



สถิติวิเคราะห์ของการแพร่ระบาด COVID-19 ในภูมิภาคเอเชียตะวันออกเฉียงใต้ในช่วงปี 2020

Statistical Analysis of COVID-19 Spreading in the Southeast Asia Region during 2020

อนุสา สุวรรณวงศ์^{*}, สิทธิเดช ชูด้วง¹, โรสนี จริยะมาการ¹

Anusa Suwanwong^{*}, Sittidet Chooduang¹, Rosnee Chariyamakarn¹

(Received: January 13, 2021; Revised: March 26, 2021; Accepted: April 1, 2021)

บทคัดย่อ

การแพร่ระบาด COVID-19 เป็นประเด็นใหม่และสำคัญของโลกและจำนวนผู้ติดเชื้อ COVID-19 ได้เพิ่มขึ้นอย่างรวดเร็วก่อให้เกิดผลกระทบที่สำคัญต่อระบบสาธารณสุขและวิกฤตทางเศรษฐกิจ ในภูมิภาคเอเชียตะวันออกเฉียงใต้ มีจำนวนของผู้ติดเชื้อ COVID-19 รายวันเป็นจำนวนมากในหลาย ๆ ประเทศ ซึ่งนำไปสู่ปัญหามากมาย ประชาชนส่วนใหญ่จึงตระหนักถึงผลกระทบของการแพร่ระบาดครั้งนี้และเปิดรับข้อมูลและการศึกษาใหม่ เพื่อที่จะป้องกันและรับมือสถานการณ์ เป้าหมายของการศึกษานี้ คือ การเลือกใช้สถิติวิเคราะห์ที่เหมาะสมเพื่อศึกษาเปรียบเทียบผลต่างค่าเฉลี่ยของจำนวนผู้ติดเชื้อ COVID-19 รายเดือนของประเทศในภูมิภาคเอเชียตะวันออกเฉียงใต้ (11 ประเทศ) และศึกษาความสัมพันธ์ระหว่างลักษณะที่ตั้งทางภูมิศาสตร์และจำนวนผู้ติดเชื้อ COVID-19 รายเดือนของประเทศในภูมิภาคเอเชียตะวันออกเฉียงใต้ในช่วงปี 2020 ข้อมูลประกอบด้วยจำนวนผู้ติดเชื้อ COVID-19 สะสม และจำนวนผู้ติดเชื้อ COVID-19 รายเดือน สามารถเข้าถึงผ่านทาง Our World in Data ซึ่งเป็นแหล่งข้อมูลที่เชื่อถือได้และข้อมูลจะถูกวิเคราะห์ด้วยโปรแกรม R ผลการศึกษา พบว่าประเทศอินโดนีเซียได้รับผลกระทบจากการแพร่ระบาด COVID-19 มากที่สุดในภูมิภาคเอเชียตะวันออกเฉียงใต้ด้วยจำนวนผู้ติดเชื้อ COVID-19 สะสมประมาณ 600,000 คนในช่วงปี 2020 ผลจากการทดสอบด้วยการทดสอบไครสคาล-วัลลิส พบว่าผลต่างค่าเฉลี่ยของจำนวนผู้ติดเชื้อ COVID-19 รายเดือนของทั้ง 11 ประเทศในช่วงปี 2020 แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) และจากการทดสอบด้วยการทดสอบไควสแควร์ พบว่าลักษณะที่ตั้งทางภูมิศาสตร์และจำนวนผู้ติดเชื้อ COVID-19 รายเดือนในภูมิภาคเอเชียตะวันออกเฉียงใต้มีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติ ($p < 0.05$)

คำสำคัญ: COVID-19 เอเชียตะวันออกเฉียงใต้ การทดสอบไครสคาล-วัลลิส การทดสอบไควสแควร์

¹คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏวราชนครินทร์

¹Faculty of Science and Technology, Princess of Naradhiwas University,

^{*}Corresponding Author: anusa.s@pnu.ac.th



Abstract

COVID-19 spreading has become a novel and serious issue globally, and the number of COVID-19 infection cases has been rapidly risen, causing a crucial impact on the healthcare system and economic crisis. In Southeast Asia, there are many of daily COVID-19 infection cases in many countries leading to several problems. A large number of people are aware of the pandemics and exposed to new information and studies to prevent and tackle the situation. The aim of this study was to adopt an appropriate statistical analysis to compare the mean differences of the monthly number of COVID-19 confirmed cases in the Southeast Asia countries and the association between the geography and the monthly number of COVID-19 infections cases in the Southeast Asia region during 2020. The data, the total cumulative number of COVID-19 confirmed cases and the monthly number of COVID-19 infection cases, can be accessed by the free resource of Our World in Data, which is a reliable data source. These data are analyzed by R program. The results showed that Indonesia was affected dramatically by COVID-19 spreading with approximately 600,000 cumulative infection cases as a result of the descriptive analysis method. The Kruskal Wallis test and Dunn's test revealed there were significant mean differences in the monthly number of COVID-19 infection cases in the Southeast Asia countries during 2020 ($p < 0.05$). In addition, the Chi-square test has shown that the geography was associated with the monthly number of COVID-19 infection cases ($p < 0.05$).

Keywords: COVID-19, Southeast Asia, Kruskal wallis test, Chi-square test

Introduction

Coronavirus disease or COVID-19 is a new virus causing an infectious disease across the world. Many people who infected with the COVID-19 virus will encounter respiratory illness and recover without any special treatment. The elderly and those with underlying disease, namely hypertension, diabetes, cardiovascular disease, respiratory disease, and kidney disease tend to have a higher mortality rate (Shahid et al., 2020). The crucial means to prevent and control transmission is clarified about the COVID-19 to cause and spreads. In general, most the people clean their hands with an alcohol based hand sanitizer to protect themselves from infection. The effective preventive approach of COVID-19 transmission in the workplace or organizations is detecting the temperature and flu-like symptoms of the workers (Cirrincione et al., 2020). The COVID-19 virus had erupted in



