CHAPTER III

RESEARCH METHODOLOGY

3.1. Research Design

The research model used in this study uses descriptive research using quantitative research. According to Syafrida (2022), quantitative research when viewed from a theoretical perspective, researchers who use this method adhere to the positivism study, namely their attention to the facts of the research. Quantitative research looks for the relationship of other variables with other variables, with the aim of answering the problem formulation of the initial hypothesis by means of statistical techniques.

This research will apply both descriptive and causal studies. According to Sumanto (2014) in Fadjarajani et al (2020) that descriptive research is research that tries to describe symptoms, events and events that occur at the present time, where researchers try to photograph events and events that are the centre of attention and then describe them as they are.

According to Sugiyono (2020), quantitative research is cause and effect, which means that everything occurs because of a cause. Causality refers to the connection between two variables, where a modification in one variable leads to a corresponding alteration or consequence in another variable. The initial variable is known as the independent variable, and it exerts an influence on the second variable known as the dependent variable. The research examines the relationship between 3 independent variables and 1 dependent variable. According to Fadjarajani et al

(2020), causal relationship is a relationship that is causal. effect. so here there are independent variables (variables that influence) and dependent (influenced). A causality between the independent variable (X), the effect of (X_1), (X_2) (X_3) and the dependent variable (Y).

3.2. Population and Sample

3.2.1. Population

According to Daniel and Agnes (2020), "A population refers to the entire group of members, while a sample represents only a portion or percentage of that group." The population in this research is all customers of Cobek Ayam Penyet whose size is unidentified because nature of business.

3.2.2. Sample

According to Syafrida (2022), sample refers to the collective score of individuals whose attributes are being examined. These individuals are referred to as units of analysis, which might include persons, institutions, or objects. The sampling that will be used in this research is convenience sampling. According to Fadjarajani et al (2020), convenience sample selection is a way of determining the sample by looking for subjects based on things that are pleasant or pleasing to the researcher. This sampling is chosen when lack of approach and does not allow for control bias. Subjects were selected based on their coincidental presence at the same location and time during data collection. The sample is taken in a non-systematic manner, which means it cannot be considered representative of either the source

population or the target population. The Lemeshow formula is utilized to ascertain the required sample size for this study, given the indeterminate or indefinite nature of the population. This study uses the Lemeshow formula to determine the sample size because the study population is not known with certainty. This formula is considered appropriate in conditions of unlimited or very large populations, and is able to provide a reliable minimum sample estimate by considering the level of confidence and the predetermined error limit. The Lameshow formula is as follows:

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

Information:

N = number of samples

Z = score at 95% confidence = 1.96

p = maximal estimation = 0.5

d = alpha (0.10) or sampling error = 10%

Through the formula above, the number of samples to be taken is

$$n = \frac{1.96^2 . 0.5(1-0.5)}{0.1^2}$$

$$n = \frac{3.8416 \, . \, 0.25}{0.01}$$

$$n = 96.04 = 97$$

From the results of the following formula, the writer will use 68 people as a sample of the study. In this research, the researcher used data from 105 respondents collected from the surveys answered by the customers at Cobek Ayam Penyet restaurant at Sun Plaza Medan.

3.3. Data Collection Method

There are a few data collection techniques used in this study. The researcher uses two types of data, namely:

a. Primary data

According to Fadjarajani et al (2020), primary data refers to information that researchers collect directly, firsthand. In contrast, secondary data typically refers to existing data that is already available.

1) Observation

According to Syafrida (2022), observation is a method of gathering data where researchers directly visit the field and observe the symptoms under study. Subsequently, the researcher can describe the problems that arise, which can be compared to data collected through other techniques like questionnaires or interviews. The results obtained from observation are then linked to existing theories and previous research.

2) Interview

According to Syafrida (2022), interview is a data collection technique by giving a number of questions related to research to predetermined sources.

3) Questionnaire

According to Syafrida (2022), the questionnaire is a compilation of question instruments designed to measure study variables. Data

collection using questionnaires is highly efficient, as respondents are just required to select responses provided by researchers.

b. Secondary Data

According to Sugiyono (2020), secondary sources are those that don't give data collectors information directly. For example, in the use of other persons or records. Documentation studies are used to collect secondary results. The business biography, number of employees, and details gleaned from theories, books, magazines, and the internet are only a few examples of the data. Secondary data in this study using data obtained from internet, journal, books and reviews from Google.

