CHAPTER III

RESEARCH METHODOLOGY

3.1 Research Design

The research design serves as the architectural blueprint for conducting the study, guiding the selection of research approaches, data collection methods, and analysis techniques (Creswell, 2020). In this study, a mixed-methods design is proposed, integrating both quantitative and qualitative elements to gain a comprehensive understanding of the research problem (Creswell, 2020). The quantitative approach involves the collection and analysis of numerical data, allowing for statistical testing and the identification of patterns in the data (Creswell, 2020). In this study, a structured questionnaire will be administered to a large sample of customers who have made purchases at Honda IDK 2. This method makes it possible to quantify and compare factors, which makes it easier to develop connections between client purchase decisions, word-of-mouth, personal selling, and promotions.

The quantitative approach offers several advantages. Firstly, it provides a broad perspective by capturing data from a large number of respondents, enhancing the generalizability of findings (Creswell, 2020). Secondly, it allows for statistical analysis, enabling the testing of hypotheses, identification of significant relationships, and quantification of effects (Creswell, 2020). Additionally, the structured nature of quantitative data collection facilitates efficient data analysis

and the ability to make inferences about larger population (Creswell, 2020). However, the quantitative approach also has limitations. It may overlook the depth and richness of individual experiences, as it primarily focuses on aggregate data and statistical patterns (Creswell, 2020).

3.2 Population and Sample

3.2.1. Research Location and Time

The research will be conducted at Honda IDK 2 dealerships located in Medan. Data collection will take place over a period starting from January 2025 to May 2025. This timeframe aims to capture a diverse range of customer experiences and perspectives within a defined period.

3.2.2. Population

The target population for this research is customers who are relevant and had ever purchased a Honda brand car. This includes individuals who have engaged in transactions involving vehicle purchases, after-sales services, or other dealership interactions. By focusing on recent customers, the study aims to gather timely and relevant insights.

3.2.3. Sample

According to Sujarweni (2023), a sample is a subset of the population that is used to represent the characteristics of the entire population. When the population is large, it is often impractical to collect data from every individual, due to

limitations such as time, cost, and resources (Ghozali, 2022). In such cases, to choose a representative subset of the population in these situations, researchers employ sampling procedures. The sample should fairly represent the population's characteristics, and the inferences made from it should be generalizable to the entire population. (Ferdinand, 2022).

Sampling techniques are crucial in research to collect data and draw conclusions (Ghozali, 2022). Probability sampling and non-probability sampling are the two primary categories of sampling procedures (Ghozali, 2022):

1. Probability Sampling

   With probability sampling, samples are chosen so that each person in the population has a known, non-zero chance of being chosen (Ghozali, 2022). This approach is based on random selection, which helps guarantee that the sample is representative of the overall population. Among the primary varieties of probability sampling are:

   a. Simple Random Sampling: Every individual has an equal chance of being selected (Ghozali, 2022).
   b. Stratified Sampling: The population is divided into distinct subgroups based on specific characteristics, and random samples are drawn from each stratum (Ghozali, 2022).
   c. Cluster Sampling: The population is divided into clusters, and entire clusters are randomly selected for inclusion in the sample (Ghozali, 2022).

2. Non-probability sampling

   Non-probability sampling involves selecting samples where not all members of the population have a chance of being selected, and the selection is often based on subjective judgment (Ghozali, 2022). Non-probability sampling comes in the following primary forms:

   a. Convenience Sampling: Samples are taken from a group that is easily accessible to the researcher (Ghozali, 2022).
   b. Purposive Sampling: Participants are selected based on specific characteristics or criteria relevant to the research question (Ghozali, 2022).
   c. Snowball Sampling: Existing study subjects recruit future subjects from among their acquaintances (Ghozali, 2022).

