How to Cite:
Choi, M.- hyuk, & Masatsugu, N. (2022). The combination of factors influencing COVID-19 infections in Korea, by region: Performing neural network analysis and qualitative comparative analysis. *International Journal of Health Sciences*, *6*(S2), 13311–13333. https://doi.org/10.53730/ijhs.v6nS2.8527

The combination of factors influencing COVID-19 infections in Korea, by region: Performing neural network analysis and qualitative comparative analysis

Min-hyuk Choi
MPH Student, Mailman School of Public Health, Columbia University, USA
Corresponding author email: jesmondmh@naver.com

Nemoto Masatsugu
Research Professor, IDI, Chungbuk National University, Korea

***Abstract*---**Given the fact that the number of COVID-19 infections in Korea shows a large variation by region, this study attempted to analyse what combination of factors might lead to a decrease in the number of infections. As the dependent variable, the proportion of the local population infected with the virus (i.e. the coronavirus infection rate) was used. Sixteen independent variables were then selected, and five of them which have an important effect were selected from among these. Neural network analysis and Qualitative Comparative Analysis (QCA) were the analysis methods used. In selecting the analysis period, from January 2020 to December 2021, it was found that COVID-19 in Korea could be divided into three stages, each of which could be explained by a different combination of conditions.

***Keywords*---**corona in Korea, neural network analysis, QCA, corona infection rate, combination condition.

Introduction

On 20 January 2020, the first Covid case was reported in Korea. Around two years later, as of 31 January 2022 the cumulative number of coronavirus patients stood at 845,646 (http://ncov.mohw.go.kr/). When looking at Korea's seventeen metropolitan authorities, we see that there is large regional variation in the number of Covid patients per 1,000 population: some regions have a large number of patients relative to the population and some only a small number of cases. At the national level too, some countries have a high number of cases expressed as a percentage of their population, and some have only a relatively

Manuscript submitted: 09 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 27 May 2022

small number of cases. It is clear that several factors are influencing the number of Covid patients, but what factors are these?

The ease with which individuals get Covid depends on personal factors such as their health or immune status. It may also be assumed that regional environmental factors play a part (Poudel et al., 2020; Torales et al., 2020; bakar and Rosbi, 2020). More research is needed to determine which of these two broad factors has the greater influence. In any event, if we can establish the impact of various factors on the number of Covid patients, we will be able to combat infections not only in the current Covid situation, but also in the case of a similar pandemic in the future. With Covid, it is very difficult to analyse the influence of a given person's individual factors on communicable disease infection. This is because it is difficult to analyse a patient's personal immune system, underlying disease(s), dietary habits and personal factors (e.g. genetic factors) unless sufficient data are provided. Conversely, if regional factors are analysed in relation to the number of patients in each region after the characteristics of the region have been identified, meaningful research results can be derived (Blundell et al., 2020; Johansson et al., 2020; Reinsdorf, 2020).

Against this background, this study aims to analyse why the incidence of patients differs by region, excluding individual/physical factors and focusing purely on regional factors. The unit of analysis chosen is the Korean metropolitan authority. Thus the subject of the analysis is the seventeen metropolitan authorities in Korea. In other words: if regionally characteristic factors that can be assumed to have an impact on the incidence of Covid patients are identified and analysed in relation to the number of patients in specific metropolitan authorities, it is possible to analyse why a particular region has a large number of patients. By drawing on these research results, local authorities should in the future be able to prevent and control infectious diseases such as coronavirus by appropriately controlling local conditions.

Theoretical Discussion and Research Questions

Research on COVID-19 has been conducted in various fields over the past two years. Nevertheless, it is difficult to ascertain the temporal and spatial characteristics of the spread of the virus in a situation where thousands, or tens of thousands, of new confirmed cases occur every day due to the rapid spread of the virus. Consequently, studies on major infection routes and group infections are relatively inadequate. Understanding the main transmission route of COVID-19 and the temporal and spatial characteristics of group infection is very important (Serafine et al., 2020; Reinsdorf, 2020), because it is directly related to the quarantine policy of individuals and governments. As outlined above, the aim of this study is, first, to analyse whether on the basis of regional data for confirmed Covid cases the number of people infected differs by region, and then to analyse the factors affecting infection. The objective is to improve the ability to respond to infectious diseases.

One important point to observe here is that the 'shape' of the influence of regional conditions at each stage of the spread of an infectious disease will be different. In other words, it can be assumed that some local conditions may have a particular impact when an infectious disease is experiencing a lull, whereas other local conditions may exert a strong influence when the spread is rapid. This is because certain local conditions may be suitable for controlling the initial stage of an infectious disease, while other such conditions may effectively contribute to its spread. Therefore, even in a specific country or region it is somewhat problematic to use the number of infected people per 1,000 population as an average measure from the beginning to the end of the pandemic. Instead, we divide the progress of the pandemic into stages, such as the initial stage, the spread stage and the deep stage, and attempt to discover the factors affecting the number of infections at each stage.

Certainly, it is important to attempt to lower infection rates. However, in previous studies (Want et al., 2020; Johansson et al., 2020), analytical efforts using regional influencing factors proved insufficient because it was difficult to identify the infectious disease path. As a result, studies of the factors affecting the increase or decrease in the number of infected persons by region are quite omissive (Rebecca and Ziyang, 2020; Lee, 2020; Diewert, 2020; Want et al., 2020). This situation also applies to Korea, and as a result there are few studies analysing the relative differences in the number of infected persons by regional government in Korea.

Of course, the factors affecting the spread of COVID-19 may include seasonal factors (Choi, 2022a; Choi, 2022b). Various other influencing factors can be included, such as whether individuals engage in activities mainly in closed rooms, degrees of personal health, and the preventive measures and counter-measures adopted by local authorities. The main infection routes have been changing, depending on the periods of the first, second and third epidemics; and strong quarantine measures, such as restrictions on behaviours enforced by local authorities, can be effective in some cases in reducing the number of confirmed cases. Nevertheless, it is not easy to analyse the factors affecting the increase or decrease in the number of infections at the regional level. Against this background, this study set the following research questions:

What are the major factors affecting the increase or decrease in the number of infected persons in a particular region?
What are the conditions for combining the variables that affect reduction of the local infection rate?

Survey Design

Unit of analysis

The unit of analysis in this study is a metropolitan authority. Local authorities in Korea are made up of a basic self-governing body and a metropolitan self-governing body. In Korea, the metropolitan authorities number seventeen, and the total number of local authorities is 226. In order to study the increase or decrease in the number of infected persons by region on the basis of regional differences, it is appropriate to draw on data at the level of local authorities. Obtaining these data is difficult, however, and so this study takes the seventeen metropolitan authorities in Korea as the basic focus of analysis.

Analysis method

In order to answer Research Question 1, neural network analysis is used. This method imitates the learning techniques of the human brain, and aims to discover patterns and features contained in data by repeatedly learning via a machine learning algorithm presented with given data. It may be termed an analysis algorithm that is used to classify and predict future events by generalising the results derived through this process. By applying this neural network analysis to Research Question 1, we intend to derive five major influencing factors that affect the reduction of the infection rate (Wen et al., 2018).

Meanwhile, in order to solve Research Question 2, Qualitative Comparative Analysis (QCA) is applied. This is a research methodology that analyses the qualitative and quantitative data collected in order to identify the phenomenon that the researcher aims to capture. The researcher constructs a coding frame, and classifies the collected data into each category in the frame so as to understand the meaning of the data (Cooper, 2005; Cronqvist, 2003; Goertz, 2020). Most studies using QCA analyse the qualitative characteristics of text, but quantitative data can also be analysed using binary data (Cronqvist, 2009; Glasesser and Cooper, 2012; Goertz, 2020; Goldthorpe, 1998).

There are several ways in which quantitative data can be converted into binary data, but in this study a threshold was derived using the Tool for Small-N Analysis (TOSMANA) program, and the values of all variables were converted into 1 and 0 on the basis of this threshold and analysed (Lieberson, 1998; Liebersoun, 1985; Killewald et al., 2017). By using the values of the dependent and independent variables converted into 1 and 0, we attempt to derive the overlapping conditions for the variables that affect the proportion of infected people by region considering the size of the population. In other words, the values of the independent variables derived by applying neural network analysis are converted into 1 and 0 using a threshold and analysed, and then the dependent variable, the combination of conditions that affect the reduction of the infection rate in specific local authorities, is analysed.

