# CHAPTER 4
# METHODOLOGY

This chapter outlines the methods used in this research to analyse the proposed conceptual framework and the impacts of the Farmer Info application. First, the research design is discussed followed by a description of the study area and the process of data collection. In the second section, the methods for the qualitative and quantitative data analysis are described in detail.

## 4.1 Research Design

Numerous studies are focusing on the use and impact of ICTs in agriculture. However, research regarding the impact of smartphones or smartphone applications is limited. In Thailand studies on ICTs in agriculture are particular rare and for smartphones non-existing. Therefore this thesis can be seen as an exploratory study. According to Zikmund, Babin, Carr, and Griffin (2009), an exploratory research provides new insights and can provide the ground for further investigations. Exploratory research is not providing conclusive evidence but is the first step for further studies. In this case, research data and literature are limited and do not allow to generalise the findings, but they can contribute to the future development of the application and further studies. Although exploratory research mainly bases on qualitative methods to better understand certain situations or problems (Zikmund et al., 2009, pp. 54–55), a mixed method approach was chosen to provide a better analyzation of the impact of the Farmer Info application. "Impact" can thereby be defined as "[all] positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended" (OECD, 2010, p. 24).

Mixed methods combine the strengths of both, quantitative and qualitative methodologies. Bamberger (2012) identifies five reasons to use mixed methods, based on Greene (2005):

- Triangulation of evaluation findings**:**

*Enhancing the validity or credibility of evaluation findings by comparing information obtained from different methods of data collection (for example comparing responses to survey questions with what the interviewer observes directly). When estimates from different sources converge and agree this increases the validity and credibility of findings or interpretation. When different estimates are not consistent, the researcher explores further to understand the reason for the inconsistencies*

- Development:

*Using results of one method to help develop the sample or instrumentation for another*

- Complementarity:

*Extending the comprehensiveness of evaluation findings through results from different methods that broaden and deepen the understanding reached*

- Initiation:

*Generating new insights into evaluation findings through results from the different methods that diverge and thus call for reconciliation through further analysis, reframing or a shift in perspective*

- Value diversity:

*Incorporating a wider diversity of values through the use of different methods that themselves advance difference values. This encourages greater consciousness about the value dimensions of the evaluation (Bamberger, 2012)*

The analysation of a mobile application for agricultural information dissemination requires a comprehensive understanding of the design of the service as well as the surrounding environment of the service. As a result, a mixed method approach can guarantee the best results for a comprehensive evaluation.

The research, therefore, used qualitative and quantitative methods for data collection in a sequential design. The principal methodology for the study will be a survey using a structured questionnaire to collect data from farmers. In addition to the survey, qualitative methods will help to generate a broad understanding of the program and the environment in which it takes place.

### 4.2 Research Area and Participants

The quantitative part of this research took place in Chiang Rai Province in the northern, mountainous region of Thailand, while the qualitative interviews were divided between experts in the capital, Bangkok, and farmers in Chiang Rai Province (see Figure 11).

Thailand is a country in the heart of South-East Asia bordering Lao, Myanmar, Cambodia and Malaysia. Broadly it can be divided into four natural areas, the south, central and north-eastern areas as well as the northern region where the data collection took place. The north is characterised by a series of north-south ridges and narrow valleys with flat river basins (Aditto, 2011, p. 11). Chiang Rai is the northernmost province of Thailand and is divided into 18 districts with a flat region in the eastern part and a mountainous in the west (see figure 10). It covers a total area of 11,678 $km^2$ with an average elevation of 550 m and the highest point at 1998 m (Akber & Shrestha, 2013, pp. 110–111). The area is characterised by a wet season from May to October and a dry season from November to April with temperatures varying from 12.2°C to 34.9°C and 1704.8 mm annual rainfall (World Meteorological Organization, 2014). The total population of Chiang Rai was according to the National Statistical Office 1,194,933 in 2009 with 145,177 holdings engaged in agriculture covering an area of 2,598,525 rai resulting in an average farm size of 17.9 rai (2.864 hectare) (Akber & Shrestha, 2013; National Statistical Office of Thailand, 2014a). Rice production is dominating within the agricultural sector followed by field crops (National Statistical Office of Thailand, 2014a).

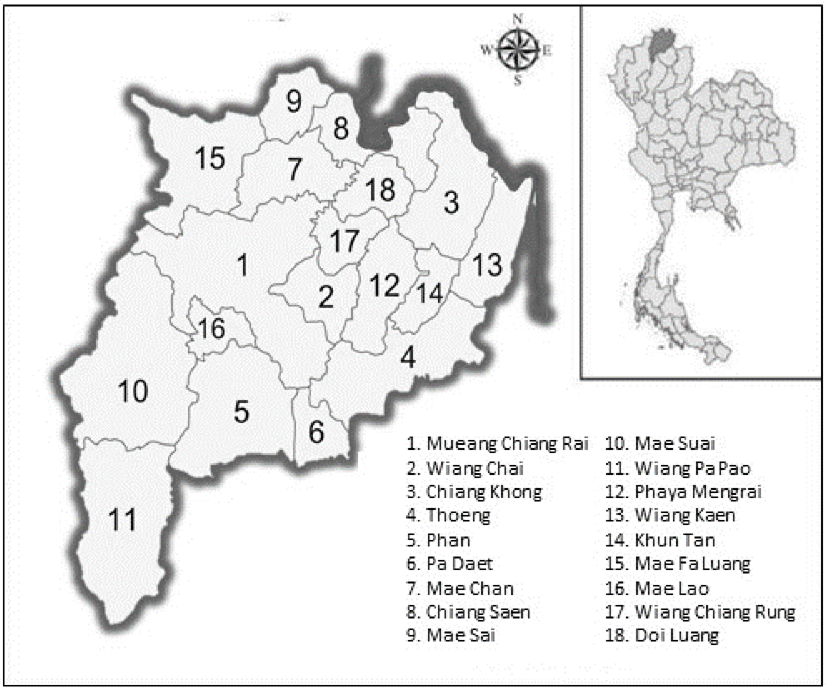


**Figure 11** Administrative Districts of Chiang Rai Province

*Source: Own illustration based on Wikipedia (2016) and Akber and Shrestha (2013)*

