Chapter 1

Introduction

What is your reaction to a headline such as: "Cancer mortality rate appears to be higher along downstream of contaminated rivers in Taiwan area"? If the statement is true and you neglect it, it may endanger your health; however, if it is not true and you overanxious about it, you are receiving unnecessary false alarm.

For the sake of personal health, there will be no doubt that the general public has a great concern about the locations where people are easier to be infected with diseases of lethality such as cancers. The spatial pattern of diseases is always of interest to both epidemiologists and the general public, because it can link up disease incidence with suspected agents or environment factors. For instance, Openshaw (1988) investigated cases of childhood leukemia in the Northern, Northwestern and part of the Mersey Regional Health Authority Districts of England in the period 1975-85 and inferred that the causes of leukemia are probably to be environmental factor, perhaps pollution related, other than radiation.

Specifically, areas of high disease incidence are always referred as "clusters." According to Knox (1988), a cluster is defined as "a bounded group of occurrences related to each other through some social or biological mechanism, or having a common relationship with some other event or circumstance." In general, we divide the study for spatial pattern of diseases into two categories: tests of clustering and tests for detection of clusters. In tests of clustering, we are focus on investigating whether one or more geographical regions have a tendency of observing more disease cases than they are supposed to. There are three types of clustering: clustering in space, clustering in time or clustering in both. On the other hand, in tests for detection of clusters, we aim at determining whether disease cases gather together in some regions, such as along downstream of contaminated river in Taiwan area in the first paragraph.

In the past, several methods have been proposed to detect clusters. However, most of them use “overlapping circles” or “circular windows” to detect clusters, so it may have lower power to detect non-circular clusters. In most of the studies, they use Type I and Type II errors to measure the susceptibility of cluster detection methods.

In this study, first of all, we proposed an approach based on the idea by Choynowski (1959) for tests of clustering. Then, we conduct Monte Carlo simulations to detect possible clusters based on the information of clustering. The main difference between our method and previous methods is that the proposed method of cluster detection does not require the shape of clusters close to circles. We will use simulation to check if our method, as well as other method, is effective in detecting (both circular and non-circular) clusters.

In Chapter 2, we review past work for test of clustering and detection of clusters. In Chapter 3 we provide the main idea of the thesis that shall give readers a full picture of our method. In Chapter 4, we present testing results from Monte Carlo simulations to check the validity of the proposed method. Then, we compare our method with Nagarwalla’s Spatial Scan Statistic on synthetic data sets in Chapter 5. Next, in Chapter 6, we apply our method on the 2001 Cancer Mortality Data in Taiwan as a demonstration to check if there are unusually high mortality rates of cancers in Taiwan.

Note that in spatial analysis, neighborhood information often plays an important role. In this study, because we deal with aggregate data in each township, we define two townships that share a common border as neighbors. The term “cluster” in the empirical studies will be defined as a region consisting townships whereby cancer mortality rates are excessive compared with the rest of the study area. Discussion of our proposed approach from this study and potential extensions of our method are given in Chapter 7.