The Influence of Meteorological Factors on the Effect of SARS-CoV-2 Transmission and Its Delayed Effect

Zeyu Tong *

Department of Animal and Plant Sciences, University of Sheffield, Sheffield, The United Kingdom

* Corresponding author: Zeyu_Tong@stu.qhnu.edu.cn

Abstract. The influence of meteorological factors in the COVID-19 epidemic has been a popular topic in recent years. Since the emergence of Omicron, data on COVID-19 outbreaks unaffected by control measures have begun to accumulate. This paper uses data on the COVID-19 epidemic in Korea after the lifting of control measures on 18 April 2022. This paper uses infection rates corrected for antibody failure function as the response variable and correlates meteorological factors. A delay interval of 40 days was set to investigate the delayed effect of meteorological factors on infection rates. In the logistic regression analysis, temperature difference, barometric pressure difference, mean relative humidity, relative humidity difference and mean wind rating appeared significant. In the ANOVA, mean temperature, temperature difference, barometric pressure difference, mean relative humidity and mean wind rating were found to be significant. This study also shows that there are varying degrees of delay in the effects of the meteorological data.

Keywords: COVID-19, SARS-CoV-2, Omicron, Korea, meteorological factors.

1. Introduction

Since the emergence of COVID-19, the world has been watching the development of the largest outbreak in almost a century. In order to better control the spread of SARS-CoV-2, the impact of meteorological factors on the spread of SARS-CoV-2 needs to be fully investigated. During the three years that the world has been in an epidemic, there have been many attempts by researchers to exploit the value of data on the number of new coronavirus infections and meteorological data. These data can be used, through a number of statistical methods, to analyse the association between the transmission effects of SARS-CoV-2 and meteorological factors. However, the results of these studies have been mixed and even contradictory.

There is a wide range of studies on the effect of temperature and humidity on the spread of the new crown virus. An analysis of Wuhan COVID-19 epidemic data at the beginning of the new crown epidemic in 2020 by Ma et al. concluded that diurnal temperature difference was positively associated with the number of deaths and relative humidity was negatively associated with the number of deaths [1]. Nottmeyer et al. used data from 455 cities in 20 countries, a negative correlation was analysed between temperature and absolute humidity and the effect of new coronavirus transmission [2]. Examining data from Seoul from 1 January 2020 to 31 March 2021, Choi and Kim found that temperature and relative humidity were negatively correlated with the rate of new coronavirus infections [3]. In addition, a research team have conducted animal model tests using guinea pigs as hosts for the new coronavirus, showing that both cold and dry conditions are conducive to the spread of the new coronavirus [4]. However, other studies have produced contradictory results in terms of temperature factors. Using data from several cities around the world, Sajadi et al. found that the spread of the new coronavirus was hindered by low temperatures [5]. Liu et al. showed that heat waves have a negative impact on the immune system, which indirectly leads to an increased effect on the spread of the new coronavirus [6]. Comparing data from the Seoul Metropolitan Region (SMR) and the Daegu-Gyeongbuk Region (DGR) in Korea, Lim et al. found that temperature showed contradictory effects on the effect of new coronavirus transmission [7].

Relatively few studies have been conducted on the effects of atmospheric pressure and wind speed on the spread effectiveness of the new coronavirus. Using data from the Wuhan epidemic in 2020, Lin et al. analysed that both wind speed and atmospheric pressure were positively correlated with the prevalence of New Coronavirus infection [8]. Using data from the COVID-19 epidemic in Italy, Coccia analysed that low wind speeds lead to high air pollution and thus to increased efficiency in the transmission of the new coronavirus [9].