3.4. Operational Variable Definition and Variable Measurement

The process of measuring and classifying each variable for analysis is known as operationalization. The variables used in the study, together with their operational definitions, indicators, and measurement scale, are listed in the following table:

Table 3.1 Operational Variable Definition and Variable Measurement

Variable	Operational Definition	Indicator	Scale
Restaurant Authenticity (X1)	According to Levitt (2020), restaurant authenticity is the perception that a restaurant is truly representative of a given tradition or culture. It is an important concept to the restaurant industry as there are indications that it positively influences key outcome variables such as satisfaction and behavioral intentions.	Continuity, credibility, relevance, originality (Zhifeng and Ziwei, 2023)	5 point Likert Scale

Staff Knowledge (X2)	According to Sanz and Hovell (2021), staff knowledge is defined as the ability of an organization or team to maintain unique and critical knowledge, whether tacit (undocumented), explicit (written), or implicit (implied).	product knowledge, Restaurant knowledge, service quality, training and expertise (Rahmawati, 2021)	5 point Likert Scale
Menu Design (X3)	Mujahidin et al (2022), menu design is a condition created by the availability of goods in very varied quantities and types, resulting in many choices in the consumer shopping process. product completeness is the availability of all types of products offered to b	Visual appeal, clarity, Restaurant integration, menu organization. (Alotaibi & Alshiha, 2021),	5 point Likert Scale
Customer Loyalty (Y)	According to (Rahmawati et al., 2023), loyalty is the unwavering resolve to regularly buy a product or service again in the future, leading to recurring brand purchases, even in the face of external factors and marketing campaigns that could potentially alter behavior.	Customer satisfaction, repeat patronage, word-of-mouth recommendations, emotional connection, brand trust. (Y. Española et al., 2024)	5 point Likert Scale

3.5. Data Analysis Method

In order to evaluate the data, the author utilizes an array of analytical tools, which encompass statistical tests, validity assessments, reliability evaluations, descriptive statistics, classical assumption testing, regression equations, determination coefficients, and hypothesis testing.

3.5.1. Descriptive Statistics

Through data distribution analysis, descriptive statistics are utilized to explain each variable's properties. After adding together all of the data and

dividing it by the total number of observations, the mean is the average value. When data is arranged in order, the median is the the middle value, and the mode is the value that occurs most frequently in the dataset.

1. Mean

   The mean, which can be calculated by adding together all of the data and dividing by the total amount of data observed, is the average of the data. The following is the mean formula:

$$\bar{x} = \frac{\sum x}{n}$$

   Where:

   $\bar{x}$ = Mean

   x = specific values

   n = total amount of observations

2. Median

   The median, either in descending or ascending order, is the midpoint value of a dataset. With 50% of the values below the median and 50% above, it splits the data into two equal sections. Due to its greater resistance to extreme values than the mean, the median is frequently used when data contains outliers. These are the formula:

- For odd data:

  Me = $\frac{(n+1)}{2}$

- Even data:

  Me = $\frac{\frac{n}{2}+\left(\frac{n}{2}+1\right)}{2}$

Where:

x = specific values

Me = Median

N = total amount of observations

3. Mode

   The value that shows up the most in a dataset is called the mode. If no value repeats, a dataset may have multimodal (many modes), unimodal (one mode), or no mode at all. When determining the most prevalent value in a collection of observations, the mode might be helpful.

4. Standard deviation

   A dataset's degree of variation or dispersion is measured by its standard deviation. When the standard deviation is low, the data points are around the mean; when it is high, the data points are dispersed over a larger range. It is a crucial sign of how consistent or inconsistent the data is.

5. Variance

   The amount that a dataset's data deviates from the mean is measured by its variance. It indicates that every data point has the same value when the variance value is zero. When the variance is low, the data points are closely grouped around the mean; when the variance is high, the data points are widely dispersed from the mean.