the Huanan seafood market, which is located in Wuhan, Hubei, China. The first 27 patients from 41 patients had the direct revelation to the Huanan seafood market, and only 9 patients were admitting in ICU care. The first infection case had a fever for a week with cough and dyspnea as the symptom. However, there was no epidemiological link between the first patient and later cases (Huang et al., 2019). The World Health Organization (WHO) has declared this disease as "COVID-19" on 11 February 2020 due to a rapid spreading causing many infections and deaths throughout the world. There were 42,708 confirmed cases and 1,000 deaths in China, while 393 infection cases and one death were found outside China (World Health Organization, 2020). The virus is well-known as severe immediate respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel discovered virus relatively associates with bat coronaviruses (Perlman, 2020), Pangolin coronaviruses (Cyranoski, 2020) and SARS-CoV. (Nkwoemeka, Okwelogu, & Chiedozi, 2020). The scientist has conceded COVID-19 as a natural origin (Andersen, Rambaut, Lipkin, Holmes, & Garry, 2020). Some people involved with bat carrion and guano in the manufacture of traditional Chinese medicines cause bat-to-human infection. Even through the earliest known person who had COVID-19 infection was discovered to ill with symptoms on the first date of December 2019, there was no interaction with the later wet market group. However, an earlier infection case could have arisen on November 17, and 66% of infection was found to have a link with the market (Joseph, 2020; Chan et al., 2020; Duarte, 2020).

In previous studies, the Chi-square test was frequently used to study an association related to COVID-19 spreading. The recovered cases were associated with different symptoms based on the patient history with a large information of 13,174 patients from different countries (Al-Najjar, Al-Najjar, & Al-Rousan, 2020; Czeisler et al., 2020) suggested that age is significantly associated with safety feeling without community alleviation strategies in the United States, Los Angeles, and New York City on 5–12 May 2020. In India, the Chi-square test represented that the gender was independent of current status, while age group was associated with upon current status (Sharma & Sharma, 2020). The variables of sex and the number of recovered cases were used to indicate associated study. The geographical variable was used to examine the proportion or the independence with the COVID-19 dataset through the Chi-square test. The number of the COVID-19 cases was found more in urban areas than in Maharashtra (Mohanty, Doke, & Patil, 2020). Moreover, the



COVID-19 data with Chi-square test was used in cross-sectional study (Karaaslan, Dikilita, & Aydin, 2020), in Prevention Behaviors among Health Sciences Students (Singkun, Patiwikriwong, Chainapong, & Weerakhachon, 2020), and in environmental studies (Livadiotis, 2020). In addition, analysis of variance or ANOVA is a powerful statistical technique to examine mean differences in 3 or more groups of data, and some studies have been used this technique to check the means differences of the recovery rate, perception and anxiety, and attitudes between COVID-19 infection groups (Nayak, Arora, Sinha, & Poonia, 2020; Shiina et al., 2020; Mohammed et al., 2020). The Kruskal Wallis test, a non-parametric method, was used to compare the change of exporting values between these agricultural exports COVID-19 pandemic (Lin & Zhang, 2020). Moreover, comparing means between two groups was adopted in many COVID-19 studies (Piroth et al., 2020; Chew et al., 2020; Wang et al., 2020), and these results were advantageous to further studies.

Objectives

The objectives of this study was to adopt appropriate statistical analysis, namely Shapiro-Wilk Test, Bartlett's Test, Kruskal Wallis test, Dunn's test, and Chi-square test to compare the mean differences in the monthly number of COVID-19 infection cases and association between the geography and the monthly number of COVID-19 infection cases from January to December 2020 in the Southeast Asia countries with 3074 samples.

Research methodology

1. Study Area

The study area was in Southeast Asia countries such as Indonesia, Philippines, Myanmar, Malaysia, Singapore, Thailand, Vietnam, Cambodia, Brunei, Timor-Leste, and Laos as in Figure 1. The region was divided into 2 vital zones, which were mainland and island. The mainland zone consists of five countries such as Myanmar, Thailand, Cambodia, Vietnam, and Laos. While the island zone comprised of Indonesia, Philippines, Malaysia, Singapore, Brunei, and Timor-Leste (Asia Society, 2020). Southeast Asia ranked as number 3 in Asia among subregions in terms of population. The total land area is 4,340,700 Km², and 50 % of the population is living in the urban area.



Figure 1 The map of Southeast Asia countries (Southeast Asia Map, 2020)

2. Dataset and Sample Size

In these studies, the data were provided by Our World in Data or <https://ourworldindata.org/coronavirus-source-data> with several variables such as the total cumulative number of COVID-19 confirmed cases, and the monthly number of COVID-19 infection cases. The dataset of COVID-19 was selected from January to December 2020, in total of 336 days. The values as NA mean there was no occurred case in the day, and the most of NA values were found at the beginning of COVID-19 spreading among Southeast Asia countries. The first country of COVID-19 spreading in the ASEAN region is Thailand. The population of 1522764 COVID-19 infection cases was between January and December 2020, which is indicated as the total cumulative number of COVID-19 confirmed cases. The 3074 samples were determined from January to December 2020, which are defined as the monthly number of COVID-19 infections.

3. Statistical Analysis

In this study, R program, R version 4.0.0, were used to analyze the COVID-19 data by using appreciative statistical analysis methods for efficiency objectives. Firstly, the monthly data is examined with the Shapiro-Wilk test and Bartlett's test regarding to the assumption of the analysis of variance. The Kruskal Wallis test and the Dunn's test then are used to investigate the mean differences of the monthly number of COVID-19 cases among 11 countries. Consequently, the Chi-



square test is applied for the independence test between the geography and the number of COVID-19 infection.

3.1 Shapiro-Wilk Test

Shapiro-Wilk test or W test is one of the well-known alternatives of normality test. The test handles sample sizes as a large observations and represents an accurate value to adopt future methods. In this study, the data of monthly COVID-19 infection cases were examined normality distribution by 2 techniques: graphical and statistical techniques. The graphical procedure is presented by Normal Q-Q plot which plotted two sets of quintiles and represent values between two different numeric variables. The Shapiro-Wilk Test provides by a probability value of data. These approaches were aimed to check the assumptions of mean differences between groups. The W test is given by

$$W = \frac{(\sum_{i=1}^n a_i X_{(i)})^2}{\sum_{i=1}^n (X_i - \bar{x})^2} \quad (1)$$

where $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ are the ordered values of a sample $X_1, X_2, \dots, X_n$, and a_i are tabulated coefficients. A lower tail of W indicates nonnormality (Zofia, 2014).