Although there have been many studies related to meteorological factors and the effects of the spread of the new coronavirus, few of the relevant studies published so far have dealt with countries or regions where the restrictions on the control of the new coronavirus epidemic have been completely lifted. This paper will use data from the COVID19 epidemic in Korea after the lifting of control measures on 18 April 2022 for this study, with the aim of filling a gap in this research. In addition, the delayed effects of the effects of meteorological factors have rarely been discussed in depth in relevant studies. The study by Choi and Kim et al. established a 2-day delay, but did not elaborate [3]. The study by Ma et al. established a delay interval of 1 to 5 days, but the interval was clearly too small [1]. Therefore, this paper also deals with an in-depth study of the delayed effect of meteorological factors on the effect of new coronavirus transmission.

2. Methods

2.1. Data Source

The COVID-19 epidemic data for Korea used in this paper were obtained from the official website of the World Health Organization (https://covid19.who.int/data). The Korean meteorological data used in this paper were obtained from the rp5.ru website (https://rp5.ru). The website states that the data was obtained from the Seoul weather station. Both tow group of data span the period from 1 April 2022 to 12 January 2023.

2.2. Data Preprocessing

The COVID-19 data for Korea obtained from the WHO website are daily data and contain variables such as the number of new infections per day. Weather data for Korea obtained from the rp5.ru website is hourly in precision and contain variables such as temperature, relative humidity, sea level pressure and wind level. In order to match the precision of the epidemic data, the weather data needs to be pre-processed to obtain daily data. In this paper, the mean and difference of each variable between 0:00 on the current day and 0:00 on the following day are calculated from the source data. The epidemic data and meteorological data were compiled into the same CSV file after unifying the precision.

This paper uses the rate of same-day infection as the response variable. A study published by the European CDC in September 2022 showed that antibody protection dropped to 51% three months after infection with Omicron and 25% six months later [10]. This paper sets up a simple linear function (proportion of those infected-on a given day in history whose antibodies failed on today = 0.25 + number of days from today * 5 / 18000) based on this data, which is used to calibrate the infection rate to some extent. The daily infection rate is the number of confirmed cases divided by the number of infectable cases. The number of infectable persons is the total population of Korea (approximately 5,174,000) minus the total number of historical antibody acquisitions plus the total number of historical antibody failures. The historical total number of acquired antibodies is the sum of the number of infections in the past 270 days, and the historical total number of antibody failure is the number of confirmed cases per day in the past 270 days multiplied by the proportion of those infected on that day whose antibodies failed on that day and then added up.

For in-depth analysis of delay effects, a delay buffer interval of size 40 days was set up in this paper. This is achieved by deferring the number of infections backwards by one day to obtain a delay of one day, and so on. Each day's delay is saved as a CSV file, for a total of 41.

2.3. Variable Description

T represents air temperature (degree Celsius) at 2 metre height above the earth's surface. *P* represents atmospheric reduced to mean sea level (millimeters of mercury). *U* represents relative humidity (%) at a height of 2 metres above the earth surface. *Ff* represents mean wind speed at a height of 10-12 metres above the earth's surface over the 10-minute period immediately preceding the observation (meters per second). Suffix *mean* represents the average value, and suffix *dv* represents the difference value. Delay represents the number of days new cases deferred (day).

2.4. Model Description

This paper uses R studio software for data analysis. This paper uses a generalised linear model (GLM) and the pre-processed data are analysed separately by logistic regression analysis and analysis of variance (ANOVA). As the model built using the standard binomial distribution was analysed to be excessively off-potential, it was adjusted to a quasibinomial distribution. ANOVA was performed on the logistic regression model. Due to the presence of a 40-day delay buffer interval, the above analyses were performed separately for each data set.

P-values from the ANOVA, P-values from the regression analysis and regression coefficient from the regression analysis were collected for each model for each delay day, and then separate CSV tables were created for each variable. Using the plotting function of ggplot2 and ggpubr, images were generated of the results of the various analyses for each meteorological variable as a function of the magnitude of the delay days. The images were combined to determine the first significant and most significant days of delay for each meteorological variable in the ANOVA and regression analysis, and to determine positive and negative correlations.