The research participants consisted of 150 farmers from different districts of Chiang Rai province growing longan and rice predominantly. Out of the 150 farmers, 50 farmers were users of the Farmer Info application while 100 were not. Additional data in Chiang Rai was gathered from three farmers, whereby two were using the application and one did not. Outside of the study area, two officials from Rak Ban Kerd and dtac have been participating in the research.



### 4.3 Data Collection

Data collection was done using two methods, in-depth interviews and a survey among app users and non-users, with supporting field observations and unrecorded, unofficial discussion. Additionally, success stories about the SMS-service of RBK were provided by officials from the foundation.

Furthermore, data was collected from various online sources, such as Twitter or Google Play Store (see Appendix 9). This includes also a CSR promotion video by dtac which has been extracted from YouTube (see Appendix 8).

The data collection was carried out during January 2016 and July 2016 in several stages. First, in-depth interviews provided a deeper understanding of the concept before a survey collected the quantitative data. Additional in-depth interviews were carried out at the final stage of the study.

4.3.1 Interview

In-depth interviews are used by researchers to get more detailed and thorough information as well as better insights into unknown topics (Boyce & Neale, 2006, p. 3). Interviews are among the most commonly used qualitative research methods. There are three main types of interviews. First, a fully structured interview follows strictly predefined questions in a given order. The advantage thereby is that questions can be easier evaluated and distractions are slightly to occur. However, structured interviews also do not allow for further investigation besides the given question and can, therefore, miss relevant content. Another form is the completely unstructured interview which follows no pre-defined questions. Often those interviews start with a simple question and develop over time. This can, on the one hand, provide a thorough understanding of the interviewee's opinion, on the contrary, it can easily lead to distractions and losing track of the topic (Gill, Stewart, Treasure, & Chadwick, 2008).

Both of the suggested forms are valid instruments and can be useful in certain research designs. However, the interview design most often used is the semi-structured interview, which lies in between. Semi-structured interviews, in this research also referred to as in-depth interviews, are using several key questions that provide a degree of guidance but still allow the interviewer or interviewee to follow up their thoughts (Gill et al., 2008, p. 291).



Successfully conducting interviews requires certain rules and steps, which have to be followed to provide valid results. In the beginning, it is important to identify key stakeholders and actors as well as relevant content for the interviews. Thereupon an interview guide can be developed. According to Boyce and Neale (2006), there should be no more than 15 questions, and they should rather be open-ended than closed-ended (Boyce & Neale, 2006, p. 5). Furthermore, the wording of the questions has to be considered carefully. Questions should be short and simple, while being understandable to every interviewee in the same way, and should not contain pre-defined implications leading to bias in the response (Cairns & Cox, 2008, pp. 19–20). Then, before the interview is conducted the participants should be informed about the schedule, records and confidentiality (Adams & Cox, 2008, pp. 21–22). A general introduction and a short background to the study provide a good start. During the interview, it is crucial that the interviewer provides guidance but only contributes slightly to the conversation and also does not lead the interviewee in a certain direction with own opinions (Adams & Cox, 2008, pp. 23–24). Transcription is the final step, which makes the content more readily accessible, before analysing the interview. Several tools are available for transcribing and analysing in-depth interviews.

The overall advantage of semi-structured, in-depth interviews is the detailed information they provide to the researcher. They can also significantly contribute to widening the scope of the study. However, such interviews can be prone to bias as the participant follows an own agenda trying to influence the research. Furthermore conducting, transcribing and analysing interviews can be time intensive and often requires training to take full advantage of the approach. Also, results often cannot be generalised due to the unique nature of each interview (Boyce & Neale, 2006, pp. 3–4).

All interviews in this research were considering the instructions and guidelines provided by the literature. The interviews were either conducted in the private homes or offices of the interviewees and followed a semi-structured approach. The guiding questions thereby differed between the key stakeholders of RBK and dtac and the farmers. The farmer interviews were conducted in Thai language with the support of a local interpreter, while the other interviews were carried out in English. All interviewees were informed about the background of the study and the confidentiality with which their data and statements will be handled. The interviews in this research were transcribed using the program f4[1], which was also used to transcribe the CSR promotion video from dtac. The interviews were thereby transcribed as close as possible to a pure verbatim transcript. A pure verbatim protocol includes slang, utterances like "uhms" or ahs" as well as decorating words such as "you know" or yeah" (Mayring, 2014, p. 45). This approach was chosen to highlight the limitations and difficulties aligned with the

[1] https://www.audiotranskription.de/english

in-depth interviews which arise due to language barriers and average English skills of some interviewees. The transcripts of the interviews are attached to thesis (Appendix 2-7). The analyzation of the interviews was subsequently done with QCAmap[2], an open-access program developed by Prof. Dr Philipp Mayring and Dr Thomas Fenzl. The analyzation with QCAmap thereby included also the data derived from Twitter, Instagram, YouTube and Google Play Store.