Analysis procedure

The two years from January 2020, when the first coronavirus infection in Korea occurred, to December 2021 was first set as the analysis period. Table 1 shows infections by month for the two years at national level, allowing the trend in the numbers of infections to be judged.

Table 1Number of infected persons by month
(January 2020–December 2021)

Period	confirmed	Period	confirmed
Jan. 20	11	Jan. 21	17,494
Feb. 20	2,920	Feb. 21	11,495
Mar. 20	6,855	Mar. 21	13,430

Apr. 20	979	Apr. 21	18,938
May 20	703	May 21	17,902
Jun. 20	1,334	Jun. 21	16,109
Jul. 20	1,506	Jul. 21	41,401
Aug. 20	5,642	Aug. 21	53,121
Sep. 20	3,865	Sep. 21	59,899
Oct. 20	2,714	Oct. 21	53,450
Nov. 20	7,768	Nov. 21	82,565
Dec. 20	26,564	Dec. 21	183,680

Source: https://coronaboard.kr/

Figure 1 is a graph showing the monthly number of coronavirus infections as presented in Table 1. The number of infections over the two years are shown divided into three stages. The period from January 2020 to November 2020 is a stage showing a relatively low number of infections. December 2020 to October 2021 represents the second stage. On the other hand, from November 2021 the number of infected people increases explosively, and this marks the third stage. Although the stage 3 time-period is relatively brief, it is assumed that the increase in the numbers of infected persons applying there will continue for the time being. In the Table 2, the dependent variable, the infection rate, represents the average number of infected persons for each time-period, corresponding to each stage.

Since it is necessary to derive a combination of conditions affecting the ratio of the number of infected persons at each stage, the overall analysis can be divided into three stages so as to identify the combined conditions of the influencing factors at each stage.

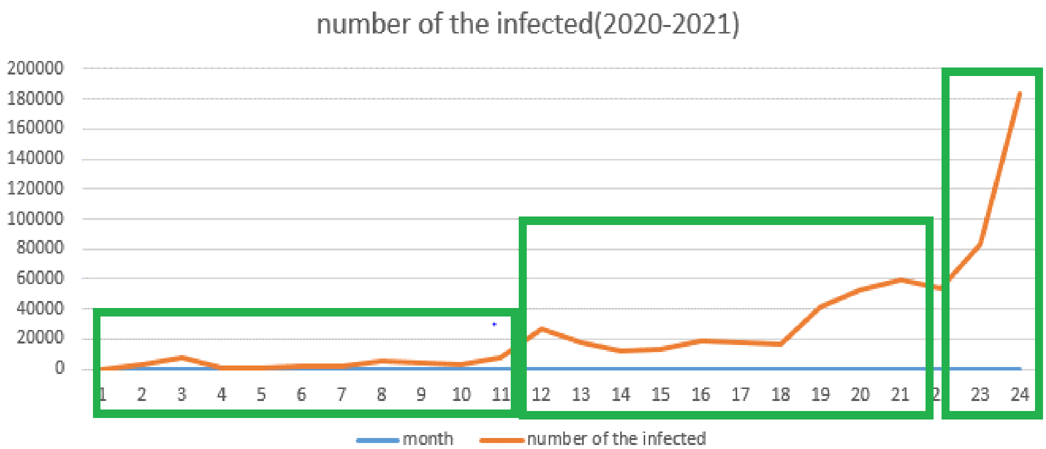


Figure 1 Classification of monthly confirmed cases (January 2020–December 2021)
Note. The periods of the three phases are shown in green.

Analysis variable

Although regional factors that affect coronavirus infection vary, this study intends to select sixteen regional influencing variables on the basis of the results of a number of previous studies. The dependent variable is the proportion of the number of infected persons per 1,000 people in a given region. Here, 'region' means a metropolitan authority, and statistics for the basic local authorities included in this regional government are collected and used. By way of statistics for the proportion of the number of infected people at each stage, the average value of the number of persons infected per month during each stage is used. Explanations of the sixteen independent and dependent variables are shown in Table 2.

Table 2 Descriptions and abbreviations of the analysis variables

Variable name	*Explanation*	*Abbreviation*	*Remarks*
population	population	population	
Number of infected persons in stage 3	Average number of persons infected between November 2021 and December 2021	*confirmed(21-12)*	dependent variable
Number of infected persons in stage 2	Average number of infected persons December 2020–October 2021	*confirmed(21-10)*	dependent variable
Number of infected persons in stage 1	Average number of infected persons January–November 2020	*confirmed(20-11)*	dependent variable
Vaccination rate (primary)	Percentage of people who received the primary coronavirus vaccine	*vaccinated(1st)*	
Percentage of single-person households	Proportion of households with one head of household	*singleperson*	
Healthy living practice rate	Percentage of residents who practised the rules of healthy living set by the Government	*healthyliving*	
Ratio of elderly population	Proportion of the population over 65	*elderly(65over)*	
Ratio of elderly households living alone	Proportion of elderly living alone households	*elderlyalone*	
Unmet medical rate	Proportion of people living without minimum medical conditions	*unmetmedical*	
Obesity prevalence	Percentage of people	*obesity*	

	determined to be obese		
Stress awareness	Percentage of people who think they are under serious stress	*stressperception*	
Drinking rate	Percentage of people who drink more than a certain standard amount per day	*drinkingrate*	
Number of hospital beds per 1,000 population	Number of hospital beds per 1,000 population	*nobed1000*	
Number of doctors employed in medical institutions per 1,000 population	Number of doctors working in medical institutions per 1,000 people	*doctor1000*	
Influenza vaccination rate	Influenza vaccination rate	*influenzavaccination*	
Financial independence	Financial independence of local authorities	*financialindep*	
Financial discretion	Proportion of resources over which local authorities have discretion	*financialdescre*	
Subjective health awareness	Percentage of people who consider themselves healthy	*healthawareness*	
Smoking rate	Percentage of people who smoke above a certain threshold	*smokingrate*	

Analysis Result

Result for stage 1

Neural network analysis was performed to derive the important variables that affected the number of infected persons relative to the population in the first stage. The results are shown in Figure 2.

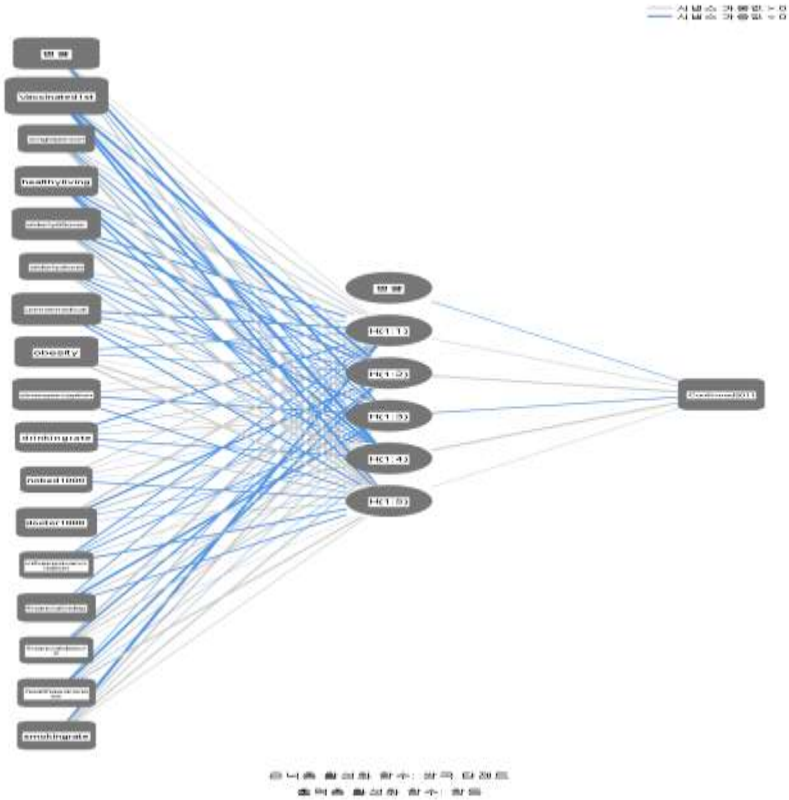


Figure 2 Neural network analysis result for stage 1

Table 3 shows the importance of each independent variable affecting the dependent variable. The variable that had the greatest influence on the ratio of infected persons to the whole population in the first stage is the ratio of primary vaccine receipt (*vaccinated(1st)*). In other words, the more vaccinated the population, the lower the number of infected persons. This means that the proportion of people who received the vaccine is really important in explaining why some regions are high in terms of the infected people and others are not. The next most important variables were: the proportion of people not meeting the minimum medical condition required for healthy living (*unmetmedical*); the proportion of persons aged 65 and over in the total population of the region (*elderly65over*); the proportion of persons in the population who feel they are under serious stress (*stressperception*); the proportion of persons judged to be obese (*obesity*); etc. It may be said that the five independent variables presented above are those that have a significant impact on the overall proportion of coronavirus-infected persons in the local population.

