

## Chapter 2

### Methodology

This chapter describes the methods used in the study. These methods include the following components.

- (a) the study design;
- (b) the data management;
- (c) the methods used for the statistical analysis;
- (d) the methods used for creating the geographical maps.

#### 2.1 Study Design

This research is to study the spatio-temporal distribution of rainfall in Pattani province over the last 20 years, from 1982 to 2001 inclusive. The study design was of cross-sectional type, based on the sampled observations of daily totals (in millimeters) recorded at 14 meteorological stations in the 12 districts of Pattani Province. The data for the first 19 years were used to develop the model specification using the one-way analysis of variance method applied to transformed five-day totals. The model was mapped using a geographical information system (GIS) package to create range, grid, and contour maps. The data for the year 2001 were also mapped using these methods.

#### 2.2 Management of Data

The following computer programs were used for data analysis and thesis preparation.

##### *Microsoft Excel*

This program was mainly used to manage the data for this research. It has functions that are helpful in managing data.

##### *EcStat*

EcStat is an add-in to Microsoft Excel. It is a suite of routines for graphing and analysing statistical data using an IBM-compatible PC.

EcStat was developed because there is a need for both specialists and non-specialists in statistics to be able to analyze data easily and to display the relevant results in a concise, appropriate way showing the information content of the data.

EcStat was mainly used in 'Comparison' and 'Relation' commands. The result of 'Comparison' is a one way analysis of variance (anova), and the result of 'Relation' is a scatter plot graph with a fitted linear relation shown on the plot.

### 2.3 Methods used for the Statistical Analysis

The graphical methods were presented and obtained by using EcStat 2002, as follows:

1. Histograms with statistical summaries of raw data for all variables represent the distribution and summaries including the sample size, mean, standard deviation, minimum and maximum of a set of data.
2. Two-sample t-test , one-way analysis of variance and multiple regression analysis of the variables described by scatter plot and 95% confidence.

#### *Univariate and Bivariate Summaries*

The mean and standard deviation (SD) are used to summarise the data for a single variable. They are calculated from the formulas

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \text{ and } S.D = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

#### *One-way Analysis of Variance*

One-way analysis of variance is the method used for the analysis of data in which the outcome is continuous and the determinants is categorical. The null hypothesis states that the samples have arisen from the same population. This null hypothesis can be tested by computing a statistic called the  $F$ -statistic and comparing it with an appropriate distribution to get a  $p$ -value. Suppose that there are  $n_j$  observation in sample  $j$ , denoted by  $y_{ij}$  for  $i=1, 2, 3, \dots, n_j$ . The  $F$ -statistic is defined as follows (McNeil, 1996, page 67).

$$F = \frac{(S_0 - S_1)/(c - 1)}{S_1/(n - c)},$$

where

$$S_0 = \sum_{j=1}^c \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$$

$$S_1 = \sum_{j=1}^c \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

and

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^c \sum_{i=1}^{n_j} y_{ij}$$

$$n = \sum_{j=1}^c n_j$$

$S_0$  is the sum of squares of the data after subtracting their overall mean, while  $S_1$  is the sum of the squares of the residuals obtained by subtracting each sample mean. If the population means are the same the numerator and the denominator in the  $F$ -statistic are independent estimates of square of the population standard deviation and the  $p$ -value is the area in the tail of the  $F$ -distribution which  $c-1$  and  $n-c$  degrees of freedom. The statistical assumptions are that the data arise from normally distributed population.

#### *Two-way Analysis of Variance*

The response variable is classified by each of the two factors (station and year). Thus  $y_{ij}$  denotes an observation at station  $i$  on year  $j$ .

Two-way analysis of variance is used to compare the means of a continuous outcome variable between different values of a categorical determinant, after adjusting for the effect of another categorical determinant. The  $p$ -value is based on an  $F$ -statistic defined as follows (McNeil, 1996, page 73).

$$F = \frac{(S_2 - S_{12})/(c-1)}{S_{12}/(n-c-r+1)},$$

where  $S_2 = \sum_{j=1}^c \sum_{i=1}^r (y_{ij} - \bar{y}_{i+})^2$ ,  $S_{12} = \sum_{j=1}^c \sum_{i=1}^r (y_{ij} - \bar{y}_{i+} - \bar{y}_{+j} + \bar{y})^2$ , and  $\bar{y}_{i+}$ ,  $\bar{y}_{+j}$  are

estimates of row and column means, and  $\bar{y}$  is the overall mean of the data.

These formulas assume that there is the same number of observations in each cell of the data table. If this assumption is not met, some adjustment is needed, and the E-M algorithm is used to estimate the row and column means (Dempster et al, 1977, page 1-38). The assumptions required for two-way analysis of variance are as follows:

- (1) The errors have constant variation;
- (2) The errors are independent and normally distributed

If assumption (1) is not satisfied it may be necessary to transform the data before the analysis. If assumption (2) is not met it may be necessary to remove outliers

## 2.4 Methods for Creating Geographical Maps

*Software used:* MapInfo 6.5, Vertical Mapper.

To begin with, you can use the data you already have, in the form it's already in, such as a spreadsheet or database or other GIS application. If your data are stored in a remote database, you can access it directly from MapInfo. If you have data that is not already online, you can create database files inside MapInfo, or use data supplied by MapInfo, such as census data.

Similarly, you can use maps available from MapInfo, such as street and highway maps or world maps. You can also create your own maps, either in MapInfo or with a drawing package. Anything that can be diagrammed – floor plans, flow charts, even brain anatomy – can be treated as a map and entered into MapInfo.

After you've organised your data visually, you can save the results to files, or send them to a printer.

### *Range and grid maps*

A thematic map is a type of map that uses in a variety of graphic styles (e.g., colours or fill patterns) to graphically display information about the map's underlying data.

Thus, a thematic map of sales territories might show one region in dark red (to

indicate the region has a large number of customers), while showing another region in very pale red (to indicate the region has relatively few customers).

A range map is a type of thematic map that displays data according to ranges set by the users. The ranges are shaded using colour or patterns.

A grid map is a type of thematic map that displays data as continuous colour gradations across the map. This type of thematic map is produced by an interpolation of point data from the source table. A grid file from the data interpolation is generated and is displayed as a raster image in a map window. The steps in creating a range or grid map are as follows.

1. Register the raster map by connecting it to the corresponding MapInfo map.
2. Draw the district or region boundaries for the map and thus create a MapInfo table containing these boundaries.
3. Join the boundary table to the data.
4. Create thematic maps (range maps and grid maps).
5. Add any extra features required, such as point of station and name of station.

#### *Contour map*

A contour map is a way of graphing a variable that takes values in a two-dimensional region. It's not the only way of graphing such data: you could also use a 3D wire-frame map or a prism map.

MapInfo can produce grid maps, but not contours. The third-party plug-in, Vertical Mapper, provides the tool. In Vertical Mapper, contour maps have to be created from a grid map. The software has two types of grid, **.grd** and **.grc** files. It can also import a **.mig** file, which is the grid format created by MapInfo.