### 4.3.2 Survey

In quantitative research, surveys are a common tool for data collection using standardised questionnaires. Similar to structured interviews, standardised questionnaires follow a strict line with limited space for further investigation. The benefits of questionnaires in surveys over structured interviews are that they can provide quantifiable and reliable data covering a broad range of topics (Jalil, 2013). A questionnaire can cover a larger number of people producing statistically analysable data highlighting relation between actors and characteristics. Different question types thereby can produce different results. The most common questions are:

1) *Simple factual questions – requiring a yes/no responses*
2) *Complex factual questions – requiring some interpretation or analysis*
3) *Opinion and attitudinal questions – requiring more alternatives and deeper concentration*
4) *Open ended questions – requiring participants' full concentration. (Adams & Cox, 2008, pp. 20–21)*

Depending on the question type surveys can produce different results and therefore each type requires careful consideration. However, they always should be phrased as straightforward and short as possible to increase the understanding for the participants and the accuracy of their answers (Adams & Cox, 2008, p. 20). Two key concepts related are validity and reliability. Reliability thereby refers to "the consistency of a measure" while validity to "its ability to measure what it is supposed to be measuring" (Cairns & Cox, 2008, pp. 18–19).

In general survey can be divided into two categories:

[2] https://www.qcamap.org/

1) *Cross-sectional surveys where data is collected at one point in time*
2) *Longitudinal surveys where data is collected over a time period (Jalil, 2013)*

In the context of agriculture and developing countries surveys are generally conducted face-to-face, although online and phone surveys provide an alternative. Similar to the interviews, questionnaires require an introduction and a brief background of the study. Additionally, participants have to be informed about the confidentiality with which their data will be treated. This is necessary to receive truthful answers by the participants, particular for sensitive questions such as money or politic related.

Although a valuable tool to access statistical data, questionnaires have certain limitations. Even if questions are phrased carefully, the meaning can be different for respondents and literacy can be a limiting factor. Furthermore, surveys lack personal insight and do not provide additional information apart from the questions. Also, it is challenging and takes training to develop useful questionnaires which provide valuable data (Jalil, 2013). A pre-test can be used to test questionnaires and adopt them to the received feedback (Adams & Cox, 2008, p. 19).

Questionnaires in this research have been used to collect statistically valid data from two groups, application users (treatment group) and non-users (control group). Simple factual questions, complex factual questions, opinion questions were combined to provide a broad picture of the participants and their attitude (see Appendix 11 & 12). While straightforward and complex factual questions provide an easily understandable way for data collection, opinion questions require a more thoughtful approach. The most common method used, is the Likert Scale. There are different versions of the Likert Scale, however, generally checklists are used, including several options such as "very useful" or "not very useful". It is still debated if a middle category is helpful and provides a true scale or if it just gives the participants a simple way to avoid a clear decision (Adams & Cox, 2008, p. 21). In this study, Likert Scales were used including middle categories.

Questions in both surveys were grouped in common themes for better guidance, increasing the usability and effectiveness (Cairns & Cox, 2008, p. 19). Critical to the success of a survey is, besides good questions, the length. Long questionnaires can lead to decreasing motivation and accuracy (Cairns & Cox, 2008, p. 19). In this context, both questionnaires took between 30-60 minutes to finish, which can already be seen as critical. However, a broad database was required as additional data was limited.

Surveys for this research were conducted by phone as well as face to face. The phone survey was carried out by dtac's call centre. The face to face survey was executed with help from officials from local agricultural extension offices and an interpreter. Language and illiteracy were among the barriers which increased the level of complexity.

### 4.4 Data Analysis

In science, nothing contains more disagreement than the appropriate research methodology used. In this research, the qualitative data is analysed by a summative content analysis, while the quantitative data is analysed with the methods based on descriptive and inductive statistics. The following two subchapters will describe the methods in more detail and also highlight the advantages and disadvantages of the methodology.

#### 4.4.1 Qualitative Content Analysis

Qualitative data analysis is diverse, with a variety of different focuses and purposes related to the researcher's perspective. Content analysis is a flexible method for text analyzation depending on the theoretical background and interest of the investigator (Hsieh & Shannon, 2005, p. 1277). Qualitative content analysis can be defined as:



*"an approach of empirical, methodologically controlled analysis of texts within their context of communication, following content analytical rules and step by step models, without rash quantification." (Mayring, 2000, p. 2)*

or

*" […] a research method for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns. (Hsieh & Shannon, 2005, p. 1278)*

Content analysis was developed to evaluate large textual corpuses. Historically content analysis is a relatively new method. According to Mayring (2014), the development can be divided into three steps. The preliminary phase dates backs centuries and uses mainly quantitative approaches for text analysis, such as word frequency. The first period with less quantitative and more qualitative approaches started around the turn of the 20$^{th}$ century. It followed simple comparison and analysis, often in hermeneutic contexts or newspaper analysis and also included the dream analysis by Sigmund Freud. The first book on qualitative content analysis as a research method was published by Berelson in 1952 after preparatory work by Paul F. Lazarsfeld and Harold D. Lasswell throughout the consolidation phase in the 1920's and 30's. Following the first publications, the methodology was developed further in the sixties, predominant in research areas such as psychology, linguistics or sociology. Thereby the procedures had been clarified and adapted to models and computer applications. After years of stagnation and criticism of the superficiality of the analysis, new approaches had been developed (Mayring, 2014, pp. 18–21).