3. Results and Discussion

In this paper, one image for each of the eight meteorological variables is plotted in the same format, with each image containing two secondary images on the left and right. The left image shows the variation of the P-value of the meteorological variables in ANOVA analysis and logistic regression analysis with the number of days of delay. The image on the right shows the variation of the regression coefficient of the weather variable in logistic regression analysis with the number of days of delay. The black horizontal line in the left-hand image represents P = 0.05 and is used as an aid to determine whether the meteorological variable is significant at a given number of days of delay. The black horizontal line in the right-hand image represents the regression coefficient = 0, which aids in identifying the positive and negative correlation between the meteorological variable and the number of infections at a given number of days of delay. Note that the Y-axis in the left-hand image has been exponentiated to make it easier to distinguish.

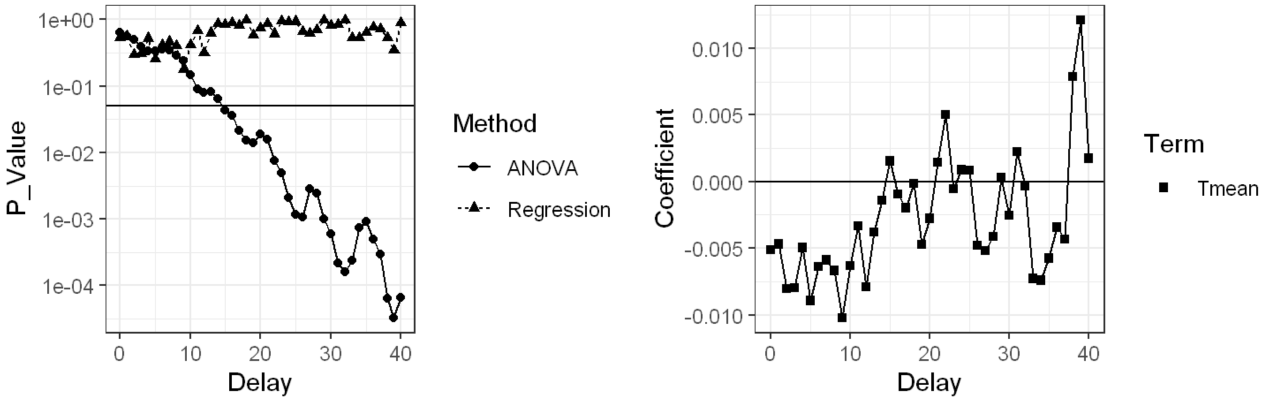

Figure 1. Mean temperature (Tmean)

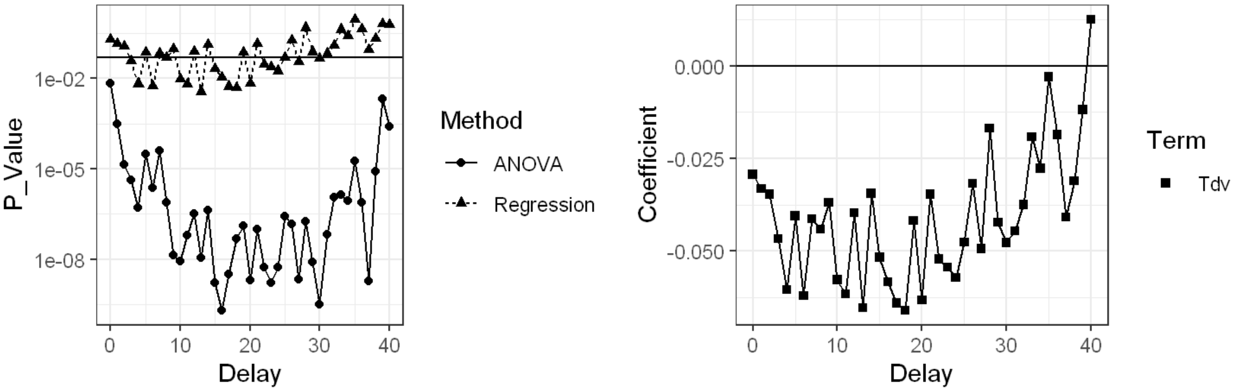

Figure 2. Temperature difference (Tdv)

