Investigation and Analysis of People Preferences in Choosing Public Transportation During COVID-19 Pandemic: Statistical Approach and Case Study in Malang City

Ivan Fadhil Amrullah, Sri Wiwoho Mudjanarko
Department of Civil Engineering, Universitas Narotama, Surabaya, INDONESIA
E-mail: ivanamrullah14@gmail.com

Submitted: March 16, 2023 | Revised: April 01, 2023 | Accepted: October 05, 2023
Published: October 07, 2023

ABSTRACT

The transportation condition in Malang City during COVID-19 pandemic has been the subject of this research, with one study proposing to support information about public transportation services in the city. This policy encourages people to participate in activities such as working from home and school during the COVID-19 pandemic. Another concern is that the availability of public transportation facilities is being limited by a lack of conventional transportation as a result of this policy. The greatest impact on this variable is the appropriateness of the price, as customers of the service are typically users on a daily basis. Online transportation is considered more effective in terms of distance because there is a mapping application or location that is provided by service providers or drivers in specific areas. The coefficient of determination is a statistical measure used to evaluate the accuracy of a model in predicting future outcomes. The model is generally accurate because all of the variables are significant and have a direct effect on the dependent variables. The F-test value associated with the coefficient was 33.927, and the level of significance was >0.0001. The following describes the results of the t-test regarding the effect of variables on variables:

Reliability The value of the t-count is 2.358 > 1.68, and the significance value is 0.013 < 0.05.
Responsiveness The t-count of 9.513 is greater than 1.68, and the associated significance level is 0.0001 which is less than 0.05.

Keywords: public transportation; COVID-19 pandemic; SPSS analysis; preferences; statistical modeling

INTRODUCTION

According to the World Health Organization (WHO), as of 1 January 2023 there have been 6,736,046 confirmed cases of COVID-19 in Indonesia, with 160,914 deaths reported (Wirastri et al., 2023). It is important to note that the actual death toll from COVID-19 may be higher than the number of confirmed deaths due to limited testing and challenges in accurately tracking the virus. As for entry requirements into Indonesia, travelers are currently required to comply with the Indonesian Government’s health protocols, which includes using the PeduliLindungi application and having proof of a negative COVID-19 test result (Ekawati et al., 2023). It is crucial for individuals to stay informed about the COVID-19 situation in Indonesia and take necessary precautions to protect themselves and others. The Indonesian government has made various efforts to prevent the spread of the virus, including the Large-Scale Social Restriction Policy (PSBB), the implementation of restrictions on community activities (PPKM), and the Adaptation to the New Normal (AKB). This policy encourages people to coordinate their behavior according to health protocols set by the Government (Pratama et al., 2023). In addition, the COVID-19 pandemic changed people's lifestyles, and aspects of people's lives changed rapidly. As a provider of high-quality statistical data, the Central Statistics Agency (BPS) is responsible for providing information on how to deal with the pandemic, provides information about compliance awareness and effectiveness of health protocols, awareness and assessment of the COVID-19 pandemic, the role of the media in providing information in efforts to prevent COVID-19 (Hartantri et al., 2023).

During a pandemic situation it is interesting to assess effect of COVID-19 condition upon the public transportation preferences, especially in Malang City. Malang is a landlocked city in the Indonesian province of East Java (Purwanto et al., 2022). The transportation condition in Malang City has been
the subject of research, with one study proposing a draft proposal of designed to support information about public transportation services in the city. Another study suggests that the existing transport conditions in Malang can be maximized by adding more trips to vehicles, allowing them to carry more waste (Siregar, 2014). While these studies provide some insight into transportation conditions in Malang, it is important to note that the city's transportation system is complex and multifaceted, with various modes of transportation available, including buses, taxis, and ojeks (motorcycle taxis). Overall, the transportation conditions in Malang City are a crucial aspect of the city's infrastructure, which plays a significant role in the daily lives of its residents. This development has brought Malang City to a balance between supply and demand of transportation, the low level of service and the performance of public transportation has caused an increase in the number of private car users. Malang City public transportation, including city transportation, is one of the transportation facilities used for the activities of the people of Malang City (Siregar, 2014). Although the number of public transportation passengers is decreasing day by day, the people of Malang City still need local public transportation passengers in their daily activities.