Mayring identifies four main aspects which are necessary for qualitative content analysis and have to be considered:

• *"Fitting the material into a model of communication: It should be determined on what part of the communication inferences shall be made, to aspects of the communicator (his experiences, opinions feelings), to the situation of text production, to the socio-cultural background, to the text itself or the effect of the message.*

• *Rules of analysis: The material is to be analysed step by step, following rules of procedure, devising the material into analytical content units.*

• *Categories in the centre of analysis: The aspects of text interpretation, following the research questions, are put into categories, which were carefully founded and revised within the process of analysis (feedback loops).*

• *Criteria of reliability and validity: The procedure has the pretension to be intersubjectively comprehensible, to compare the results of other studies in the sense of triangulation and to carry out checks for reliability. For estimating the inter-coder reliability we use in qualitative content analysis (in contrary to quantitative content* *analysis) only trained members of the project team, and we reduce the standard of coder agreement" (Mayring, 2000, pp. 2–3)*

In general Mayring (2014) purposes eleven steps for the general qualitative content analysis (see Figure 12). The first steps, including the definition of the material, the situation of origin and formal characteristics were mainly discussed above. Therefore, in the following the development of the category system, using a summative approach, is described in more detail.

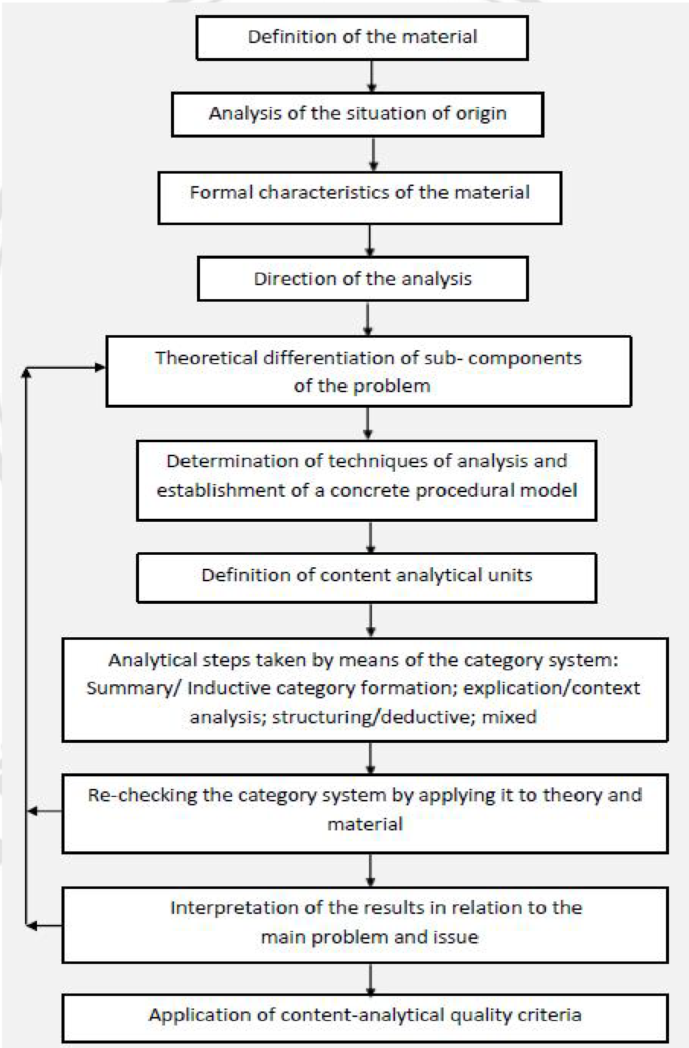


**Figure 12** General Content-Analytical Procedure Model
*Source: Mayring (2014, p. 54)*

A major concern in the development of a category system for the collected data was that different aspects were covered by the qualitative methods. While the interviews with the key stakeholders were conducted for a better understanding of the project, the interviews with the farmers and the secondary data were collected to gain more information about the projects impact and people’s opinion.

The best approach to cover all the aspects highlighted in the interviews is the “Summarizing” as suggest by Mayring (2014). This form of qualitative analysis tries to identify key aspects in the material and summarises them under a previously defined level of abstraction (see Figure 13). In the first step, after determining the units of analysis, the context is minimised to the essential parts by eliminating irrelevant statements. The relevant parts of the material are then transformed into a consistent, abbreviated style (Mayring, 2014, pp. 66–67). Following the step of “Paraphrasing”, the now rewritten statements are generalised following the predefined level of abstraction. Thereby the content of the paraphrases should be implied in the generalisation (Mayring, 2014, p. 67). In the next step, the generalised statements are then reduced by combining similar statements or cutting irrelevant ones under a new level of abstraction. If necessary, a second phase of reduction can be carried out, in case a more general level of abstraction is required (Mayring, 2014, p. 67)