Table 3 Importance of each variable in stage 1

Variable	*Importance*	*Normalised importance*
Vaccinated(1st)	0.174	1
unmetmedical	0.102	0.587

elderly(65over)	0.1	0.572
stressperception	0.095	0.545
obesity	0.073	0.417
healthyliving	0.071	0.407
drinkingrate	0.067	0.383
doctor1000	0.058	0.335
smokingrate	0.047	0.27
financialindep	0.046	0.266
healthawareness	0.044	0.253
singleperson	0.038	0.219
elderlyalone	0.026	0.148
influenzavaccination	0.024	0.136
financialdescre	0.022	0.127
nobed1000	0.013	0.075

Note. Normalised importance shows the relative importance of each variable.

Figure 3 shows the normalised importance of the variables. Assuming that the value of the most important variable is 1, this gives us a value for the relative importance of the other variables.

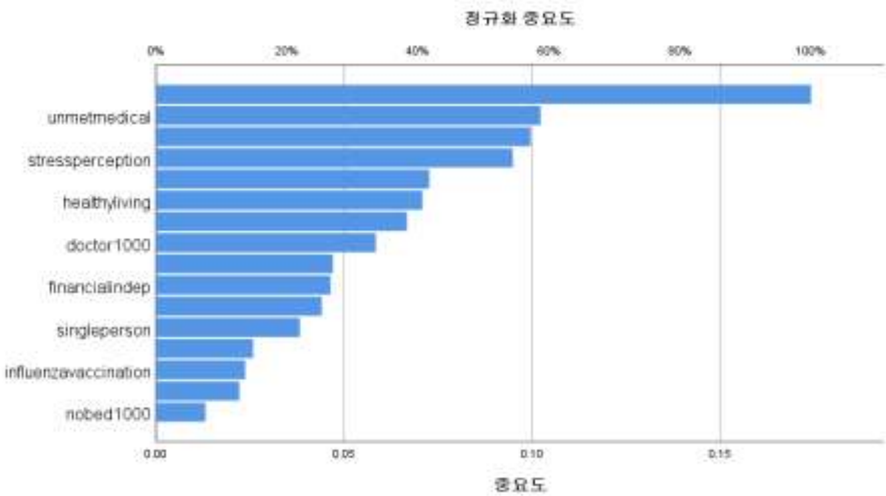
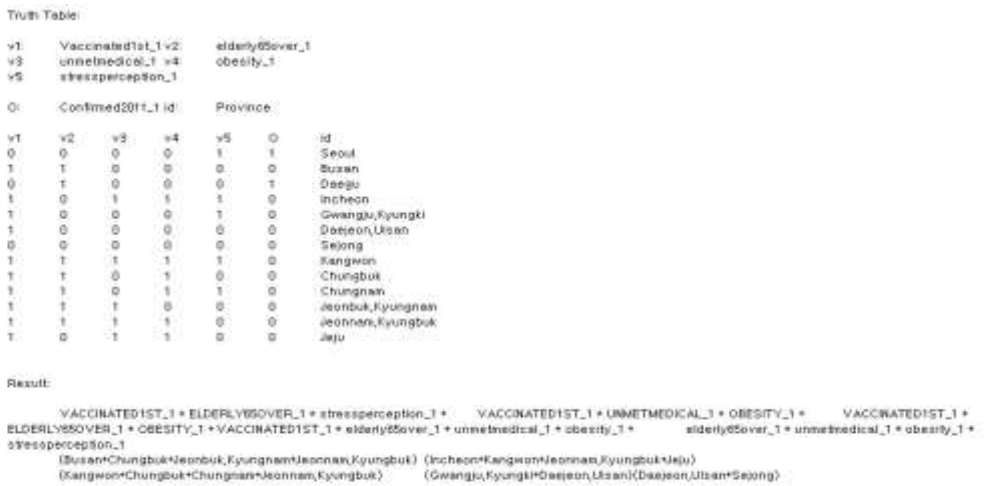


Figure 3 Importance of normalisation of variables in stage 1

As mentioned above, it is necessary to analyse whether the dependent variable, the number of infected persons, decreases when the five independent variables affecting the ratio of infected persons have any binding conditions as discussed above. For this purpose, the QCA method is employed as described above. However, in order for QCA analysis to be performed, the values of the independent and dependent variables must be expressed as binary data. To this end, the threshold of the values of each variable was obtained using the TOSMANA program. Figure 4 shows the result of deriving the threshold for the ratio of infected persons in stage 1. The median of this variable is 0.075, but the threshold is 0.18.

Figure 4 Threshold for confirmed cases in 2011 (confirmed 2011)

The thresholds of the five variables included in the independent variable were also derived in the same way as above, and these are presented in the Appendix. A truth table was obtained using QCA analysis, and a combination of the five conditions that brought about a reduction in the local infection rate was derived. Figure 5 represents the truth table derived from QCA analysis.



Truth Table:

v1: Vaccinated1st_1 v2: elderly65over_1
v3: unmetmedical_1 v4: obesity_1
v5: stressperception_1

O: Confirmed2011_1 id: Province

v1	v2	v3	v4	v5	O	id
0	0	0	0	1	1	Seoul
1	1	0	0	0	0	Busan
0	1	0	0	0	1	Daegu
1	0	1	1	1	0	Incheon
1	0	0	0	1	0	Gwangju,Kyungki
1	0	0	0	0	0	Daejeon,Ulsan
0	0	0	0	0	0	Sejong
1	1	1	1	1	0	Kangwon
1	1	0	1	0	0	Chungbuk
1	1	0	1	1	0	Chungnam
1	1	1	0	0	0	Jeonbuk,Kyungnam
1	1	1	1	0	0	Jeonnam,Kyungbuk
1	0	1	1	0	0	Jeju

Result:

VACCINATED1ST_1 * ELDERLY65OVER_1 * stressperception_1 + VACCINATED1ST_1 * UNMETMEDICAL_1 * OBESITY_1 + VACCINATED1ST_1 * ELDERLY65OVER_1 * OBESITY_1 + VACCINATED1ST_1 * elderly65over_1 * unmetmedical_1 * obesity_1 + elderly65over_1 * unmetmedical_1 * obesity_1 * stressperception_1

(Busan+Chungbuk+Jeonbuk,Kyungnam+Jeonnam,Kyungbuk) (Incheon+Kangwon+Jeonnam,Kyungbuk+Jeju) (Kangwon+Chungbuk+Chungnam+Jeonnam,Kyungbuk) (Gwangju,Kyungki+Daejeon,Ulsan)(Daejeon,Ulsan+Sejong)

Figure 5 Stage 1 truth table

As Figure 5 shows, there are five conditional expressions that result in a low rate of Covid infection in the region. The first condition consists of high *VACCINATEDIST* (↑), high *elder65over* (↑) and low *stressperception* (↓), and this category includes Busan, Chungbuk, Jeonbuk, Kyungnam, Jeonnam and

Kyunbuk. If a specific region meets this condition, the number of infected people will decrease. The second condition consists of high *VACCINATEDIST* (↑), high *unmetmedical* (↑) and high *obesity*(↑), and this condition includes Incheon, Kangwon, Jeonnam, Kyungbuk and Jeju.

The third conditional expression consists of high *VACCINATEDIST* (↑), high *elderly 65over* (↑) and high *obesity* (↑). Kangwon, Chungbuk, Chungnam, Jeonnam and Kyungbuk are included in this conditional expression. The fourth conditional expression consists of high *VACCINATEDIST* (↑), low *elderly 65over* (↓), low *unmetmedical* (↓) and low *obesity* (↓), and this category includes Gwangju, Daejeon, Kyungbuk and Ulsan. The fifth condition consists of low *elderly 65over* (↓), low *unmetmedical* (↓), low *obesity* (↓) and low *stressperception* (↓). This category includes Daejeon, Ulsan and Sejong.