The COVID-19 pandemic has had a significant impact on transportation systems worldwide, including in Malang City. In response to the crisis, the Malang City Government has implemented various policies to ensure the safety of its citizens while also maintaining essential transportation services. These policies include reducing the capacity of public transportation to comply with physical distancing guidelines, requiring the use of masks for all passengers, and conducting regular disinfection of public transport vehicles and stations to prevent the spread of the virus. Additionally, the government has encouraged the use of alternative modes of transportation such as bicycles and walking to reduce the risk of infection (Dirgahayani, 2013; Suryani et al., 2022). The Department of Transportation has played an active role in the United States Government's response to the pandemic, implementing measures to ensure the safety of passengers and transportation workers. Similarly, in China, existing policies and impacts were reviewed to analyze the impact of the pandemic on the urban transportation sector and propose measures to mitigate its effects (Delclo- Alió et al., 2022).

The government encourages citizens to participate in activities such as working from home and school during the COVID-19 pandemic. However, not everyone is capable of working from home, some people have been forced to continue working in the office during this pandemic because of the demands of their jobs (Deveci et al., 2022; Shabani et al., 2022). Another concern is that the availability of public transportation facilities is being limited by a lack of conventional transportation as a result of this policy. One means of transportation that has increased in popularity during the pandemic is online transportation. The component that affects customer satisfaction is the quality of service. The quality of service intended should be satisfactory if the experience of the service is equivalent to or exceeds the expected quality of service during the pandemic of COVID-19. Excellent customer service that matches customer expectations will lead to a sense of customer satisfaction and trust in the products or services they use. Conversely, the lack of quality of service that is incompatible with the customer's expectations will lead to dissatisfaction with the customer.

This investigation was intended to determine the people's decisions regarding the selection of conventional and online transportation for daily activities based on customer satisfaction during the COVID-19 pandemic. The research conducted can provide an overview of the conventional and online transportation consumers in Malang City, this can contribute to science in the form of the application of transportation theories to the needs of the larger community particularly in the pandemic/ epidemic condition.

**RESEARCH METHODS**

**Location of research**

The importance of sampling location in research cannot be overstated. The location from where a sample is collected can have a significant impact on the outcome of a research study. This is because sampling techniques, such as random and systematic sampling, ensure that the sample is representative of the population being studied, thus increasing the external validity of the research findings (Guerrieri & Parla, 2022). Moreover, the number of people a researcher contacts is directly
related to the cost of a study, and obtaining samples from the right location saves money by allowing researchers to gather the same answers and results as they would by surveying a larger group (Pan & Ryan, 2022). In research design, population and sampling are two important terms. A population is a group of individuals that share common connections (Guerrieri & Parla, 2022). Therefore, choosing the right sampling location is a crucial step in conducting research, as it can impact the accuracy, reliability, and generalizability of the study findings. This study took place in Malang City, East Java, Indonesia, from October to November of 2021. The topic of investigation is the user's perception of the AG (Arjosari-Gadang) public transportation system and online transportation services including Gojek and Grab. The location of the research was chosen with a deliberate consideration of the AG (Arjosari-Gadang) Line through the center of the crowd in Malang City, including schools and universities, as shown in Figure 1. The decision to select Gojek and Grab as online transportation mode is because they are the largest shareholders of ride hailing in Indonesia.