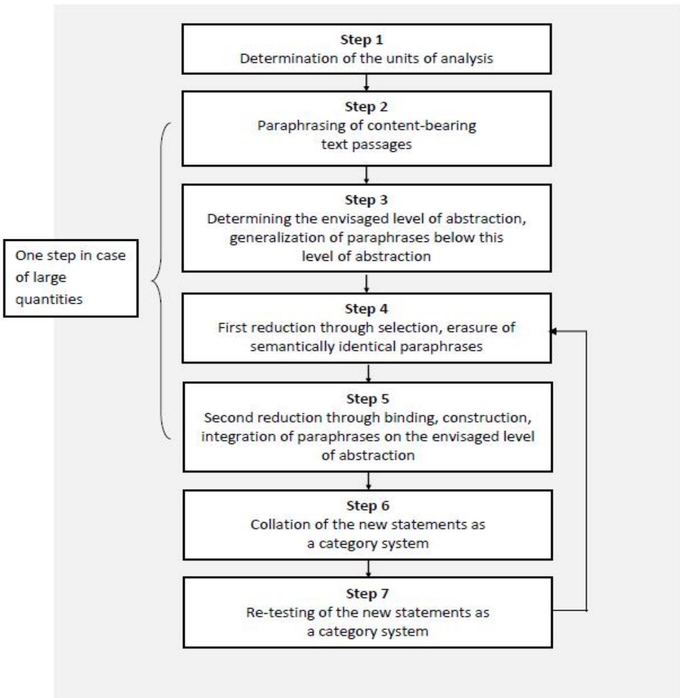


**Figure 13** Step-by-step Model of Summarising in Qualitative Content Analysis

*Source: (Mayring, 2014, p. 66)*

Following the steps suggested by Mayring (2014), eight main categories were identified after the second reduction, with a total number of 24 sub-categories which were the result of the first reduction (see Figure 14). All categories cover the essential parts of the interviews and can be used additionally to the quantitative results answering the research questions. Appendix 13 will provide a more detailed overview of the single steps, including paraphrasing and generalising. The single categories will not be discussed independently in detail in chapter five, but they will be used in combination with the quantitative results to support or weaken the statistical numbers provided through the survey. This follows the suggestion of the triangulation of evaluation findings as highlighted by Bamberger (2012).

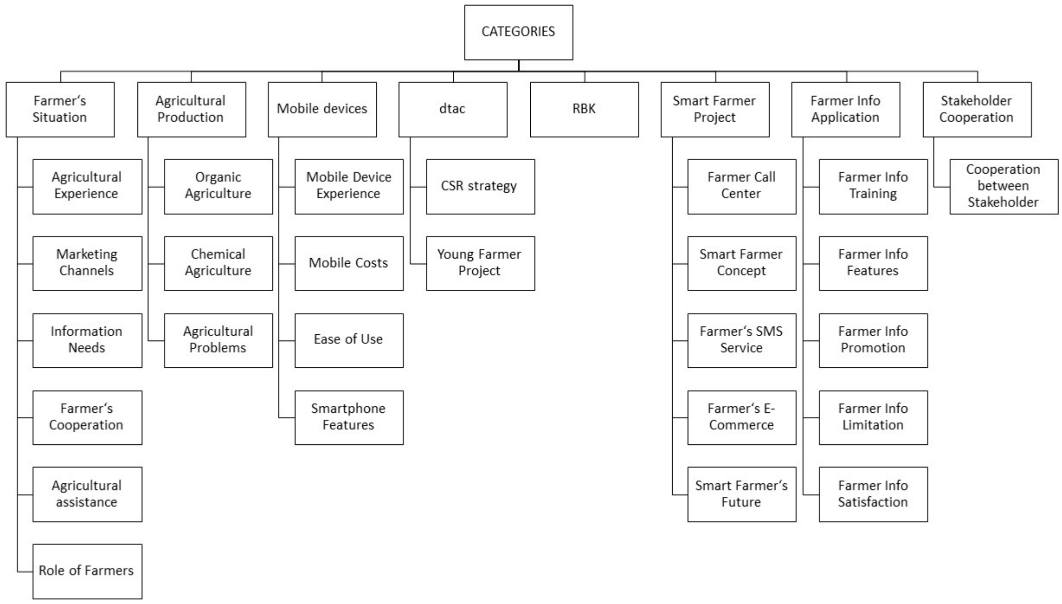


**Figure 14** Category System
*Source: Own illustration*

4.4.2 Quantitative Analysis

Although the qualitative research was an important and necessary part of the research to understand the system and views of stakeholders, the main focus of this thesis lies on the quantitative analysis of the survey data, collected in Chiang Rai Province. In this chapter, the single steps of the analysis are pointed out.

The main challenge in the evaluation of impacts remains thereby to identify what would have happened to the beneficiaries without the project, the counterfactual (Diaz & Handa, 2014, p. 11).

The counterfactual can be defined as “the situation or condition which hypothetically may prevail for individuals, organizations, or groups were there no development intervention” (OECD, 2010, p. 19). The general problem of understanding the impact of a project and finding a counterfactual is that the same group cannot be analysed with and without the treatment or intervention at the same time. Therefore it is required to find a counterfactual either by comparing two groups, treatment and non-treatment group which are as similar as possible, or by comparing baseline data with data after the treatment was implemented (Khandker, Koolwal, & Samad, 2010, p. 23). In the case of this explorative evaluation there is no baseline data available and therefore a before-after comparison cannot be conducted. As a result this evaluation will use the so called with-and-without comparison to find an appropriate counterfactual. This will include a treatment group (App-Users) and a control group (Non-Users).

Simple descriptive statistic is used to highlight the characteristics of the participants as well as the needs and opinions of farmers regarding the application and their agricultural activities. Frequencies and means will be dominating the analysis. The variables thereby include among others, the age, gender and education of the farmer as well as their agricultural problems and information needs or opinions on the usefulness of mobile phones as well as the Farmer Info application.



