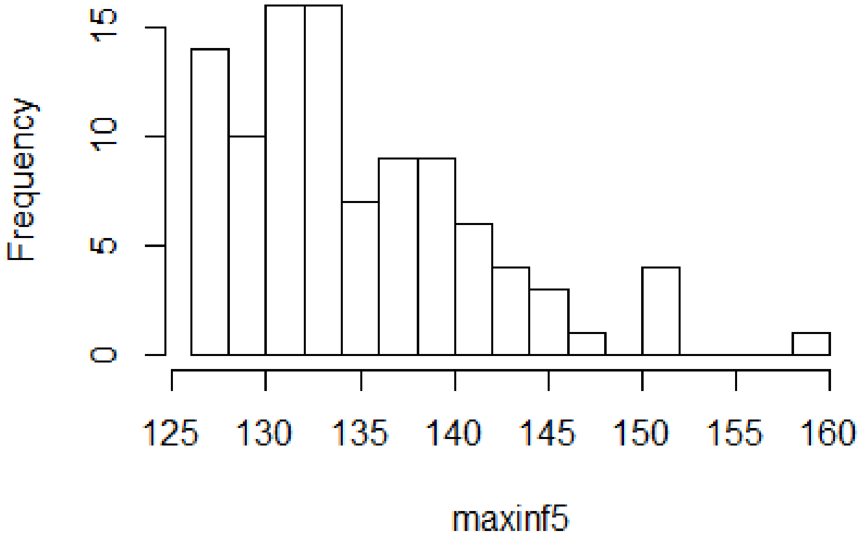

Figure 6.51. Maximum infected days of İstanbul.

In Samsun, maximum infected days are distributed around 120 and 140 days. When we compare the results with İstanbul, we see much higher maximum infected days in İstanbul than Samsun.

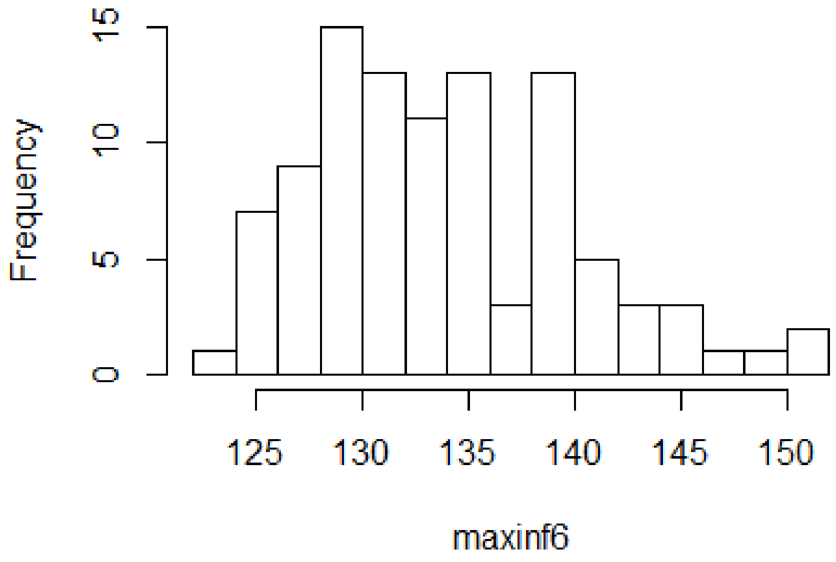

Figure 6.52. Maximum infected days of İzmir.