Method of sampling

Sampling is a crucial technique used in research studies to draw valid conclusions about a larger population. It involves selecting individual members or a subset of the population to make statistical inferences and estimate the characteristics of the entire population (Shabani et al., 2022). The process of sampling allows researchers to avoid bias and make accurate inferences about a larger population. One common sampling method is random sampling, which involves using randomization to ensure that every element of the population has an equal chance of being part of the selected sample (Shabani et al., 2022). Other sampling methods include stratified sampling, cluster sampling, and convenience sampling. It is important to carefully choose a sampling method to ensure that the sample accurately represents the population and leads to valid conclusions. The investigation was often conducted on a population as a whole, so a specific sample size was chosen that would allow this study to be completed. The sample consisted of 100 respondents that met several requirements, including utilizing the AG Line (Arjosari-Gadang) or online transportation. Convenience sampling is based on the accessibility of objects and the ease of acquiring them. Samples are drawn because the sample is located at the correct location and at the correct time.

Method of collecting data

Data collection is a crucial component of the research process, as it allows researchers to gather information related to their research question or topic that can be analyzed and interpreted (Sherry & Wang, 2022). The systematic process of gathering observations or measurements in research is
known as data collection, and it can be carried out using various methods. These methods can be broadly classified into two categories: qualitative and quantitative. Qualitative data collection methods, such as interviews, focus groups, and observations, involve gathering non-numerical data that provides in-depth insights into the research topic. In contrast, quantitative data collection methods, such as surveys and experiments, involve the collection of numerical data that can be analyzed using statistical methods to identify trends and relationships (Zhang et al., 2023). Regardless of the method used, it is crucial to ensure that the data collected is accurate, reliable, and relevant to the research question or topic at hand. In this study, it was employed a descriptive, analytical method with a quantitative approach. The data was collected via an open-ended interview that utilized a questionnaire that included five variables' dimensions: Reliability, Responsiveness, Assurance, Tangibility, and User Experience.

**Data analysis**

Multiple linear regression is a powerful statistical method used to predict the value of a dependent variable based on two or more independent variables (Singh et al., 2023). This method assumes that there is a linear relationship between the dependent variable and the independent variables, which means that changes in the independent variables are associated with proportional changes in the dependent variable. It is a useful tool for analyzing complex datasets with multiple variables and can provide valuable insights into the underlying relationships between variables. In summary, multiple linear regression is an essential statistical method that has numerous applications in various fields and is a valuable tool for data analysis and predictive modeling (Nasseri, 2018). In this study, the analysis of the data was conducted using the SPSS v25 program, which is used to perform Multiple Linear Regression. Before regression analysis was conducted, the data were first processed through instrumental tests (validity and reliability tests) and classical assumptions. The data analysis method employs multiple linear regression analysis in SPSS v25 (Shao et al., 2019). The mathematical expression of the regression equation is written in Equation 1.

\[
Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5
\]

(Eq. 1)

Where, \( Y \) represent the consumer decision, \( b \) symbolize the coefficient of each variable, and \( X \) represents the dependent variables. Sub-scripts 1, 2, 3, 4, and 5 shows the variables of reliability, responsive, assurance, tangibility, and user experience, respectively.

**RESULTS AND DISCUSSION**

**Respondent’s characteristics**

Respondent characteristics play a crucial role in survey research and data analysis. Understanding these characteristics is essential for survey designers to ensure that their questionnaires are comprehensible and effective (Šebjan & Tominc, 2015). For instance, survey designers should consider the mapping of questions and responses to minimize interactional problems and increase the quality of responses. Additionally, investigating the differences in respondent characteristics who use different devices for online surveys can help in better designing and targeting surveys that cater to the needs of the respondents (Shao et al., 2019; Zhang et al., 2023). Furthermore, knowing the demographic characteristics of the respondents, such as their age, gender, educational level, and health status, can provide valuable insights into their responses to specific questions. Therefore, understanding respondent characteristics is essential for survey designers to develop effective questionnaires that ensure accurate data collection and analysis. The characteristics of respondents that was employed in this study are summarized in Table 1.