Figure 6 presents the results of the above truth table analysis in a Venn diagram. Here, the pink cell shows the conditions under which the dependent variable has a value of 0. There are ten conditional expressions in which the dependent variable has a value of 0, but the number can be reduced to five by summarising them.

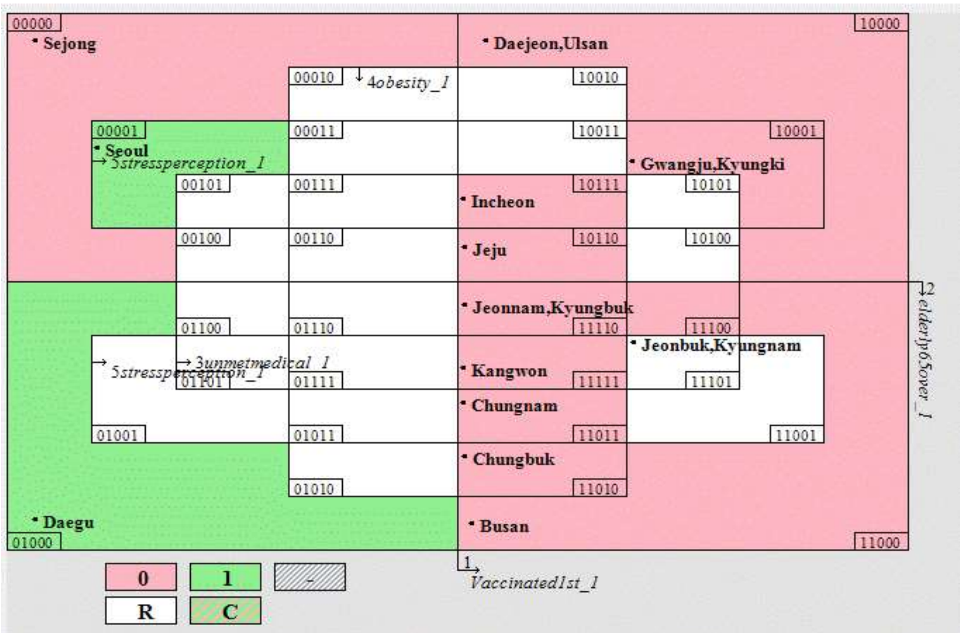


Figure 6 Venn diagram in stage 1

Result for stage 2

As at stage 1, neural network analysis was performed to derive the important variables that affected the number of infected persons relative to the population in stage 2. The results are shown in Figure 7.

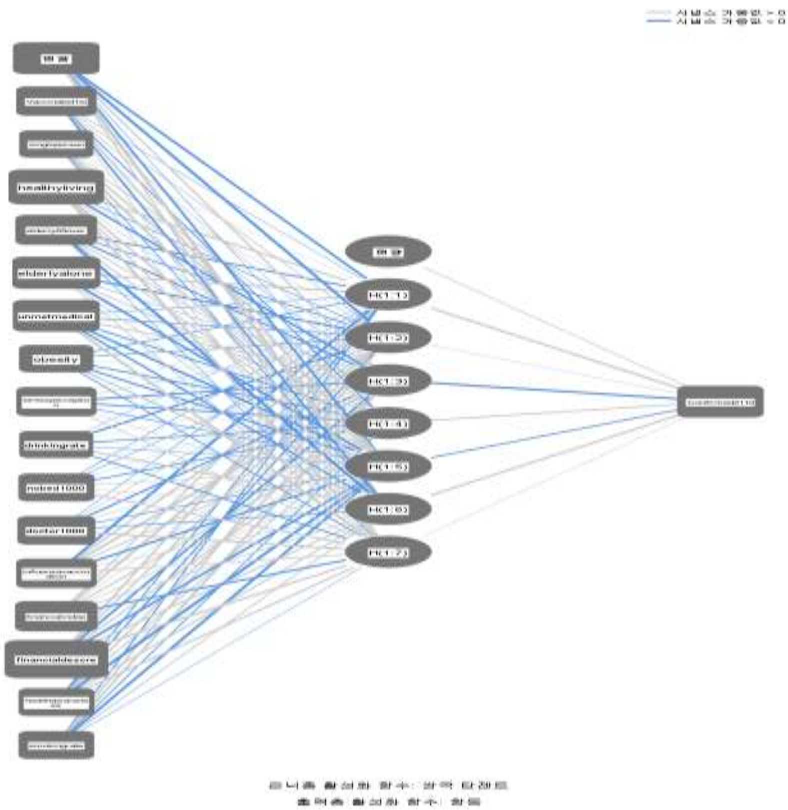


Figure 7 Results of neural network analysis in stage 2

Table 4 shows the results of the analysis in stage 2, and shows the importance of each independent variable affecting the dependent variable. The variable that had the greatest influence on the ratio of infected persons to the population in stage 2 is the ratio of freely available resources to the total resources of local authorities, unlike in Stage 1. In other words, local authorities with numerous financial resources with which to take various medical measures have an advantage in lowering the infection rate due to their high levels of financial independence. The second most important variable is *healthyliving*, followed by the *elderlyalone*, *unmetmedical* and *financialindep* variables. It may be said that the five independent variables presented above are those that had a strong influence on the rate of coronavirus infections during the second stage of spread.

Table 4 Importance of each variable in stage 2

Variable	*Importance*	*Normalised importance*
financialdescre	0.176	1
healthyliving	0.133	0.759
elderlyalone	0.094	0.537
unmetmedical	0.089	0.509
financialindep	0.066	0.376

elderly(65over)	0.063	0.359
influenzavaccination	0.055	0.315
Vaccinated(1st)	0.054	0.31
stressperception	0.054	0.305
doctor1000	0.043	0.245
nobed1000	0.034	0.195
healthawareness	0.032	0.185
smokingrate	0.032	0.182
obesity	0.025	0.141
singleperson	0.024	0.137
drinkingrate	0.023	0.132

Figure 8 shows the normalised importance values of the independent variables in stage 2. Assuming that the value of the most important variable is 1, this gives us a value for the relative importance of other variables.

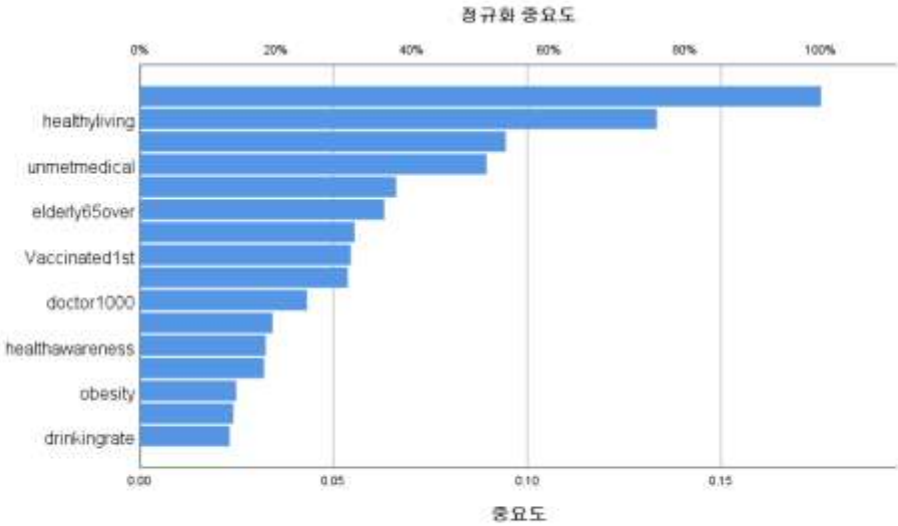


Figure 8 Importance of normalisation of variables in stage 2

Figure 9 shows the results of the truth table analysis in stage 2. There are seven combinations with a value of 0, and they can be reduced to the combination of all four conditional expressions.