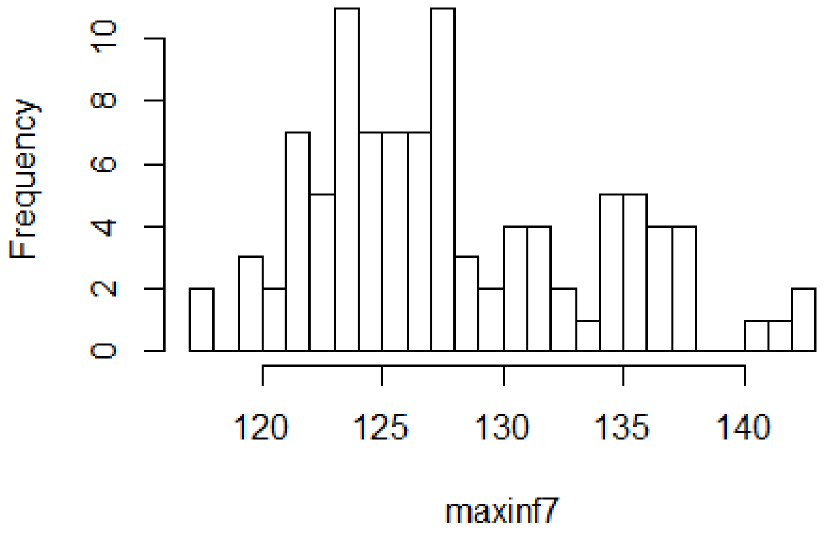

Figure 6.53. Maximum infected days of Antalya.

In İzmir, maximum infected days are distributed around 125 and 145 days. When we compare the results with Antalya, we see much higher maximum infected days in İzmir than Antalya.

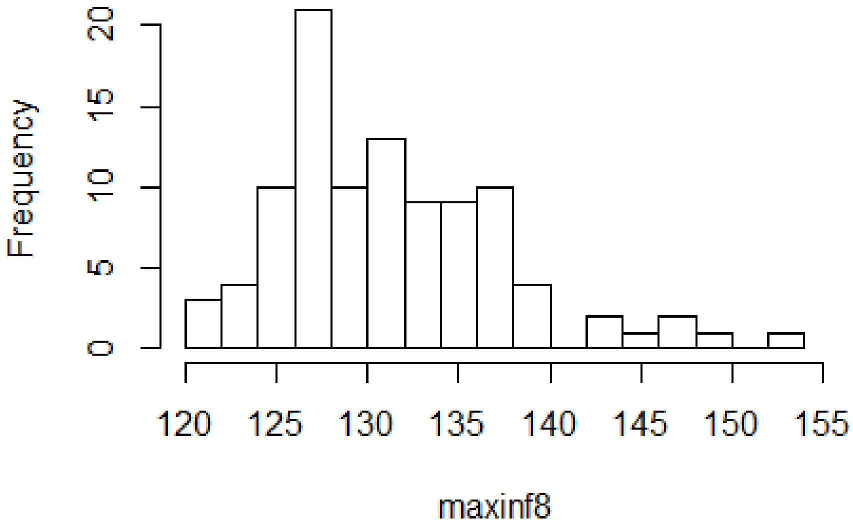

Figure 6.54. Maximum infected days of Adana.

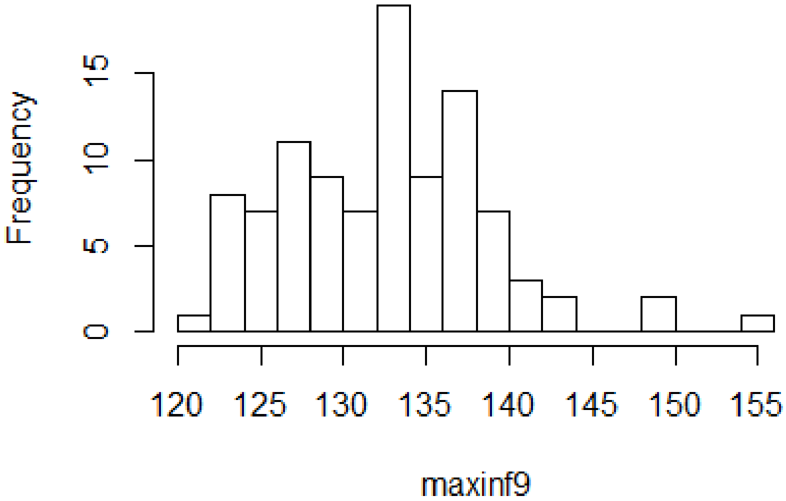

Figure 6.55. Maximum infected days of Ankara.

In Adana, maximum infected days are distributed around 125 and 140 days. When we compare the results with Ankara, we see much higher maximum infected days in Ankara than Antalya.