**Table 1.** Summary of the characteristics of respondents that participated in this study.

<table>
<thead>
<tr>
<th>No</th>
<th>Characteristics</th>
<th>Number of respondents</th>
<th>Distribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;20</td>
<td>33</td>
<td>33</td>
</tr>
</tbody>
</table>
Classical assumptions test from SPSS results

Classical assumption test is a statistical method used in linear regression analysis to test the assumptions underlying the Ordinary Least Squares (OLS) model. The OLS model is widely used in regression analysis to estimate the relationship between a dependent variable and one or more independent variables. To test this assumption, a residual versus fitted value plot is created to visually inspect the distribution of residuals. Another important assumption is that the relationship between the dependent variable and the independent variables should be linear (Šebjan & Tominc, 2015). The classical assumption test aims to produce an F-count, which is then compared with a critical value to determine whether the linear relationship assumption holds. Classical test theory also assumes that an examinee's test score is the composite of a true score and random error, and validity addresses the true score by examining the relationship between the test scores and the criterion variable (Shvetsov et al., 2023).

The normality test using the Kolmorogov-Smirnov test in SPSS produced a result that was 0.079 greater than the p (sig) value of 0.05, as a result, it can be concluded that the data are normally distributed. The Kolmogorov-Smirnov test is a statistical tool used to determine whether a given set of data is normally distributed (Levashev et al., 2023). When conducting this test, the first step is to calculate summary measures to describe the data, such as the mean and standard deviation. Once this is done, the next step is to test the normality of the data using the Kolmogorov-Smirnov test. This test compares the distribution of the data to a normal distribution and calculates a p-value, which indicates the likelihood that the data is normally distributed. Conclusively, the Kolmogorov-Smirnov test is a powerful tool in determining whether a set of data is normally distributed, and it is widely used in various fields such as finance, engineering, and healthcare.

The SPSS program's multicollinearity test indicated that all of the tolerance values were greater than the default value of 0.10. Regarding the VIF value, it is below 10. As a result, it can be concluded that all of the variables satisfied the threshold for tolerance and the VIF value, thus there was no multicollinearity. Multicollinearity is a statistical phenomenon that occurs when two or more independent variables in a multiple regression model have a strong correlation with each other. In other words, it is a condition where the predictor variables are highly linearly related. This can lead to problems in the statistical analysis of the data, as it can make it difficult to determine the individual effects of each variable on the dependent variable. To check whether multicollinearity is present in a regression model, one can plot the correlation matrix of all the independent variables. Another way is to use statistical tests such as the Variance Inflation Factor (VIF) test, which measures the extent to which each independent variable is explained by the other independent variables in the model. If the VIF value is greater than 5 or 10, then multicollinearity may be present (Chen et al., 2014). Hence, detecting and dealing with multicollinearity is an important aspect of regression analysis to ensure accurate and reliable results.

The heteroskedasticity test with the Glejser test using the SPSS program yielded the result that the coefficient of each of the independent variables was insignificant (p>0.05) against the residual, as a result, it can be concluded that the data did not have heteroskedasticity issues. The Glejser test for heteroskedasticity is a statistical tool developed by Herbert Glejser in 1969 (Ilbeigi, 2019). This test is commonly used to detect the presence of heteroskedasticity in a regression model. Heteroskedasticity refers to a situation where the variance of the residuals is not constant across the range of the explanatory variable. The Glejser test works by regressing the residuals on the explanatory variable and transforming the residuals to get their absolute value.

The autocorrelation test using the run test with the SPSS v25 program yielded an asymptotic significance value of 0.841 > 0.05, as a result, it can be concluded that the data is free of autocorrelation. Autocorrelation test is a statistical analysis that measures the relationship between
observations at different points in time. It seeks to identify patterns or trends in the data by examining the correlation of the observations over successive time periods (Paikun et al., 2021). Autocorrelation is a characteristic of data in which the correlation between the values of the same variable is based on related objects. In other words, it measures the degree of correlation between variables over successive time periods. Most statistical tests assume the independence of observations, but autocorrelation shows that this assumption may not always hold true (Rakhmatulloh et al., 2021).