Truth Table

v1: healthyliving_1 v2: elderlyalone_1
v3: unmetmedical_1 v4: financialindep_1
v5: financialdescre_1

O: Confirmed2110 id: Province

v1	v2	v3	v4	v5	O	id
1	0	0	1	1	1	Seoul
0	1	0	0	0	0	Busan,Chungbuk,Chungnam
0	0	0	0	0	0	Daegu,Gwangju,Daejeon,Ulsan
1	0	1	1	0	0	Incheon
0	0	0	1	0	0	Sejong
0	0	0	1	1	1	Kyungki
0	1	1	0	1	0	Kangwon
0	1	1	0	0	0	Jeonbuk,Jeonnam,Kyungbuk,Kyungnam
0	0	1	0	0	0	Jeju

Result

healthyliving_1 * financialindep_1 * financialdescre_1 + healthyliving_1 * elderlyalone_1 * unmetmedical_1 * financialdescre_1 + healthyliving_1 * ELDERLYALONE_1 * UNMETMEDICAL_1 * financialindep_1 + HEALTHYLIVING_1 * elderlyalone_1 * UNMETMEDICAL_1 * FINANCIALINDEP_1 * financialdescre_1
(Busan,Chungbuk,Chungnam+Daegu,Gwangju,Daejeon,Ulsan+Jeonbuk,Jeonnam,Kyungbuk,Kyungnam+Jeju) (Daegu,Gwangju,Daejeon,Ulsan+Sejong)
(Kangwon+Jeonbuk,Jeonnam,Kyungbuk,Kyungnam) (Incheon)

Figure 9 Truth table in stage 2

As Figure 9 shows, in the second stage, when one of the four conditions is created, the infection rate is lowered. The first condition is that the *healthyliving* variable is low (↓), *financialindep* is low (↓) and *financialdescre* is low. This conditional expression includes Busan, Chungbuk, Chungnam, Daegu, Gwangju, Daejeon, Ulsan, Jeonbuk, Kyungbuk, Kyungnam and Jeju. The second condition is that *healthyliving* is low (↓), *elderlyalone* is low (↓), *unmetmedical* is low (↓) and *financialdescre* is low (↓). This includes Daegu, Gwangju, Daejeon, Ulsan and Sejong. The third condition is a conditional expression of low *healthyliving* (↓), high *elderlyalone* (↑), high *unmetmedical* (↓) and low *financialindep* (↓). This includes Kangwon, Jeonbuk, Jeonnam, Kyungbuk and Kyungnam. The fourth condition is that *healthyliving* is high (↑), *elderlyalone* is low (↓), *unmetmedical* is high (↑), *financialindep* is high (↑) and *financialdescre* is low (↓). Incheon is included in this conditional expression. Figure 10 shows the conditional expression of the independent variables for which the dependent variable becomes 0 in stage 2.

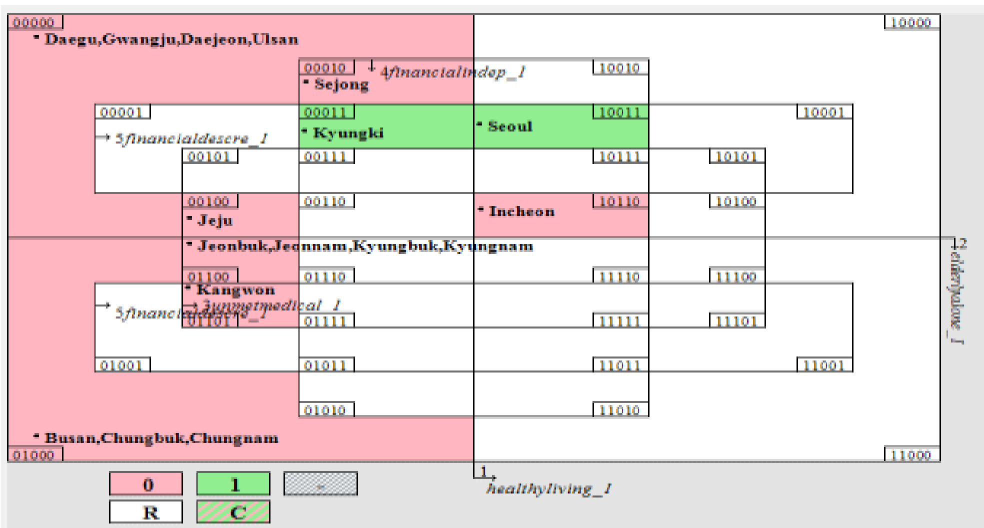


Figure 10 Venn diagram of conditional expression in stage 2

Result for stage 3

As in stages 1 and 2, neural network analysis was performed to derive the important variables that affected the number of infected persons relative to the population. The results are shown in Figure 11.

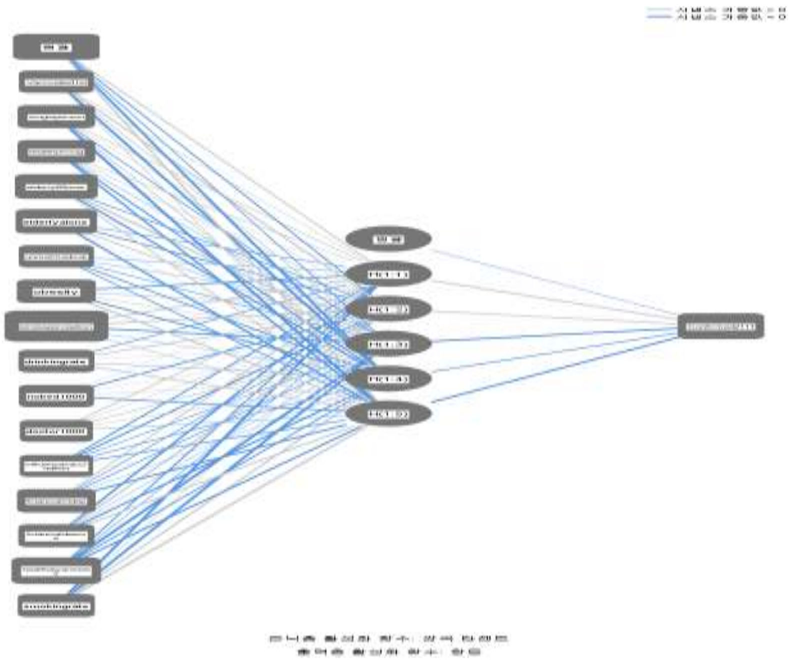


Figure 11 Neural network analysis result in stage 3

Table 5 shows the importance of each independent variable affecting the dependent variable in stage 3. The variable that had the greatest influence on the ratio of infected persons relative to the population in stage 3 is *stressperception*. Next, the *healthawareness*, *elderly65over*, *elderlyalone* and *obesity* variables appear to have an important effect on the dependent variable.

Table 5 Importance of each variable in stage 3

Variable	*Importance*	*Normalised importance*
stressperception	0.218	1
healthawareness	0.126	0.58
elderly(65over)	0.084	0.384
elderlyalone	0.076	0.35
obesity	0.059	0.273
financialindep	0.058	0.266
healthyliving	0.05	0.232
singleperson	0.049	0.227
smokingrate	0.047	0.215

financialdescre	0.042	0.195
drinkingrate	0.04	0.185
unmetmedical	0.035	0.161
nobed1000	0.034	0.157
Vaccinated(1st)	0.033	0.15
influenzavaccination	0.025	0.117
doctor1000	0.023	0.106

Figure 12 shows the normalised importance values of the independent variables in stage 3. Assuming that the value of the most important variable is 1, this gives us a value for the relative importance of other variables.

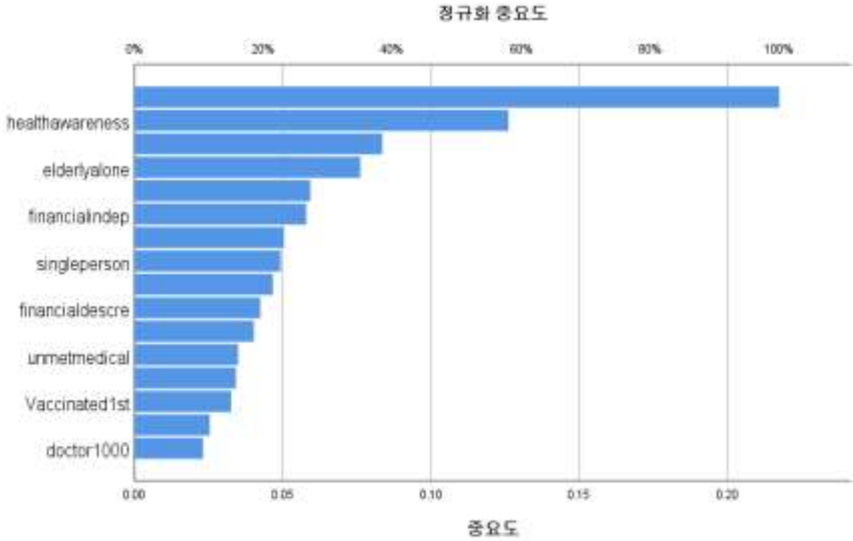


Figure 12 Importance of normalisation of variables in stage 3

Figure 13 presents the truth table for the three-stage analysis. There are four cases in which the dependent variable has a value of 0.