Besides a simple descriptive analysis, inductive statistical methods are used to test the hypotheses proposed and their corresponding null hypotheses:

<u>Impact on livelihood</u>

$H_{1a}$: *The FARMER INFO application affects the farmer’s livelihood.*

$H_{0b}$: *The FARMER INFO application does not affect the farmer’s livelihood.*

<u>Impact on agricultural practices</u>

$H_{1b}$: *The FARMER INFO application influences the use of chemical fertilisers and pesticides.*

$H_{0b}$: *The FARMER INFO application does not influence the use of chemical fertilisers and pesticides.*

The livelihood of the farmers is thereby covered by five variables, including:

- *Average monthly income from agricultural activities (in Baht)*
- *Average monthly income from other activities (in Baht)*
- *Average selling price (in Baht per kilo)*
- *Highest selling price (in Baht per kilo)*
- *Lowest selling price (in Baht per kilo)*

The agricultural practices of the farmers are represented by the following variables:

- *Pesticide application (per week)*
- *Monthly spending on pesticide (in Baht)*
- *Pesticide used*
- *Fertiliser application (per month)*
- *Monthly spending on fertiliser (in Baht)*
- *Fertiliser used*

An explorative approach is needed to test the data on normal distribution and variance homogeneity, to identify valid tools for the analysis.

The explorative statistic is often seen as a form between the descriptive and the inductive statistic. It is often used to formulate hypotheses or to develop further surveys, questionnaires. However, explorative statistical methods are also necessary to assist in choosing the most appropriate tool or technique for further analysis. In particular, the tests for normal distribution and variance homogeneity are determining the use of proper tools.

There are several techniques to check data sets for normal distribution and variance homogeneity. Some of the approaches have been criticised due to weak results and therefore untrue statements. In the case of the normality test, most recently a study conducted by Razali and Wah (2011) found out that the Shapiro-Wilk test provides the most accurate results identifying the normal distribution. Therefore it will be used in this thesis to determine which tools can be utilised for further analysis. This test, as well as the following tests were conducted using the IPM SPSS Statistics Software, version 23. Furthermore, all tests assumed a significance level of α = 0.05.

*Shapiro-Wilk Test*

The Shapiro-Wilk test is a technique to check for normal distribution among samples. The normal distribution is a requirement for certain test such as the t-test. The Shapiro-Wilk test was first published by Samuel Sanford Shapiro and Martin Wilk in 1965. It assumes a null hypotheses which states that the sample is normally distributed and uses the following test statistics to either prove or disprove it:

$$W = \frac{\sum_{i=1}^{n} a_i y_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Equation 1: Shapiro-Wilk Test of Normality

Source: Smith, 2011, p. 458

where,

- $y_i$ are the data from the sample, sorted by size
- $a_i$ are constants to be evaluated (Smith, 2011, p. 458)

The assumption underlying this test is that it should be possible to express the data with a simple linear regression if the sample data is a random sample from a normal distribution.

**Table 1** Shapiro-Wilk Test of Normality for Livelihood Variables of Longan and Rice Farmers

| | Shapiro-Wilk[1] | | | Shapiro-Wilk[2] | | |
|---|---|---|---|---|---|---|
| | Statistic | df[3] | Sig. | Statistic | df | Sig. |
| **Average monthly income from agriculture (Baht)** | .851 | 65 | .000 | .833 | 68 | .000 |
| **Average monthly income from other activities (Baht)** | .808 | 65 | .000 | .397 | 68 | .000 |
| **Average price (Baht per kilo[1,2])** | .950 | 65 | .010 | .900 | 68 | .000 |
| **Highest price (Baht per kilo[1,2])** | .941 | 65 | .004 | .682 | 68 | .000 |
| **Lowest price (Baht per kilo[1,2])** | .880 | 65 | .000 | .954 | 68 | .014 |

*Source: Own Calculations*

1) *Longan*
2) *Rice*

**Table 2** Shapiro-Wilk Test of Normality for Input Variables

| | Shapiro-Wilk | | |
|---|---|---|---|
| | Statistic | df | Sig. |
| **Pesticide application per week** | 0.694 | 145 | 0.000 |
| **Monthly spending on pesticide** | 0.622 | 145 | 0.000 |
| **Fertilizer application per month** | 0.695 | 145 | 0.000 |
| **Monthly spending on fertilizer** | 0.798 | 145 | 0.000 |

*Source: Own Calculations*

The Shapiro-Wilk test shows that no variable follows a normal distribution. All of the variables show a significance level between $p < .000$ and $p < .014$. Therefore the null hypotheses that the variables are normally distributed can be refuted



[3] df refers to the term “Degrees of Freedom”. A brief definition of degrees of freedom is given by Healey (1990), cited in Pandey and Bright (2008): "Degrees of freedom are the number of values in a distribution that are free to vary for any particular statistic"

*Levene's Test*

A second test to determine valid tools for the inductive statistic is the Levene's test for equality of variances. The Levene test is a method to assess the equality of variances between two groups. This test provides another indicator on which methods can be used for hypotheses testing:

$$W = \frac{(N-k)}{(k-1)} \frac{\sum_{i=1}^{k} N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2}$$

Equation 2: Levene Test for Equality of Variances

Source: NIST/SEMATECH, n.d.

where,

- $W$ is the result of the test
- $k$ is the number of different groups to which the sampled cases belong
- $N$ is the total number of cases in all groups
- $N_i$ is the number of cases in the i-th group
- $Z_{ij}$ is the value of the measured variable of the j-th case from the i-th group

