Chapter 2

Methodology

This chapter describes the methods used in this study. The methodology consists of the data source, variables and path diagrams, data collection, data management, and statistical method.

2.1 Data source

A retrospective study was applied to this study that uses secondary data, which were obtained from the Deep South Coordination Centre (DSCC), Prince of Songkla University (PSU), Pattani Campus.

Figure 2.1 Map of the three southern provinces of Thailand

Figure 2.1 represents the unrest area of this study covering the province of Pattani, Yala, Narathiwat, and Songkhla (including the district of Chana, Na Tawi, The Pa, and Saba Yoi).

The data were recorded as a daily unrest event, which was occurred in several areas in the three southern provinces of Thailand. The type of the reported unrest event was formed of explosives recovered, firecracker, kill and burn, motorcycle bomb, army base attack, waylay, gun robbery, raid blocked, firing the government offices and schools, found a dead body, sharp weapon attack, and some persecution events.

2.2 Data management

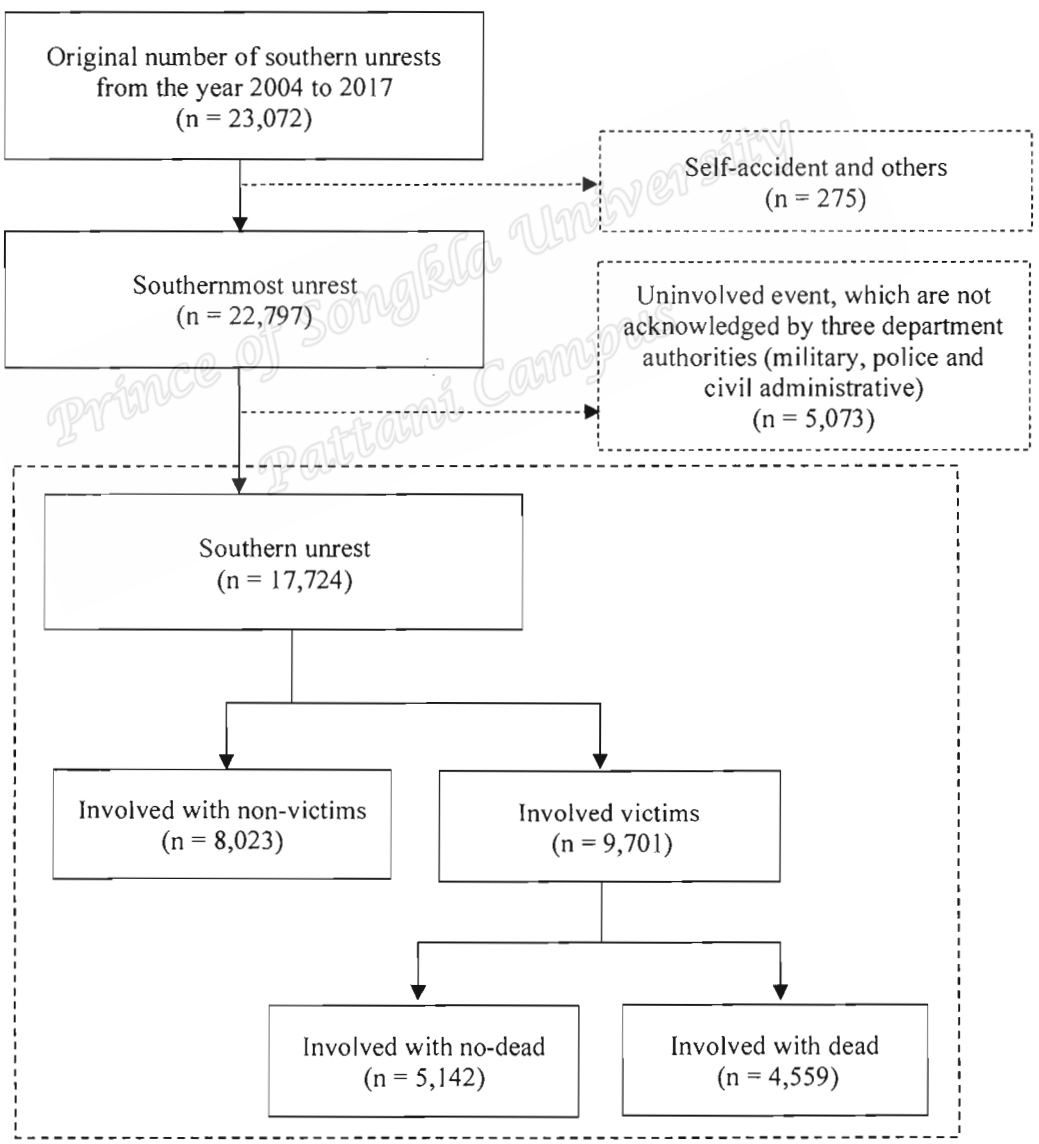

Figure 2.2 The structure of the southern unrest

Figure 2.2 shows the 23,072 events of the southern unrest of the origin recorded number from the year 2004 to 2007. The data have filtered out 275 events, which is involved with self-accident and others. Next, they have remained 22,797 events and 5,073 events were filtered out the unacknowledged events. Finally, this study was finalized the unrest events number at 17,724 events. Furthermore, the 17,724 of the southern unrest were categorized into two groups. The 8,023 of events involved non-victims and 9,701 events involved victims. Moreover, the events involved victims are divided into two groups; 5,142 events of no-dead victims and 4,559 events of dead victims.