Even though a targeted population is the customer who had ever purchased a Honda brand car, but the research sampling method are still using Lemeshow formula because there is no data regarding the customer of Honda provided. Therefore, the population could determined as unknown. According to Lwanga & Lemeshow (1991), if the population of the research is hardly estimated, it is the safest choice to put P = 0.5. The precision of the true proportion used in this research is 10% as the number of populations is unknown. By using 95% of confidence level, the sample size for this research would require 96 respondents. The following Lemeshow formula can be used to calculate the necessary sample size.:

$$\boldsymbol{n = \frac{z^2 \times P(1 - P)}{d^2}}$$

In this formula:

N = Number of samples

Z = The confidence level, typically set at 95% (Z = 1.95).

P = Anticipated proportion of the target population (P = 0.5).

d/Moe = Absolute precision of a proportion (d = 0.1).

There are 3 levels of confidence that can be used, namely 90% (1.645), 95% (1.960), and 99% (2.576). In this research, the writer chooses 0.5 for P in the formula is sufficient to meet requirements to determine the sample size. The absolute precision used is 0.1 (d/Moe). For example, assuming a 95% confidence level and a 5% margin of error, the computation might look like this::

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

$$\boldsymbol{n = \frac{3.8416 \times 0.25}{0.01} = \frac{0.9604}{0.01} = 96.04 \approx 96}$$

Purposive sampling, also known as non-probability sampling, will be the sample technique employed. The research sample will be chosen based on the following criteria:

1. Domiciled in Medan
2. Aged Between 17 – 60 years old
3. Have owned a Honda car or have ever purchased a Honda car at Honda IDK 2 or outside Honda IDK 2

3.3 Data Collection Method

According to Jatmiko (2021), The first phase in the data collection process is the literature review. One technique for gathering data is library study, which focuses on looking through documents for information and data. The writer gathered an information that mention that there are two kind of data that is being used to support the research process:

1. Primary Data

   Primary data is a data that could be gathered through a questionnaire survey, direct observations or also data from researchers. In this Study, the data collection related to the problems studied was carried out by:

   a. Questionnaires

   A structured questionnaire was designed to collect quantitative data from the target population. The questionnaire included closed-ended questions with a five-point Likert scale, covering aspects such as sales promotions, personal selling, word-of-mouth influence, and customer purchase decisions. The questionnaire was pre-tested on a group of respondents to ensure reability and validity. It was then distributed to a larger sample of customers through online platforms and in-person interactions.

   b. Observations

   Direct observations were conducted in retail stores to gather insights into customer behaviour and interactions with sales promotions and personnel. Observational data provided contextual understanding and complemented the quantitative data collected through the questionnaire.

2. Secondary Data

In addition to primary data, Relevant secondary data was examined in addition to primary data in order to support the goals of the study and obtain new insights. Industry reports, market research studies, scholarly journals, the internet, and business websites were examples of secondary data sources. These resources included insightful data on industry trends in customer behavior, personal selling methods, and sales promotion tactics, all of which might serve as references for this research.

3.4 Operational Variable Definition and Variable Measurement

3.4.1 Operational Variable Definition

According to Sugiyono (2020), operational definitions of variables are definitions that explain how variables are measured or calculated. The measurement scale of the variable is an important aspect to consider. An operational definition of a variable is a concrete and specific way used to measure or define a variable in a research study. This definition specifies how an abstract variable is transformed into observable and measurable data.

According to Setiana (2021), operational definitions involve defining variables through observable characteristics to allow for accurate observation and measurement of phenomena. This type of definition provides a clear description of how theoretical concepts are translated into measurable and testable forms for research, linking abstract theories to concrete measurement methods.

There are two variables that the writer uses for this research:

a. Independent Variable (X)

A variable that is unaffected by another variable is called an independent variable. Word-of-mouth (WOM), personal selling (PS), and sales promotions (SP) are the independent variables (X) in this study.

b. Dependent Variable (Y)

A variable that is impacted by one or more independent variables is called a dependent variable. Customer Purchase Decision (CPD) is the dependent variable (Y) in this study.