**Table 3** Test for Equality of Variances for Livelihood Variables for Longan and Rice Farmers

| | Levene Statistic[1] | df1[1] | df2[1] | Sig.[1] | Levene Statistic[2] | df1[2] | df2[2] | Sig.[2] |
|---|---|---|---|---|---|---|---|---|
| **Average monthly income from agriculture (Baht)** | .872 | 1 | 73 | .353 | .007 | 1 | 73 | .934 |
| **Average monthly income from other activities (Baht)** | 29.529 | 1 | 73 | .000 | 4.731 | 1 | 73 | .033 |
| **Average price (Baht per kilo[1,2])** | .020 | 1 | 72 | .887 | 8.218 | 1 | 72 | .005 |
| **Highest price (Baht per kilo[1,2])** | 1.331 | 1 | 63 | .253 | 9.954 | 1 | 67 | .002 |
| **Lowest price (Baht per kilo[1,2])** | 6.564 | 1 | 63 | .013 | 10.994 | 1 | 66 | .001 |

*Source: Own Calculation*

1) *Longan*
2) *Rice*

The results for the livelihood variables show that for only for three variables the test is not significant, for the rest, it shows a significance level between p < .000 and p < .033. Therefore the null hypotheses can be refuted for those variables indicating that there not homogenous. The other three variables are assumed to be homogenous due to the significance level.

**Table 4** Test for Equality of Variances for Input Variables

| | Levene Statistic | df1 | df2 | Sig. |
|---|---|---|---|---|
| **Pesticide application per week** | 21.179 | 1 | 145 | 0.000 |
| **Monthly spending on pesticide** | 28.639 | 1 | 147 | 0.000 |
| **Fertilizer application per month** | 33.508 | 1 | 145 | 0.000 |
| **Monthly spending on fertilizer** | 6.898 | 1 | 146 | 0.010 |

*Source: Own Calculation*

The results for the practice variables show that all variables have a significance level between p < .000 and p < .010. Therefore the null hypotheses can be refuted for all variables indicating that they are heterogeneous.

As a result of the preliminary tests, two methods were chosen to analyse the differences between application users and non-users and thereby evaluate the impact the application can have on the livelihood and agricultural practices of farmers. Namely, these tests are the “Mann-Whitney U Test” for metric variables and the “Chi-Square Test” for nominal variables.



*Mann-Whitney U Test*

The Mann-Whitney U test bases on the idea from Frank Wilcoxon who developed a test based on data comparing the effectiveness of two preparations of fly spray (Smith, 2011, p. 487). For each sample, he conducted eight tests and assigned a rank to each result based on the effectiveness. All ranks were then summed up for each sample (see Table 5).

Table 1: Fly Spray Test Conducted by Frank Wilcoxon

| **Sample 1** | | **Sample 2** | |
|---|---|---|---|
| **% kill** | Rank | % kill | Rank |
| **68** | 12.5 | 60 | 4 |
| **68** | 12.5 | 67 | 10 |
| **59** | 3 | 61 | 5 |
| **72** | 15 | 62 | 6 |
| **64** | 8 | 67 | 10 |
| **67** | 10 | 63 | 7 |
| **70** | 14 | 56 | 1 |
| **74** | 16 | 58 | 2 |
| | R1 = 91 | | R2 = 45 |

*Source: Smith, 2011, p. 487*

Wilcoxon was then interested in the probability of getting a rank sum of 45 within the two samples. Therefore he calculated all possible arrangements for eight ranks out of the total 16, resulting in a total number of 12870 (Smith, 2011, p. 487). This number he divided through the possible arrangements which provide the exact rank sum of 45 or less. As a result, he got a probability level on which he was able to make a decision on the null hypotheses that the samples have different means.

According to Smith (2011) there are several aspects indicating the usefulness of the test. The first is that the test is non-parametric as it has no expectation on the distribution of the data and is also robust against outliers as ranks are calculated. Secondly, it is efficient as the t-test and therefore provides a valid alternative. Furthermore, a z-transformation can lead to a z-test, in particular for larger samples. The only limitation of the method is that it is based on equal sample sizes.

However, briefly, after the Frank Wilcoxon developed the test, Henry B. Mann and Donald R. Whitney published an almost identical test which can be used with different sample sizes. The Mann-Whitney test thereby also uses the rank sums R1 or R2 to calculate the statistics (see Equation 3 and 4).

$$U_1 = R1 - \frac{n_1(n_1 + 1)}{2}$$

Equation 3: Mann-Whitney U Test

for U1

Source: Smith, 2011, p. 487

$$U_2 = R2 - \frac{n_2(n_2+1)}{2}$$

Equation 4: Mann-Whitney U Test for U2

Source: Smith, 2011, p. 488

If the sample size is large enough, a z-transformation of the results is suggested:

$$z = \frac{U - m_U}{\sigma_U}$$

*Equation 5*: Z-transformation of Mann-Whitney Test Results

Source: Smith, 2011, p. 488

where,

$m_U = \frac{n_1 n_2}{2}$

and,

$\sigma_U = \sqrt{\frac{n_1 n_2 (n+1)}{12}}$

Additionally to the test for the significance the effect size can provide an additional indicator for the results:

$$r = |\frac{z}{\sqrt{n}}|$$

*Equation 6*: Effect Size Calculation

Source: Universität Zürich, 2016



Cohen identified three levels of the effect size, which can be transformed to r, when a sample size is divided into a treatment and control group :

$r =$ .10 indicates the starting point for weak effect

$r =$ .30 indicates the starting point for an average effect

$r =$ .50 indicates the starting point for a strong effect (Universität Zürich, 2016)

*Chi-Square Test*

Beside the Mann-Whitney U test for the metric variables, the Chi-Square test is applied for the nominal variables. In this test the variables are arranged in a table of r rows and c columns, where A and B represent certain characteristics which are under evaluation (see Table 6).