2.3 Variables and path diagrams

Outcome variable

Outcome variables are the unrest that consisting of

- The unrest involved no-victim or victims
- The severity of unrest that less than 75 percent or greater than or equal to 75 percent

The unrest severity scale is stable and there was no fundamental difference between the large and small event (Clauset et al., 2007). The definition of unrest severity is the proportion of deaths in a particular event. In this study, the severity is categorized into groups where the proportion of dead less than 0.75 or greater than or equal to 0.75 or in term of the percentage that is less than 75 percent or greater than or equal to 75 percent.

Determinants

The characteristic of southern unrest is consisting the years of the event, the months of the event, days of the event, times of the event, provinces of the event, the zones of the event, the areas of the event, and type of unrest.

Variable definition

The definition of all eight determinant variables are described in the Table 2.1:

Table 2.1 Variables definition

Variable	Description
Year	Year of unrest: 2004 - 2017
Month	Month of unrest: January - December
Day	Day of unrest: Monday - Sunday
Time	Time of unrest, categorized into eight time slots: 06.01 pm to 09.00 pm, 09.01 pm to 12.00 am, 12.01 am to 03.00 am, 03.01am to 06.00 am, 06.01 am to 09.00 am, 09.01 am to 12.00 pm, 12.01 pm to 03.00 pm, and 03.01 pm to 06.00 pm
Province	Province where unrest happened, includes Pattani, Yala, Narathiwat and Songkhla
Zone	Zone where unrest happened, includes business zone, check point, public place, residential zone, road and others
Area	Area where unrest happened: rural or urban
Type of unrest	Type of occurred unrest, includes aggravate, arson, bomb, shooting and others (assault and nail trapping)
Unrest involved victims	Yes or No

Variable	Description
Severity of unrest	The proportion of dead categorized into two groups; < 0.75 and ≥ 0.75 or in terms of percentage; < 75 percent and ≥ 75 percent

Path diagrams

Unrest involved victims

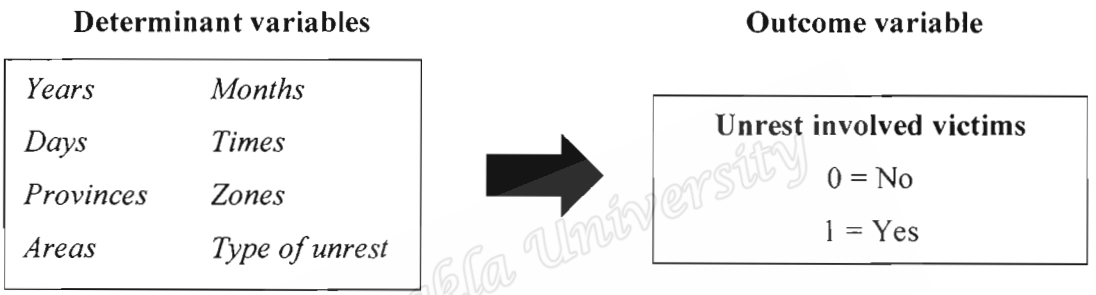

Figure 2.3 (a) Path diagram of unrest involved victims

Severity of unrest

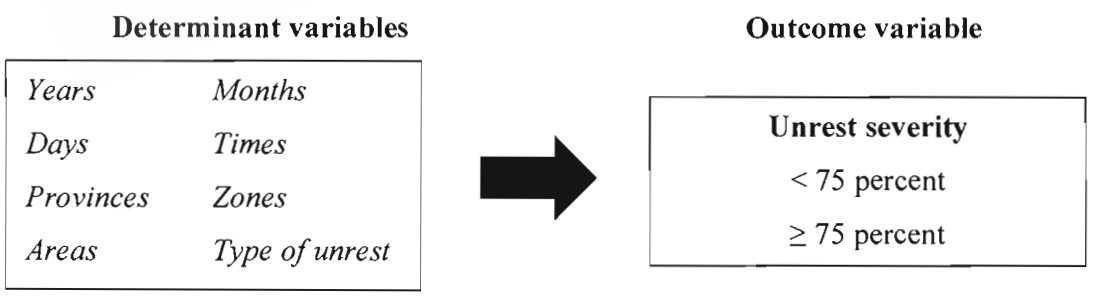

Figure 2.3 (b) Path diagram of unrest severity

Figure 2.3 (a) and (b) illustrate the conceptual framework of the association between the determinants and the unrest.