The meteorological variable in Figure 1 is the mean temperature (Tmean) and in Figure 2 the meteorological variable is the temperature difference (Tdv). As can be seen from the P value figure in Figure 1, Tmean showed significance from day 15 in the ANOVA, but was not significant throughout the delay interval in the regression analysis. The regression coefficient figure in Figure 1 shows that the value of the regression coefficient for Tmean hovers around zero and is not informative. The P value figure in Figure 2 shows that Tdv is significant from day 4 in the regression analysis and is significant throughout the delay interval in the ANOVA. As can be seen from the regression coefficient figure in Figure 2, the regression coefficient for Tdv is negative for the first 39 days of the delay interval.

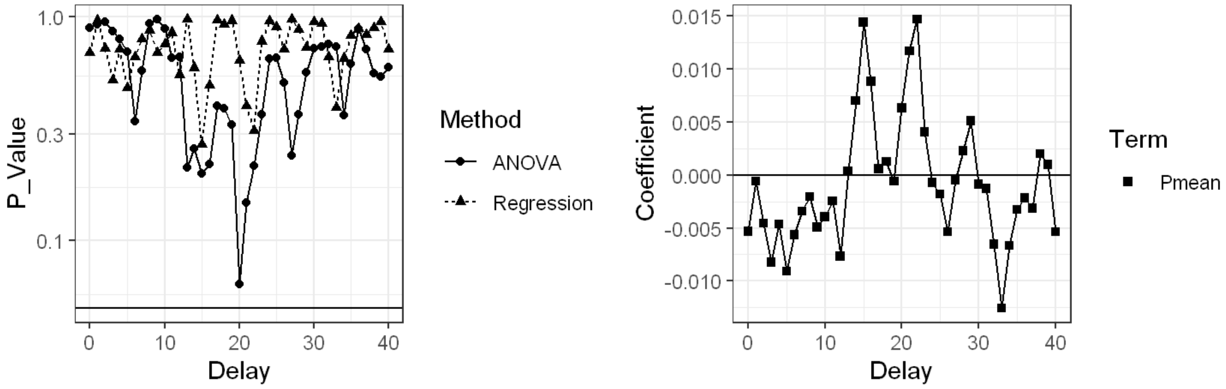

Figure 3. Mean barometric pressure (Pmean)

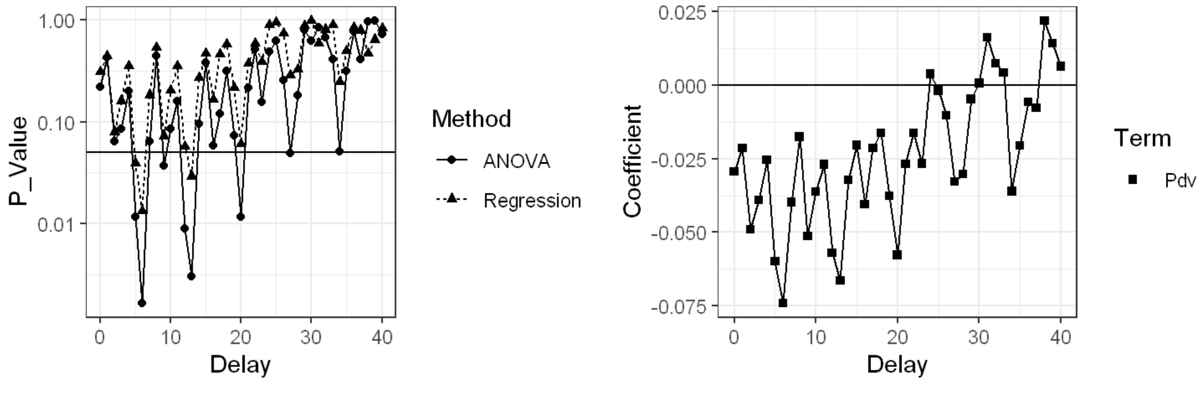

Figure 4. Barometric pressure difference (Pdv)