Truth Table:

v1: elderly65over_1 v2: elderlyalone_1
v3: obesity_1 v4: stressperception_1
v5: healthawareness_1

O: Confirmed2111 id: Province

v1	v2	v3	v4	v5	O	id
0	0	0	1	1	1	Seoul,Gwangju
1	1	0	0	0	0	Busan,Jeonbuk,Kyungnam
1	0	0	0	1	0	Daegu
0	0	1	1	0	1	Incheon
0	0	0	0	1	1	Daejeon
0	0	0	0	0	C	Ulsan,Sejong
0	0	0	1	0	1	Kyungki
1	1	1	1	0	0	Kangwon,Chungnam
1	1	1	0	0	C	Chungbuk,Kyungbuk
1	1	1	0	1	0	Jeonnam
0	0	1	0	1	1	Jeju

Result:

ELDERLY65OVER_1 * ELDERLYALONE_1 * obesity_1 * stressperception_1 * healthawareness_1 + ELDERLY65OVER_1 * elderlyalone_1 * obesity_1 * stressperception_1 * HEALTHAWARENESS_1 + ELDERLY65OVER_1 * ELDERLYALONE_1 * OBESITY_1 * STRESSPERCEPTION_1 * healthawareness_1 + ELDERLY65OVER_1 * ELDERLYALONE_1 * OBESITY_1 * stressperception_1 * HEALTHAWARENESS_1
(Busan,Jeonbuk,Kyungnam) (Daegu) (Kangwon,Chungnam) (Jeonnam)

Figure 13 The truth table in stage 3

A case in which a value of 0 is obtained from the dependent variable is one in which one of the four conditions is met. The first condition was that *elderly 65over* was high (↑), *elderlyalone* was high (↑), *obesity* was low (↓), *stressperception* was low (↓) and *heathawareness* was low (↓). This conditional expression includes Busan, Jeonbuk and Kyungnam. The second condition was high *elderly65over* (↑), low *elderlyalone* (↓), low *obesity* (↓), low *stressperception* (↓) and low *healthawareness* (↓). Daegu is included in this conditional expression. The third conditional expression is that *elderly 65over* is high (↑), *elderlyalone* is high (↑), *obesity* is high (↑), *stressperception* is high (↑) and *healthawareness* is low. This conditional expression includes Kangwon and Chungnam. The fourth conditional expression is that *elderly65over* is high (↑), *elderlyalone* is high (↑), *obesity* is low (↓), *stresspreception* is low (↓) and *healthawareness* is high (↑). This conditional expression includes Jeonnam. Figure 14 shows the above conditional expressions represented in a Venn diagram.

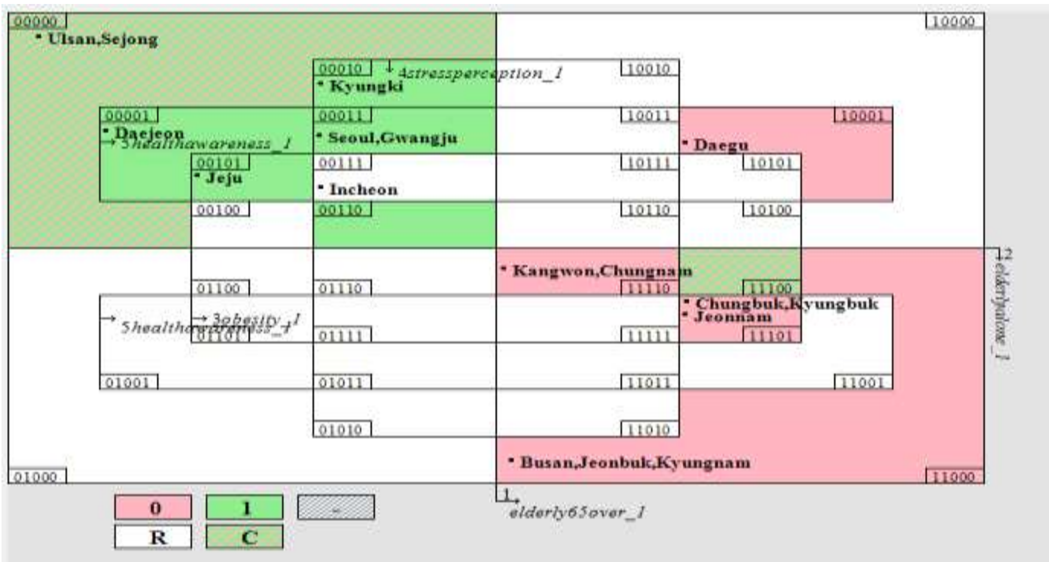


Figure 14 Venn diagram of conditional expression in stage 3

Conclusion

This study attempted to analyse why the numbers of persons in Korea infected with coronavirus showed regional variation. To this end, the period from January 2020 to December 2021 was divided into three and analysed by dividing each of these three stages, the reason being that the numbers of infected persons varied greatly at each stage. The period from January to November 2020, which corresponds to the first stage, is a period showing a relatively low number of infections. The period from December 2020 to October 2021 is one of spread, showing a different trend from the first stage. And in the period encompassing November and December 2021, the spread is characterised by an explosive increase. Of course, since each stage represents a relatively brief time-period, there is a limit to the analysis. Nevertheless, a number of factors set this study apart from other studies. First, it divides up the analysis period for coronavirus-infected patients in relation to the trend in the number of patients. Second, its predictive power was improved by using the neural network analysis method in the process of finding the variable that has the greatest influence on the dependent variable. And third, when the number of cases for analysis was not large (currently, seventeen cases), analysis was attempted by applying the methodology called Qualitative Comparative Analysis (QCA).

If the aim is use a quantitative method to analyse social phenomena, the number of cases for analysis must be large. In addition, many prerequisites for the analysis must be met. However, in real life, it often happens that not many cases present themselves for analysis, and at the same time some social phenomena can in many cases be explained by the combining of several variables. This study attempted to analyse regional differences in the number of coronavirus-infected persons under these realistic constraints.

The analysis results of this study may be summarised as follows. In the first stage, where the number of infected persons remained relatively low, five binding conditions were found. The first condition consists of a case where *VACCINATEDIST* is high (↑), *elder65over* is high (↑) and *stressperception* is low (↓). This conditional expression includes Busan, Chungbuk, Jeonbuk, Kyungnam, Jeonnam and Kyunbuk. If this condition is achieved locally, it means that the number of infected persons will be lower. The second condition consisted of high *VACCINATEDIST* (↑), high *unmetmedical* (↑) and high *obesity* (↑). This condition includes Incheon, Kangwon, Jeonnam, Kyungbuk and Jeju.

The third conditional expression consisted of high *VACCINATEDIST* (↑), high *elderly 65over* (↑) and high *obesity* (↑). Kangwon, Chungbuk, Chungnam, Jeonnam and Kyungbuk were included in this conditional expression. The fourth conditional expression consisted of high *VACCINATEDIST* (↑), low *elderly 65over* (↓), low *unmetmedical* (↓) and low *obesity* (↓). This conditional expression includes Gwangju, Daejeon, Kyungbuk and Ulsan. The fifth condition consisted of low *elderly65over* (↓), low *unmetmedical* (↓), low *obesity* (↓) and low *stressperception* (↓). This conditional expression includes Daejeon, Ulsan and Sejong.

In the second stage, when one of the four conditions is created, the local infection rate will be lowered. The first condition is that *healthyliving* is low (↓), *financialindep* is low (↓) and *financialdescre* is low. Busan, Chungbuk, Chungnam, Daegu, Gwangju, Daejeon, Ulsan, Jeonbuk, Kyungbuk, Kyungnam and Jeju are included in this case. The second condition is that *healthyliving* is low (↓), *elderlyalone* is low (↓), *unmetmedical* is low (↓) and *financialdescre* is low (↓). This includes Daegu, Gwangju, Daejeon, Ulsan and Sejong. The third condition is that *healthyliving* is low (↓), *elderlyalone* high (↑), *unmetmedical* is high (↓) and *financialindep* is low (↓). This includes Kangwon, Jeonbuk, Jeonnam, Kyungbuk and Kyungnam. The fourth condition is that *healthyliving* is high (↑), *elderlyalone* is low (↓), *unmetmedical* is high (↑), *financialindep* is high (↑) and *financialdescre* is low (↓). Incheon is included in this conditional expression.