3.2 Bartlett's Test

Bartlett's test is used to investigate the homogeneity of variances in different populations, and the test is identical to Chi-square when the k factor samples are from independent normal populations. In this study, the dataset of the monthly number of COVID-19 infection cases were examined by the Bartlett's test in order to verify the assumption of the equality of variance in different populations. The Bartlett's test has the following expression as in equation 2.

$$B = \frac{(N-k) \ln \left(\frac{\sum_{i=1}^k (n_i - 1) s_i^2}{N-k} \right) - \sum_{i=1}^k (n_i - 1) \ln(s_i^2)}{1 + \frac{1}{3(k-1)} \left[\left(\sum_{i=1}^k \frac{1}{n_i - 1} \right) - \frac{1}{N-k} \right]} \quad (2)$$

where N is the total sample size, k is the number of groups, n_i is the sample size of the i^{th} group, and S_i^2 is the variance of the i^{th} group (Arsham & Lovric, 2011).

3.3 Kruskal Wallis Test

Kruskal Wallis test is one of nonparametric methods for the comparison of more than two independent samples. The method is used when the assumptions of ANOVA, normality distribution and the homogeneity of variances, are not met (Andras & DelaneyHarold, 1998). In this study, the Kruskal Wallis test was adopted to analyze with the monthly number of COVID-19 infection cases to



compare the medians in 11 countries based on Southeast Asia region. The Kruskal Wallis test is applied to the following equation 3.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^c \frac{R_i^2}{n_i} - 3(N+1) \quad (3)$$

where N indicates sum of sample sizes for all samples, c implies number of groups, R_i is sum of ranks in the i^{th} sample, and n_i defines as a size of the i^{th} sample (Kruskal, 1952).

3.4 Dunn's Test

Dunn's test is the effective nonparametric pairwise multiple-comparison method that is used when a null hypothesis of Kruskal Wallis test is rejected. In this study, Dunn's test was applied with the monthly number of COVID-19 infection cases to perceive mean differences of multiple pairwise comparisons in Southeast Asia countries. Dunn's test is adopted to assess certain rank-sum test by the mean ranking from Kruskal Wallis test $\bar{W}_i$ is W_i/n_i where W_i implies as the sum of ranks, and n_i indicates the sample size for the i^{th} group. To compare group A with group B, it is estimated as equation 4.

$$Z_i = \frac{y_i}{\sigma_i} \quad (4)$$

where i implies one of 1 to n multiple comparisons, y_i is $\bar{W}_A - \bar{W}_B$, and σ_i defines as the standard deviation of y_i (Dinno, 2015).

3.5 Chi-squared Test

Chi-squared test of Independence is a nonparametric statistical method which is often used to determine a significant association between two categorical variables. In this study, the Chi-square test of independence has been investigated an association between the monthly number of COVID-19 infection cases and the geography of the Southeast Asia region. The formula of the Chi-square is calculated as equation 5.

$$X^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (5)$$

where the observed frequency indicates as O_{ij} , the expected frequency is E_{ij} , and the Chi-square expected value are calculated as equation 6.

$$E = \frac{M_r * M_c}{n} \quad (6)$$



where E is the cell expected value, M_r implies the row marginal for that cell, M_c indicates the column marginal for that cell, and n is the total sample size (Zibran, 2007).

4. R Programming Flowchart

The flowchart has demonstrated the process of this study by using R program software. The total cumulative number of COVID-19 confirmed cases and the monthly number of COVID-19 infection cases have been used as input data. Figure 2 shows that the procedure of statistical analysis divided into two crucial sections regarding the objectives of this study. Firstly, the population of 1522764 were applied with the descriptive analysis to describe the general information of COVID-19 pandemics in Southeast Asia countries from January to December 2020 with the total cumulative number of COVID-19 confirmed cases. Secondly, 3074 samples, the monthly number of COVID-19 infection cases, was implied with the Shapiro-Walk test, the Barttle's test, the Kruskal Wallis test, and the Dunn's test to perceive the mean differences in the monthly number of COVID-19 infection cases in Southeast Asia countries during 2020. In addition, the test of independence by using Chi-square test was adopted to investigate the association between the geography and the monthly number of COVID-19 infection cases during 2020.

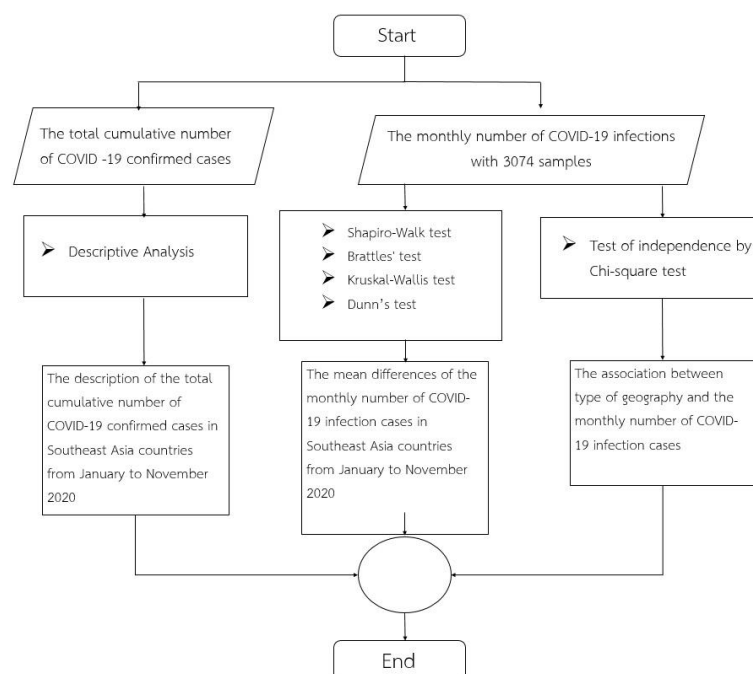


Figure 2 Flowchart for the R program



Results

According to descriptive analysis, the data was plotted from day 1 to 366 to illustrate the total cumulative number of COVID-19 confirmed cases by 11 countries in the Southeast Asia. In Figure 3, the COVID-19 spreading in the Indonesia, the Philippines, and Myanmar rang the highest level as the 3 most-affected countries by COVID-19 among Southeast Asia countries, while there are low levels of COVID-19 spreading in Thailand, Vietnam, Cambodia, Brunie, Timor-Leste, and Laos. The Indonesia's COVID-19 infection cases has been increasing steadily before surpassing the Philippines' total COVID-19 infection cases around day 298 (October), meanwhile there are similarities of the total cumulative number of COVID-19 confirmed cases from day 277 (October) to 366 (December) in the Myanmar and the Malaysia.