The meteorological variable in Figure 3 is the mean barometric pressure (Pmean) and the meteorological variable in Figure 4 is the difference in barometric pressure (Pdv). As can be seen from the P value figure in Figure 3, Pmean was not significant throughout the delay interval in both the regression analysis and the ANOVA. The regression coefficient figure in Figure 3 shows that the value of the regression coefficient for Pmean hovers around zero and is not informative. As can be seen from the P value figure in Figure 4, from day 5 onwards, Pdv appears to be significant in the regression analysis and ANOVA. As can be seen from the regression coefficient figure in Figure 4, the regression coefficient for Pdv is negative for the first 24 days of the delay interval.

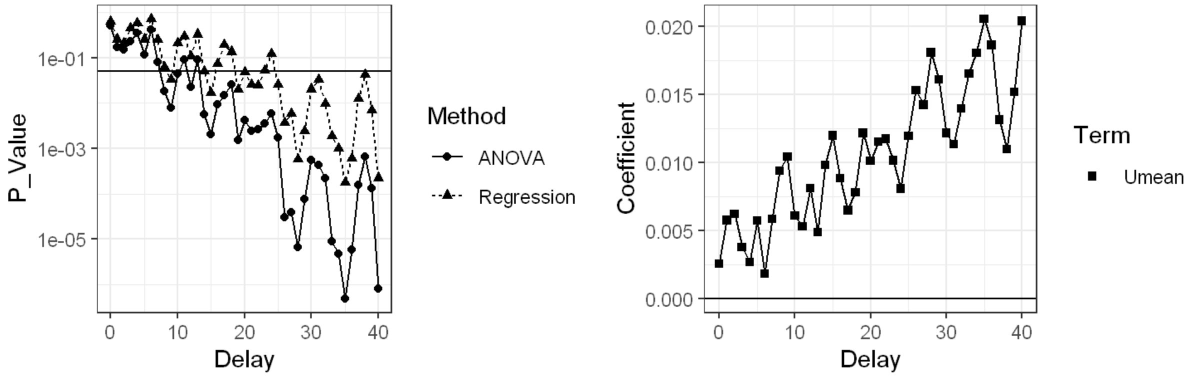

Figure 5. Mean relative humidity (Umean)

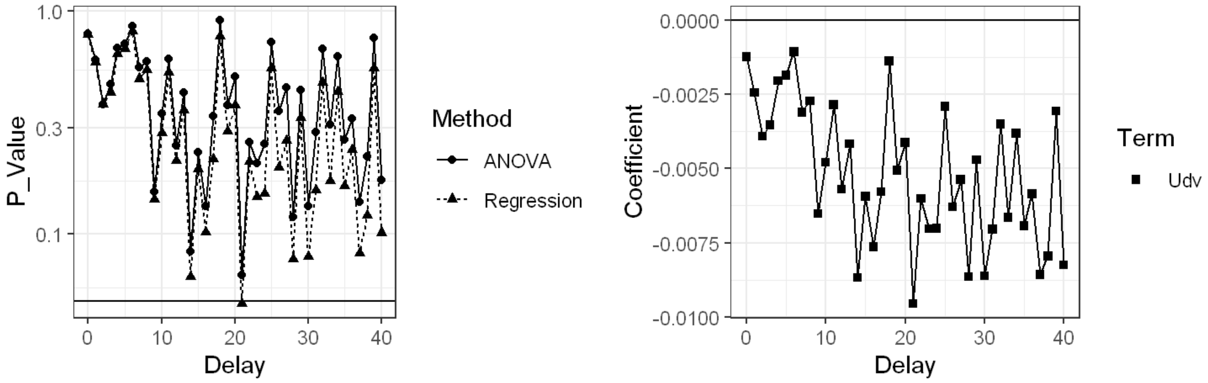

Figure 6. Relative humidity difference (Udv)

