UNIVERSITI TEKNOLOGI MARA

FACTORS INFLUENCING AIR QUALITY IN SEREMBAN

FARAH IZZATY BINTI MAHZAN
IZZAH HUSNA BINTI HAZRIN
UMMUL KHALIESA NAJWA BINTI ROZMAN

BACHELOR OF SCIENCE (HONS.) STATISTICS
FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

FEBRUARY 2022

UNIVERSITI TEKNOLOGI MARA

FACTORS INFLUENCING AIR QUALITY IN SEREMBAN

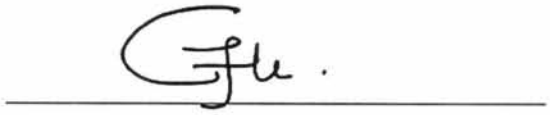
FARAH IZZATY BINTI MAHZAN **(2020993241)**
IZZAH HUSNA BINTI HAZRIN **(2020993327)**
UMMUL KHALIESA NAJWA BINTI ROZMAN **(2020979215)**

Final Year Project Submitted in Partial Fulfillment of the Requirements for the Degree of
BACHELOR OF SCIENCE (HONS.) STATISTICS

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

FEBRUARY 2022

APPROVED BY:



(HASLINDA BINTI AB MALEK)
Supervisor
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA Negeri Sembilan
Seremban Campus

ABSTRACT

Air pollution is a form of particles and gases in the air. Air pollution might be happening in any surrounding such as the outdoor or the indoor buildings. The air pollutant chemicals may cause health issues where it would affect more on older people, children, and patients with heart or lung attacks. Several areas in Malaysia recorded readings of unhealthy air quality, followed by the increased cases of open burning in the year 2021. Therefore, this study is conducted to find the factors that influence air quality in Seremban from June to December 2019. There are five factors that the researcher has studied such as relative humidity, temperature, wind speed, wind direction, and solar radiation. This study used several methods to analyze the research objectives. A stacked bar chart was used to identify the level of air quality in Seremban where the result is shown as a moderate level. In addition, Ordinal Logistic Regression helps the study to identify the significant factors that influence air quality in Seremban. It is concluded that wind direction (p-value= 0.000) and wind speed (p-value= 0.006) are shown as significant factors. Moreover, based on the Chi-Square Test of Independence, it is found that there is no association between time and air quality levels. It is hoped that this research would increase awareness among the society on how unhealthy air quality could affect the environment including people's well-being as it would result in long-term and short-term health problems.

ACKNOWLEDGEMENTS

In the name of Allah S.W.T, the Most Loving and Merciful, and let there be from you a nation inviting to good, enjoining what is right and forbidding what is wrong and those will be successful.

An award we would like to express is our very great appreciation to our intelligent supervisor, Madam Haslinda Binti Ab Malek and our knowledgeable lecturer, Madam Che Norhalila Binti Che Mohamed for the endless help, guidance and support in order for us to complete this whole final year project throughout the semester as well as for giving this golden opportunity for this research to be carried out by us.

In addition, we would like to congratulate every and each of our group members, Farah Izzaty Mahzan, Izzah Husna Hazrin and Ummul Khaliesa Najwa Rozman for their cooperation and dedication in preparing this Final Year Project. Last but not least, we would like to seek for forgiveness if this research project does not conform to or may not satisfy everyone's expectations. For its existence, we would love to receive any comments which can help us in the upcoming research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

Air pollution is a form of particles and gases in the air. The particles may consist of car emissions which are from bad engines, chemicals that came out from factories, mold spores, dust, and pollen which comes from flowers. Air pollution that happened in cities is mainly because of a gas named ozone (Rossi, Ceccato, & Gastaldi, 2020). Ozone can be smog when it is formed and termed air pollution. The chemicals of some air pollutants might cause people to have health problems as those people inhale the poisonous pollutants without knowing the effect that can have on their bodies. Air pollution might be happening in any kind of surrounding such as the outdoor or maybe indoor buildings. Air pollution can cause a huge risk towards older people, people that have heart or lung attacks, and also it can be among children. In a study done by A.Aarthi (2020), elderly people and young children could be much affected by pollution as it can attack both long-term and short-term health. Some examples of the attack towards short-term health are eye, nose, throat pain, headaches give a reaction on the allergies and the presence of infections on the upper respiratory. As for long-term health, it may be affecting people's lungs that causing lung cancer, damaging people's brain, liver, and kidney, and also the incoming of heart and respiratory disease. According to Liyuan Zhang (2017), around 3 to 7 million people were dying every year as the particulate matter pollution damaging the disease of cardiorespiratory.

In Malaysia, air quality was measured by the Air Pollution Index which was measured by the Department of Environment (DOE). The purpose of these indicators of the Air Pollution Index is to know the current level of air quality hourly throughout Malaysia. The indicator level of air quality consists of good, moderate, unhealthy, very unhealthy, and hazardous. API scale 0-50 indicates a good air pollution index while greater than 100 is unhealthy. Table 1.1 shows the guidelines of the Air Pollution Index which was provided by the Department of Environment in Malaysia:

Table 1.1: Air Pollution Index Indicator

API	Status
0-50	Good
51-100	Moderate
101-200	Unhealthy
201-300	Very Unhealthy
>301	Hazardous

Source: DOE (1996)

Malaysia has always faced the same issue through the year of unhealthy air quality. According to DOE (2015), unhealthy air quality was recorded for the first time in 34 areas in Malaysia on September 15, 2015. It is known for breaking Malaysia's haze history after the haze happened in the country in 1997. Hence, the government released an official statement to close all schools for the affected states such as Putrajaya, Kuala Lumpur, Selangor, Negeri Sembilan, Melaka. Kuching and Samarahan Divisions in Sarawak also closed all schools later on 18 September 2015.

1.2 PROBLEM STATEMENT

There are three kinds of diseases monitored seriously during the presence of air pollution which are upper respiratory tract infection (URTI), conjuctivitis and asthma. For URTI, the symptoms that this disease commonly have are cough, sore throat and runny nose (Thomas & Bomar, 2021). Meanwhile, conjunctivitis, also called pink eye, is an inflammation on the conjunctiva, which is a thin layer of tissue on the eyes, due to the allergic reaction from pollen, dust, or smoke (Begum, 2021). As for asthma, it is a long-term disease for human's lungs as the bronchial tubes in the lungs were inflamed. Symptoms may happen for this disease are pressure felt in the chest, breath getting shorter and hard to sleep due to the breathing problems (Melinda Ratini, 2021). Children are to be in a great risk as they are easily affected by these diseases and some of the effects can be highly contagious towards them such as URTI and conjunctivitis.

According to Jamal (2019), three stations in Negeri Sembilan recorded unhealthy API readings on the year 2019. Nilai with readings of 159, Seremban (132) and Port Dickson (102). As for that Sidek (2019) mentioned that there are 2605 cases of URTI disease on the epi week of 37 during Malaysia's air pollution in 2019. The cases increased compared to 2280 cases on the epi week of 36 with an increased percentage of 14.3%. Also, a total of 283 conjunctivitis cases recorded an increase of 3.1% compared to the previous case.

Hence, the Negeri Sembilan Department of Environment conducted 24-hour monitoring in all districts, especially in agricultural areas. Steps were taken to ensure that the air quality is at a good level to ensure fewer disease cases that could worsen the people's health problems. Hence, this study aimed to determine the significant factors that influence air quality in Seremban.

1.3 RESEARCH OBJECTIVES

1. To analyse the air quality level in Seremban
2. To find an association between time and air quality level in Seremban
3. To identify the significant factors that influence air quality in Seremban

1.4 RESEARCH QUESTIONS

1. What is the level of air quality in Seremban?
2. Is there any association between time and air quality level in Seremban?
3. Are there any significant factors that influence air quality in Seremban?

1.5 RESEARCH HYPOTHESES

H_1 : There is an association between time and air quality.

H_2 : Wind direction is the significant factor to the air quality.

H_3 : Wind speed is the significant factor to the air quality.

H_4 : Relative humidity is the significant factor to the air quality.

H_5 : Solar radiation is the significant factor to the air quality.

H_6 : Temperature is the significant factor to the air quality.

1.6 SCOPE AND LIMITATION OF STUDY

During the year 2019, air pollution began to hit Malaysia after the worst haze in history in 2015. According to Noraina (2019), Negeri Sembilan was attacked with air pollution on September 10, 2019, when three stations in the district recorded unhealthy API readings with Nilai recording API readings of 151, while Seremban (121) and Port Dickson (105). As for that, the study will only focus on Seremban as the researchers wanted to study only on that station, compared to Nilai and Port Dickson.

Malaysia's Department of Environment (DOE) has provided five factors that may influence air quality in Seremban towards the researchers. The factors include wind direction, wind speed, relative humidity, solar radiation, and temperature. The data was acquired hourly for Seremban from June to December 2019.

A few factors limiting to this research which the researcher is completely unaware of how the data was obtained as it is gained from the department of government. In addition, the researcher must learn to read between the observations and investigate any problems that have been faced throughout the data collection process. Aside from that, the variables may be defined or classified differently than the researcher would have preferred.

