CHAPTER III
RESEARCH METHODOLOGY

3.1 Research Design

In this research, a quantitative methodology is chosen because it offers a clear, organized framework for exploring hypotheses and analyzing data. The study is categorized as causal research, also known as explanatory research, which focuses on determining whether a change in an independent variable can result in a change in a dependent variable.

Descriptive Research aims to provide a comprehensive account of a particular situation or phenomenon by systematically describing its characteristics and patterns. This type of research focuses on documenting what is happening, such as describing the demographics of a population, identifying trends, or detailing the features of a phenomenon. It does not attempt to influence outcomes or determine causality but instead seeks to present an accurate picture of the present state of affairs (Sugiyono, 2020).

Quantitative Research, on the other hand, is centered on collecting and analyzing numerical data to uncover patterns, test hypotheses, and make predictions. This method involves the use of structured tools such as surveys, experiments, or statistical analysis to gather measurable data, which is then used to quantify relationships and phenomena. It is characterized by its emphasis on objective measurements and statistical techniques to ensure that findings are reliable and can be generalized to larger populations (Sugiyono, 2020).

Causal Research also known as Explanatory Research, seeks to identify and establish cause-and-effect relationships between variables. Unlike descriptive research, which describes phenomena, or quantitative research, which quantifies them, causal research aims to determine whether a change in one variable (the independent variable) directly causes a change in another variable (the dependent variable). This type of research often employs experimental designs, such as randomized controlled trials, to isolate variables and assess causation, rather than just correlation. Together, these research methodologies offer a range of approaches for understanding different aspects of phenomena, from describing current conditions and quantifying variables to exploring causal relationships and predicting future outcomes (Sugiyono, 2020).

3.2 Population and Sample

3.2.1 Research Object

The location of the research object that will be used is Lazada E-Commerce which is in Medan City, North Sumatra. The research period is predicted starts from January 2025 to April 2025.

3.2.2 Population

According to Qamar and Rezah (2020), population is a group of research objects that are the target of research observations. However, because sometimes the population is too large, the method requires a technique that does not need to

be observed or observed in its entirety, by using a sample or sampling technique. In the context of research, the population consists of all individuals who are eligible to participate in a study based on specific inclusion and exclusion criteria established by the researcher. The research population that will be used in this research is all consumers who using Lazada E-commerce application which total is unknown.

3.2.3 Sample

According to Endra (2019), sample is a portion of the number and characteristics possessed by the population or a small portion of the population members taken according to certain procedures so that it can represent the population. In experimental research, a sample refers to the group of subjects who participate in the study's experimental procedures. This group is selected to test hypotheses and measure outcomes in a controlled setting, with the findings used to infer results for the larger population.

Because of the unknown population, Lemeshow formula are being used to determine the number of samples in the research. The research sampling method are using Lemeshow formula because the population is unknown, that's why in this research where the population number is unknown, the sampling using the Lemeshow formula, therefore the Lemeshow formula is used:

$$n = \frac{Z^2{}_{1-\propto/2}\ P(1-P)}{d^2}$$

Information:

n = Total sample

z^2 1-α/2 = Z is the score at the 1-α/2 confidence level

P = Proportion Estimation

d = Precision used

Because the population is unknown. Then a table is needed level of confidence to determine the size of the research sample. There are 3 levels Consumer Attitude / confidence that can be used, namely 90% (1.645), 95% (1960), and the highest is 99% (2.576). Next, the researcher chose a P value of 0.5 in determining the number of samples. Choosing 0.5 for P in the formula for sample size will always provide enough observations. Using a P value of 0.5 is sufficient to meet the requirements for determining the sample size. The precision used is 0.1 (d). Based on the Lemeshow formula, we get calculation as follows:

$$n = \frac{1{,}96^2 \cdot 0{,}5\,(1 - 0{,}5)}{0{,}1^2}$$

$$n = \frac{1{,}96^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}1^2}$$

$$n = \frac{3{,}8416 \cdot 0{,}25}{0{,}01}$$

$$n = 96{,}04$$

(Revision) This study uses accidental sampling, a non-probability sampling method in which respondents are selected based on availability and willingness to participate. Although it was mentioned during the explanation that respondents should be Lazada users, no formal selection or filtering process was applied. The questionnaire was distributed openly, and responses were accepted from anyone willing to participate. The mention of “criteria” was a misstatement and does not reflect the actual sampling process. Because the result of the population is 96,04 since there population cannot be decimal so the 0,4 converted it into 1, and the final result for the population will be 97.

3.3 Data Collection Method

According to Syawaludin (2019), documentation study is a data collection technique that is not directly aimed at research subjects and is needed as a separate stage, namely a preliminary study to understand various theories and better capture new phenomena that are developing in the field or in society. According to Sholihah (2020), library study activities include collecting library sources and are often the longest in research. The most important research library sources are data sources from research reports which are usually published in scientific journals, theses, dissertations, or in scientific seminars. Other library sources such as textbooks, the internet, encyclopedias, and others can help as additional information. According to Magdalena (2020), a questionnaire is a collection of written questions to be answered by a group of research respondents.