The meteorological variable in Figure 5 is the mean relative humidity (Umean) and the meteorological variable in Figure 6 is the difference in relative humidity (Udv). The P value figure in Figure 5 shows that Umean is significant from day 9 in the regression analysis and from day 8 in the ANOVA. The regression coefficient figure in Figure 5 shows that the regression coefficient for Umean is positive throughout the delay interval. As can be seen from the P value figure in Figure 6, Udv only appears significant in the regression analysis on day 21. However, as can be seen from the regression coefficient figure in Figure 6, Udv has a stable negative regression coefficient throughout the delay interval.

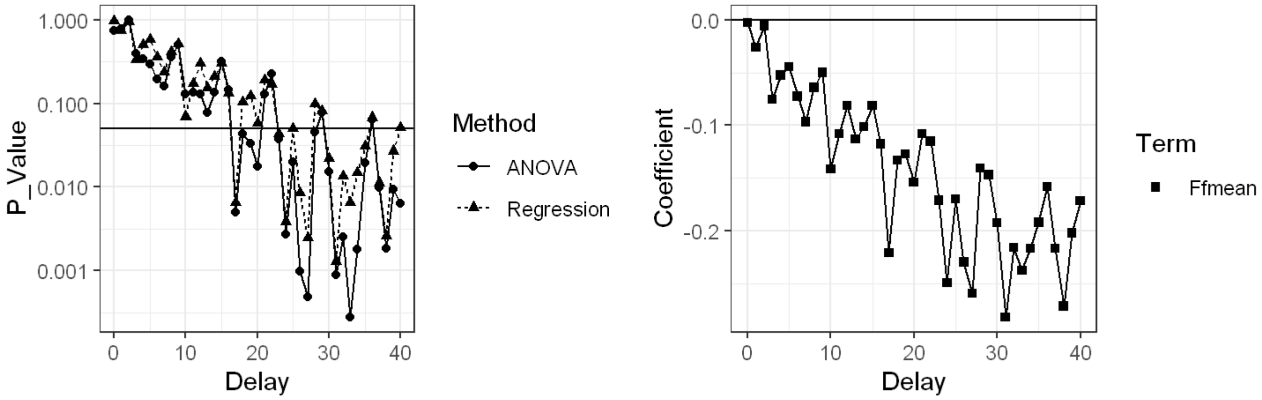

Figure 7. Mean wind rating (Ffmean)