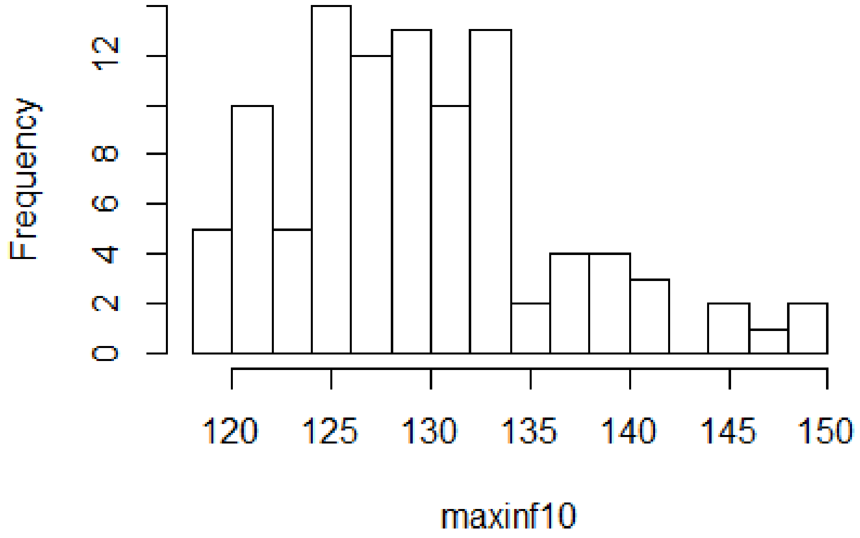

Figure 6.56. Maximum infected days of Kayseri.

Table 6.19. Average of maximum infected days.

Gaziantep	Van	Trabzon	Samsun	İstanbul	İzmir	Antalya	Adana	Ankara	Kayseri
155	150	149	152	159	152	143	153	155	150

In Figure 6.57, the total number of infected people in Gaziantep, Van, Trabzon, Samsun, İstanbul, İzmir, Antalya, Adana, Ankara and Kayseri are shown. Approximately more than 12,000,000 people in İstanbul, between 800,000 and 5,000,000 people in other cities are being infected. Also, in this plot it is seen that an outbreak begins in İstanbul earlier than other cities. It shows us that the results that we see in ths plot support the results that we see in Figure 6.44.

In addition, peak times of infection in Gaziantep, Van, Trabzon, Samsun, İzmir, Antalya, Adana, Ankara and Kayseri are lower than peak time of infection in İstanbul. Also, disease spreads fast between large cities such as İstanbul, Ankara and İzmir due to high number of people who are travelling. After peak time in all cities, number of infected cases are starting to be constant as there are no new infections.

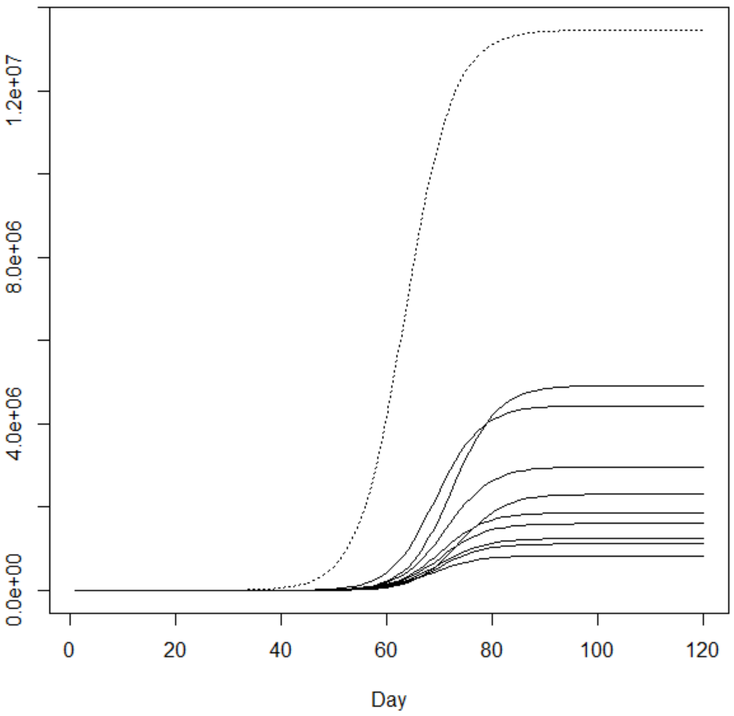

Figure 6.57. Total number of infected in cities.