1.7 SIGNIFICANCE OF STUDY

The findings of this research would increase awareness among the society on how a high level of air quality could affect the environment including someone's health as it would result in long-term and short-term health problems. Furthermore, the research results would provide sufficient information on the pattern of air quality as well as verification of specific air quality parameters. This will assist the Ministry of Education in Malaysia to allow all schools in Seremban to close whenever the reading of API is extremely unhealthy. This investigation will also guide the Ministry of Tourism to be cautious throughout the surrounding of Seremban to take care of all tourists that come to Seremban, Negeri Sembilan.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter looks at previous research on the factors that affect air quality. This section also assesses an overview of research carried out in a specific country that examined the study's measures and components. The study will also evaluate studies that look at characteristics connected to air quality-causing factors.

2.2 AIR QUALITY

According to Ghorani-Azam, Riahi-Zanjani, and Balali-Mood (2016), the definition of air pollution is interpreted as toxic, meaning that one needs to be aware since it will affect the environment and human well-being. Some vital air pollutants that have been detected by the World Health Organization (WHO) are ground-level ozone, sulfur oxides, particle pollution, nitrogen oxides, and carbon monoxide. How and Ling (2016) stated that the impurity of air might occur when the atmosphere holds any unhealthy gases, dust, or smoke and can happen either indoor or outdoor. Plants, animals, and humans which are considered living things would eventually be affected by it. Moreover, Dong et al. (2019) mentioned that the people in Shanghai, China had always been concerned about the issue of air pollution and encouragement from other people including the government are urgently needed to make a move towards better air quality. This is because 91% of the world's population lives in areas where air pollution exceeds the safety limits according to the World Health Organization. Moreover, human's emotional and physical health including the sustainability of economic growth, such as labor productivity and tourism might get struck by the presence of pollution. A recent study done by Mohammad and

Basheer Hannon (2020) shows that many researchers have been attracted to forecast the air pollution along with the air quality around these years as they want to investigate its impact on people's health, control the quality index, and find ways to lessen the damage of the pollution on the environment and human health. China who is an example of the developing countries had continuously controlled its air pollutant levels since the government has made it a primary task to conduct Wei (2014).

2.3 RELATIVE HUMIDITY

Humidity is one of the contributing elements to air pollution, which is becoming a serious concern nowadays. Humidity levels differ from one location to the next, as well as from one season to the next. It is necessary to be aware of relative humidity levels, as both high and low relative humidity levels can have significant health consequences. Humidity levels indicate how much water vapor is present in the air. Absolute and relative humidity are two different ways to measure humidity. The amount of water vapor in the air is referred to as absolute humidity, whereas relative humidity is the ratio of absolute humidity to the highest attainable absolute humidity given the current temperature. According to Wolkoff and Kjærgaard (2007), relative humidity was searched in major databases and combined with air quality. Humidity is a key component of the environment that affects weather and is closely tied to air quality in the region, particularly in enclosed spaces. This also indicates that humidity affects the constant consistency of the air we breathe. Many people all over the world suffer from physical discomfort, showing that humidity is a crucial element that can simply influence particle motion in the troposphere due to particle electrical charge distribution (L. Zhang et al., 2017). The normal accumulation of particulate matter in the air is influenced by humidity. The size of particulate matter grows in proportion to the rise in humidity until it is too heavy to stay in the air. A study on a weather station situated in Damak, Nepal reported that relative humidity is inversely related to air temperature. The relative humidity drops when the temperature rises, and vice versa. The variation of the

relative humidity can be explained as follows whereas, at nighttime, the mass of water vapors present in the atmosphere is relatively high, but at higher temperatures in daytime, the evaporation rate is high enough to convert water into water vapor and mass of the water vapor present in the atmosphere of a given volume is relatively low and as a result, relative humidity is also low (Shrestha, Thapa, & Gautam, 2019).

2.4 TEMPERATURE

Based on studies from Kalisa et al. (2018) temperature has long been known to cause pollutants to become more airborne, it is unknown how an increase in temperature during heatwaves influences air pollution. In the meantime, according to Hansel, McCormack, and Kim (2016) found that temperature rises and extreme weather occurrences are the impacts of heat exposure have received more attention as a result of climate change. Heat and cold extremes have been linked to higher respiratory morbidity in patients with chronic obstructive pulmonary disease. Although the findings are not clear, it is critical to understand characteristics that may change susceptibility to air pollution in people with chronic obstructive lung disease. Next Lou et al. (2019) conclude that previous temperature variations and air pollution caused by climate change have had a significant impact on the mortality or morbidity of human diseases and that climate change would increase these impacts. A study conducted by Buckley, Samet, and Richardson (2014), said it has become normal practice to include air pollution in temperature research, this method may not always be suitable. In this opinion, researchers use directed acyclic graphs (DAGs) to show potential short-term links between temperature, air pollution, and disease. Examine the causal structures for estimates of temperature-related health consequences.

2.5 WIND SPEED

One of the contributing factors that influence air quality is wind speed. Wind speed and direction must be assessed to monitor air quality. It can help locate pollution sources and provide a more accurate picture of what is going on in the atmosphere. Air quality and emission levels are a problem not only in industrial and highly urbanized regions but also in typical rural areas. According to a study by Coccia (2021), it indicates that cities in low-wind regions have more days of air pollution than cities in high-wind regions, with 88 contaminated days compared to 65 polluted days surpassing PM10 or ozone every year. High wind speeds increase gaseous and particulate matter dispersion, according to this preliminary result. In terms of health and present environmental issues, a study found that low wind speeds, as well as often high levels of air pollution that exceeded permissible limits of ozone or particle matter, resulted in a higher number of COVID-19 affected persons and deaths Coccia (2020). It indicates that atmospheric stability, because of low wind speed, reduces the dispersion of gaseous and particulate matter, which can act as a carrier of SARS-CoV-2 in the air, sustaining COVID-19 diffusion and producing public health problems. According to a finding done by Y. Zhang (2019) reported that severe wind speed, rather than average wind speed, was the most important meteorological component impacting the change in Beijing's air quality index, as it created greater atmospheric motion and turbulence, boosting air pollution diffusion and dilution. Nonetheless, average wind speed and rainfall capacity were shown to have no substantial impact on air quality (Y. Zhang, 2019). Apart from that, a study conducted in Krakow, Poland indicated that wind speed and mixing-layer height are in at least moderate agreement with concentration values for various pollutants such as PM10, PM2.5, NO2, NOx, CO, and C6H6 with a negative correlation coefficient which indicates that low wind speed and low mixing-layer height may be the primary cause of high concentrations of these compounds in the air, particularly during the winter months (Oleniacz et al., 2016).

2.6 WIND DIRECTION

According to a conducted research made by Huang et al. (2019), seven different wind directions with the angle of 0, 15, 30, 45, 60, 75, and 90 were studied in a long secluded canyon. The numerical results of the study show that ambient wind direction gives an effect to the wind flow pattern and the pollutants caused by traffic. Henry, Chang, and Spiegelman (2002) also stated in their research where wind direction is a variable that is suitable to check its relationship with the air pollutants by using Gaussian kernel. These nonparametric techniques used are adequate to study the air quality and the science of atmospheric. As for study research in Seoul, Korea by Kim and Guldmann (2011), wind direction during the afternoon which is affected by the solar radiation produces a composite form of ozone, which is one of the air pollutants. They were mostly scattered in the direction of the wind blowing hence, increasing the gathering amount of the pollutions. Other researchers such as Demirci and Cuhadaroglu (2000) conducted a study in the northeast of Turkey in Trabzon to analyze the relationship between wind direction and pollutions which the result appears as a weak relationship. It was stated in the study that those kinds of results may happen due to the impact of the existing winds that were overlooked during the urban planning on that located area. However, Guerra et al. (2006) found that there is a significant relationship for the direction of the wind in the Southern Kansas region as it produces a high-level concentration of pollutant matters throughout the whole study.

2.7 SOLAR RADIATION

Pollution and extreme weather events can alter how much and what kinds of light can reach the Earth's surface. Electromagnetic radiation emitted by the sun is known as solar radiation. Using several technologies, solar radiation may be absorbed and converted into useful forms of energy such as heat and electricity. Hourly variations in sunlight are also caused by the Earth's rotation. The sun is low in the sky in the early morning and late

afternoon. At midday, when the sun is at its highest position, its rays travel further through the atmosphere. According to a study by Su et al. (2021), air pollution has a significant weakening effect on solar radiation. In the study, relative humidity and air quality index are used to modify the beam and diffuse solar radiation models. The results show that the adjusted models' calculated value is closer to the measured data and the order in which the correction factors are applied has an impact on the results. According to the findings, the best corrective effect is to apply relative humidity first, followed by an air quality index. The layer-by-layer decreasing solar radiation models can be more precise and reliable using this strategy. Furthermore, a study by Khodakarami and Ghobadi (2016) attempts to analyze solar radiation in terms of air quality to enhance living conditions in metropolitan cities, as well as examines and recommends solutions to air pollution and inversion problems in metropolitan areas. The findings of this study show that solar energy, as a clean energy source, can help to mitigate the effects of urban pollution. Furthermore, the obtained results show that solar energy can be used as an engine to induce turbulence in static air masses and influence air layers on a city scale, resulting in significant pollution reduction.