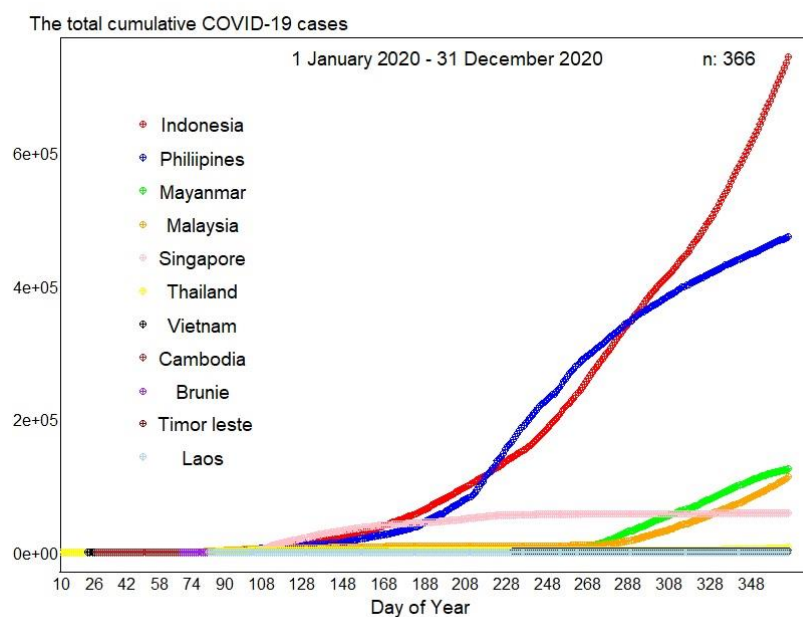


Figure 3 The total cumulative number of COVID-19 confirmed cases in the Southeast Asia during 2020

In figure 4, the highest cumulative number of COVID-19 infection cases are approximately 600,000 cases in Indonesia. The trend is likely to increase gradually to the beginning of 2021, making Indonesia the most affected COVID-19 spreading country in the Southeast Asia region. The Philippines' cumulative cases and the Myanmar's cumulative cases rank as the second level, and the third level in Southeast Asia countries with just under 500,000 cases and 150,000 cases, respectively.



However, there is the smallest cumulative number of COVID-19 confirmed cases in Laos with only 41 cases from January to December, and the trend maintains stable from the beginning of April. In addition, the total cumulative number of COVID-19 level remains the same in Brunei with approximately 200 cases. The monthly number of COVID-19 infection cases are plotted as line graphs categorized into 2 zones which are mainland and island. The number of COVID-19 infection cases surge from approximately 50,000 cases to 200,000 cases from June to August and rocket to reach a peak of just under 300,000 cases at the end of the year in the island zone. Whereas there is a slight level of monthly COVID-19 infection in the mainland before climbing to approximately 50,000 cases in October and remaining just under 50,000 cases toward the end of December as Figure 5.

