**The Association Between Urbanization and Water Chemistry Parameters**

G11 AP Statistics

Cheng (Paul) Ai

**Abstract**

**Content**

1. Introduction

Since the end of two Industrial Revolutions, the development of human technology reaches to a peak in around 20 years. This development produce lots of convenience for human activities on excavate and use of natural resources, and also brings a completely new life for human in that time. However, the negative by-product of Industrial Revolution soon comes to people’s mind, which is the defection on environment. The deterioration on the environment thus soon become a very serious problem in development of technology, and the progress of urbanization or agriculturalization and some other humanize progresses.

Long in the end of last century, American and Canada firstly revised environmental protection law for restrain the further deterioration of natural environment brought by industrialization and urbanization, and investigated some supervising methods to quantify the pollution by some certain index, in order to solve the problems brought by human activities.

This passage aims to analyze a certain case of urbanization affecting on nearby river system by using statistical evidence and simple regression model, and thus popularize the result to the normal river system in the world, therefore introduce a way to protect the environment.

2. Background Information

Urban Stream Syndrome stands for the “illness” of near-water environment due to the pollution of urbanization or agriculturalizing or some other natural effect of land use which defect the quality of the river system, such as eutrophication, or direct pollution to the river [1]. These defections will influence the whole water system which spread to the nearby urban city or residential area. Therefore, supervising quality of water system becomes an important task for every country. For a particularly water system, in order to quantify its quality, we choose some measurable quantity and create a certain mechanism to evaluate the quality of water, and give a final response of how those land use patterns may influence the river system, in order to make the best use of land without causing lots of serious environmental issue.

In the recent study, scientists have found that the effects of urbanization on river system based on system structure, peak flows or some other visible quantities. To illustrate, research done around Taihu Region, China indicated the relation between water surface area and population growth [2], another research done around San Antonio River Basin in Texas showed the relation between peaks flow and urbanization degree [3] … These researches firstly introduce the analysis of urbanization into account of river system defection in macroscopic view. However, there is lack of analysis from micro-view of water system, which called water chemistry parameters. These parameters include chemical properties such as metal or ion concentration, which create a complete evaluation system or index to help government supervise water quality [4].

3. Experimental Design

This passage mainly focuses on determine relation between urbanization degree, using environmental usage to predict, and water quality, which indicate by water chemistry parameters. Therefore, I create an observational study, from collecting data to data analysis and result summary.

**3.1 Data Collection**

Data was adapted from the UK Government River Water Quality Monitoring 1990 to 2018[5]. In this database, it provides a large amount of raw data collected from a certain river system in England, specific shows in the area below in Figure 1.

**3.2 Variable Analysis**

I choose land use patterns as urbanization index, thus set it as explanatory variable, and chemistry parameters, such as gases dissolved, ion concentration, as response variable. However, other aspects, natural environment, particularly states as primary basin, and time difference will be the confounding variable, which will be discuss specifically in method and procedure.