In stage 3, the cases that result in a low infection rate are all the cases where one of the four conditions is satisfied. The first condition was that *elderly 65over* was high (↑), *elderly alone* was high (↑), *obesity* was low (↓), *stressperception* was low (↓) and *heathawareness* was low (↓). This conditional expression includes Busan, Jeonbuk and Kyungnam. The second condition was that *elderly 65over* was high (↑), *elderlyalone* was low (↓), *obesity* was low (↓), *stressperception* was low (↓) and *healthawareness* was low (↓). Daegu is included in this conditional expression. The third conditional expression was that *elderly65over* is high (↑), *elderlyalone* is high (↑), *obesity* is high (↑), *stressperception* is high (↑) and *healthawareness* is low. This conditional expression includes Kangwon and Chungnam.

The fourth conditional expression is that *elderly65over* is high (↑), *elderlyalone* is high (↑), *obesity* is low (↓), *stresspreception* is low (↓) and *healthawareness* is high (↑). This conditional expression includes Jeonnam. As may be seen in the various combination conditions mentioned above – when the number of infected persons is low, when the virus spreads, and when the spread becomes more severe and explodes – different conditions in the region combine to determine the number of infected persons. The results of this study may be said to have certain policy implications in terms of coping with pandemics that are expected to occur in the future. Measures such as vaccination must be effective in the early stages of a pandemic, and when a pandemic spreads over time, infrastructure such as the financial power of the relevant local authorities can affect reduction in the numbers of those infected.

Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5C2A03082775).

References

Bakar, N. A., & Rosbi, S. (2020). Effect of coronavirus disease (COVID-19) to tourism industry. *International Journal of Advanced Engineering Research and Science*, 7(4), 189-193.

Blundell, R., Griffith, R., Levell, P., O'Connell, M. (2020), Could COVID-19 infect the consumer prices index?, NBER Working Paper 27352.

Choi, Young-C. (2022a). Comparative Analysis of the Role of Medical Staff in Korea Following the Coronavirus Pandemic: Comparisons after One and Two Years, *International Journal of Health Sciencs*, 6(s1): 2190-2219.

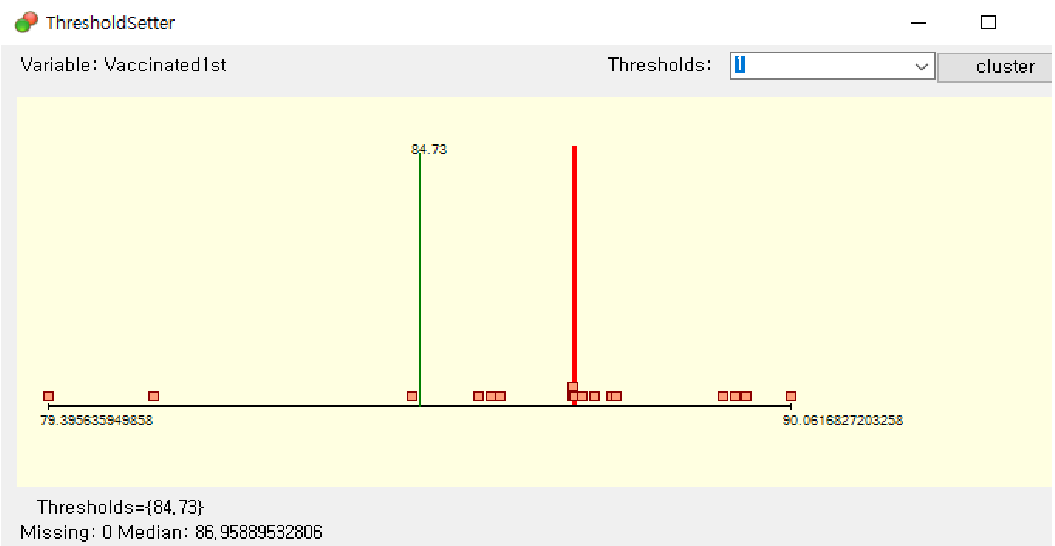
Choi, Young-C. (2022b). Identification of Social Issues Following the Corona Virus Pandemic in Korea:Topic Analysis and Ego Network Analysis of Public Health. *International Journal of Special Eduction*, 37(3): 12782-12798.
Cooper, B. (2005). Applying Ragin's crisp and fuzzy set QCA to large datasets: Social class and educational achievement in the national child development study. *Sociological Research Online* 10(2): 1–20.
Cronqvist, L. (2003). Presentation of TOSMANA. In: COMPASSS launching conference, Leuven, 16–17 September.
Cronqvist, L. (2009). Multi-value QCA (mvQCA). In: Rihoux, B. and Berg-Schlosser D. (eds) Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques. Thousand Oaks, CA: SAGE, pp. 69–86.
Diewert, E. (2020). Measuring real consumption and CPI bias under lockdown conditions, NBER Working Paper 27144.
DOI : 10.5392/JKCA.2020.20.07.342
Glaesser, J. and Cooper, B. (2012). Gender, parental education, and ability: Their interacting roles in predicting GCSE success. *Cambridge Journal of Education* 42(4): 463–480.
Goertz, G. (2020). *Social Science Concepts and Measurement*. Princeton, NJ: Princeton University Press.
Goldthorpe, J. H. (1998). The quantitative analysis of large-scale datasets and rational action theory: For a sociological alliance. In: Blossfeld H-P and Prein G (eds) *Rational Choice Theory and Large-scale Data Analysis*. Boulder, CA: Westview Press, pp.31–53.
http://ncov.mohw.go.kr
https://coronaboard.kr/
Johansson, J., Lof, M., Tysklind, O., (2020). *Changed consumption during the pandemic affects inflation*. Economic Commentaries, Sveriges Riksbank.
Killewald, A, Pfeffer F. T and Schachner J. N. (2017). Wealth inequality and accumulation. *Annual Review of Sociology* 43: 379–404.
Lee, J. M. (2020). An Exploratory Study on Effects of Loneliness and YouTube Addiction on College Life Adjustment in the Distance Education During COVID-19. *The Journal of the Korea Contents Association*, 20(7), 342-351.
Lieberson, S. (1985). *Making It Count. The Improvement of Social Research and Theory*. Berkeley, CA; Los Angeles, CA; London: University of California Press.
Lieberson, S. (1998). Causal analysis and comparative research: What can we learn from studies based on a small number of cases? In: Blossfeld H-P and Prein G (eds) *Rational Choice Theory and Large-Scale Data Analysis*. Boulder, CO: Westview
Poudel, K., & Subedi, P. (2020). Impact of COVID-19 pandemic on socioeconomic and mental health aspectsin Nepal. *International Journal of Social Psychiatry*. 66-8. 748-755.
Press, pp. 129–145.
Rebecca, Liao and Ziyang Fan. (2020). Supply chains have been upended. Here's how to make them more resilient. *WORLD ECONOMIC FORUM*, April 6.
Reinsdorf, M., (2020), COVID-19 and the CPI: Is inflation underestimated?, IMF Working Paper/20/224.
Serafini, G., Parmigiani, B., Amerio, A., Aguglia, A., Sher, L., & Amore, M. (2020). The psychological impact of COVID-19 on the mental health in the general population. QJM: *An International Journal of Medicine*, 113(8), 531-537.