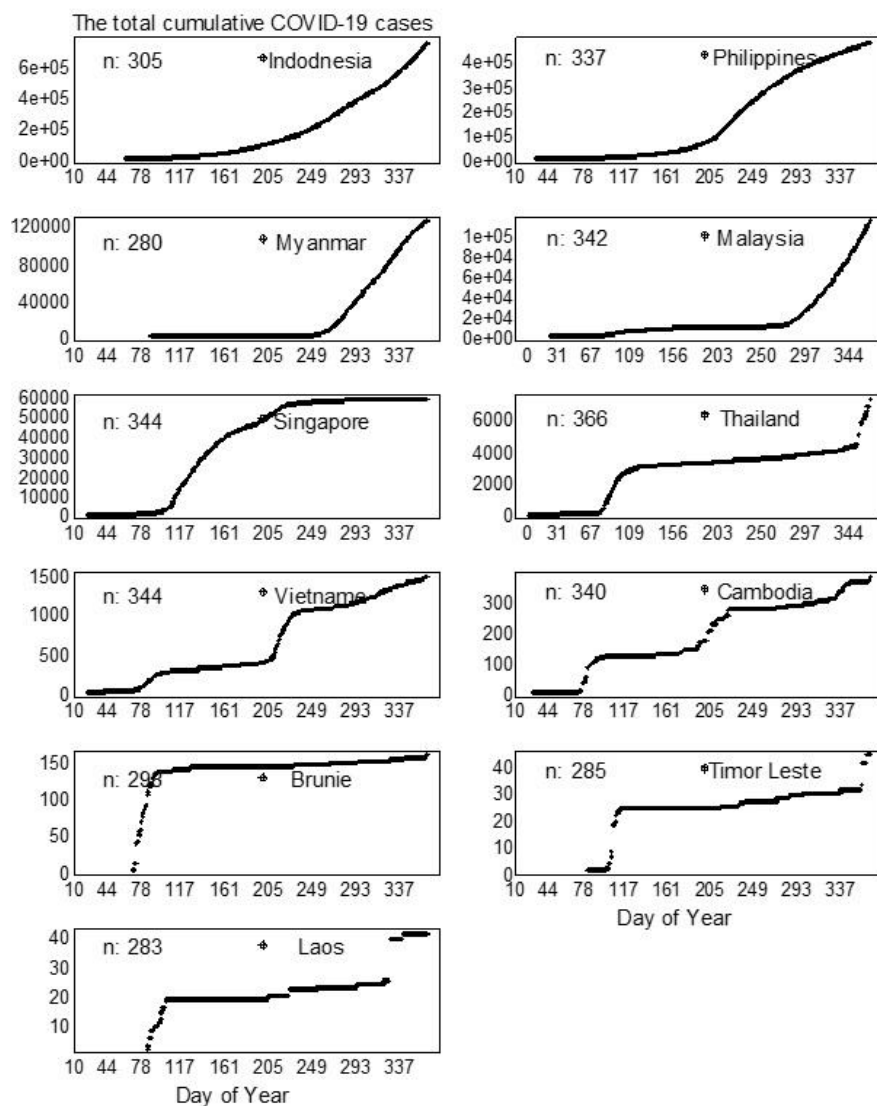


Figure 4 The total cumulative number of COVID-19 confirmed cases in 11 countries

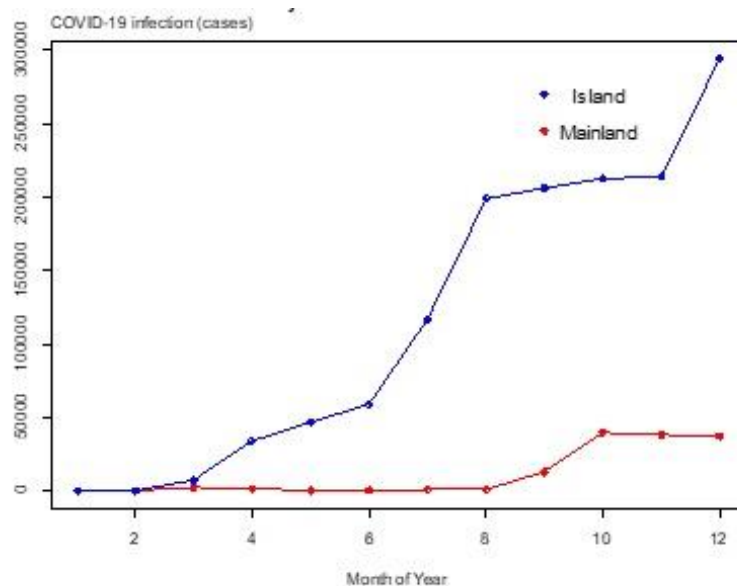


Figure 5 The cumulative COVID-19 confirmed cases in 11 countries

Normal Q-Q plot and Shapiro-Wilk test are adopted to examine normality distribution with the monthly COVID-19 confirmed cases among 11 countries in the Southeast Asia. Even though the monthly COVID-19 confirmed cases track well on the normality line in the Indonesia, the probability value of the Indonesia is 0.0291 by using the Shapiro-Wilk test. These results have been considered as a non-normality distribution. While the normal Q-Q plots remain at the low level before tracking the normality line at the end in the Brunei, Laos, and Timor-Leste with the identical probability value. The monthly number of COVID-19 infection cases are not normally distributed among 11 countries as in Figure 6.

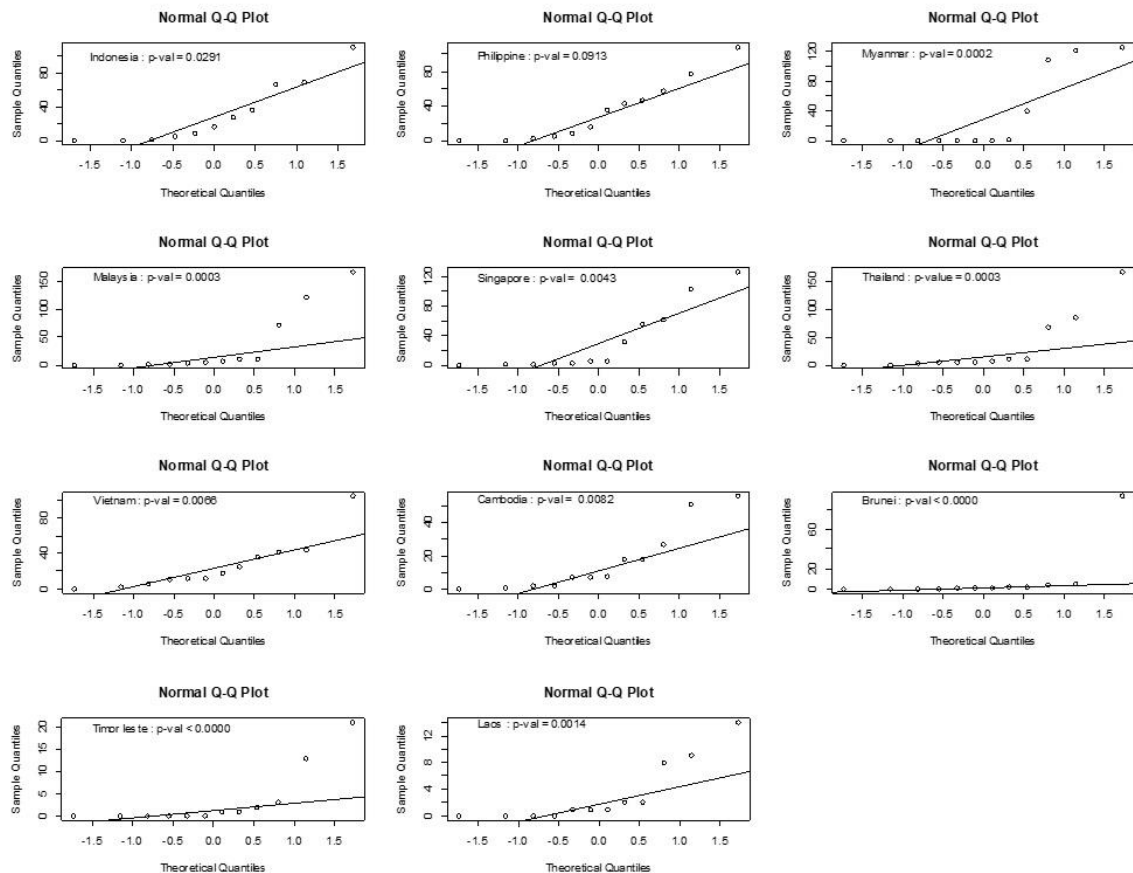


Figure 6 The normality distribution of the monthly number of covid-19 infection cases in 11 countries

The monthly number of COVID-19 infection cases have been investigated relevant to the homogeneity of variances based on 11 countries. Table 1 shows the Bartlett's test is 81.6580, and the probability value is less than 0.0000. Therefore, the level of variance for the monthly number COVID-19 infection cases is not constant across the dataset. As results of the normality distribution and the homogeneity of variances presented to check the assumption of one-way ANOVA so that these dataset is analyzed with effective statistical method. From Table 2, the Kruskal Wallis test is assessed for significant differences in the monthly number of COVID-19 infection cases in Southeast Asia countries. The Kruskal Wallis test is 25.4850 with 10 degrees of freedom. In addition, the probability value of Kruskal Wallis test is 0.0045, so it means the monthly number of COVID-19 infection cases are significant different among 11 countries from January to December 2020 with 3074 samples.