There are some operational variable definitions of research that will be used in this study:

Table 3. 1 Operational Variable Definition

Variable	Indicator	Statement
Sales Promotion (X1)	Sales Volume and Growth	I am satisfied with the promotions and discounts offered by Honda IDK2.
		The sales promotions offered by Honda IDK2 are better than those offered by competitors.
		I have taken advantage of the sales promotions offered by Honda IDK2 in the past.
	Website Traffic	I frequently visit the Honda IDK2 website to check for new products and promotions.
		The Honda IDK2 website is easy to navigate and provides useful information.
		I have made a purchase from the Honda IDK2 website in the past.
	Customer Satisfaction	I am satisfied with the overall quality of the products and services offered by Honda IDK2.
		The customer service provided by Honda IDK2 is excellent.
		I have experienced any problems with the products and services offered by Honda IDK2.

Personal Selling (X2)	Sales Targets and Revenue Growth	The sales team at Honda IDK2 is knowledgeable about the products and services they offer.
		I have been satisfied with the sales experience at Honda IDK2.
		The sales team at Honda IDK2 is friendly and courteous.
	Long-Term Relationships	I think Honda IDK2 values their customers and strives to build long-term relationships with them.
		I have recommended Honda IDK2 to friends and family because of their excellent customer service.
		I think the customer service provided by Honda IDK2 is better than that of their competitors.
	Customer Feedback and Testimonials	I think Honda IDK2 values customer feedback and uses it to improve their products and services.
		I have seen testimonials from other customers about their positive experiences with Honda IDK2.
		I would be willing to provide a testimonial about my positive experience with Honda IDK2.
Word of Mouth (X3)	Social Media Engagement	I have interacted with Honda IDK2 on social media.
		I think the social media content provided by Honda IDK2 is useful and informative.
		I would share content from Honda IDK2 on my own social media channels.
	Customer Testimonials and Reviews	I think the reviews from other customers are accurate and reflect my own experience with Honda IDK2.
		I would recommend Honda IDK2 to friends and family based on the positive reviews I have read.
		I think the reviews from other customers are helpful in making a purchasing decision.
	Brand Reputation and Perceived Quality	I think Honda IDK2 has a strong brand reputation.
		I think the products and services offered by Honda IDK2 are of high quality.

		I would pay a premium for products and services from Honda IDK2 because of their strong brand reputation
Customer Purchase Decision (Y)	Purchase Frequency	The quality of products and services offered by Honda IDK2 influences my decision to make frequent purchases.
		I consider Honda IDK2 as my preferred choice for automotive needs due to their excellent customer service.
		I am satisfied with the overall experience of purchasing from Honda IDK2, which motivates me to make repeat purchases.
	Purchase Pathways	The sales team at Honda IDK2 is knowledgeable and helpful, which makes me feel comfortable purchasing from them in person.
		I consider the reviews and ratings from other customers when deciding which purchase pathway to use for Honda IDK2 products.
		The availability of multiple purchase pathways, including online and offline channels, makes it easy for me to purchase from Honda IDK2.
	Customer Acquisition Sources	A friend or family member recommended Honda IDK2 to me, which influenced my decision to purchase from them.
		I attended an event or promotion hosted by Honda IDK2, which gave me the opportunity to learn more about their products and services.
		I saw an advertisement for Honda IDK2 on television or print media, which sparked my interest and led me to visit their website or store.

Source: Prepared by The Writer (2025)

3.4.1 Variable Measurement

This research will be using the five-level Likert Scale in the questionnaire to be distributed. Respondents will be required to respond to given statements based on the five options based on the rate shown below. It is to know the perception of the respondents towards the statements given.

Table 3. 2 Five-Level Likert Scale

Response	Scale
Strongly Agree	5
Agree	4
Neutral	3
Disagree	2
Strongly Disagree	1

Source: Prepared by the Writer (2025)

3.5 Data Analysis Method

3.5.1 Test of Research Instrument

According to Dewi, et al. (2023), Before the questionnaire is given or distributed to respondents either directly or online in the form of a google form, it is essential to establish its validity and reliability through a pre-test that will be used in the study. For instance, if the research sample consists of 100 participants, a pre-test can be conducted with 30 participants or a separate group with similar characteristics to the research sample. If the questionnaire demonstrates satisfactory validity and reliability, it can then be distributed to the entire sample needed in the study.