In this section, we carried out simulation experiments and analyzed the results in order to understand the behaviour of a metapopulation model in Turkey. We constructed three models respectively including three cities, six cities and ten cities. In three cities model, we also focused on using vaccination to observe the effect on the disease spread. In some experiments we observed a herd immunity in local, but due to travelling between cities, it was still possible to see infected people in each city. In other words, herd immunity in one city is not enough. In order to prevent the spreading of the disease, herd immunity must be in all cities.

During the experiments we also measured the performance of the code. We observed that for 1000 repetitions and with large cities runtime duration is approximately two hours. The code works stable and can be modified according to different number of repetitions and cities.

7. CONCLUSION

We defined a stochastic SIR metapopulation model. We succeeded to implement a metapopulation simulation of disease spread models in R. R was enough to get useful results because it has a good performance and is easy to interpret results. This model is not only useful Turkey, it is also useful for other countries. In other words, it is flexible and can be modified for different cities. We observed that for 1000 repetitions and with large cities the runtime is approximately two hours. The code was tested and works correctly and stable.

We run the simulation code for a model with two large populations. We then conducted simulation experiments and considered various expected number of people who travel. We concluded that when R_0 is equal to 1.5, although the number of travelling is decreased considerably, it is difficult to prevent an outbreak in the second city. If an outbreak occurs in a city, the disease spreads to the other one almost surely unless the travelling rate is extremely low.

In order to carry out simulation experiments for Turkey, we needed some parameters such as the expected number of people who travel between cities. We were able to find data in the internet concerning main airlines and bus companies to get rough estimate of the number of people who travel between cities. For identifying daily number of flights and bus services, the related firm’s reservation system were examined in detail for all city pairs. It was difficult to find the exact number of people who travel between cities each day in the internet, therefore we did some assumptions on these information gathered from the internet to get rough estimates of the total travelling rates.

Right after investigating and defining parameters for Turkey, simulation experiments are conducted. We wanted to see how fast the epidemy will spread in Turkey. According to the results, due to the high rate of flows between cities and R_0 equals to 1.5, an outbreak occurs in all cities. We also considered vaccination in the largest three cities to see how it will affect the behaviour of the disease spread. We tried to find out what is the appropriate percentage of vaccination that we should use. We saw that as the vaccination rate increases, the probability of an outbreak decreases. In some experiments we observed a local herd immunity, but due to travelling between cities, it was still possible to see infected people in each city. In other words, herd immunity in one city is not enough. In order to prevent the spreading of the disease, herd immunity must be reached in all cities.

REFERENCES

1. Weiss, H., “The SIR Model and the Foundations of Public Health”, *Materials Mathematics,* No. 3, pp. 2, 2013.

2. Brauer, F., J.Wu, P. Van Den Driessche, *Mathematical Epidemiology*, Springer, pp.6-82, 1945.

3. Romulus B., R.Vardavas, S. Blower, “Theory versus Data: How to Calculate R_0?”, *Plos One Journals,* 2007

4. Ridenhour B., J. M. Kowalik, D. K. Shay, “Unraveling R_0: Considerations for Public Health Applications”, *Am J Public Health*, 2014

5. Alam, M.,Wei,W., *Agent-Based Models*, pp. 2-3

6. Frias-Martinez, E., G. Williamson, V. Frias-Martinez, “An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information*”,* University College Dublin, Ireland, Telefonica Research, Spain.

7. Ajelli, M., B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J.J. Ramasco, S. Merler and A. Vespignani, "Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models", *BMC infectious diseases*, Vol. 10, No. 1, pp. 1, 2010.

8. Longini, I.M., M.E. Halloran, A. Nizam and Y. Yang, "Containing pandemic influenza with antiviral agents", *American journal of epidemiology*, Vol. 159, No. 7, pp. 623- 633, 2004.