Multiple linear regression analysis by SPSS results

Multiple regression analysis is a statistical tool used to analyze data and make predictions. The objective of multiple regression analysis is to use independent variables, whose values are known, to predict the value of a single dependent variable. In other words, multiple regression analysis is used to determine the relationship between two or more independent variables and a dependent variable. This technique works by considering the values of multiple independent variables and predicting the value of one dependent variable (Silva et al., 2013). The use of multiple regression analysis allows businesses to make informed decisions by providing insights into the relationships between different variables. Multiple regression analysis is a statistical technique that allows researchers to examine the relationship between a dependent variable and multiple independent variables. In SPSS, carrying out multiple regression analysis is a straightforward process, and there are several resources available to help users interpret and report the results of their analysis. For example, a "quick start" guide is available that shows how to carry out multiple regression using SPSS Statistics and interpret the results of the test (Šebjan & Tominc, 2015). Additionally, step-by-step example analyses are available that cover SPSS output, checking model assumptions, and reporting. One example multiple regression analysis in SPSS examines the relationship between scores obtained by students on a test and their level of study, age, and gender. This analysis includes detailed footnotes to explain the output, making it easy for researchers to understand and report their results accurately. Table 2 contains the detailed results of multiple linear variable regression parameters. The regression equation can be mathematically expressed as written in Equation 2.

\[
Y = 7.26 + 0.026X_1 + 0.983X_2 + 0.033X_3 + 0.989X_4 - 0.005X_5
\]

(Eq.2)

With a \( R^2 \) value of 0.690, the equation above indicates that 69% of the variation in responses to the dependent variable (whether or not people in Malang prefer conventional or online transportation) is explained by the variables in the regression model, while the remaining 31% is attributed to other factors. The coefficient of determination is a statistical measure used to evaluate the accuracy of a model in predicting future outcomes (Silviana et al., 2019). More specifically, the coefficient of determination, commonly denoted as \( R^2 \), is used to analyze how differences in one variable can be explained by a difference in a second variable. In other words, \( R^2 \) represents the proportion of variance in the dependent variable that can be predicted from the independent variable (Dalanta et al., 2021). This measure is particularly useful in regression analysis to assess the goodness of fit of a model and to compare the accuracy of different models. A high value of \( R^2 \) indicates that the model can explain a large proportion of the variation in the dependent variable, while a low value suggests that the model may not be a good fit for the data.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>b</th>
<th>t-count</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant (b₀)</td>
<td>7.260</td>
<td>-8.677</td>
<td>&gt;0.0001</td>
</tr>
<tr>
<td>Reliability (X₁)</td>
<td>0.026</td>
<td>2.385</td>
<td>0.013</td>
</tr>
<tr>
<td>Responsive (X₂)</td>
<td>0.983</td>
<td>9.513</td>
<td>&gt;0.0001</td>
</tr>
<tr>
<td>Assurance (X₃)</td>
<td>0.033</td>
<td>1.921</td>
<td>0.049</td>
</tr>
<tr>
<td>Tangibility Asset (X₄)</td>
<td>0.989</td>
<td>2.154</td>
<td>&gt;0.0001</td>
</tr>
<tr>
<td>User Experience (X₅)</td>
<td>-0.005</td>
<td>-0.525</td>
<td>0.601</td>
</tr>
</tbody>
</table>
The F-test value associated with the coefficient was 33.928, and the level of significance was >0.0001. The model is generally accurate because all of the variables are significant and have a direct effect on the dependent variables. The following describes the results of the t-test regarding the effect of variables on variables:

**Reliability**

The value of the t-count is 2.358 > 1.68, and the significance value is 0.013 < 0.05. Based on this number, it can be inferred that the reliability variable has a significant impact on the decision of residents of Malang to choose transportation (conventional or online) during the COVID-19 pandemic. The greatest impact on this variable is the appropriateness of the price, as customers of the service are typically users on a daily basis, this knowledge facilitates the price of transportation. Online transportation is considered more effective in terms of distance because there is a mapping application or location that is provided by service providers or drivers in specific areas that are spread across all important areas (Li et al., 2021).