3.5.1.1 Validity test

One essential tool for guaranteeing the precision and consistency of the data gathered for research is the validity test. According to (Sulistiyani & Widyatama, 2023), the validity test is conducted by comparing the calculated correlation coefficient (r_{count}) with the critical value (r_{table}). The r_{count} value can be obtained through the Pearson Correlation analysis using SPSS (Kusuma, 2022). This research conducts a pre-test to 30 respondents with 5% of significance level (α). The accepted value of r_{table} for 30 respondents is 0.361 (Saraswati & Wirayudha,

2022). The instrument may be regarded as legitimate if the r_{count} value exceeds the r_{table} value. In this particular study, the instrument is considered legitimate if the r_{count} number is more than 0.361, which means that it measures the intended construct successfully.

3.5.1.2 Reliability test

When assessing the regularity and consistency of measurement tools, especially those that use questionnaires, reliability tests are crucial. According to Kusuma (2022), the reliability test assesses whether the measuring instrument yields consistent measurements when repeated. The Cronbach Alpha test is a frequently used technique to assess reliability; an instrument is considered reliable if its value is greater than 0.6. By conducting a reliability test, researchers can ensure that their measuring instrument is consistent and accurate, providing trustworthy results.

3.5.2 Descriptive statistic

According to Cohen et al. (2020), descriptive statistics are the type of statistics that only describe, present, and summarize data, with no attempt to do prediction or inference. Measures of dispersion, central tendency, and frequencies are examples of descriptive statistics. The mean, median, and mode are the three ways to analyze data around the center of a data collection.

a. Mean

   By adding up all of the values and dividing by the total number of values in the set, the mean—the average of a set of values—is determined. It gives

each value equal weight and is sensitive to excessive values. The most common way to measure central tendency is with the mean.

b. Median

When a dataset is arranged from smallest to greatest, the median is the midway value. It works well with skewed distributions or ordinal data and is less impacted by extreme values than the mean.

c. Mode

The value that occurs most frequently in a dataset is called the mode. One mode (unimodal), two modes (bimodal), or multiple modes (multimodal) can be present in a dataset. For nominal or categorical data, the mode is helpful.

d. Variance

The average of the squared deviations between each value and the mean is known as variance. It is helpful for comprehending the variability around the mean and offers insight into the distribution of the data. Variance is measured in square units and is never negative.

e. Standard Deviation

The variance's square root, or standard deviation, is expressed in the same units as the original data. It shows the average separation between the mean and each data point. Because it is simpler to understand and uses the same units as the data, standard deviation is frequently chosen over variance.

3.5.3 Classical assumption test

One essential stage in guaranteeing the precision and dependability of regression analysis is the classical assumption test. According to Gunawan (2020), this classical assumption test is to provide certainty that the regression equation obtained has accuracy in estimation, is unbiased, consistent and precise. This test includes a number of essential elements that together ensure the validity of the regression model, such as the residual normality test, multicollinearity test, heteroscedasticity test, linearity test, and autocorrelation test. By carrying out these tests, researchers can verify that their regression equation satisfies the required presumptions, giving them a strong basis on which to base their deliberations and findings. This study employs five traditional assumption tests: the heteroscedasticity test, the multicollinearity test, and the normality test.

3.5.3.1 Normality test

According to Santoso (2022), The normality test is a crucial step in ensuring that the residuals are distributed normally and independently, which is a fundamental assumption in regression analysis. To verify data normality, this study will use both graphical and statistical methods through regression calculations. Researchers can use the histogram or normal probability plot, which shows a bell-shaped curve or a diagonal line to suggest a normal distribution, to evaluate the normality of the data. According to Sulistiyani (2022), the Kolmogorov-Smirnov test also can be used to evaluate normality, with the following criteria:

1. 1. The data is deemed regularly distributed if the asymptotic significance (Asymp. Sig.) value is higher than 0.05..
2. 2. The data is deemed non-normal if the value is less than 0.05..