9. Longini Jr, I.M. and J.S. Koopman, "Household and community transmission parameters from final distributions of infections in households", *Biometrics,* Vol.38, No.1, pp. 115-126, 1982.

10. Khalil, K.M., et al., *An agent-based modeling for pandemic influenza in Egypt, in Handbook on Decision Making,* Springer, pp. 205-218, 2012.

11. Bektaş, A.B., *Simulation of Agent Based Disease Spread Models,* MSc, Boğaziçi University, 2017.

12. Martcheva M., *An Introduction to Mathematical Epidemiology,* Springer, Vol. 61, pp. 388-389, 2015.

13. Vazquez A., “Epidemic Outbreaks on Structured Populations”, *Journal of Theoretial Biology,* Vol. 245, pp. 125, 2007.

14. Apolloni A., C. Poletto, J. J. Ramasco, P. Jensen, V. Colizza, “Metapopulation Epidemic Models with Heterogeneous Mixing and Travel Behaviour”, *Theoretical Biology and Medical Modelling,* Vol.11, No. 3, pp. 1-26, 2014.

15. Colizza V., A. Vespignani, “Epidemic Modeling in Metapopulation Systems with Heterogeneous Coupling Pattern: Theory and Simulations”, *Journal of Theoretial Biology,* Vol. 251, pp. 450-467, 2008.

16. Siska V., *Disease Dynamics in Metapopulations,* University of Gothenburg, 2013.

17. Eriksson H., M.Morin, J.Ekberg, J.Jenvald, T.Timpka, “Assumptions Management in Simulation of Infectious Disease Outbreaks”, *AMIA Annual Symposium Proceedings,* pp. 173-177, 2009

18. Juher D., J. Ripoll, J. Saldana, “ Analysis and Monte Carlo Simulations of a Model for the Spread of Infectious Diseases in Heterogeneous Metapopulations”, *Physical Review,* Vol. 80, 2009.

19. Longini, I.M., A. Nizam, S. Xu, K.Ungchusak, W. Hanshaoworakul, D.A. Cummings and M.E. Halloran, "Containing pandemic influenza at the source", *Science,* Vol. 309, No. 5737, pp. 1083-1087, 2005.

20. Brauer F., *Deterministic Compartmental Disease Transmission Models,* pp. 1-66, 2012.

21. Turkish Airlines, *Turkish Airlines, Reservation System, Manage Booking,* https://www.turkishairlines.com, accessed at April 2017.

22. Pegasus Airlines, *Pegasus Airlines Reservation System, Manage Booking,* https://www.flypgs.com, accessed at April 2017.

23. Ulusoy, *Ulusoy Bus Company Reservation System, Manage Booking,* https://www.ulusoy.com, accessed at April 2017.

24. Metro, *Metro Bus Company Reservation System, Manage Booking,* https://www.metroturizm.com, accessed at April 2017.

25. Kamilkoç, *Kamilkoç Bus Company Reservation System, Manage Booking,* https://www.kamilkoc.com, accessed at April 2017.

APPENDIX A: R CODES

```
RepSIRdi <-
function(n=1000,d=20,npopV=c(100,100,100),R0=3,I0=c(0,0,0),pmfdi=c(0.2,0.5,0.3),pmfdexc=c(0.2,0.5,0.3),excV=c(1,2,3),facpasym=1/2,propsympt=2/3,pmf1day=c(0.2,0.3,0.5),pmfsend=c(1,1,1,0.6,0.7,1,1,1),pmfstayhome=c(0.1,0.2,0.1),pimm=c(0.2,0.4,0.3)){

  nc<-length(npopV)

  dimax<-length(pmfdi)

  dvmax<-length(pmfdexc)

  pinfV<-numeric(nc)

  excm<-matrix(0,nc,nc)

  excm[lower.tri(excm,diag=FALSE)]<-excV

  excm<-t(excm)

  excm[lower.tri(excm)]<-t(excm)[lower.tri(excm)]

  ratematrix<-function(excm,npopV,nc){

  #This function returns the exchange rate matrix between cities

    rate<-matrix(0,nc,nc)

    for(r in 1:nc){

     rate[r,]<-excm[r,]/npopV[r]

      }

return(rate)

}

pexc<-ratematrix(excm,npopV,nc)
```