Figure 1. Data Covered Area

4. Procedure & Method

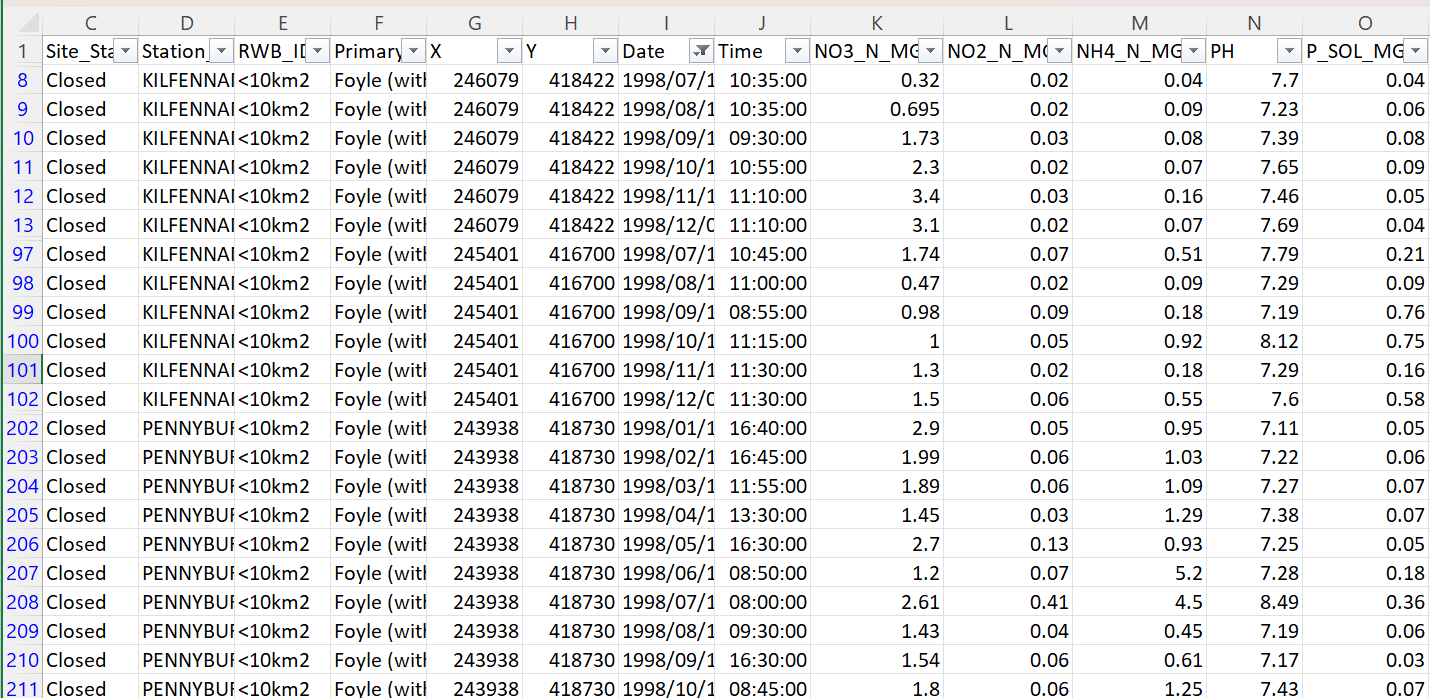
**4.1 Random Sampling Methods**

As stated above, since there are two confounding variables in the raw data, which is the “Primary Basin” and time difference. Therefore, here I firstly conduct a stratify sampling method in order to eliminate the effect from confounding variable, which is the difference between each different natural environment, since it will affect concentration of certain substance in the river, which will influence the determination of water quality based on water chemistry parameters.

Based on the raw data, I assign 56 stratums, from 1-56, which determined by how many “Primary Basin” has been supervised. After that, I filtered them by how many data in it in order to satisfy the condition of scope of research, the result is showed in Table 1. Since the scope of inference need to be larger enough for popularizing the result, and some invalid data in each stratum, I only choose strata that has larger than 1000 data, which is enough for analysis even regard with invalid data due to some unexpected reasons.

Table 1. Numbers of Data in Each Stratum

Therefore, there are 39 stratums that satisfy the condition. After assign and choose suitable stratums, a cluster sampling method will be again to use in order to eliminate the effect of time. I choose timeline from 1998 to 2005, each year as a cluster, example cluster, year of 1998 is showed below in table 2, and analysis data in it.

****Table 2. Example Cluster: Year 1998

**4.2 Parameters Selection**

In the raw data, it provides 14 water chemistry parameters, for example, *Pondus Hydrogenii* (pH), ammonia concentration and some other water quality index, showed in Table 2 above. By choosing 5 available parameters (available means there is data for every choosing sample in the cluster), the response variables are set.