**Table 6** Assumptions of the Chi-Square

| **Characteristics** | | **B** | | | | **TOTAL** |
|---|---|---|---|---|---|---|
| | | 1 | 2 | … | c | |
| **A** | 1 | $n_{11}$ | $n_{12}$ | | $n_{1c}$ | $n_{1.}$ |
| | 2 | $n_{21}$ | $n_{22}$ | | $n_{2c}$ | $n_{2.}$ |
| | 3 | | | | | |
| | … | | | | | |
| | r | $n_{r1}$ | $n_{r2}$ | | $n_{rc}$ | $n_{r.}$ |
| **TOTAL** | | $n_{.1}$ | $n_{.1}$ | | $n_{.c}$ | $n_{.1}$ |

*Source: Smith, 2011, p. 495*

The Chi-Square test compares the observed frequency in each cell (O) with the expected frequencies (E). The expected values are calculated from the marginal probability when assumed that the characteristics are independent. The result of the subtraction is then squared to achieve a positive sign (see Equation 7). Finally, the result is divided by the expected values to standardise the measurement (Smith, 2011, p. 495).

$$\chi^2_{(r-1)(c-1)} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

*Equation 7: Chi-Square Calculation*

*Source: Smith, 2011, p. 488*



Where,



$O_{ij} = n_{ij}$ & $E_{ij} = \frac{n_{i.}n_{.j}}{n_{..}}$



A requirement of the Chi-Square test is thereby that no entries in the cells are less than 5. In case one cell has a count less than five an exact test is required, and the Chi-Square test is not suitable anymore. One test which is then normally used is called Fisher's exact test.

Associations between two nominal variables can be further investigated by analysing the "Phi Coefficient", which can be applied to a 2x2 table. Other alternatives for tables large than 2x2 are “Cramer’s V” or the “Contingency Coefficient”. Thereby, all measurements indicate no association between two variables, when the given value is equal to zero, and a complete association with a value equal to one. Latter applies when two identical variables are compared. As a result, an association is strong as closer it is to one.

*Selection Bias*

Beside the problem of the counterfactual, overcoming selection bias, which can be caused by the statistical analysis of non-random data, is an issue which always needs to be addressed in such evaluation. There are several ways to deal with selection bias. Most recently “Propensity Score Matching” has received growing attention within the scientific community (Stuart, Elizabeth, A. & Rubin, Donald, B., 2007, p. 159). In this research non-random data is analysed with the methods described in this chapter, however, selection bias might be distort the results. Therefore, additional to the general evaluation, the data is analysed again after propensity score matching has been applied, using the nearest neighbour approach with a 1:1 matching.

The propensity score matching method uses observable covariates to identify a single or multiple matches from the control data for each treated, based on a propensity score (Stuart, Elizabeth, A. & Rubin, Donald, B., 2007, p. 159). It is important, when conducting propensity score matching, to choose proper covariates on which the model bases. Thereby, the covariates should be related to the treatment assignment and possibly to the outcome while not being affected by the treatment (Stuart, Elizabeth, A. & Rubin, Donald, B., 2007, pp. 161–162). The literature suggests that It is better to include more covariates than too less as highlighted by Stuart, Elizabeth, A. and Rubin, Donald, B. (2007, pp. 161–162). However, personal consultation with statisticians has shown that up to five covariates are sufficient for propensity score matching. This will give close matches between treatment and control individuals.

The covariates for the propensity score matching in this research will cover the following:

- *Age*
- *Sex*

- *Gender*
- *Farm Size*
- *Time since first purchase of a mobile phone*

A problem arising when using propensity score matching is missing data within the collected data. Missing data is a common problem in research and can have several causes, such as people not wanting to answer certain questions or not remembering. With missing data it is not possible to conduct propensity score matching, as it is the case in this research. However, there are certain ways to overcome the problem of missing data in propensity score matching. Two types have been used in this study to provide valid results. On the one hand the missing data was replaced through simple imputation, or mean substitution. The biggest advantage of this method is that it keeps the complete sample size for analysation, but it decreases the variances with results getting closer to the mean and underestimates the variability (Zhu, 2014, p. 935). On the other hand cases containing missing data were excluded. It is a simple method to apply and does not influence the data. However, the main disadvantage of this method is the limitation of the sample size reducing the statistical power (Zhu, 2014, p. 935). In the case of this research the sample size was already small and has been limited even more by the elimination of cases containing missing data. The results of both matching processes are shown in Figure 36 to 41, attached in Appendix 14. Unlike the previous quantitative analysis, IPM SPSS Statistics 23 was not used for the matching due to limitation within the software. Instead the software R was used, which provides more functions but is also less user-friendly and requires deeper understanding of the statistical processes.



The original data was analysed and later controlled by comparing the results to the estimates provided by both matching techniques. Differences regarding the significance of the impact where highlighted accordingly. As propensity score matching with missing data does also not provided exact results, the original data is used in the discussion and only significant differences with the matched data are highlighted and discussed.