2.4 Statistical analysis

Descriptive analysis

Descriptive analysis is the conversion of raw data into a new format by providing useful information. It represents a frequency number, which is summarized and described by percentage.

Chi-squared test

Univariate analysis is applied by using a chi-squared test. The comparison among determinants and outcomes, non-stratified r × c table was applied to see the association among the determinants and outcomes. For example, x is defining the years of the unrest events and y is defining the binary outcome of southern unrest events involving non-victim or victims (0, 1).

		y	
		1	*2*
x	*1*	a_{11}	a_{12}
	2	a_{11}	a_{12}
	...	..	..
	r	a_{r1}	a_{r2}

Pearson's chi-squared statistic for the independence (no association) in an $r \times c$ table is defined as

$$\chi^2_{(r-1)(c-1)} = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(a_{ij} - \hat{a}_{ij})^2}{\hat{a}_{ij}} \qquad (2.1)$$

where a_{ij} = Observed frequency, $\hat{a}_{ij}$ = Expected frequency

Logistic regression

Logistic regression is an analysis method that represents the odds ratio logarithm, which is presenting the relationship between determinants and binary outcome. It provides the estimated odd ratio and confidence interval for a specific combination of risk factor (McNeil, 1996). Predictor variable sets of $x_1, x_2, \ldots, x_n$ and the outcome y. The logistic regression model is shown below:

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \sum_{i=1}^{k} \beta_i x_i \qquad (2.2)$$

where

P = The southern unrest probability

x = The determinants set

α = The constant coefficient

β = The regression coefficient set

k = The predictor variables' number

The outcomes' probability $y = 1$ as defined by

$$P[y = 1] = \frac{e^{\alpha + \sum_{i=1}^{k} \beta_i x_i}}{1 + e^{\alpha + \sum_{i=1}^{k} \beta_i x_i}} \qquad (2.3)$$

The measurement of the association arising from the contingency table 2×2, given that both of victims and dead victims of the unrest events had only two possible outcomes of x, which 0 is referred to non-victims or no-dead and 1 is referred to victims or dead of the unrest events. Therefore, the logistic regression model that given the outcome is equal to ($x = 1$) then

$$\ln = \left(\frac{P(Y = 1 \mid X = 1)}{1\text{-}P(Y = 1 \mid X = 1)}\right) = \alpha+\beta \quad (2.4)$$

While the logistic regression model of the presented outcome is equal to ($x = 0$) then

$$\ln = \left(\frac{P(Y = 1 \mid X = 0)}{1\text{-}P(Y = 1 \mid X = 0)}\right) = \alpha \quad (2.5)$$

After the exponential in the Equation 2.4 and 2.5, odds ratio for victims or dead and non-victims or no-dead can be applied as $e^{\alpha+\beta}$ and e^{α}, respectively. Thus, the odds ratio is represented then

$$OR = \frac{e^{\alpha+\beta}}{e^{\alpha}} = e^{\beta} \quad (2.6)$$

Comparing models

Adequate checking of the logistic model after removing a subset of the variable from the model will use the value of the likelihood function of the model (Woodward, 2005). The representation of adequate maximum likelihood value of sub-model and full model that defined as L_{sub} and L_{full}. The theory of standard likelihood implies that if the sub-model is adequate, then the difference $2\log L_{full} - 2\log L_{sub}$ will have an approximately distribution of χ^2_d where d is denoted the number of the removed variables, if the removed variable is a categorical variable with multiple categories, d is denoted the number of categories. The difference of deviance can be defined as the following:

$2\log L_{full} - 2\log L_{sub}$ = deviance of sub model – deviance of full model (2.7)

The increasing of the deviance after removing the d terms from the model that expresses the difference and it will always be positive. If the increase is small, the

removed extra terms will not increase much the deviance then. Thus, the variables can be removed to get model simpler. If the removed variables are not needed in the model, the distribution of χ_d^2 has approximately difference of the deviance. Hence, the comparison of this difference in the deviance to a χ_d^2 distribution, a p-value is calculated to examine the hypothesis of the removed variables. A small p-value gives an evidence alongside the sub-model that designates that all d of the variables should not be removed from the models.

Receiver operating characteristic (ROC) curve

The ROC curves represent the correct and incorrect model prediction by calculating the proportion of positive outcomes (Fan et al., 2006). The area under the curve (AUC) is a standard method that used to test model accuracy. The AUC value that near to 1 represents the good separability measurement. The value of AUC near to the 0 shows the worst separability measurement (Narkhede, 2018).