2.8 TIME

A study by EurekAlert (2001) which consists of teamwork of researchers, the consequences of air quality towards nighttime and daytime was conducted in Phoenix, Arizona. The results found that the pollutants which come from the use of automobiles such as commuters will be captured during the cold temperature in the nighttime. After that, as the sunrises, the heat of the sun will warm the ground causing the pollutants to react quickly inside the layer of itself. Hence, the high concentration of ozone is found to be significantly linked to time. Next, Talhelm (2017) made a study on what time the lowest air pollution happened in a day by using the data of Beijing that was gained by the US Embassy. It is found that the lowest air pollution occurred between midday and the

end of the evening which is around 6 o'clock. The outcome was a bit different from people's opinionated thinking since most people assume that good air would happen during the nighttime. Meanwhile, according to Askariyeh, Vallamsundar, and Farzaneh (2018), the concentration of the pollutants was increasing inside the molecule of the worst air quality during nighttime as the state of the atmospheric was solid, the heights of reaction were low and the speed of the wind was low. The estimation also found that air pollution mostly appeared in rural areas during the period of the night.

2.9 CONCLUSION

The literature review is based on the previous studies that were conducted by other researchers. For this study, the researchers used six independent variables which are wind direction, wind speed, relative humidity, solar radiation, temperature and and time. These independent variables were tested to see if they have any relationship with the dependent variable, which is air quality. Previous studies can also be used as guidance to determine the factors that lead to air pollutants in Seremban. Furthermore, previous studies stated that when analyzing air quality, it is necessary to measure the following factors as it can help to understand the chemical reactions that occur in the air. As a result, all of the predictors listed are appropriate to be used in this study.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

This chapter focuses on the methodological part of this research. Hence, this methodological part contains information about the method, a process for the data analysis and archive the objectives. Source of data, theoretical framework and method of analysis are elaborate in this part.

3.2 SOURCE OF DATA

The data used in this research is secondary data. It referred to historical data structures of variables that have already been collected, assembled, and interpreted at least once. This data can be a form of newspaper reports, the annual reports of a particular department, or the published report by a specific organization.

For this research, the secondary data was obtained from Malaysia's Department of Environment (DOE) containing air quality (API), temperature, wind speed, relative humidity, wind direction, solar radiation, and time. A list of hourly data in Seremban was taken from June to December 2019 with 5135 samples. Based on the observation, the lowest air quality reading was 19 (good) and the highest was 178 (unhealthy). Hence, the air quality data will be categorized into three levels such as unhealthy, moderate, and good.

3.2.1 Description of Variables

Based on Table 3.1, the description of variables displays a table of dependent variable and predictor factors that may influence air quality, along with their descriptions.

Table 3.1: Description of Variables

Variables	Description
Air Quality	0: API more than 100 and less than or equal to 200 (Unhealthy) 1: API more than 50 and less than or equal to 100 (Moderate) 2: API less than or equal to 50 (Good)
Temperature	Calculated every 1 hours in a day Unit of measurement: Degree Celsius
Wind Speed	Calculated every 1 hours in a day Unit of measurement: Meters per second (m/s)
Relative Humidity	Calculated every 1 hours in a day Percentage of moisture in the air compared to the maximum allowable level of moisture in the air at a given temperature
Wind Direction	Calculated every 1 hours in a day Unit of measurement: Cardinal direction or in degrees
Solar Radiation	Calculated every 1 hours in a day Unit of measurement: Watt per square metre (W/m2) in SI units
Time	Unit of measurement: Day time (am) / Night time (pm)

3.3 THEORETICAL FRAMEWORK

Based on Figure 3.1, the theoretical framework shows the air quality as the dependent variable and it depends on relative humidity, temperature, wind speed, wind direction and solar radiation

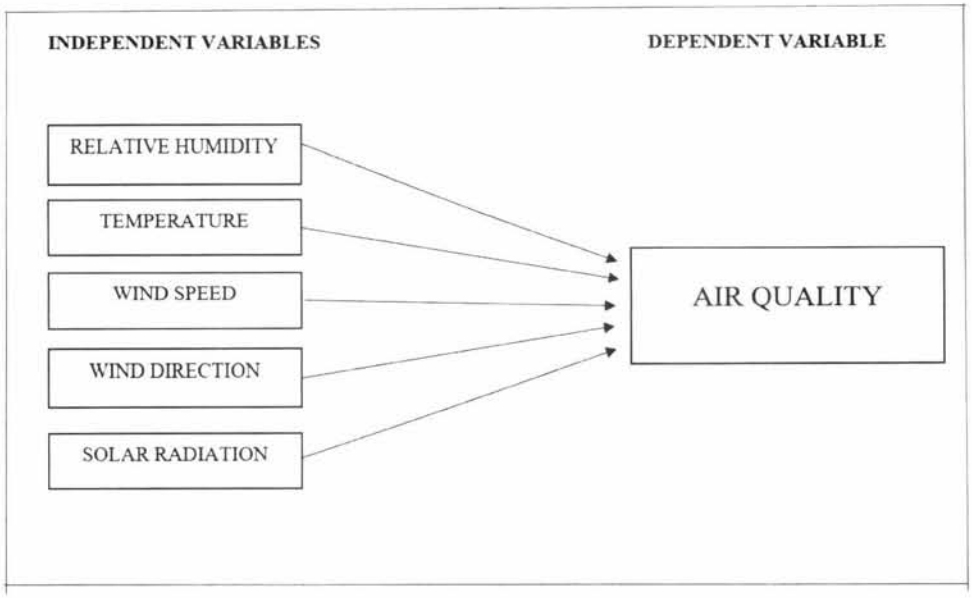


Figure 3.1: Theoretical Framework of Factors Influence Air Quality

3.4 DATA ANALYSIS

3.4.1 Descriptive Analysis

In this research, a descriptive statistic is to describe basic features such as measures of central tendency (mean, minimum, and maximum) and measures of variability (standard deviation) of the study. Descriptive statistics is the study dealing with method and data is summarized, organized, and present in a suitable and instructive way. As an example, graphical techniques are into measures of central tendency and measures of variability, it provides simple summaries about the data and estimates together with simple graphical analysis.

Hence for this study, a stacked bar chart would be suitable to check on the air quality level in Seremban from June to December 2019. It would define whether, during that day, the air quality level is either unhealthy, moderate, or good. Thus, descriptive statistics are used to get a clearer vision of the data.

3.4.2 Chi Square Test of Independence

The chi-square test of independence, also known as Pearson's chi-square test or chi-square test of association, is used to determine whether there is an association between categorical variables. It is a non-parametric test that can only compare categorical variables and cannot make comparisons between continuous variables or between categorical and continuous variables. This test also produce result of a contingency table which analyzed the data. A contingency table, also known as a cross-tabulation is a table where data is categorised using two categorical variables, in which one variable appears in the rows and the other in the columns.

For this study, variable time and air quality is used to achieve the second objective. Using chi square test of independence, it would determine whether there is an association between both variables or not. Before that, there are four assumptions must be met in order to use this test and trust its results, two of it will be analyzed using SPSS Statistics;

1. Two variables should be measured at an ordinal or nominal level.
2. Both variables should consist of two or more categorical, independent groups.
3. Independence of observations. This means that the subjects in each group have no relation with one another.
4. Expected value of cells in the contingency table should be 5 or greater in at least 80% of cells.

3.4.3 Ordinal Logistic Regression Analysis

Ordinal logistic regression, often known as ordinal regression, is a type of regression that uses one or more independent variables to predict an ordinal dependent variable. It can be considered as a generalization of either multiple linear regression or binomial logistic regression. Ordinal regression, like other types of regression, can predict the dependent variable through interactions between independent variables.

For this study, dummy variables are created for the dependent variable which is air quality, and it follows the guidelines from the Department of Environment. The air quality was constructed in order with categories of "Unhealthy", "Moderate" and "Good".

This model refers to which factors influence the most on the response variable which contains temperature, wind speed, relative humidity, wind direction and solar radiation. By using ordinal logistic outcomes, it will help to achieve the third objective which is to find the significant factors that influence air quality in Seremban.

a) Assumptions of Ordinal Logistic Regression

These assumptions should be evaluated in order, because if an assumption violation is not correctable, ordinal regression will no longer work. If these assumptions are not met, the findings of ordinal regression may not be valid.

1. Response variable is measured on an ordinal level. Ordinal level is a variable that has two or more categories, where only categories can be ordered or ranked. For this study, air quality, the dependent variable is measured in an ordinal level that has three categories. It was constructed in order with categories of "Unhealthy", "Moderate" and "Good".

2. One or more of the independent variables are either continuous, categorical, or ordinal. When using SPSS statistics to do an ordinal logistic regression, it cannot be considered as ordinal variables. Therefore, the independent variables that will be used in ordinal logistic regression to meet the third objective which are wind direction, wind speed, relative humidity, solar radiation and temperature are continuous and calculated in hourly.