Table 1 The homogeneity of variances on the monthly number of COVID-19 infection cases in South East Asia

Test	df	P-value
Bartlett's test = 81.6580	10	< 0.0000

Table 2 The mean difference of the monthly number of COVID-19 infection cases in 11 countries

Test	df	P-value
Kruskal Wallis test = 25.4850	10	0.0045

The monthly number of COVID-19 infection cases is subsequently applied to perceive the certain conditions by the multiple comparison test. From Table 3, the Dunn's test is used to conduct pairwise multiple comparisons in the monthly number of COVID-19 infection cases of 11 countries with probability value. It is shown there are 55 pairs, which have significant mean different of the monthly number of COVID-19 infection cases with 0.05 significant level.

Table 3 Multiple pairwise comparison using Dunn's test among 11 countries in the Southeast Asia

	1	2	3	4	5	6	7	8	9	10
2	1.0000									
3	1.0000	1.0000								
4	1.0000	1.0000	1.0000							
5	1.0000	1.0000	1.0000	1.0000						
6	1.0000	1.0000	1.0000	1.0000	1.0000					
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000			
9	1.0000	0.7833	1.0000	1.0000	1.0000	1.0000	0.3969	1.0000		
10	0.6755	0.2760	1.0000	1.0000	0.8633	0.3628	0.1294	1.0000	1.0000	
11	1.0000	0.7601	1.0000	1.0000	1.0000	0.9716	0.3842	1.0000	1.0000	1.0000

Note: The numbers from 1 to 11 indicate as countries in the ASEAN, which are Indonesia, Philipines, Myanmar, Malaysia, Singapore, Thailand, Vietnam, Cambodia, Brunie, Timor Leste, and Laos.



The monthly number of COVID-19 infection cases and the geography have been used to test independence by the Chi-square test. The geography is divided into two groups are mainland and island. There are the highest number of COVID-19 infection cases in the mainland and the island approximately 350 cases at the end of year, meanwhile the slightly numbers of COVID-19 infection cases hit only 5 cases in two zones especially on January and February as in the Figure 7(a). From Figure 7(b), even though the expected number is less than 5 on January and February in the mainland and island, it would not affect the result as the number took into account less than 20% of cells.

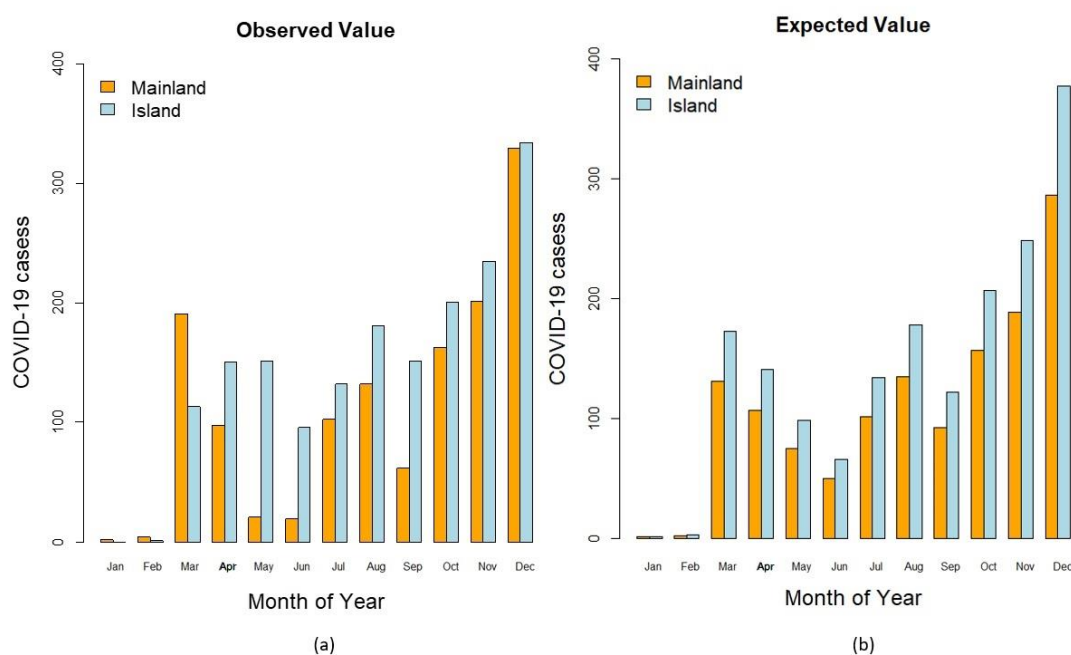


Figure 7 (a) The frequency of observed value in the mainland and the island during 2020 and
(b) The frequency of expected value in the mainland and the island during 2020.

The Table 4 has shown that the Pearson's Chi-squared test is 187.5700, and the probability value is less than 2.2×10^{-16} with 11 degrees of freedom from the independence test. It means that the geography, Mainland and Island, was associated with the monthly number of COVID-19 infection cases and independent ($p < 0.05$).



Table 4 The Pearson's Chi-square test between the geography and the monthly number of COVID-19 infection cases.