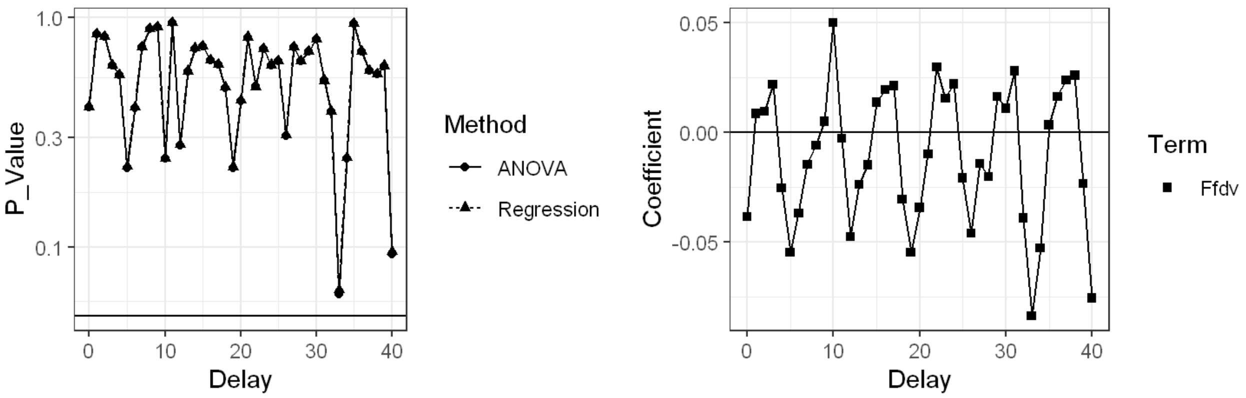

Figure 8. Wind rating difference (Ffdv)

The meteorological variable in Figure 7 is the mean wind rating (Ffmean) and the meteorological variable in Figure 8 is the wind rating difference (Ffdv). As can be seen from the P value figure in Figure 7, Ffmean was significant in the regression analysis and ANOVA from day 16 onwards. The regression coefficient figure in Figure 7 shows that Ffmean has a negative regression coefficient throughout the delay interval. The P value figure in Figure 8 shows that Ffdv was not significant in the regression analysis and ANOVA throughout the delay interval. The regression coefficient figure in Figure 8 shows that the value of the regression coefficient for Ffdv hovers around zero and is not informative.

Based on the data that have been calculated, together with the information given in the images above, the following tables can be made. Table 1 shows the number of days of delay between the first and highest significance of each meteorological variable in the ANOVA analysis, and the P value for each meteorological variable for that number of days of delay. Table 2 shows the number of days delayed when each meteorological variable first became significant in the logistic regression analysis, and when the highest significance occurred, as well as the P value and regression coefficient for each meteorological variable at that number of days delayed. In the following tables, 'First' represents the first occurrence of significance and 'Highest' represents the highest significance. The blank rows mean that the variables in those rows are not significant.

Table 1. ANOVA analysis

Term	First		Highest	
	Delay	P value	Delay	P value
Tmean	15	0.043*	39	<0.001***
Tdv	0	0.007**	16	<0.001***
Pmean	-	-	-	-
Pdv	5	0.012*	6	0.002**
Umean	8	0.018*	35	<0.001***
Udv	-	-	-	-
Ffmean	17	0.005**	33	<0.001***
Ffdv	-	-	-	-

Table 2. Logistic regression analysis

Term	First			Highest		
	Delay	P value	Coefficient	Delay	P value	Coefficient
Tmean	-	-	-	-	-	-
Tdv	4	0.007**	-0.060	13	0.004**	-0.065
Pmean	-	-	-	-	-	-
Pdv	5	0.039*	-0.060	6	0.013*	-0.074
Umean	9	0.032*	0.010	35	<0.001***	0.021
Udv	21	0.049*	-0.010	21	0.049*	-0.010
Ffmean	17	0.006**	-0.221	31	0.001**	-0.282
Ffdv	-	-	-	-	-	-

After listing the findings of this study, this paper also needs to discuss the limitations of this study. Korea is a relatively small country with a high population density, and the proportion of the population in the metropolitan area cities centred on Seoul accounts for nearly half of the whole of Korea. Therefore, the author considers the daily data for the city of Seoul to be representative. Since the author was unable to obtain historical daily meteorological data for the whole of Korea, data for the city of Seoul were used as a proxy. The time period between the lifting of COVID-19 control in Korea and the start of this study was only 8 months, which makes the amount of data somewhat inadequate compared to other studies. However, these 8 months cover the highest and lowest temperatures of the year. Due to the inherent inadequacy of the time span of the data, this paper could not set sufficient delay days to obtain complete significant intervals for all meteorological variables.

4. Conclusion

Logistic regression analysis and analysis of variance both suggest that the difference in temperature, the difference in barometric pressure, the average relative humidity and the average wind rating are related to the infection rate, while the difference in average barometric pressure and wind rating are not related to the infection rate. Analysis of variance suggests that the average temperature is related to the infection rate. Logistic regression analysis suggests that the difference in relative humidity is related to infection rate.