3.5.2. Data Quality Test

1. Validity Test

According to Syafrida (2022), validity refers to the evaluation of research questions to determine the degree to which respondents comprehend the inquiries posed by researchers. If the results are invalid, it is possible that the respondents do not understand the questions. To test the validity of each question, the value of the question is correlated with the total value. The value of each question is expressed as X score and the total value is expressed as Y score. The following is a valid test calculation formula:

$$r = \frac{n\Sigma XY - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (x)^2)(n\Sigma Y^2 - (\Sigma Y)^2}}$$

Which:

r = Coefficient of Correlation

n = Amount of Sample

y = Customer Loyalty

x = restaurant authenticity, staff knowledge, menu design.

The t_{table} value used as a reference in this test can be seen in the appendix, which contains calculations and r_{table} value references based on the appropriate degrees of freedom. According to Haryono (2020), the criteria in determining validity of questionnaire are as follows:

a. If $r_{count} > r_{table}$, the statement is valid.

b. If $r_{count} < r_{table}$, the statement is invalid.

2. Reliability test

According to Syafrida (2022), reliability is testing the consistency of respondents' answers. Reliability is expressed in the form of numbers, usually as a coefficient, the higher the coefficient, the reliability or consistency of the respondent's answer is high. To measure the reliability of research data, namely by means of the Cronbach Alpha test. The Cronbach Alpha formula in question is as follows:

$$\alpha = [\frac{k}{k-1}][1-\frac{\sum \alpha b^2}{\alpha t^2}]$$

k = amount of the question

$\sum \alpha b^2$ = sum of the variance question

αt^2 = total of variance

According to Haryono (2020), this reliability test can be said pass the test when the minimum value of Cronbach Alpha at least 0.6., Questionnaires are reliable if:

1) Cronbach's alpha > 0.60 is reliable.
2) Cronbach's alpha < 0.60 is not reliable .

3.5.3. Classical Assumption Test

The following traditional assumption tests must be performed in order for the multiple linear regression model to generate unbiased estimates:

1. **Normality Test**

   According to Ghozali (2021), the goal of the normality test is to ascertain if the distribution of a dataset adheres to or approximates a normal distribution. A dataset is considered to follow a normal distribution if its values form a linear pattern when plotted on a graph,

and the residual data can be assessed by comparing it to this linear pattern using a normal probability plot. The normality test can also be done with the One-Sample Kolmogorov-Smirnov test. For the One-Sample Klomogorov-Smirnov test, residuals are normally distributed if the significance value is more than the level of confidence coefficient.:

1) Graph Analysis

   A histogram graph that contrasts observation data and distributions that are like the normal distribution is one of the simplest ways to see residual normality. However, this can be deceiving if you just look at the histogram, particularly if there are a limited number of samples. Standard Probability Plots, which compare the cumulative distribution of a normal distribution, are a more accurate tool. The diagonal line can be equivalent to the regular distribution. The foundation on which decisions are made is:

   a) When the data is distributed along the diagonal line and follows the path of the diagonal line or the histogram graph, demonstrating a normal distribution pattern, the regression model satisfies the assumption of normality.

   b) If the data considerably deviates from the diagonal line, does not follow the diagonal line's expected path, or the histogram does not show a consistent distribution pattern, the regression model violates the normality assumption.

2) Statistical analysis

The normality test with graphs can be deceiving if it does not appear normal physically, but objectively it is not. As a result, it is suggested that the graph test should be followed by statistical analysis. Kolmogorov-Smirnov is a non-parametric mathematical procedure that can be used to determine residual normality (K-S). The K-S test is performed by forming a hypothesis with a significant level > 0.05 so it was normal distribution for the residual.

2. **Multicollinearity Test**

   According to Syafrida (2022), multicolinearity test is to see whether there is a high relationship between independent variables. To detect multicolinearity using the Variance Inflation Factor (VIF) and Tolerance (TOL) methods. VIF is the variance inflation factor. If the VIF value is getting bigger, it is suspected that there is multicolinearity between independent variables or if VIF exceeds 10 then it can be concluded that there is multicolinearity. Thus TOL is getting closer to 0 then it is suspected that there is Multicolonierity and vice versa the value of TOL value is closer to 1, it is suspected that there is no multicolinearity.