3.4 Operational Definition and Variable Measurement

3.4.1. Operational Definition

According to Ardyan et al. (2023), research variables are the focus of a study, which can be either independent or dependent variables. When determining research variables, factors such as alignment with research problems, research objectives, potential benefits, and resource availability must be considered. The operational definition of a variable is a detailed procedure for measuring or manipulating variables, explaining how theoretical concepts will be translated into empirical data.

1. Independent Variable (X)

   Based on this research the independent variables are Purchase Experience (X_1), Ease of Use (X_2), and Application Quality (X_3).

2. Dependent Variable (Y)

   Based on this research the dependent variable is purchase decision (Y).

The following are some operational variable definitions of research variables that will be used in this study:

Table 3.1. Operational Variable Definition

Variable	Indicator	Measurement
Purchase Experience (X_1)	Previous purchasing experience	Likert Scale
	Decision to repurchase from an online shop	
	A sense of comfort when using a shopping app	
Ease of Use (X_2)	Easy to use	Likert Scale
	Easy to learn	
	Eary to find what is being searched for	
Application Quality (X_3)	Reliability	Likert Scale
	Performance	
	Durability	
Repurchase Intention (Y)	Consumer satisfaction	Likert Scale
	Product and service quality	
	Previous purchasing experience	
	Value obtained in previous transaction	

Source: Prepared by the writer (2025)

3.4.2 Variable Measurement

According to Riyanto and Hatmawan (2020), the Likert scale is a scale model that is widely used by researchers to measure attitudes, opinions, perceptions or other social phenomena. The Likert scale that is often used is the Likert scale with 5 categories, namely:

1. Strongly Disagree
2. Disagree
3. Neutral

4. Agree

5. Strongly Agree

3.5 Data Analysis Method

3.5.1 Research Instrument Test

3.5.1.1. Validity Test

According to Santi and Sudiasmo (2020), the validity test is used to test research instruments in questionnaires on variables with the aim of making the statement items suitable for use in research. The validity test is carried out by comparing rcount with rtable. Meanwhile, the value of rcount can be seen in the Pearson Correlation in the output with SPSS. Then, to make a decision, if rcount > rtable then the discussion variable can be said to be valid.

According to Priyatno (2020), before the questionnaire is distributed to the actual research sample, a pretest is usually conducted involving 30 respondents. This pretest is essential to identify whether the items in the questionnaire are understandable, consistent, and able to accurately capture the dimensions of each variable. The use of 30 respondents follows the general rule of thumb in instrument testing, as this number is considered sufficient to detect potential weaknesses in the instrument design. The feedback and data collected from this pretest are then analyzed for validity, and necessary revisions or improvements can be made before the final questionnaire is disseminated to a larger population. This step ensures higher reliability and validity of the instrument in the main study.

3.5.1.2. Reliability Test

According to Priyatno (2020), reliability tests are used to determine the regularity or consistency of measuring instruments which usually use questionnaires. This means whether the measuring instrument will get consistent measurements if the measurements are repeated. The method often used in research to measure range scales is Cronbach Alpha > 0.6 the it be concluded reliable.

3.5.2 Descriptive Statistics

According to Sugiyono (2020), descriptive statistics provide a description or description of data seen from the average value (mean), standard deviation, variance, maximum, minimum, sum, range, kurtosis and skewness (skew of distribution). This analysis is a descriptive technique that provides information about the data held.

1. Mean

   The mean is the average value of a data. The average value is the quotient between the total value and the amount of data processed.

2. Median

   The median is a value that divides the area of the frequency histogram into equal parts. The median is used to find the middle value of the total score of all the answers given by the respondents, which are arranged in the distribution of the data. How to determine the median on a single data is quite easy. We can sort it from smallest to largest data.

3. Mode

   To determine the mode, usually for single data, a single data frequency table is made to make it easier to determine the frequency of each data, then find the data with the largest frequency.

4. Variance

   The variance is the average of the squared differences, also known as the standard deviation, of the mean. Simply put, variance is a statistical measure of how scattered the data points are in a sample or data set. In addition to the mean and standard deviation, the variance of a sample set allows statisticians to understand, organize, and evaluate the data they collect for research purposes.

5. Standard Deviation

   Standard deviation or standard deviation is usually taught in statistics to measure the degree of similarity or closeness in a group. The standard deviation is generally used by statisticians or people involvedin statistics to find out whether the sample data taken is representative of the population.