3.5.3.2 Multicollinearity test

When two or more independent variables in a regression model have a strong or perfect correlation, this is known as multicollinearity. According to Sulistiyani (2023), If perfect multicollinearity exists, the regression coefficients of the independent variables become indeterminate, and the standard error values become infinite, rendering the model unstable. Even if the multicollinearity is imperfect but high, the regression coefficient of the independent variable can still be estimated, but with a high standard error value, which compromises the precision of the estimate. Researchers frequently employ a tolerance cutoff value of less than 0.1 or a Variance Inflation Factor (VIF) value greater than 10 to detect multicollinearity.

3.5.3.3 Heteroscedasticity test

When the dependent variable's variability varies across the values of the independent variables, this is known as heteroscedasticity. It violates the assumption of homogeneity of variance in regression analysis. According to Widyatama, A. (2020), heteroscedasticity can be detected using scatter plots image, residual plots, or formal statistical tests like the Glejser test, Park test, Breusch-Pagan test or White's test.

1. Scatterplot Image

   A scatterplot image can be used to visually inspect for heteroscedasticity, with the following criteria indicating the absence of heteroscedasticity: There should be no discernible patterns or concentrations among the data points, which should be dispersed randomly above and below the horizontal axis. The regression criteria that do not occur heteroscedasticity if:

   a. Data points are spread above and below or around the number 0
   b. Data points do not gather only above or below.
   c. The distribution of data points should not form a wavy pattern that widens then narrows and widens again.
   d. The distribution of data points is not patterned.

2. Park Test

   The Park test is another statistical test used to detect heteroscedasticity in regression analysis. According to Kusuma (2023), the Park test involves regressing the logarithm of the squared residuals on the independent variable. The absence of heteroscedasticity in the regression model is indicated if the independent variable's significance level is higher than 0.05. There is heteroscedasticity present, however, if the significance level is smaller than 0.05.

3.5.3.4 Linearity Test

According to Santoso (2022), the linearity test is a statistical method used to evaluate whether the relationship between the independent and the dependent variables is linear. This test is crucial because it uses particular criteria to determine whether the relationship between two variables is linear or non-linear:

1. The relationship is deemed linear if the computed significance of linearity is less than the selected significance threshold of 0.05.
2. The relationship is considered non-linear if the computed significance of linearity is higher than 0.05.

3.5.3.5 Autocorrelation Test

According to Ghozali (2021), the autocorrelation test is a statistical method used to detect the presence of autocorrelation in residuals. The Durbin-Watson test is the more popular of the two approaches used to conduct the autocorrelation test. The Run test is also frequently employed. The Run Test is also a helpful tool for detecting autocorrelation, particularly when the Durbin-Watson test is equivocal. This is because the Durbin-Watson test has a flaw in that it cannot definitively determine whether autocorrelation is present. The following serves as the foundation for the decision-making process for the Run Test, a non-parametric statistical technique for determining if residuals exhibit autocorrelation:

1. If the significance level (Sig) value is less than 0.05, it indicates the residuals are not random, meaning that there is an autocorrelation between the residual values.

2. Where if the significance level (Sig) value is greater than 0.05 can be assumed that the residuals are random, which indicates there is no autocorrelation between the residual values.

3.5.4 Multiple Linear Regression Analysis

A statistical method for predicting a dependent variable's value based on the values of several independent variables is multiple linear regression analysis. By incorporating more than one predictor, it expands upon the basic linear regression model. The model's general shape is:

$$\boldsymbol{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + e}$$

where β0 is the intercept, β1, β2, ..., βp are the regression coefficients, ε is the error term, Y is the dependent variable, and X1, X2, ..., Xp are the independent variables.