3. There is no multicollinearity. Happen when two or more independent variables are significantly connected with each other. In ordinal regression, determining if there is multicollinearity is an essential part. There are three ways to detect a multicollinearity existence. Firstly, determine which pair of independent variables is strongly related by get the correlation matrix. Next, identify correlations between independent variables using variance inflation factors (VIF) and assess the strength of the correlation. A value of 1 denotes that there is no relationship between independent variable and any other variables. Value between 1 and 5 indicate a moderate correlation while more than 10 indicate critical levels of multicollinearity with poorly estimated coefficients. Lastly, checking from the value of tolerance which is below than 0.2, it shows multicollinearity problem exist.

4. Proportional odds, where each independent variable has an identical effect at each cumulative split of the ordinal dependent variable, and it is a fundamental assumption for this type of ordinal regression model. The assumption score test is insignificant at the five percent level of significance, assuming that data satisfies the proportional odds assumption. In this study, by looking at the value of predicted probabilities in the test of parallel lines will identify how well the ordinal regression model predicts the dependent variable. When the proportional odds assumption gives non-significant results, this means the assumption is satisfied.

b) Ordinal Logistic Function

This research interested in unhealthy air quality and SPSS Statistics will treat good air quality as the reference category. The purpose of the function below is to find the best fit line for ordinal logistic regression with valid for proportional odds assumption.

$$logit[p(y \geq 3)] = \alpha_j + (\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5) \tag{3.1}$$

where x_1 : Temperature

x_2 : Wind speed

x_3 : Relative humidity

x_4 : Wind direction

x_5 : Solar radiation

Y : 0: more than 100 and less than or equal to 200 (Unhealthy)

: 1: more than 50 and less than or equal to 100 (Moderate)

: 2: less than or equal to 50 (Good)

c) Model Fitting

This section will improvise the model to help the research on forecasting the outcome value. It will be considering a model that does not include any explanatory variables with a model which includes all the explanatory variables. Then, the final model will differentiate from the baseline to conclude whether there are significant improvements in the model fit of data.

The significant value from the model fitting will be used and compared to the level of significance 0.05. The result of the final model considered fits the data as the value of the significance is less than 0.05. As can be seen, accurate predictions would be generated by the model rather than estimating the probabilities of the outcome values.

d) Goodness of Fit

The 'Goodness of Fit' table defines the value of Pearson Chi-Square and Deviance Chi-Square statistics which help discover the fitness data of the model. In this analysis, the p-values will be compared to 0.05 and they need to be larger than the level of significance to conclude the data as a good fit.

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected} \tag{3.2}$$

Hence that, as the p-values are not significant, the study would fail to reject the null hypothesis and suggest that the model is a good fit.

e) Pseudo R-Square

In the regression model, R-Squared is used to estimate the coefficient of determination. It would indicate the proportion or percentage of the total variation in the air quality explained by variation in the temperature, wind speed, relative humidity, wind direction and solar radiation. The function of R squared is to adjust the scale of the statistic cover from 0 to 1.

Three approximate statistics of R-Squared to be computed such as Cox and Snell, Nagelkerke and McFadden. As in logistic model, a small value of R-Squared is considered poor since it shows the performance of predictor variables to the air quality. Other than that, R squared value cannot be negative; the value of R squared will increase when more predictors are added to the model.

f) Parameter Estimates

The function for parameter estimates is to analyze the effect of each predictor variable. For quantitative terms, it represents the change in response affected by a one-unit change in the predictor while all other predictors remain constant. The coefficient units of measurement are the predictor units of response per unit. Meanwhile, for the constant variable, when all predictors are zero and the units of measurement are the same as the response variable, the coefficient for the constant term equals the response.

Thus, reject null hypothesis to conclude the parameter is not equal to zero and contributes to the model when the p-value is less than significant value 0.05. Furthermore, a coefficient specifies the strength of predictors contribution, for more likely near to zero the coefficient shows that variable has small impact on the result. Other than that, the direction of the relationship is indicated by the sign of the coefficient while confidence interval indicates the range in the estimate for the assumption of normally distributed errors. As a result, it concludes the factors is significant.

3.4.4 Summary of Data Analysis

Table 3.2 summarises the data analysis for each objective, along with the method of analysis used.

Table 3.2: Summary of Data Analysis

Objectives	Method of Analysis
To analyse the air quality level in Seremban	Descriptive Analysis by performing Stacked Bar Chart
To find an association between time and air quality level in Seremban	Chi Square Test of Independence
To identify the significant factors that influencing air quality in Seremban	Ordinal Logistic Regression

3.4.5 Conclusion

In conclusion, there are three objectives to be met in this study. The descriptive analysis would present the data in an informative way with graphical technique. It is useful in showing the air quality level in Seremban by using a stacked bar chart. Next, the chi-square test of independence helps find any association between time and air quality levels in Seremban. It also produces a contingency table to study the correlation between the two variables and to show the frequency distribution of air quality levels given the occurrence of time. Lastly, ordinal logistic regression will identify the significant factor influencing air quality in Seremban. Procedures that need to measure in the analysis are multicollinearity, proportional odds, model fitting, the goodness of fit, Nagelkerke R-squared, and parameter estimates. A prediction model of the probability factor influencing the air quality will eventually help identify the cause of air quality and make Seremban a healthy environment with good air quality.

CHAPTER 4

RESULT AND DATA ANALYSIS

4.1 INTRODUCTION

In this chapter, the findings of this study were displayed and briefly discussed the outcomes. Descriptive analysis, chi-square test of independence, and ordinal logistic regression were used in this study to analyze the research objectives.

4.2 DESCRIPTIVE ANALYSIS

In this analysis, five independent variables were used such as wind direction, wind speed, relative humidity, solar radiation and temperature. It contains the measures of central tendency (mean, median and mode), variability of descriptive statistic (standard deviation) and a stacked bar chart.

4.2.1 Measure of Central Tendency and Variability

Table 4.1: Measure of Central Tendency and Variability

	Wind direction	Wind speed	Relative humidity	Solar radiation	Temperature
Mean	159.42	1.01	79.54	150.07	27.29
Minimum	.000	.000	42.22	.000	21.21
Maximum	359.8	6.89	96.64	964.85	34.66
Standard Deviation	103.68	.98	11.37	225.42	2.93

From the Table 4.1, there are five independent variables used to measure central tendency and variability of descriptive statistics. Firstly, the variables wind direction contained 159.46 for mean, .000 for minimum value, 359.8 for maximum value, and 103.66 for standard deviation. Next, for variables wind speed, the average value is 1.01, the minimum is .000, the maximum is 6.89, and the standard deviation is 0.99. Other than that, variables relative humidity obtained mean at 79.54, minimum value at 42.22, maximum 96.64 and 11.34 for standard deviation. Furthermore, the average value for solar radiation was 150.09, with .000 for minimum, a maximum of 964.85 and the standard deviation at 225.45. Temperature was the last variable, with a mean of 27.29, a minimum of 21.21, a maximum of 34.66, and a standard deviation of 2.93. In conclusion, the most influential variables for measuring central tendency are wind direction and solar radiation, however for the variability of descriptive statistics, solar radiation had the largest standard deviation of all the variables and wind speed had the lowest.

4.2.2 Stacked Bar Chart

Table 4.2: Frequency of Air Quality Level

	Frequency	Percentage
Good	1350	26.29%
Moderate	3475	67.67%
Unhealthy	310	6.04%

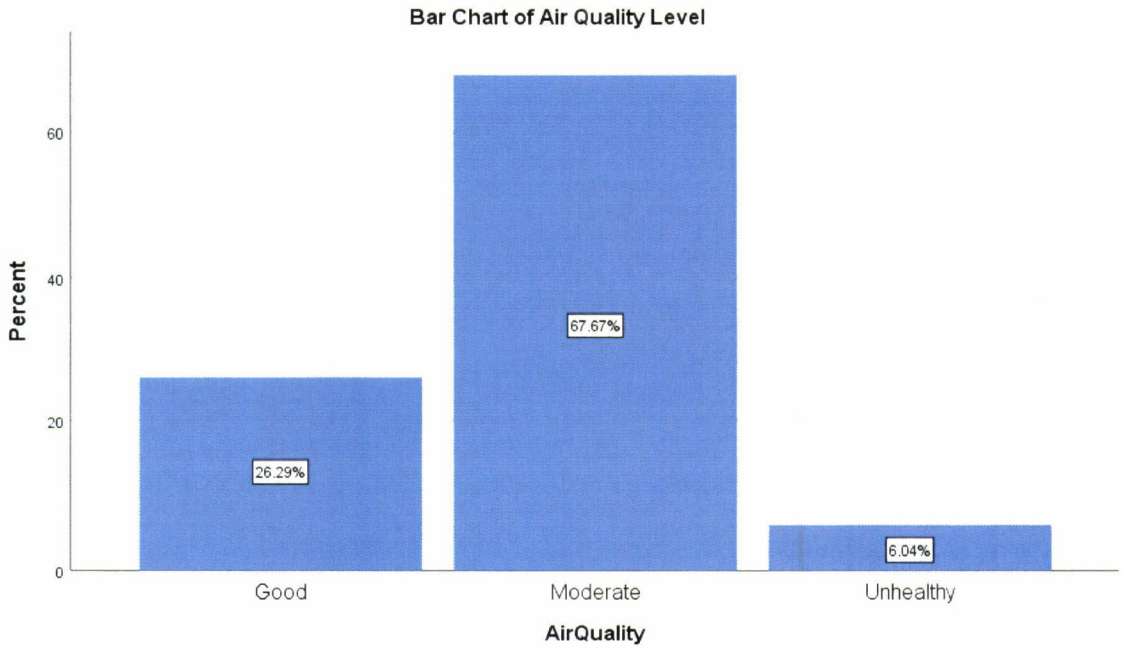


Figure 4.1: Bar Chart of Air Quality Level

The frequencies of each air quality level are shown in Table 4.2. The highest frequency was 3475 at 67.67 percent, which indicates a moderate air quality level. The second highest frequency was good air quality at 1350 frequencies with 26.29 percent. Meanwhile, the lowest frequency was 310 frequency with 6.0% unhealthy air quality level.