The following conclusions are supported by logistic regression analysis. The temperature difference is negatively correlated with the infection rate, with its influence starting on day 4 and reaching a maximum on day 13. Differential barometric pressure was negatively correlated with infection rate, with its influence starting on day 5 and reaching a maximum on day 6. Average relative humidity was positively correlated with the prevalence of infection, with its influence beginning on day 9 and reaching a maximum on day 35. The average wind level was negatively correlated with the infection rate, with its influence beginning on day 17 and reaching a maximum on day 31.

The following conclusions are supported by the analysis of variance. The temperature variance is related to the infection rate and its influence starts to be seen on the same day and reaches its maximum on day 16. Differential barometric pressure is related to infection rate, with its influence beginning on day 5 and reaching a maximum on day 6. Mean relative humidity is related to the prevalence of infection, with its influence beginning on day 8 and reaching a maximum on day 35. The average wind level is related to the infection rate and its influence starts to be seen on day 17 and reaches its maximum on day 33.

Analysis of variance suggests that mean temperature is related to infection rate, with its influence beginning on day 15 and reaching a maximum on day 39. However, logistic regression analysis showed that mean temperature was not related to infection rate and that the regression coefficients fluctuated around the zero value. Analysis of variance concluded that the relative humidity difference was not related to the infection rate. However, logistic regression analysis suggests that the relative humidity variance is related to the infection rate and that its influence starts on day 21, reaching a maximum on the same day, and has a stable negative regression coefficient.

The COVID-19 epidemic continues with meteorological factors and their delayed effects still playing a role. It can be expected that future relevant studies will be able to use more adequate data that are not affected by control measures. With sufficiently large delay intervals, the significant intervals for each meteorological factor can be obtained in full.

References

[1] Ma Y, Zhao H, Wei X. Changes of air pollutants and simulation of a heavy pollution process during COVID-19 in Shenyang. Environ Monit Assess, 2022, 194 (10): 723.

[2] Nottmeyer L, Armstrong B, Lowe R, et al. The association of COVID-19 incidence with temperature, humidity, and UV radiation – A global multi-city analysis. Science of The Total Environment, 2023, 854: 158636.

[3] Choi W, Kim KY. Correlation between COVID-19 Infection Rate and Atmospheric Environmental Factors in Seoul Metropolitan City, South Korea. In Review, 2022.

[4] Lowen AC, Mubareka S, Steel J, Palese P. Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. PLOS Pathogens, 2007, 3 (10): e151.

[5] Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature, Humidity, and Latitude Analysis to Estimate Potential Spread and Seasonality of Coronavirus Disease 2019 (COVID-19). JAMA Network Open, 2020, 3 (6): e2011834.

[6] Liu K, Yuan Y, LIU Z, et al. The impact of meteorological and environmental conditions on the spread of COVID-19. Journal of Public Health and Preventive Medicine, 2020, 9 - 13.

[7] Lim YK, Kweon OJ, Kim HR, Kim TH, Lee MK. The impact of environmental variables on the spread of COVID-19 in the Republic of Korea. Sci Rep, 2021, 11 (1): 5977.

[8] Lin Q, Ou G, Wang R, Li Y, Zhao Y, Dong Z. The Spatiotemporal Characteristics and Climatic Factors of COVID-19 in Wuhan, China. Sustainability, 2021, 13 (14): 8112.

[9] Coccia M. How do low wind speeds and high levels of air pollution support the spread of COVID-19? Atmospheric Pollution Research, 2021, 12 (1): 437 - 445.

[10] Ma Y, Zhao H, Wei X. Changes of air pollutants and simulation of a heavy pollution process during COVID-19 in Shenyang. Environ Monit Assess, 2022, 194 (10): 723.