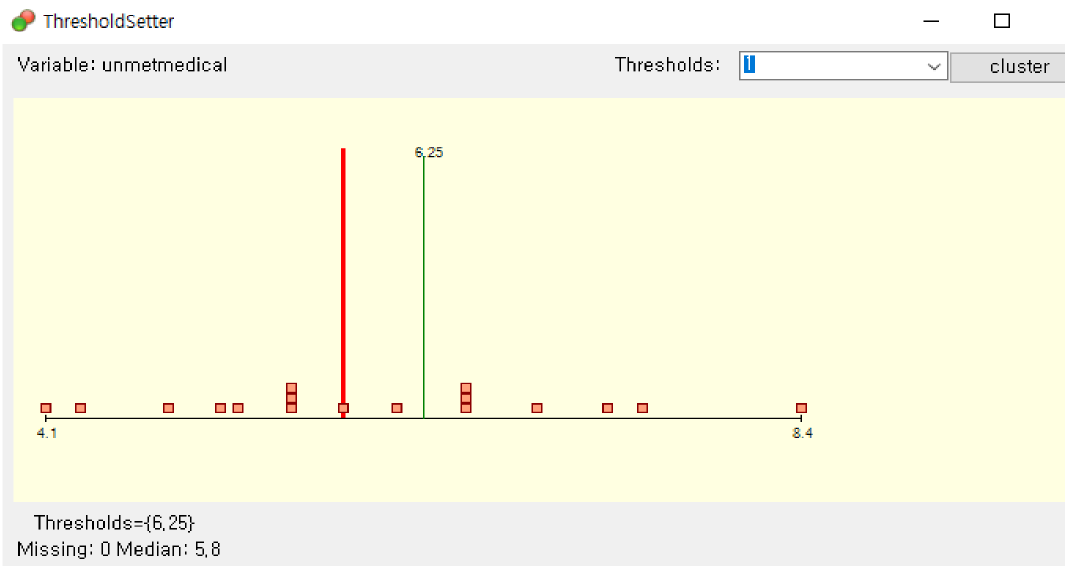
Torales, J., O'Higgins, M., Castaldelli-Maia, J. M., & Ventriglio, A. (2020). The outbreak of COVID-19 coronavirus and its impact on global mental health. *International Journal of Social Psychiatry*.

Wang, C, Pan R, Wan X, Tan Y, Xu L, Ho CS (2020). Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *International Journal of Environmental Research and Public Health* 17(5):1729-1735.

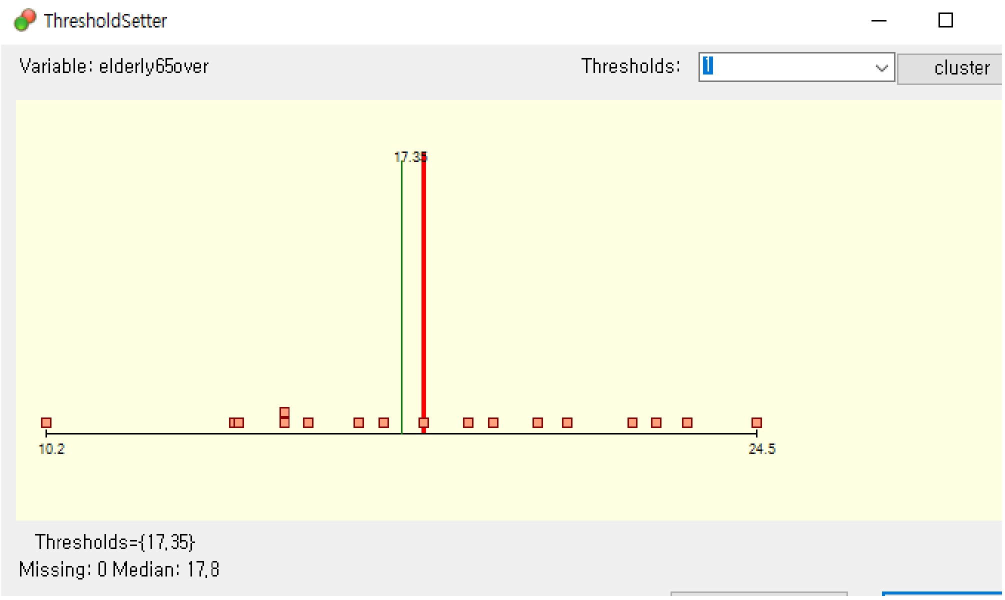
Wen, L, Li, X, Gao, L (2018) A new BP neural network-based data-driven fault diagnosis method. *IEEE Trans Industr Electron* 65(99):5990–5998

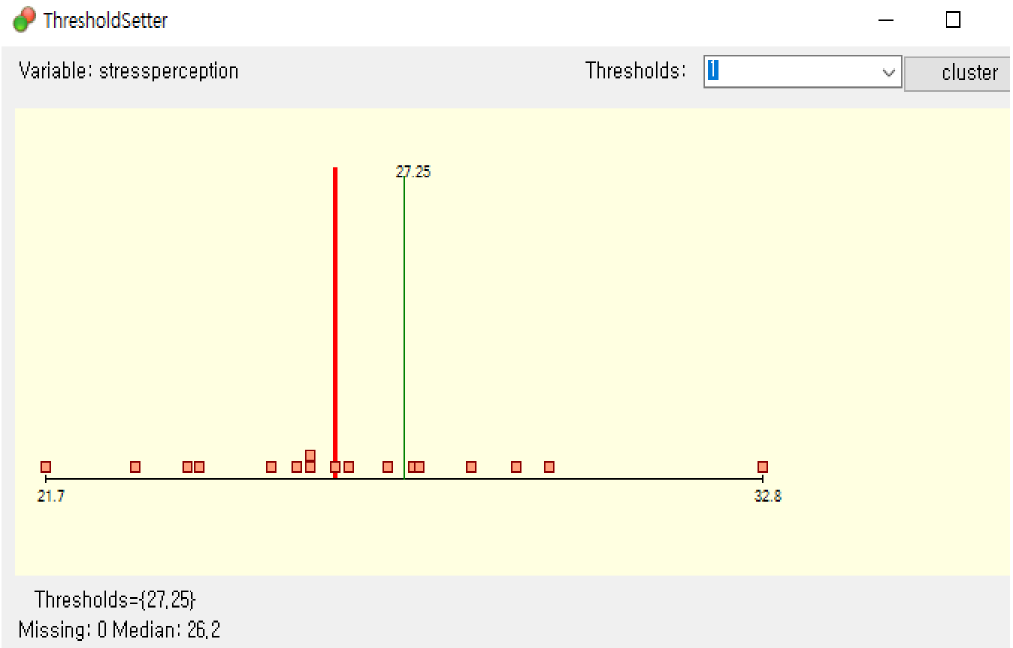
Appendix: Thresholds of Independent Variables (Derived using TOSMANA)

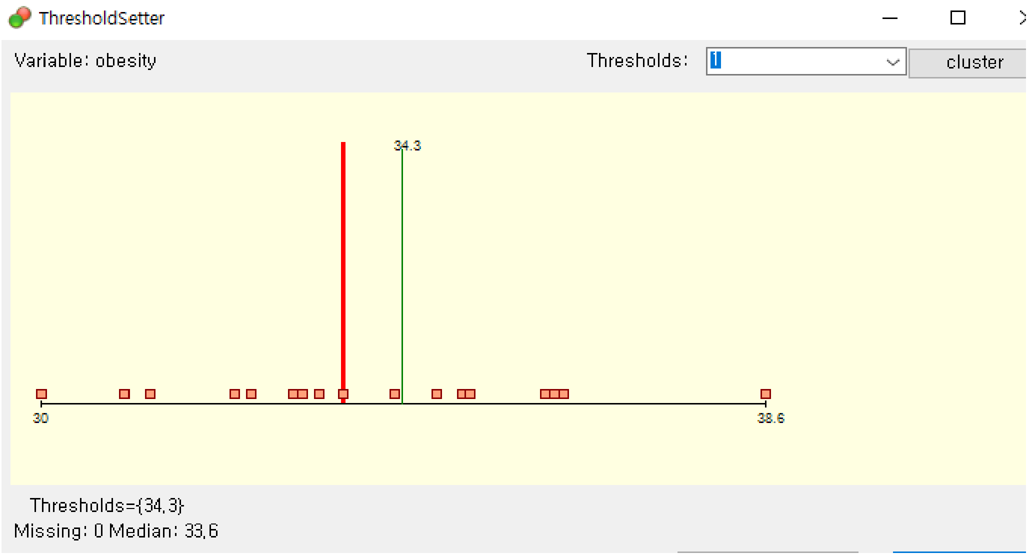




13332


ThresholdSetter
Variable: elderly65over
Thresholds:
cluster
17.35
10.2
24.5
Thresholds={17.35}
Missing: 0 Median: 17.8


ThresholdSetter
Variable: stressperception
Thresholds:
cluster
27.25
21.7
32.8
Thresholds={27.25}
Missing: 0 Median: 26.2


ThresholdSetter
Variable: obesity
Thresholds:
cluster
34.3
30
38.6
Thresholds={34,3}
Missing: 0 Median: 33,6