Test	df	P-value
Pearson's Chi-squared test = 187.5100	11	$< 2.2 \times 10^{-16}$

Discussion

From the descriptive analysis, COVID-19 spreading had erupted on the last week of January 2020, and Thailand is the first country of the first case appearance among Southeast Asia countries. The excessive number of total cumulative COVID-19 confirmed cases are in Indonesia, while Laos has the smallest number of total cumulative COVID-19 confirmed cases corresponding to COVID-19 pandemic study in Southeast Asia countries (Center for Strategic & International Studies, 2020). However, Thailand is one of the most countries in the Southeast Asia region with effective capacities for case detection with daily recorded information (336 days) and a successful healthcare management: a daily zero case during the era of February (World Health Organization Thailand, 2020). The line graph has presented the total cumulative number of COVID-19 confirmed cases sustained at the same level between April and July in Laos and Timor-Leste. The monthly number of COVID-19 infection cases have been climbing gradually in the island zone, while there were a rocketing number of COVID-19 infection cases starting September in the mainland area regarding to the risk assessment of COVID-19 spreading in the ASEAN Region (ASEAN Biodiaspora Virtual Center, 2020). Due to the Shapiro-Wilk test and the Bartlett's test, the dataset of COVID-19 confirmed cases relevant to the Southeast Asia region from January to December 2020 is available for nonparametric statistics. Therefore, Kruskal Wallis test and Dunn's Test are appropriate methods for the differences in the monthly number of COVID-19 infection cases in Southeast Asia countries during 2020. There are 55 pairs of significant mean differences in the monthly number of COVID-19 infection cases among 11 countries based on the Southeast Asia region. According to Chi-square test for independence, there is an association between the monthly number of COVID-19 infection cases and the type of geography in the Southeast Asia region relevant to the effect of factors on the COVID-19 cases (Al-Rousan & Al-Najjar, 2020).



Conclusions

The Kruskal Wallis test, the Dunn test, and the Chi-square test are effective and appropriate for these datasets as those are met with assumptions, making the consequences reliable and thoroughly for these objectives. The results of this study could be advantageous to new studies of COVID-19 spreading based on statistical analysis. It would raise people's awareness of COVID-19 pandemics, and encourage the researchers to conduct valuable studies related to COVID-19 spreading in the future.

Suggestion

Majority of people get the information from COVID-19 studies, and researches provided valuable consequences by using several statistical methods. Nonparametric Statistics is one of the popular statistical method and it is used significantly when the dataset does not meet the assumptions (Song et al., 2020; Baldassarri et al., 2021; Wei, Hu, Cheng, Zhao, & Ge, 2020). As this data is non-normality distributed, therefore the data of total cumulative COVID-19 confirmed cases, and the monthly number of COVID-19 infection cases are suitable for nonparametric statistics. The objective of further study is to study the COVID-19 spreading among 11 countries by using nonparametric method to perceive factors, and association of COVID-19 spreading based on environmental or demographic variables in the Southeast Asia region. It would be beneficial knowledge to prevent and reduce the proportion of COVID-19 spreading in the future.

Acknowledgments

This study was funded by Faculty of Science and Technology, Princess of Naradhiwas University. We also acknowledge the website of Our World in Data for advantageous data resource of COVID-19 spreading. The authors wish to thank the reviewer for their valuable contribution to this work.

References

- Al-Najjar, D., Al-Najjar, H., & Al-Rousan, N. (2020). CoVID-19 symptoms analysis of deceased and recovered cases using Chi-square test. *European Review for Medical and Pharmacological Sciences*, 24(21), 11428-11431.



- Al-Rousan, N., & Al-Najjar, H. (2020). Data analysis of coronavirus COVID-19 epidemic in South Korea based on recovered and death cases. *Journal of Medical Virology*, 92(9), 1603-1608.
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature medicine*, 26(4), 450-452.
- Andras V., & DelaneyHarold D. (1998). The Kruskal Wallis Test and Stochastic Homogeneity. *Journal of Educational and Behavioral Statistics*, 23(2), 170-192.
- Arsham, H., & Lovric, M. (2011). *Bartlett's Test*. Retrieved November 9, 2020 from https://www.researchgate.net/profile/Dr-Hosseini-Arsham/publication/344731676_BARTLETT%27S_TEST/links/5f8ce5cb299bfb1b53e324a18/BARTLETTS-TEST.pdf
- ASEAN Biodiaspora Virtual Center (2020). *Risk Assessment for International Dissemination of COVID-19 to the ASEAN Region*. Retrieved Nov 17, 2020 from https://asean.org/storage/2020/02/COVID-19_Report-of-ASEAN-BioDiaspora-Regional-Virtual-Center_30September2020.pdf.
- Asia Society (2020). *Introduction to Southeast Asia*. Retrieved Dec 30, 2020 from <https://asiasociety.org/education/introduction-southeast-asia>.
- Baldassarri, M., Picchiotti, N., Fava, F., Fallerini, C., Benetti, E., Daga, S., ... & Study, G. C. M. (2021). Shorter androgen receptor polyQ alleles protect against life-threatening COVID-19 disease in European males. *EBioMedicine*, 65(1), 103-246.
- Center for Strategic & International Studies (2020). *The latest on Covid-19 in Southeast Asia: December 17, 2020*. Retrieved December 31, 2020 from <https://www.csis.org/blogs/latest-covid-19-southeast-asia/latest-covid-19-southeast-asia-december-17-2020>.
- Chan, J. F. W., Yuan, S., Kok, K. H., To, K. K. W., Chu, H., Yang, J., ... & Yuen, K. Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The lancet*, 395(10223), 514-523.
- Chew, N. W., Ngiam, J. N., Tan, B. Y. Q., Tham, S. M., Tan, C. Y. S., Jing, M., ... & Khan, F. A. (2020). Asian-Pacific perspective on the psychological well-being of healthcare workers during the evolution of the COVID-19 pandemic. *BJPsych Open*, 6(6), 1-11.