Figure A.1. Implementation Code.

```
expectdi <- sum((1:dimax)*pmfdi)

pinfV<-R0/(expectdi*npopV*(propsympt+(1-propsympt)*facpasym))

if(length(I0)==1){

 I0V<-rep(0,nc)

 I0V[sample(1:nc,1)]<-I0 }

else{

 I0V<-I0}

 resm<-array(NA,dim=c(n,d,length(npopV),2))

 resmIR<-array(NA,dim=c(n,d,length(npopV),2))

OneSim<-function(){
sa<-array(0,dim=c((2+dimax*2),1+nc*dvmax,nc))
 imm<-numeric(nc)
 for(p in 1:nc) {
 sa[1:3,1,p] <- c(npopV[p]-I0V[p],0,I0V[p])
 sa[4:(dimax*2+2),1,p]<-0
 sa[1:(dimax*2+2),(p-1)*dvmax+(1:dvmax)+1,p]<-NA}
  for(f in 1:nc){
  imm[f]<-rbinom( 1, size=sa[1,1,f], prob=pimm[f])
  sa[2,1,f]<-sa[2,1,f]+imm[f]
  sa[1,1,f]<-sa[1,1,f]-imm[f]}
   resI <- array(0,dim=c(nc,d,2))
   resIR<- array(0,dim=c(nc,d,2))
```

Figure A.1. Implementation Code (cont.).

```
for(i in 1:d)

nw<-generatenewlyinfected(sa,dvmax=dvmax,nc=nc,dimax=dimax,pmfdi=pmfdi,pinfV=pinfV,facpasym=facpasym)

nwsymp<-array(0,dim=c(1,1+(nc-1)*dvmax,nc))

for(y in 1:nc){

nwsymp[1,,y]<-rbinom(1+(nc-1)*dvmax,size=nw[1,,y],prob=propsympt) }

nwasymp<-nw-nwsymp

sa[2,,]<-sa[2,,]+sa[(dimax+2),,]+sa[(dimax*2+2),,]

sa[5:(dimax+2),,]<-sa[4:(dimax+2-1),,]

sa[(5+dimax):(dimax*2+2),,]<-sa[(4+dimax):(dimax*2+1),,]

sa[4,,]<-0

sa[4+dimax,,]<-0

 after1dayS<-array(0,dim=c(dimax,1+(nc-1)*dvmax,nc))

after1dayA<-array(0,dim=c(dimax,1+(nc-1)*dvmax,nc))

 for(h in 1:nc){

 after1dayS[,,h]<-rmultinomv(n=1+(nc-1)*dvmax,size=sa[3,-((h-1)*dvmax+(1:dvmax)+1),h],prob=pmf1day)

 after1dayA[,,h]<-rmultinomv(n=1+(nc-1)*dvmax,size=sa[3+dimax,-((h-1)*dvmax+(1:dvmax)+1),h],prob=pmf1day) }
```