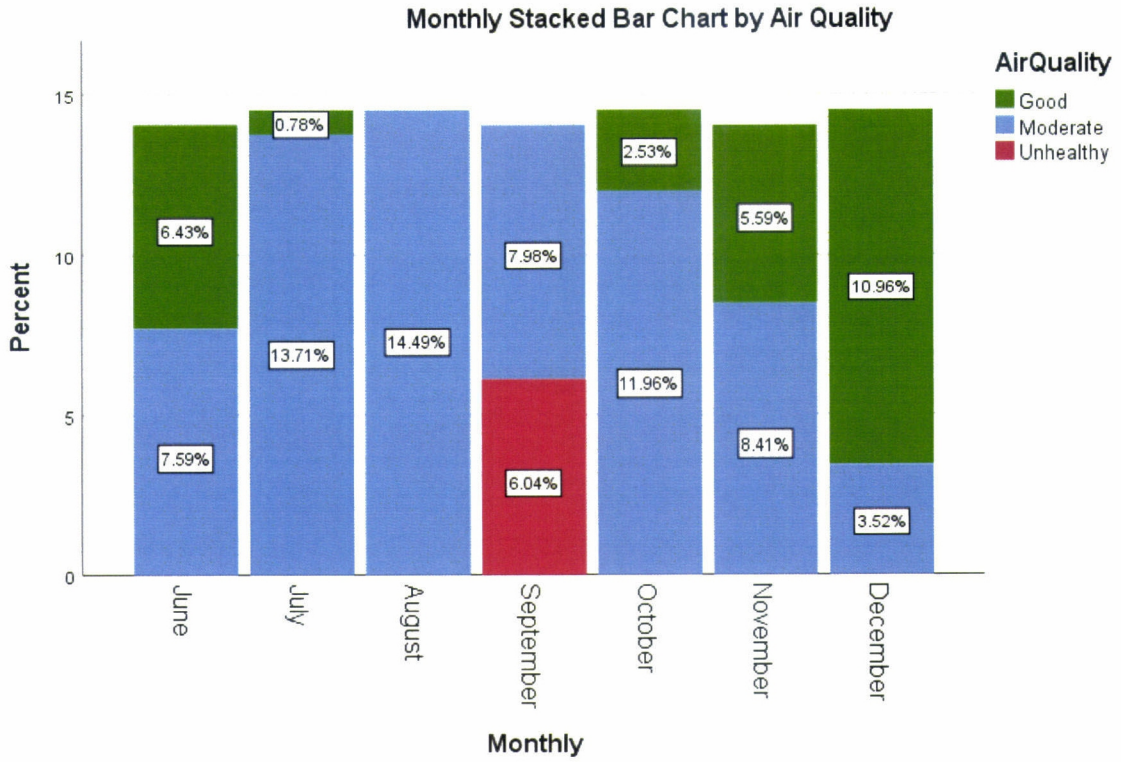


Figure 4.2: Monthly Stacked Bar Chart by Air Quality

A stacked bar chart in Figure 4.1 was used to identify the air quality in Seremban from June to December 2019. It indicate that the air quality can be categorized as good, moderate, or unhealthy. The percentage of air quality is represented on the Y-axis, while the six months of air quality is represented on the X-axis. Hence, SPSS was used to create a stacked bar chart.

Based on Figure 4.1, the good reading in June is lowest compared to moderate which is 6.43% good and 7.59% for moderate. While moderate is highest (13.71%) than good (0.78%) in July. However, there is no good reading in August, since all air quality is moderate by 14.49%. There is 6.04% for unhealthy and 7.98% moderate in September. It can see among all the months, only September has unhealthy reading. This issue has also been discussed in Berita Harian, written by (Jamal, 2019) about the forest fires in Sumatra and Kalimantan, Indonesia was caused cross border haze and affected in increased air quality readings in Seremban during September. Due to that, Haze Action Plan was activated by monitoring and conducting patrols in agricultural areas to ensure there are no open burning activities which would exacerbate more air quality during September. Other than that, for October, it shows a good reading of 2.53% and 11.95% for moderate reading.

Meanwhile, in November it was 5.59% for good and 8.43% for moderate. Furthermore, a good reading is highest (10.96%) compared to moderate (3.52%) in December. Thus, it can be concluded that the level of air quality in Seremban was in moderate level from June to December 2019.

4.3 CHI SQUARE TEST OF INDEPENDENCE

Variable time was categorized as DayTime and NightTime. Hence chi-square test of independence is used to determine whether there is an association between time and air quality level in Seremban. Other than that, a contingency table was used to shows the frequency distribution of air quality level given the occurrence of time.

4.3.1 Contigency Table

Table 4.3: Crosstabulation

			Time		
			DayTime	NightTime	Total
Air Quality	Good	Count	751	599	1350
		Expected Count	731.1	618.9	1350
		% within Air Quality	55.6%	44.4%	100%
	Moderate	Count	1859	1616	3475
		Expected Count	1882	1593	3475
		% within Air Quality	53.5%	46.5%	100%
	Unhealthy	Count	171	139	310
		Expected Count	167.9	142.1	310
		% within Air Quality	55.2%	44.8%	100%
Total		Count	2781	2354	5135
		Expected Count	2781	2354	5135
		% within Air Quality	54.2%	45.8%	100%

From the Table 4.3, The total sample size for moderate air quality was 3475, followed by good at 1350, and unhealthy at 310. As a result, during the day and night, the air quality is mostly at a moderate level. During daytime the good air quality is higher (55.6%) as compared to moderate (53.5%) and unhealthy (55.2%) while moderate air quality is the highest (46.5%) compared to good (44.4%) and unhealthy (44.8%) during nighttime.night time.

4.3.2 Chi-Square

Table 4.4: Chi Square Test

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1.916	2	.384

0 cells (0.0%) have expected count less than 5. The minimum expected count is 142.11

From Table 4.4, the chi square assumption was met since there is no cells had an expected count less than 5 and less than 20%. The value obtained for Pearson chi-square test statistics is 1.916. It is found that the p-value of test statistics (0.384) more than significant (0.05). As a result, the null hypothesis was failed to reject where there was insufficient evidence to conclude the association between time and air quality level. Hence, there is no association was found between time and air quality level in Seremban.

4.4 ORDINAL LOGISTIC REGRESSION

Ordinal Logistic Regression method was used to determine which factors affect the air quality in Seremban. Before that, there are four assumptions to be fulfilled, and two out of them will be examined by using SPSS Statistics which are the existence of multicollinearity and finding the proportional odds from the study.

4.4.1 Ordinal Dependent Variable

To meet the assumptions, firstly, dependent variable should be measured at the ordinal level. Ordinal level are variable that have two or more categories, where only categories can be ordered or ranked. This study focused in unhealthy air quality. Thus, air quality as dependent variable were constructed in order with categories of "unhealthy", "moderate", and "good" which depicts how good or bad the air quality was.

4.4.2 Independent Variables

In this ordinal logistic regression, all of the independent variables obtain are continuous. All independent variables which are wind direction, wind speed, relative humidity, solar radiation, and temperature were calculated in hourly and essential in defining air quality level in Seremban.

4.4.3 Multicollinearity

Table 4.5: Coefficients Table

Variable	Collinearity Statistics	
	Tolerance	VIF
Relative humidity	0.097	10.358
Temperature	0.093	10.720

Table 4.5 show the value of Tolerance and VIF for each independent variable. If value of tolerance is below 0.2 and VIF greater than 10, multicollinearity problem exist. As seen in the Table 4.5, relative humidity and temperature are the two independent variables with the smallest tolerance values which less than 0.2, with 0.097 and 0.093, respectively. Next, the two independent variables with the highest value of variance inflation factor are relative humidity and temperature, with 10.358 and 10.720, respectively. Thus, this indicates that multicollinearity is present. Relative humidity and temperature have a strong relationship with at least one of the model's other predictions. To meet the proportional odds assumption, both variables must be removed at the same time. If just one variable is removed, the proportional odds assumption will not be satisfied and the ordinal logistic regression' method cannot be used in this study. Therefore, these two variables has been removed from the study.