- Cirincione, L., Plescia, F., Ledda, C., Rapisarda, V., Martorana, D., Moldovan, R. E., & Cannizzaro, E. (2020). COVID-19 pandemic: Prevention and protection measures to be adopted at the workplace. *Sustainability*, 12(9), 3603.
- Cyranoski, D. (2020). Mystery deepens over animal source of coronavirus. *Nature*, 579(7797), 18-19.
- Czeisler, M. É., Tynan, M. A., Howard, M. E., Honeycutt, S., Fulmer, E. B., Kidder, D. P., ... & Czeisler, C. A. (2020). Public attitudes, behaviors, and beliefs related to COVID-19, stay-at-home orders, nonessential business closures, and public health guidance—United States, New York City, and Los Angeles, May 5–12, 2020. *Morbidity and Mortality Weekly Report*, 69(24), 751.
- Dinno, A. (2015). Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal*, 15(1), 292-300.
- Duarte F (2020). *As the cases of coronavirus increase in China and around the world, the hunt is on to identify "patient zero"*. BBC News. Retrieved Mar 22, 2021 from <https://www.bbc.com/future/article/20200221-coronavirus-the-harmful-hunt-for-covid-19s-patient-zero>.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... & Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, 395(10223), 497-506.
- Joseph A. (2020). New coronavirus can cause infections with no symptoms and sicken otherwise healthy people, studies show. Stat. Retrieved January 27, 2021 from <https://www.statnews.com/2020/01/24/coronavirus-infections-no-symptoms-lancet-studies/>.
- Karaaslan, F., Dikilitaş, A., & Aydin, E. Ö. (2020). Comparison of COVID-19 Relevant Knowledge and Attitudes of Clinical and Preclinical Dental Students in Turkey. *Balkan Journal of Dental Medicine*, 24(3), 1-7.
- Kruskal, W. H. (1952). A nonparametric test for the several sample problem. *Annals of Mathematical Statistics*, 23(4), 525-540.
- Lin, B. X., & Zhang, Y. Y. (2020). Impact of the COVID-19 pandemic on agricultural exports. *Journal of Integrative Agriculture*, 19(12), 2937-2945.
- Livadiotis, G. (2020). *Statistical analysis of the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19*. Retrieved September 18, 2020 from <https://doi.org/10.1371/journal.pone.0233875>.



- Mohanty, T , Doke, P. P., & Patil, K. H. (2020). Study of geographical distribution of number of cases of COVID-19 in different areas of Maharashtra. *Int J Community Med Public Health*, 8(1), 257-261.
- Mohammed, K., Khadijah, A., Noor, A., Ameerah, M. N., Hoda, Z. H., Yasmin, A., ... Omar, A. (2020). Knowledge, Attitude and Practice Toward COVID-19 Among the Public in the Kingdom of Saudi Arabia: A Cross-Sectional Study. *Frontiers in Public Health*, 8(1), 217.
- Nayak, S. R., Arora, V., Sinha, U., & Poonia, R. C. (2020). A statistical analysis of COVID-19 using Gaussian and probabilistic model. *Journal of Interdisciplinary Mathematics*, 24(1), 19-32.
- Nkwoemeka, N. E., Okwelogu, I. S., & Chiedozie, A. P. (2020). A Scoping Review on Epidemiology, Etiology, Transmission, Clinical Presentation, Treatment and Management of Coronavirus Disease (COVID-19). *European Journal of Biology and Medical Science Research*, 8(2), 45-54.
- Piroth, L., Cottenet, J., Mariet, A. S., Bonniaud, P., Blot, M., Tubert-Bitter, P., ... Quantin, C. (2020). Comparison of the characteristics, morbidity, and mortality of COVID-19 and seasonal influenza: a nationwide, population-based retrospective cohort study. *The Lancet Respiratory Medicine*, 9(3), 251-159.
- Perlman, S. (2020). Another decade, another coronavirus. *N Engl J Med* 2020; 382:760-762
- Rajah, A., & Acharya, A. (1999). Introduction: Reconceptualizing Southeast Asia. *Asian Journal of Social Science*, 27(1), 1-6.
- Singkun, A., Patiwikriwong, P., Chainapong, K., & Weerakhachon, P. (2020). Factors Associated with Coronavirus 2019 (COVID-19) Prevention Behaviors among Health Sciences Students of a Higher Education Institution in Yala Province, Thailand. *Walailak Journal of Science and Technology*, 17(9), 967-978.
- Shahid, Z., Kalayanamitra, R., McClafferty, B., Kepko, D., Ramgobin, D., Patel, R., ... & Jain, R. (2020). COVID-19 and older adults: what we know. *Journal of the American Geriatrics Society*, 68(5), 926-929.
- Sharma, R., & Sharma, N. (2020). Statistical Analysis of Covid-19 (SARS-Cov-2) Patients Data of Karnataka, India. Retrieved December 18, 2020 from <https://assets.researchsquare.com/files/rs-72912/v1/3d0c2300-fa47-4291-a5d2-be48d6fae514.pdf>.



- Shiina, A., Niitsu, T., Kobori, O., Idemoto, K., Hashimoto, T., Sasaki, T., ... & Iyo, M. (2020). *Relationship between perception and anxiety about COVID-19 infection and risk behaviors for spreading infection: A national survey in Japan*. Retrieved September 23, 2020 from <https://doi.org/10.1016/j.bbih.2020.100101>
- Song, J. W., Zhang, C., Fan, X., Meng, F. P., Xu, Z., Xia, P., ... & Zhang, J. Y. (2020). Immunological and inflammatory profiles in mild and severe cases of COVID-19. *Nature Communications*, 11(1), 1-10.
- Southeast Asia Map. (2020). Retrieved Dec 27, 2020 from <https://www.thinglink.com/scene/645287674243973120>.
- Wang, D., Li, R., Wang, J., Jiang, Q., Gao, C., Yang, J., ... & Hu, Q. (2020). Correlation analysis between disease severity and clinical and biochemical characteristics of 143 cases of COVID-19 in Wuhan, China: a descriptive study. *BMC Infectious Diseases*, 20(1), 1-9.
- Wei, W., Hu, X. W., Cheng, Q., Zhao, Y. M., & Ge, Y. Q. (2020). Identification of common and severe COVID-19: the value of CT texture analysis and correlation with clinical characteristics. *European Radiology*, 30(12), 6788-6796.
- World Health Organization (2020). *Novel Coronavirus (2019-nCoV) (2020)*. Retrieved January 1, 2020 from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- World Health Organization Thailand (2020). *Coronavirus disease 2019 (COVID-19) WHO Thailand Situation Report – 5 March 2020*. Retrieved May 1, 2020 from https://www.who.int/docs/default-source/searo/thailand/20200305-tha-sitrep-14-covid-19-final.pdf?sfvrsn=f7f754df_0.
- Zibran, M. F. (2007). *Chi-squared test of independence*. Retrieved October 14, 2020 from <http://pages.cpsc.ucalgary.ca/~saul/wiki/uploads/CPSC681/topic-fahim-CHI-Square.pdf>
- Zofia H. (2014). Shapiro-Wilk Test With Known Mean. *Statistical Journal*, 14(1), 89-100.