Figure A.1. Implementation Code (cont.).

```
for(k in 1:nc){

 sa[2,-((k-1)*dvmax+(1:dvmax)+1),k]<-sa[2,-((k-
1)*dvmax+(1:dvmax)+1),k]+after1dayS[1,,k]+after1dayA[1,,k]

sa[4:(dimax+2),-((k-1)*dvmax+(1:dvmax)+1),k]<-sa[4:(dimax+2),-((k-
1)*dvmax+(1:dvmax)+1),k]+after1dayS[(2:dimax),,k]

 sa[(4+dimax):(dimax*2+2),-((k-1)*dvmax+(1:dvmax)+1),k]<-sa[(4+dimax):(dimax*2+2),-((k-
1)*dvmax+(1:dvmax)+1),k]+after1dayA[(2:dimax),,k]

 sa[3,,k]<-0

sa[3+dimax,,k]<-0}

for(g in 1:nc){

sa[3,(g-1)*dvmax+(1:dvmax)+1,g]<-NA

sa[3+dimax,(g-1)*dvmax+(1:dvmax)+1,g]<-NA}

 if(i<=50){extrainfected<-rbinom(1,size=100,prob=0.01

 else {extrainfected<-0}

 for(h in 1:nc){

 sa[3,-((h-1)*dvmax+(1:dvmax)+1),h]<-sa[3,-((h-1)*dvmax+(1:dvmax)+1),h]+nwsymp[1,,h]

sa[(3+dimax),-((h-1)*dvmax+(1:dvmax)+1),h]<-sa[(3+dimax),-((h-
1)*dvmax+(1:dvmax)+1),h]+nwasymp[1,,h]

 sa[1,-((h-1)*dvmax+(1:dvmax)+1),h]<- sa[1,-((h-1)*dvmax+(1:dvmax)+1),h]-nwsymp[1,,h]-
nwasymp[1,,h]   }

 sg<-array(0,dim=c(2+dimax*2,nc*dvmax+1,nc))

 for(s in 1:nc){

 sg[1:(dimax*2+2),(s-1)*dvmax+(1:dvmax)+1,s]<-NA }
```

Figure A.1. Implementation Code (cont.).

```
for(f in 1:nc){
 for(z in 1:(2+dimax*2)){
 sg[z,-((f-1)*dvmax+(1:dvmax)+1),f]<-rbinom((nc-1)*dvmax+1,size=sa[z,-((f-1)*dvmax+(1:dvmax)+1),f],prob=pmfsend[z]) }}
sh<-array(0,dim=c(dimax,nc*dvmax+1,nc))
 for(j in 1:nc){
 sh[1:dimax,(j-1)*dvmax+(1:dvmax)+1,j]<-NA}
 for(p in 1:nc){
 for(k in 1:dimax){
 sh[k,-((p-1)*dvmax+(1:dvmax)+1),p]<-rbinom((nc-1)*dvmax+1,size=sg[k+2,-((p-1)*dvmax+(1:dvmax)+1),p],prob=pmfstayhome[p])}}
sg[3:(dimax+2),,]<-  sg[3:(dimax+2),,]-sh[1:dimax,,]
sa<-send(dvmax=dvmax,sg,pexc=pexc,pmfdexc=pmfdexc,nc)
 rst<-array(0,dim=c(nc,d,2))
 rstR<-array(0,dim=c(nc,d,2))
 for(rs in 1:nc){
  rst[rs,,1]<-sum(sa[3:(dimax+2),,rs],na.rm=TRUE)
  rst[rs,,2]<-sum(sa[3:(2*dimax+2),,rs],na.rm=TRUE)
  rstR[rs,,1]<-sum(sa[2:(dimax+2),1,rs])+sum(sa[2:(dimax+2),(rs-1)*dvmax+(1:dvmax)+1,],na.rm=TRUE)
  rstR[rs,,2]<-sum(sa[2:(2*dimax+2),1,rs])+sum(sa[2:(2*dimax+2),(rs-1)*dvmax+(1:dvmax)+1,],na.rm=TRUE)}
```

Figure A.1. Implementation Code (cont.).

```
resI[,i,1]<-rst[,i,1]
 resI[,i,2]<-rst[,i,2]
 resIR[,i,1]<-rstR[,i,1]
 resIR[,i,2]<-rstR[,i,2]}
 return(list(I=resI,IR=resIR,IM=imm)) }
 for(i in 1:n){
 inf<-OneSim()
 mm<-inf$IM
 for(y in 1:2){
 for(h in 1:length(npopV)){
 resm[i,,h,y]<- inf$I[h,1:d,y]
 resmIR[i,,h,y]<- inf$IR[h,1:d,y]}}}
 return(list(I=resm,IR=resmIR,IMM=mm))
generatenewlyinfected<-function(sa,dvmax=3,nc=3,dimax=3,pmfdi=c(0.2,0.3,0.5),pinfV=c(0.001,0.002,0.003),facpasym=1/2){
 #This function calculates and returns the newly infected cases for each city
newinf<-array(0,dim=c(1,1+(nc-1)*dvmax,nc))
for(x in 1:nc){
pinfection <- 1-(1-pinfV[x])^sum(sa[3:(dimax*2+2),,x],na.rm=TRUE)*(1-pinfV[x]*facpasym)^sum(sa[(3+dimax):(2+2*dimax),,x],na.rm=TRUE)
newinf[,,x] <- rbinom( (1+(nc-1)*dvmax), size=sa[1,-((x-1)*dvmax+(1:dvmax)+1),x], prob=pinfection)
```