Table 4.6: Correlation Matrix

		Wind direction	Wind speed	Relative humidity	Solar radiation	Temperature
Wind direction	Pearson Correlation	1	0.103	-0.083	0.137	0.108
	Sig (2-tailed)		0.000	0.000	0.000	0.000
Wind speed	Pearson Correlation	0.103	1	-0.597	0.527	0.612
	Sig (2-tailed)	0.000		0.000	0.000	0.000
Relative humidity	Pearson Correlation	-0.083	-0.597	1	-0.728	-0.946
	Sig (2-tailed)	0.000	0.000		0.000	0.000
Solar radiation	Pearson Correlation	0.137	0.527	-0.728	1	0.760
	Sig (2-tailed)	0.000	0.000	0.000		0.000
Temperature	Pearson Correlation	0.108	0.612	-0.946	0.760	1
	Sig (2-tailed)	0.000	0.000	0.000	0.000	

The existence of multicollinearity can also be detected using a correlation matrix. It identifies which pair of independent variables are strongly related. According to Krehbiel (2004), said that general rule of thumb is if the value of correlation coefficient is greater than 0.8 or less than -0.8, then multicollinearity exists. As seen in the Table 4.6, variables relative humidity and temperature are highly correlated because the correlation coefficient is -0.946 which indicates a strong negative relationship exists between the two predictor variables, but temperature does not correlate with other variables. Hence, only relative humidity will be removed according to Table 4.6.

Table 4.7: Coefficients Table

Variable	Collinearity Statistics	
	Tolerance	VIF
Wind direction	0.980	1.020
Wind speed	0.721	1.387
Solar radiation	0.715	1.399

After removing variables relative humidity and temperature from the model, the results are presented in Table 4.7. As all of the values for tolerance more than 0.2 and values for variance inflation factor (VIF) less than 10, no multicollinearity exists and the assumption is satistfied.

4.4.4 Proportional Odds

Table 4.8: Test of Parallel Lines

Model	Sig
General	.297

Based on Table 4.8, it is related to the last assumption of ordinal regression which is proportional odds. It is a fundamental assumption for this sort of ordinal regression model and is no longer valid if the assumption is violated. It can be seen that, the p-value (0.297), is greater than significant value of 0.05. Hence, it indicated that each independent variable which are wind direction, wind speed and solar radiation has an identical effect at each cumulative split of the air quality. Thus, the main assumption of the ordinal regression is satisfied.

4.4.5 Fitness of Model

Table 4.9: Fitness of Model

Model	Sig.
Model Fitting	.000
Pearson	.363
Deviance	1.000

Table 4.9 above consists of two measures in Ordinal Logistic Regression which are model fitting and goodness of fit. Both measures will be evaluated using their significant values and help the study to check whether the model is a good fit for the data or not.

According to Table 4.9, the result for model fitting (p-value = 0.000) is less than 0.05 hence, it can be concluded that the final model is a good fit for the data. Meanwhile, both significant values of Pearson (p-value = 0.363) and Deviance (p-value = 1.000) are larger than the level of significant 0.05. As for that, it can be assumed that the model is having goodness of fit.

4.4.6 Pseudo R-Square

Table 4.10: Pseudo R-Square Table

Cox and Snell	.006
Nagelkerke	.007
McFadden	.004

The nature of the outcome and the predictor variables play a role in determining a good value of R-Squared. According to Table 4.10, the Pseudo R-square values (e.g. Nagelkerke = 0.7%) indicate that there is a small amount of variation in the level of air quality. This may be due to the reason of having various factors that affect the air quality in Seremban.

4.4.7 Parameter Estimates

Table 4.11: Parameter Estimate Table

		Estimate	Sig.
Threshold	Air Quality=0	-3.038	.000
	Air Quality=1	.756	.000
Location	Wind Direction	-.001	.000
	Wind Speed	-.099	.006
	Solar Radiation	5.129E-5	.742

Table 4.11 shows the regression coefficients and significant tests for each of the independent variables in the model. Based on Table 4.11, wind direction (p-value = 0.000) and wind speed (p-value = 0.006) were statistically significant towards air quality as both significant values were less than the significance level of 0.05. However, solar radiation (p-value = 0.742) was not a significant predictor for this study since the significant value is larger than 0.05, where can be assumed as does not affect the air quality.

$$ln\theta_0 = -3.038 - 0.001x_1 - 0.099x_2 + 0.00005x_3 \tag{4.1}$$

$$ln\theta_1 = 0.756 - 0.001x_1 - 0.099x_2 + 0.00005x_3 \tag{4.2}$$

In conclusion, wind direction is a significant predictor of air quality in Seremban. One degree increase in the variable wind direction is associated with the decrease in the log odds of being in unhealthy air quality level versus good air quality level in the amount of 0.001. Based on Kim and Guldmann (2011) research study, air pollutants did produce when the wind direction is on a daytime. This situation is same as in the conducted study where the ozone did appear at the same time. The ozone was flowing together with the wind direction and increasing the amount of ozone.

In addition, wind speed also a significant predictor of air quality in Seremban. One meter per second (m/s) increase in the variable wind speed is associated with the decrease in the log odds of being in unhealthy air quality level versus good air quality level in the amount of 0.099. As in both studies of Coccia (2021) and Y. Zhang (2019), the region of the low wind often to have an impact on the air quality index.

4.5 CONCLUSION

In conclusion, the three objectives of this study have been achieved. From June through December, a stacked bar chart was used to show the readings of good, moderate, and unhealthy air quality and assists in determining which month the performance of unhealthy air quality began to increase in Seremban. Thus, objective one has been met in analysing the air quality level in Seremban. Other than that, the second objective in this study was to find any association between time and air quality level in Seremban. Since variable time was divided into "Daytime" and "Nighttime", the Chi-Square Test of Independence was applied to identify the relationship between time and air quality level. However, the p-value test statistic was greater than level of significance of 0.05, thus it concludes time change had no effect on air quality level in Seremban. Furthermore, the third objective was to identify the significant factors that influence air quality in Seremban. Since variables relative humidity and temperature gained the value of variance inflation more than 10 and tolerance value below 0.2, therefore those two variables were excluded. Wind direction and wind speed are the only two factors that influence air quality in Seremban, however solar radiation is not a significant factor that influence the unhealthy air quality.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 CONCLUSION

Air quality needs to be well taken care of, and an immense problem towards the people may occur if the air pollution happens around places. Therefore, this research was conducted based on the data obtained in Seremban from June 2019 until December 2019. The goals of study are to analyse the level of air quality in Seremban, finding an association between time and air quality level and identify the factors that lead the air quality.

The first objective of this study is to analyze the air quality level in Seremban, which was achieved using a stacked bar chart. It concluded that the highest percentage of good air quality is in December 2019 with 10.96% while the least percentage is 0.78% during July 2019. August 2019 achieved the most moderate air quality with 14.49% in a month and the least percentage was during December with 3.52%. There was also recorded unhealthy air quality which is only in September 2019 with a percentage of 6.04%. Also, a central tendency which contains mean, minimum, and maximum, and variability which has standard deviation; were measured to describe the basic features of the descriptive analysis. In summary, wind direction gained the highest value of mean with 159.42. Also, solar radiation had the highest value of maximum with an amount of 964.85. Wind direction, wind speed, and solar radiation had the lowest reading of minimum value with 0.000. Meanwhile, solar radiation obtained the amount of 225.42 as the highest value of standard deviation among the other variables.

The second objective of this study is to find an association between variable time and the air quality level in Seremban. Method used for this particular objective is the Chi-Square Test of Independence. Variable time is categorized into two categories such as "DayTime" and "NightTime". According to the result achieved with the used of SPSS Statistics, there is no enough evidence to conclude that there is an association occured among time and the air quality level in Seremban.

The third objective is to identify the significant factors that influence the air quality in Seremban where the factors are identified by using the Ordinal Logistic Regression method. It is found that, there are two significant variables which are wind speed and wind direction. Other than that, solar radiation is found as not significant, meaning that it is not the factor that influences the unhealthy air quality in Seremban. As a conclusion, all three objectives are met for this study.

5.2 RECOMMENDATIONS

There are some recommendations for the future research to be improve and more precise for conducting the study. Firstly, it is recommended to the future researchers to conduct air quality readings in the morning and prevent collecting air quality readings at night because the solar radiation is expected to be zero. Hence, the output of analyse the data will be affected. Furthermore, air pollution likely to be happen more in the morning compared at night. There is research from (Awang et al., 2001) stated during the morning peaks, the air quality is higher compared in the evening peaks. It because the traffic density is height and wind velocity are usually low in the morning. Next, the future researchers can conduct research in areas where there are more vehicles on the road, as well as an active sectors of the manufacturing industry. Hence, avoid selecting rural locations since they might reduce the outcome of unhealthy air quality. Thus, places with a high population, such as urban centers, will help to obtain more data on air quality for researchers to identify which are the most factors influenced air quality. Other factors that may influence air quality, such as pressure and rainfall, can be studied further by future researchers. Furthermore, may consider to add factors from previous study by Wei (2014) which are traffic index and the day's air quality. Lastly, future researchers may consider include additional time periods in the research since the findings will be more accurate and the data of air quality for "good", "moderate" and "unhealthy" are balance in the next research.