**Responsiveness**

The t-count of 9.513 is greater than 1.68, and the associated significance level is 0.000 less than 0.05. This suggests that the response variable has a positive and statistically significant effect on whether or not people in Malang use traditional or online modes of transportation during the COVID-19 pandemic. The greatest value is derived from the attributes of customer service that are quick responses to complaints in online transportation. The reputable company must demonstrate their ability to address customer complaints in order to provide a beneficial experience (Shao et al., 2019).

**Assurance**

The count of the t-value is 1.921 > 1.68, and the significance value is 0.049 < 0.05. This value indicates that the Assurance factor has a positive effect on the residents of Malang's decision to utilize conventional or online transportation during the COVID-19 pandemic. This effect is statistically significant. The greatest value in this variable is the guarantee of safety and health in online transportation, because the driver has received the vaccine and can be identified through the application. All of the driver's identities are known with certainty because the transportation management company has already conducted a verification process prior to partnering with transportation drivers (Zhang et al., 2023).

**Tangible asset**

The t-count is 2.154 greater than 1.68, and the significance value is 0.000 less than 0.05. The value indicates that the Tangible Asset variable has a bearing on people in Malang City in regards to their selection of traditional transportation over online transportation during the COVID-19 pandemic. The driver's appearance in accordance with the protocol's health aspects is the most significant attribute on this variable, consumers believe that the protocol's health aspects are crucial in combating the spread of COVID-19. The effects of the COVID-19 pandemic on online transportation are that all providers of the service have regulations that mandate that consumers and drivers always practice health protocols (Chen et al., 2014).

**User experience**

The t-count is -0.525, which is less than 1.68, and the significance value is 0.601, which is greater than 0.05. Based on this number, it can be inferred that the User Experience variable has no significant effect on the decision of people in Malang that prefer transportation (conventional or online) during the COVID-19 pandemic.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.690</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-count</td>
<td>33.928</td>
<td>&gt;0.0001</td>
<td></td>
</tr>
<tr>
<td>F-table</td>
<td>2.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-table</td>
<td>1.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONCLUSION

The transportation condition in Malang City has been the subject of research, with one study proposing to support information about public transportation services in the city. This policy encourages people to participate in activities such as working from home and school during the COVID-19 pandemic. These policies include reducing the capacity of public transportation to comply with physical distancing guidelines, requiring the use of masks for all passengers. Respondent characteristics play a crucial role in survey research and data analysis. Understanding these characteristics is essential for survey designers to ensure that their questionnaires are comprehensible and effective. Another concern is that the availability of public transportation facilities is being limited by a lack of conventional transportation as result of this policy. The reliability variable has a significant impact on the decision of residents of Malang to choose transportation (conventional or online) during the COVID-19 pandemic. The greatest impact on this variable is the appropriateness of the price, as customers of the service are typically users on a daily basis. Online transportation is considered more effective in terms of distance because there is a mapping application or location that is provided by service providers or drivers in specific areas. The model is generally accurate because all of the variables are significant and have a direct effect on the dependent variables. The greatest value in this variable is the guarantee of safety and health in online transportation. All of the driver's identities are known with certainty because the transportation management company has already conducted a verification process prior to partnering with transportation drivers. The effects of the COVID-19 pandemic on online transportation are that all providers of the service have regulations that mandate consumers and drivers always practice health protocols. The F-test value associated with the coefficient was 33.927, and the level of significance was >0.0001. The following describes the results of the t-test regarding the effect of variables on variables: Reliability The value of the t-count is 2.358 > 1.68, and the significance value is 0.013 < 0.05. Responsiveness The t-count of 9.513 is greater than 1.68, and the associated significance level is 0.0001 which is less than 0.05.

ACKNOWLEDGEMENT

Authors are thankful to the Department of Civil Engineering, Universitas Narotama Surabaya for the long-lasting support and assistance until this research nicely finished.

REFERENCES