5. Data Analysis

**5.1 The Relation Between Urbanization Degree and Population Growth**

This is the background data analysis in order to determine the explanatory variable, urbanization degree. Since the raw data does not provides anything about the urbanization degree, according to urbanization indicator index, I choose population growth as an essential indicator for estimating the degree of urbanization. Here, we need firstly to introduce a concept: population urbanization index, which directly manifest the urbanization degree. Although population urbanization index is calculated by the formula, , in which denominator and nominator stand for total population and non-agricultural population, respectively, in this circumstance, we ignore the effect on agricultural population in the area, since all the areas covered and data recorded are inside city, which means that , therefore population urbanization index is 1, showing that urbanized population and urbanization degree is proportional to each other in this area, thus to say they satisfy equation of  [6].

According to the population data in the shaded region (Northern Ireland), shows in the figure below [7], suggests that the population data has an approximately positive linear relationship to the year. Therefore, it can be assumed that the urbanization degree has a similar linear relationship with year.

**5.2 The relationship between Land Use Patterns and Urbanization Degree**

Since the linear relationship between population growth and urbanization degree, and the linear relationship between population number and years, we can eliminate the effect of population on urbanization degree if we assume that population indicator in each year has the same value, which stated below, as the slope of the population-year graph. Therefore, we only need to find the relation between land use patterns and urbanization degree by a certain mechanism, which based on the function of the area, for instance, roads, parks, bridges and so on.

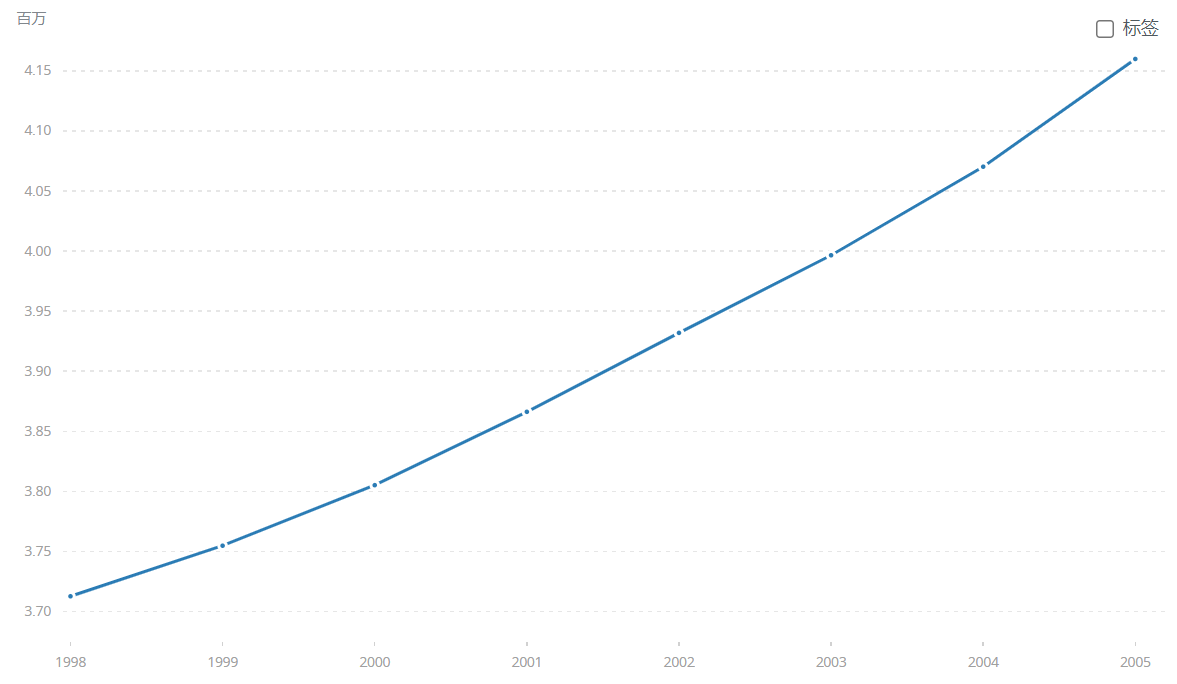


Figure 2. Population Growth Curve From 1998 to 2005 in Northern Ireland

From each cluster, we can find that the station positions are mainly composed of roads, bridges and parks, obviously, two of them are for urbanization use and another is use for greening.