3.5.3 Classical Assumption Test

3.5.3.1 Normality Test

According to Marsam (2020), the normality test can be done in 2 ways, namely with a histogram graph and normal probability plot of regression. The following is the basis for the decision:

1. If the data spreads around the diagonal line and follows the direction of the diagonal line, the histogram graph and the normal probability plot of

regression show a normal distribution pattern, so the regression model meets the assumption of normality.

2. If the data spreads far along the diagonal line and/or does not follow the direction of the diagonal line, the histogram graph of the normal probability plot of regression does not show a normal distribution pattern, so the regression model does not meet the assumption of normality.

According to Supriadi (2020), the purpose of the normality test is to test whether in a regression model, the dependent variable and independent variable or both have a normal distribution or not. A good regression model is the residual value having a normal or close to normal data distribution. One of the methods used to test normality is to use the Kolmogorov Smirnov test with a residual value of > 0.05, then the assumption of normality is met.

According to Enterprise (2020), one of the normality tests is using the Kolmogorov Smirnov with the following criteria:

1. If the value of Asymp. Sig. > 0.05, the data is normally distributed.
2. If the value of Asymp. Sig. < 0.05, the data is not normally distributed.

3.5.3.2 Multicollinearity Test

According to Priyatno (2020), multicollinearity is a condition in which a regression model finds a perfect or near perfect correlation between independent variables, where a good regression model should not have a perfect or near perfect correlation between the independent variables. According to Ghodang (2020), the multicollinearity test is used to see the

relationship between the independent variables so that the simple linear regression test does not use a multicollinearity test because the simple regression test only has one independent variable. The basis for decision making on the multicollinearity test is as follow:

1. There is no multicollinearity if the tolerance value is greater than 0.1 and the VIF (Variance Inflation Factor) value is less than 10.
2. Multicollinearity occurs if the tolerance value is less than 0.1 and the VIF (Variance Inflation Factor) value is greater than or equal to 10.

3.5.3.3 Heteroscedasticity Test

The heteroscedasticity test aims to test whether in the regression model there is an inequality of variance from the residuals of one observation to another. The measurement method uses a Scatterplot where if there is a certain pattern, such as dots that form a certain regular pattern, then identifying heteroscedasticity has occurred and vice versa if there is no clear pattern, and the dots spread above and below the number 0 on the Y axis , then there is no heteroscedasticity (Riyanto and Hatmawan, 2020).

3.5.4 Multiple Linear Regression Analysis

According to Priyatno (2020), multiple regression analysis is used to determine whether there is a significant partial or simultaneous influence between two or more independent variables on one dependent variable. The purpose of using multiple regression analysis in this research is to ascertain the extent of the

influence of independent variables on the dependent variable, which can be calculated using the following formula:

$$\mathbf{Y = a + b_1X_1 + b_2X_2 + b_3X_3 + e}$$

Information:

Y = Repurchase Intention

a = Constant

X_1 = Purchase Experience

X_2 = Ease of Use

X_3 = Product Quality

3.5.5 Hypothesis Test

3.5.5.1 T-test

According to Arvianti and Anggrasari (2020), the t-test or partial test is the hypothesis used in testing the coefficients of the partial regression model. The decision-making criteria are:

1. Ho will be rejected if $t_{count} > t_{table}$ or the significance value is < 0.05.
2. Ho will be accepted if $t_{count} < t_{table}$ or the significance value is > 0.05.

3.5.5.2 F-test

According to Mulyono (2020), the F test is used to find out whether the simultaneous independent variables have a significant effect on the dependent variable. The degree of confidence used is 0.05. If the calculated F value is greater than the F value according to the table, then the alternative hypothesis states that

all independent variables simultaneously have a significant effect on the dependent variable.

Criteria:

If Fcount > Ftable, Ho is rejected and Ha is accepted.

If Fcount <Ftable Ho is accepted and Ha is rejected.

Or:

If p < 0.05, Ho is rejected and Ha is accepted.

If p > 0.05 Ho is accepted and Ha is rejected.

3.5.5.3 Coefficient of Determination Test

According to Rachbini, et al (2020), the coefficient of assurance is one of the factual values that can be utilized to discover out whether there's a relationship between 2 factors. The esteem of the coefficient of assurance appears the percentage of variety within the subordinate variable that can be clarified by the coming about relapse condition and the rest (100% short the coefficient of determination) shows the size of the variety within the esteem of the subordinate variable which is impacted by variables exterior the condition.

Research Timeline

Table Research Timeline

No	Information	Month/Year															
		February 2025				March 2025				April 2025				May 2025			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Research Title Submission	■	■	■													
2	Proposal Preparation				■	■	■										
3	Proposal Guidance							■	■	■							
4	Data Collection										■	■					
5	Data Processing												■				
6	Skripsi Preparation													■	■		
7	Skripsi Guidance														■	■	
8	Skripsi Defence																■

Sources: Writer, 2024