For this paper, the multiple regression formula would be seen as follows:

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

Where:

Y = Consumer Purchase Decision (CPD)

β0 = Constant (α)

X1 = Sales Promotion (SP)

X2 = Personal Selling (PS)

X3 = Word of Mouth (WOM)

β1 = Regression weight between SP and CPD

β2 = Regression weight between PS and CPD

β3 = Regression weight between WOM and CPD

e = Error disturbances

3.5.5 Hypothesis test

According to Cohen et al. (2020), the activities in hypothesis testing, which include formulating hypothesis, measuring the variables and assess the relationship between them, and infer probability of finding the relationship of those variables if there is no relationship at all (null hypothesis). According to the hypothesis test, this study employs a number of tests to evaluate the ways in which sales marketing, personal selling, and word-of-mouth can all concurrently and partially impact the dependent variable of a customer's purchase choice. Three distinct forms of hypothesis testing will be used in this study: the coefficient of determination test (Adjusted R2), the partial test (T-test), and the simultaneous test (F-test).

3.5.5.1 Partial Hypothesis Test (T-Test)

According to Kusuma (2022), the t-test is a statistical method used to examine the partial influence of each independent variable on the dependent

variable. The following below are the requirement and criteria to analyze the results of a t-test, citing from Raharjo (2023):

1. Observe the significance level (α) of the independent variable:
   a. The null hypothesis (H0) is rejected and the alternative hypothesis (Ha) is accepted if the significance level is less than 0.05, indicating that the independent variable has a substantial partial influence on the dependent variable.
   b. If the significance threshold is higher than 0.05, it means that Ha is rejected and H0 is accepted, indicating that there is no discernible partial influence of the independent variable on the dependent variable.
2. Compare the value of t-count and t-table:
   a. H0 is rejected and Ha is accepted if t-count is larger than t-table or less than -t-table, indicating that the independent variable has a strong partial influence on the dependent variable.
   b. H0 is accepted and Ha is rejected if t-count is less than t-table or more than -t-table, indicating that the independent variable has no discernible partial influence on the dependent variable.

3.5.5.2 Simultaneous Hypothesis Test (F-Test)

According to Kusuma (2022), a statistical technique for assessing the concurrent relationship between all independent factors and the dependent variable is the F-test. The P-value, which shows that adding independent variables to the model enhances model fit, must be smaller than the significance level in order to

assess if the regression model fits the data more precisely. The following below are the requirement and criteria to analyze the results of a F-test, citing from Raharjo (2023):

1. Through the use of the significance level (α):
    a. The null hypothesis (H0) is rejected and the alternative hypothesis (Ha) is accepted if the significance threshold is less than 0.05, indicating that all of the independent factors have a substantial simultaneous influence on the dependent variable..
    b. If the significance threshold is higher than 0.05, then Ha is rejected and H0 is accepted, indicating that none of the independent factors significantly affects the dependent variable at the same time.
2. By comparing F-count and F-table:
    a. If F-count exceeds F-table, it means that Ha is accepted and H0 is denied, indicating that the independent factors significantly affect the dependent variable at the same time.
    b. H0 is accepted and Ha is refused if F-count is smaller than F-table, indicating that the independent factors do not significantly affect the dependent variable at the same time.

3.5.5.3 Coefficient of Determination Test (Adjusted R^2)

The coefficient of determination, also known as R-squared, is a statistical measure that estimates the proportion of the dependent variable's variation that is explained by the independent variables in a regression model (Kusuma, 2022).

According to Raharjo (2023), the R-squared value ranges from 0 to 1, with 1 being the ideal value, indicating that the independent variables perfectly explain the variation in the dependent variable. While a number around 1 shows that the independent factors nearly fully explain the variance in the dependent variable, a low R-squared value implies that the independent variables have a limited capacity to account for the variation in the dependent variable.

However, the R-squared measure has a limitation, as it can be misleading when multiple independent variables are included in the regression model (Santoso, 2022). According to Ghozali (2021), the R-squared value tends to increase when more independent variables are added to the model, even if they do not significantly contribute to explaining the dependent variable. The modified R-squared, which takes into account the sample size and the number of independent variables, might be used to address this problem. The updated R-squared value will either increase or decrease depending on how well the new independent variable explains the dependent variable. To get a more precise assessment of the model's fit in this study, the modified R-squared will be employed.