3. **Heteroscedasticity Test**

   According to Syafrida (2022), heteroscedasticity is absent when there is no discernible pattern and the data points are evenly distributed above

and below the zero point on the Y axis. According to Ghozali (2021), there are two approaches to determining heteroscedasticity's presence or absence: Heteroscedasticity has occurred if a particular pattern appears, such as the dots creating a regular pattern that is wavy, broadening, and then narrowing. When the data points are evenly distributed above and below the zero point on the Y axis and there is no visible pattern, heteroscedasticity is absent.

According to Ghozali (2021), in addition to the scatterplot graph to determine the occurrence of heteroscedasticity, the Glejser test can also be employed to assess its validity, as statistical tests provide greater assurance of result accuracy. The Glejser test is performed by regressing the independent variable with the absolute value of Ut. If the p-value for the relationship between the independent variable and the absolute residual is more than 0.05, then heteroscedasticity is not an issue.

3.5.4. Multiple Linear Regression Test

The multiple linear regression analysis method is used to evaluate the relationship between several independent variables and one dependent variable (Hair et al., 2010). The regression model used in this study is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$

Where:

Y = Customer Loyalty

X_1 = Restaurant Authenticity

X_2 = Staff Knowledge

X_3 = Menu Design

β_0= Constant

β_1, β_2, β_3 = Regression Coefficient

e = Error term

3.5.5. Hyphotesis Test

a. The T-test (Partial)

The T-test which is a partial test, assesses the relationship between each independent variable and the dependent variable. The hypothesis are as follows:

- H0: There is no significant influence between Restaurant Authenticity (X_1) and Customer Loyalty (Y).
- H_1: There is a significant influence between Restaurant Authenticity (X_1) and Customer Loyalty (Y).
- H_0: There is no significant influence between Staff Knowledge (X_2) and Customer Loyalty (Y).
- H_1: There is a significant influence between Staff Knowledge (X_2) and Customer Loyalty (Y).
- H_0: There is no significant influence between Menu Design (X_3) and Customer Loyalty (Y).
- H_1: There is a significant influence between Menu Design (X_3) and Customer Loyalty (Y).

T-test formula is as follows:

$$t = \frac{\beta_i}{SE_{\beta i}}$$

The hypothesis is accepted if sig. < 0.05 (Ghozali, 2018). SPSS will be used to perform the T-test in order to ascertain the significance of each independent variable's impact on the dependent variable.

b. The F-test (Simultaneous)

The F-test which is called a simultaneous test, assesses the effect of each independent variable on the dependent variable simultaneously.

- H0: There is no significant simultaneous influence between Restaurant Authenticity (X_1), Staff Knowledge (X_2), and Menu Design (X_3) on Customer Loyalty (Y).
- H_1: There is a significant simultaneous influence between Restaurant Authenticity (X_1), Staff Knowledge (X_2), and Menu Design (X_3) on Customer Loyalty (Y).

The F-test formula is as follows:

$$F = \frac{(R^2/k)}{((1 - R^2)/(n - k - 1)}$$

where:

R^2 = Coefficient of determination

k = Number of independent variables

n = Number of samples

c. **Coefficient of Determination (R^2)**

The degree to which the independent variables account for changes in the dependent variable is indicated by the R2 (coefficient of determination). It shows how much of the variation in customer loyalty (Y) may be attributed to menu design (X_3), staff knowledge (X_2), and Restaurant authenticity (X_1).

The formula are as follows:

$$R^2 = 1 - \frac{\Sigma(Y_i - Y_i)^2}{\Sigma(Y_i - Y)^2}$$

Criteria:

- R^2 approaching 1 → Very good model
- Low R^2 → Model is less able to explain the dependent variable (Ghozali, 2018).
- A low R^2 value suggests that other unmeasured factors (e.g., pricing, location, and marketing strategies) may also contribute to customer loyalty, while a high R^2 value indicates that Restaurant Authenticity, Staff Knowledge, and Menu Design play a significant role in influencing customer loyalty at Cobek Ayam Penyet Restaurant. The R^2 value will be analyzed using SPSS to determine whether additional factors should be considered for future research.