Figure A.1. Implementation Code (cont.).

```
}  return(newinf)   }

probforsend<-function(nc,pmfdexc,pexc){

  #By the help of this function, probability values of staying in other cities for a length of time
are calculated by using pmfdexc and pexc inputs as described in the general function of the code

 dvmax<-length(pmfdexc)

 probarray<-array(0,dim=c(dvmax,nc+1,nc))

 for(p in 1:nc){

 probarray[1:(dvmax),p,p]<-NA }

 for(i in 1:(nc-1)){

 for(j in (i+1):nc){

 probarray[,j,i]<-pmfdexc*pexc[i,j]

probarray[,i,j]<-pmfdexc*pexc[j,i] } }

 for(k in 1:nc) {

 probarray[1,nc+1,k]<-1-sum(probarray[,-k,k])

 probarray[2:dvmax,nc+1,k]<-NA }

 return(probarray)}

send<-function(dvmax,sa,nc,pmfdexc,pexc){

  #This function returns the new sa array after sending people among the cities

 prb<-probforsend(nc=nc,pmfdexc=pmfdexc,pexc=pexc)

 trv<-array(0,dim=c(length(sa[,(nc*dvmax+1),1]),nc*dvmax+1,nc))

for(p in 1:nc){

trv[,(p-1)*dvmax+(1:dvmax),p]<-NA }
```

Figure A.1. Implementation Code (cont.).

```
for(t in 1:nc){
 tr<-generatesend(sa[,1,t],prb[,,t],dvmax=dvmax,nc=nc)
 trv[,-((t-1)*dvmax+(1:dvmax)),t]<-tr #Assigning values to the related place in the travel array
 }
updtcity<-array(0,dim=c(length(sa[,(nc*dvmax+1),1]),1,nc))
for(l in 1:nc){
 updtcity[,1,l]<-trv[,nc*dvmax+1,l] }
 for(n in 1:nc){
 sa[,1,n]<-updtcity[,1,n]
 trv<-trv[,-(nc*dvmax+1),]
 for(i in 1:(nc-1)){
 for(j in (i+1):nc){
#From i to j sendings
 sa[,1,i]<-sa[,1,i]+sa[,(i-1)*dvmax+2,j]
 sa[,((i-1)*dvmax+2):((i-1)*dvmax+dvmax),j]<-sa[,((i-1)*dvmax+3):((i-
1)*dvmax+dvmax+1),j]
 sa[,(i-1)*dvmax+dvmax+1,j]<-0
 sa[,(i-1)*dvmax+(1:dvmax)+1,j]<-sa[,(i-1)*dvmax+(1:dvmax)+1,j]+trv[,(j-
1)*dvmax+(1:dvmax),i]
#From j to i sendings
sa[,1,j]<-sa[,1,j]+sa[,(j-1)*dvmax+2,i]
sa[,(j-1)*dvmax+dvmax+1,i]<-0
```

Figure A.1. Implementation Code (cont.).

```
sa[,(j-1)*dvmax+(1:dvmax)+1,i]<-sa[,(j-1)*dvmax+(1:dvmax)+1,i]+trv[,(i-1)*dvmax+(1:dvmax),j]}}

return(sa)}

generatesend<-function(s,prb,dvmax,nc){

#This function returns the matrix includes the number of people travel between each city

   prb<-prb[!is.na(prb)]

  R<-matrix(0,nrow=dvmax,ncol=length(s))

  R<-rmultinomv(n=length(s),size=s,prob=prb)

  RN<-t(R)

 return(RN)}
```

Figure A.1. Implementation Code (cont.).