REFERENCES

A.Aarthi, N. S. D., P.Gayathri. (2020). Air quality prediction through regression model. *INTERNATIONAL JOURNAL OF SCIENTIFIC TECHNOLOGY RESEARCH*, *9*(03), 923–928. Retrieved from `http://www.ijstr.org/final-print/mar2020/Air-Quality-Prediction-Through-Regression-Model.pdf`

Askariyeh, M., Vallamsundar, S., & Farzaneh, R. (2018, 09). Investigating the impact of meteorological conditions on near-road pollutant dispersion between daytime and nighttime periods. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*, 036119811879696. doi: 10.1177/0361198118796966

Awang, M., Jaafar, A. B., Abdullah, A., Ismail, M., Hassan, M., Abdullah, R., ... Hamdan, N. (2001, 12). Air quality in malaysia: Impacts, management issues and future challenges. *Respirology*, *5*, 183 - 196. doi: 10.1046/j.1440-1843.2000.00248.x

Begum, J. (2021). *Conjunctivitis (pinkeye).* Retrieved from `https://www.webmd.com/eye-health/eye-health-conjunctivitis`

Buckley, J. P., Samet, J. M., & Richardson, D. B. (2014). Commentary: does air pollution confound studies of temperature? *Epidemiology*, *25*(2), 242–245.

Coccia, M. (2020). The effects of atmospheric stability with low wind speed and of air pollution on the accelerated transmission dynamics of covid-19. *International Journal of Environmental Studies*, 1–27.

Coccia, M. (2021). How do low wind speeds and high levels of air pollution support the spread of covid-19? *Atmospheric pollution research*, *12*(1), 437–445.

Demirci, E., & Cuhadaroglu, B. (2000). Statistical analysis of wind circulation and air pollution in urban trabzon. *Energy and buildings*, *31*(1), 49–53.

DOE. (2015). *Chronology of haze episodes in malaysia.* Retrieved from `https://www.doe.gov.my/portalv1/wp-content/uploads/2015/09/`

Chronology-of-Haze-Episodes-in-Malaysia.pdf

Dong, et al. (2019). The relationship between the actual level of air pollution and residents' concern about air pollution: Evidence from shanghai, china. *International journal of environmental research and public health*, *16*(23), 4784.

EurekAlert. (2001). *Urban air quality – the difference between night and day.* Retrieved from https://www.eurekalert.org/features/doe/2001-06/ddoe-uaq061402.php

Ghorani-Azam, A., Riahi-Zanjani, B., & Balali-Mood, M. (2016). Effects of air pollution on human health and practical measures for prevention in iran. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, *21*.

Guerra, S. A., Lane, D. D., Marotz, G. A., Carter, R. E., Hohl, C. M., & Baldauf, R. W. (2006). Effects of wind direction on coarse and fine particulate matter concentrations in southeast kansas. *Journal of the Air & Waste Management Association*, *56*(11), 1525–1531.

Hansel, N. N., McCormack, M. C., & Kim, V. (2016). The effects of air pollution and temperature on copd. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *13*(3), 372–379.

Henry, R. C., Chang, Y.-S., & Spiegelman, C. H. (2002). Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction. *Atmospheric Environment*, *36*(13), 2237–2244.

How, C. Y., & Ling, Y. E. (2016). The influence of pm2. 5 and pm10 on air pollution index (api). *Environmental Engineering, Hydraulics and Hydrology: Proceeding of Civil Engineering, Universiti Teknologi Malaysia, Johor, Malaysia*, *3*, 132.

Huang, Y.-D., Hou, R.-W., Liu, Z.-Y., Song, Y., Cui, P.-Y., Kim, C.-N., et al. (2019). Effects of wind direction on the airflow and pollutant dispersion inside a long street canyon. *Aerosol and Air Quality Research*, *19*(5), 1152–1171.

Jamal, A. A. (2019). *Jerebu: Kualiti udara ns semakin pulih.* Retrieved from https://www.bharian.com.my/berita/nasional/2019/09/606114/jerebu-kualiti-udara-ns-semakin-pulih

Kalisa, et al. (2018). Temperature and air pollution relationship during heatwaves in birmingham, uk. *Sustainable cities and society*, *43*, 111–120.

Khodakarami, J., & Ghobadi, P. (2016). Urban pollution and solar radiation impacts. *Renewable and Sustainable Energy Reviews*, *57*, 965–976.

Kim, Y., & Guldmann, J.-M. (2011). Impact of traffic flows and wind directions on air pollution concentrations in seoul, korea. *Atmospheric Environment*, *45*(16), 2803–2810.

Krehbiel, T. C. (2004). Correlation coefficient rule of thumb. *Decision Sciences Journal of Innovative Education*, *2*(1), 97–100.

Liyuan Zhang, Y. Z. Y. H. Z. G. C. Y., Yan Cheng. (2017). Impact of air humidity fluctuation on the rise of pm mass concentration based on the high-resolution monitoring data. *Aerosol and Air Quality Research*, *17*, 543–552. doi: 10.4209/aaqr.2016.07.0296

Lou, et al. (2019). Health effects of climate change through temperature and air pollution. *Current Pollution Reports*, *5*(3), 144–158.

Melinda Ratini, M., DO. (2021). *asthma.* Retrieved from https://www.webmd.com/asthma/what-is-asthma

Mohammad, S. A., & Basheer Hannon, O. (2020). Using logistic regression with time-stratified method for air pollution datasets forecasting. *IRAQI JOURNAL OF STATISTICAL SCIENCES*, *17*(31), 33–48.

Noraina, R. (2019). *Tempias jerebu tiba di negeri sembilan.* Retrieved from https://www.hmetro.com.my/mutakhir/2019/09/495158/tempias-jerebu-tiba-di-negeri-sembilan

Oleniacz, et al. (2016). Assessing the impact of wind speed and mixing-layer height on

air quality in krakow (poland) in the years 2014–2015. *Journal of Civil Engineering, Environment and Architecture*, *63*, 315–342.

Rossi, R., Ceccato, R., & Gastaldi, M. (2020). Effect of road traffic on air pollution. experimental evidence from covid-19 lockdown. *Sustainability*, *12*(21), 8984.

Shrestha, A. K., Thapa, A., & Gautam, H. (2019). Solar radiation, air temperature, relative humidity, and dew point study: Damak, jhapa, nepal. *International Journal of Photoenergy*, *2019*.

Sidek, H. (2019). *Ramai sakit sebab jerebu di negeri sembilan.* Retrieved from `https://www.hmetro.com.my/mutakhir/2019/09/496881/ramai-sakit-sebab-jerebu-di-negeri-sembilan`

Su, G., Zhang, S., Hu, M., Yao, W., Li, Z., & Xi, Y. (2021). The modified layer-by-layer weakening solar radiation models based on relative humidity and air quality index. *Energy*, 122488.

Talhelm, T. (2017). *What time of day is air pollution lowest?* Retrieved from `https://smartairfilters.com/en/blog/when-is-pm-2-5-the-lowest/`

Thomas, M., & Bomar, P. A. (2021). *Upper respiratory tract infection.* StatPearls Publishing LLC.

Wei, D. (2014). *Predicting air pollution level in a specific city.* Stanford Publication.

Wolkoff, P., & Kjærgaard, S. K. (2007). The dichotomy of relative humidity on indoor air quality. *Environment international*, *33*(6), 850–857.

Zhang, L., Cheng, Y., Zhang, Y., He, Y., Gu, Z., Yu, C., et al. (2017). Impact of air humidity fluctuation on the rise of pm mass concentration based on the high-resolution monitoring data. *Aerosol and Air Quality Research*, *17*(2), 543–552.

Zhang, Y. (2019). Dynamic effect analysis of meteorological conditions on air pollution: A case study from beijing. *Science of the Total Environment*, *684*, 178–185.

APPENDICES

Research Schedule:

Activities / Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Identifying topic proposal	/	/												
Preparing and presenting proposal			/	/										
Designing and discussing instruments					/	/								
Collecting secondary sources							/	/						
Analyzing data									/					
Drafting report										/	/			
Preparing presentation												/		
Presenting report													/	
Revising final report														/

Descriptive Analysis:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
wind_direction	5135	.000	359.796	159.45091	103.661886
wind_speed	5135	.000	6.885	1.01190	.978477
relative_humidity	5135	42.217	96.644	79.53831	11.369375
solar_radiation	5135	.000	964.850	150.08522	225.445150
ambient_temperature	5135	21.207	34.658	27.29275	2.931164
Valid N (listwise)	5135				

Second Objective (Chi Square Test of Independence):

Crosstabulation

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
AirQuality * Time	5135	100.0%	0	0.0%	5135	100.0%

AirQuality * Time Crosstabulation

			Day/Night		
			DayTime	NightTime	Total
AirQuality	Good	Count	751	599	1350
		Expected Count	731.1	618.9	1350.0
		% within AirQuality	55.6%	44.4%	100.0%
	Moderate	Count	1859	1616	3475
		Expected Count	1882.0	1593.0	3475.0
		% within AirQuality	53.5%	46.5%	100.0%
	Unhealthy	Count	171	139	310
		Expected Count	167.9	142.1	310.0
		% within AirQuality	55.2%	44.8%	100.0%
Total		Count	2781	2354	5135
		Expected Count	2781.0	2354.0	5135.0
		% within AirQuality	54.2%	45.8%	100.0%

Chi-Square Tests

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1.916[a]	2	.384
Likelihood Ratio	1.918	2	.383
Linear-by-Linear Association	.780	1	.377
N of Valid Cases	5135		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 142.11.

Third Objective (Ordinal Logistic Regression Analysis):

Multicollinearity

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	wind_direction	.958	1.043
	wind_speed	.553	1.808
	relative_humidity	.097	10.358
	solar_radiation	.411	2.435
	ambient_temperature	.093	10.720

a. Dependent Variable: AirQuality

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	wind_direction	.980	1.020
	wind_speed	.721	1.387
	solar_radiation	.715	1.399

a. Dependent Variable: AirQuality

Case Processing Summary

Case Processing Summary

		N	Marginal Percentage
AirQuality	Unhealthy	310	6.0%
	Moderate	3475	67.7%
	Good	1350	26.3%
Valid		5135	100.0%
Missing		0	
Total		5135	

Model Fitting Information

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	8058.737			
Final	8029.433	29.304	3	.000

Link function: Logit.

Goodness-of-Fit

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	10304.457	10255	.363
Deviance	8026.660	10255	1.000

Link function: Logit.

Pseudo R-Square

Pseudo R-Square

Cox and Snell	.006
Nagelkerke	.007
McFadden	.004

Link function: Logit.

Parameter Estimates

Parameter Estimates

							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	[AirQuality = 0]	-3.038	.081	1407.039	1	.000	-3.197	-2.879
	[AirQuality = 1]	.756	.060	160.520	1	.000	.639	.873
Location	wind_direction	-.001	.000	16.721	1	.000	-.002	-.001
	wind_speed	-.099	.036	7.499	1	.006	-.170	-.028
	solar_radiation	5.129E-5	.000	.109	1	.742	.000	.000

Link function: Logit.

Test of Parallel Lines

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	8029.433			
General	8025.744	3.689	3	.297

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

Dataset:

	date	API	wind_direction	wind_speed	relative_humidity	solar_radiation	ambient_temperature	AirQuality	time
1	01-Jun-2019	44	32.956	498	94.000	000	25.218	Healthy	00:00
2	01-Jun-2019	45	352.809	431	94.000	000	25.120	Healthy	01:00
3	01-Jun-2019	45	45.309	656	94.000	000	24.930	Healthy	02:00
4	01-Jun-2019	46	316.614	302	94.000	000	24.883	Healthy	03:00
5	01-Jun-2019	46	322.334	071	94.000	000	24.892	Healthy	04:00
6	01-Jun-2019	46	337.969	464	93.850	000	24.774	Healthy	05:00
7	01-Jun-2019	46	23.297	307	92.683	000	24.574	Healthy	06:00
8	01-Jun-2019	44	275.307	367	92.333	000	24.273	Healthy	07:00
9	01-Jun-2019	44	318.318	293	92.267	19.558	24.505	Healthy	08:00
10	01-Jun-2019	43	66.017	279	87.267	139.847	26.598	Healthy	09:00
11	01-Jun-2019	42	333.935	100	78.667	348.356	29.392	Healthy	10:00
12	01-Jun-2019	43	210.048	632	72.234	635.031	31.069	Healthy	11:00
13	01-Jun-2019	43	237.962	1.125	70.434	547.824	30.988	Healthy	12:00
14	01-Jun-2019	43	197.405	598	65.733	422.047	31.517	Healthy	13:00
15	01-Jun-2019	45	215.688	1.639	68.569	226.801	31.533	Healthy	14:00
16	01-Jun-2019	44	220.635	2.016	81.150	62.411	28.563	Healthy	15:00
17	01-Jun-2019	44	226.846	1.602	77.850	270.612	28.859	Healthy	16:00
18	01-Jun-2019	43	240.006	1.368	75.900	152.054	29.442	Healthy	17:00
19	01-Jun-2019	42	290.207	961	77.381	78.493	29.286	Healthy	18:00
20	01-Jun-2019	42	311.367	625	83.250	6.982	28.064	Healthy	19:00
21	01-Jun-2019	43	42.686	701	86.650	000	26.967	Healthy	20:00
22	01-Jun-2019	45	66.832	747	87.517	000	26.630	Healthy	21:00
23	01-Jun-2019	46	33.828	460	87.667	000	26.329	Healthy	22:00

Reference : 100-KNS(PJI.9/19)
Date : 26 August 2021

Faculty of Computer & Mathematical Sciences
Universiti Teknologi MARA
Kampus Seremban
Persiaran Seremban Tiga/1, Seremban 3
73000 Seremban
Negeri Sembilan

Dear Sir/ Madam

ETHICS APPROVAL BY UITM RESEARCH ETHICS COMMITTEE

Thank you for submitting your research ethics application. We would like to inform that the UITM Cawangan Negeri Sembilan had deliberated your proposal.

It is our pleasure to inform you that the Research Ethics Committee has agreed to grant an ethics approval for the said study. Kindly refer to the Attachment 1 below for the approval code of the study and validity period is from 25th August 2021 – 24 August 2022.

Please submit a progress report of the study to the REC Secretariat 12 months from the date of this approval letter, and annually until the study has been completed. Amendments to the study documents are to be submitted to the REC for approval. A final report must also be submitted to the REC at the end of the said study.

The UITM Cawangan Negeri Sembilan operates in accordance to the ICH Good Clinical Practice Guidelines, Malaysia Good Clinical Practice Guidelines and the Declaration of Helsinki.

If you require further information, please contact REC Secretariat of each campus, (**Kuala Pilah** : Dr. Eddie Tan Ti Tjih (*eddietan@uitm.edu.my*) / **Seremban** : Puan Wahidah Tumijan (*wahidah06@uitm.edu.my*) / **Rembau** : Intan Nurbalzura Zainuddin (*intan422@uitm.edu.my*).

Thank you.

Yours truly,

PROF. DR. YAMIN YASSIN
Rector
UITM Cawangan Negeri Sembilan

Pejabat Penyelidikan, Jaringan Industri, Masyarakat & Alumni
Office of Research, Industrial Linkages, Community & Alumni Network

Universiti Teknologi MARA
Cawangan Negeri Sembilan
Kampus Kuala Pilah
Pekan Parit Tinggi, 72000 Kuala Pilah
Negeri Sembilan Darul Khusus
Tel : 06-483 2399
Faks : 06-484 2449/2303
https://nsembilan.uitm.edu.my/v1/

26	REC/38/2021	SELECTION SMALL RESTAURANT LOCATION USING ANALYTICAL HIERARCHY PROCESS (AHP)	ANIS AMIRA BINTI ASRI	SITI SARAH BINTI JANURI	Minimal risk research
27	REC/39/2021	AWARENESS OF FINANCIAL MANAGEMENT	ALIA NAIWA BT MOHAMAD SAID	ZAITUL ANNA MELISA BINTI MD YASIN	Minimal risk research
28	REC/40/2021	A STUDY OF STUDENT'S PREFERENCE BETWEEN PHYSICAL CLASS AND OPEN AND DISTANCE LEARNING	SHAMIRA BINTI MOHD FAIZAL	ZAITUL ANNA MELISA BT MD YASIN	Minimal risk research
29	REC/41/2021	THE APPLICATION OF THEORY OF PLANNED BEHAVIOUR ON HEALTHCARE WASTE DISPOSAL BEHAVIOUR DURING COVID-19 PANDEMIC	AHMAD DANIAL HAKIM BIN A ZAIDEY	ZURAIDA BINTI JAAFAR	Minimal risk research
30	REC/42/2021	PSYCHOLOGICAL EFFECT OF COVID-19 PANDEMIC AMONG UITM CAWANGAN NEGERI SEMBILAN STUDENTS: DEMOGRAPHIC CHARACTERISTICS AND IMPLICATION	MUHAMMAD NUR AIMAN BIN KAMARUDIN	ZURAIDA JAAFAR	Minimal risk research
31	REC/43/2021	QUALITY OF LIFE DURING COVID-19 AMONG UITM STAFF: CASE STUDY ON SEREMBAN CAMPUS	MOHAMAD SYAHMI BIN RAZAK	NORNADIAH BINTI MOHD RAZALI	Minimal risk research
32	REC/44/2021	DRIVERS BEHAVIOUR IN CAUSING CAR ACCIDENTS	NUR AINA ATHIRAH BINTI SUMERI	DR NURUL NISA' BINTI KHAIROL AZMI	EXEMPTION
33	REC/45/2021	MODELLING AND FORECASTING THE VOLATILITY OF GOLD PRICE	NUR HUDA ATHIRAH BINTI ABDULLAH	DR NURUL NISA' BINTI KHAIROL AZMI	EXEMPTION
34	REC/46/2021	TEXT MINING ANALYSIS ON CYBERBULLYING	AIZAT FARHAN BIN MUHAMMAD HARIS	DR NURUL NISA' BINTI KHAIROL AZMI	EXEMPTION
35	REC/47/2021	UNEMPLOYMENT : DOES THE UNEMPLOYMENT RATE EFFECT THE DIVORCE RATE IN MALAYSIA*	SYARIFAH AL ANISSA BINTI SYED SHAH	DR. NORYANTI NASIR	EXEMPTION
36	REC/48/2021	FACTORS INFLUENCING AIR QUALITY IN SEREMBAN	FARAH IZZATY BINTI MAHZAN	HASLINDA BINTI AB MALEK	EXEMPTION