TWO VARIABLE REDUCTION TECHNIQUES IN STUDYING THE EFFECT OF POLLUTION IN THE HEALTH OF THE POPULATION IN ONE OF THE INDUSTRIAL AREAS

Etleva (Llagami) Beliu, Kleida Haxhi, Oriana Zacaj
Department of Mathematics, Faculty of Mathematics and Physics Engineering, Polytechnic University,
e.llagami@fimif.edu.al

ABSTRACT

Air pollution is now fully acknowledged to be a significant public health problem, responsible for a growing range of conditions that are well documented as a result of extensive research conducted in many regions of the world.

The environmental pollution in the Patos-Marinza and the Kucova fields in the southern region of Albania, being the most important areas in the local petroleum industry, is the subject of our studies. The data about health problems in general, those in respirator and in dermatology is taken from five sites: Ballsh-Marinza, Sheqishte-Beline, Zharre, Patos and Kurjan; all located not far from each other.

The use of the varimax rotation (Analysis of Principal Component) and the ANOVA method reveals that pollution is responsible for about 55% of the changes occurring to the eigenvectors.

Then Factor Analysis of data, promax, shows the effect of this pollution on three specified components: 44% of the cases of medical issues can be observed as a change in general conditions, 35% in respiratory and 22% in dermatological problems.

Keywords: variable reduction , pneumonology, dermatology, air pollution

INTRODUCTION

Air pollution is now fully acknowledged to be a significant public health problem, responsible for a growing range of conditions in humans. Sulfur dioxide (SO_2) is a dangerous colorless gas, whose predominant form is found in the lower atmosphere. This gas is found in greatest amounts near industrial areas, especially those of oil tankers. In our country, areas of Kucova and Marinëz are well known oil-bearing areas (Figure 1).

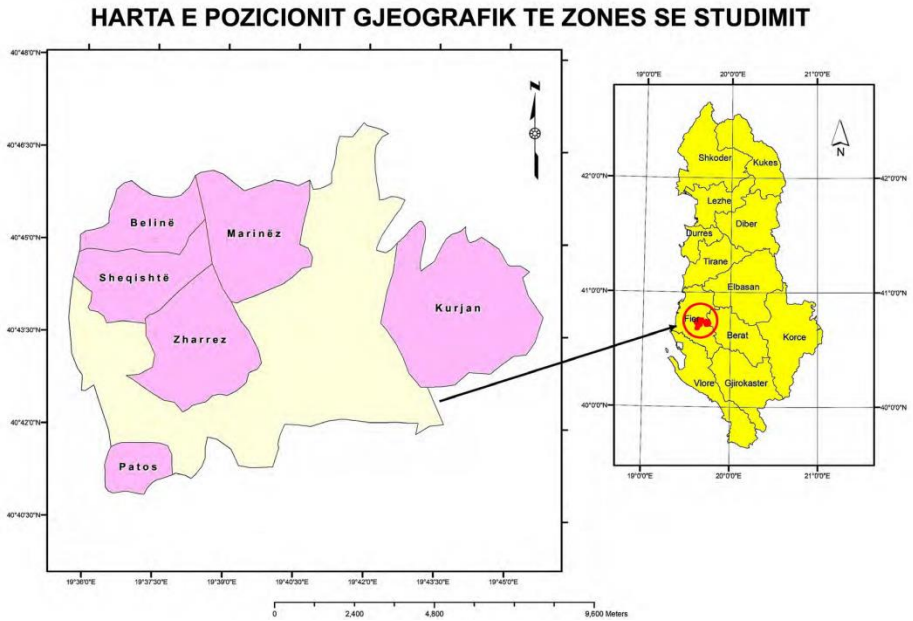

Figure 1. Geographical posicion map of sites

The environmental pollution in the Patos-Marinza and the Kucova fields in the southern region of Albania (being the most important areas in the local petroleum industry) is the subject of our studies. The data about health problems in general, those in pneumonology and in dermatology is taken from five sites: Ballsh-Marinza, Sheqishte-Beline, Zharre, Patos and Kurjan; all located not far from each other.

Using ArcGIS software, the database of all oil wells across our country helps show that the first four sites can be considered as exposed to SO2 and the last one as protected from SO2 .

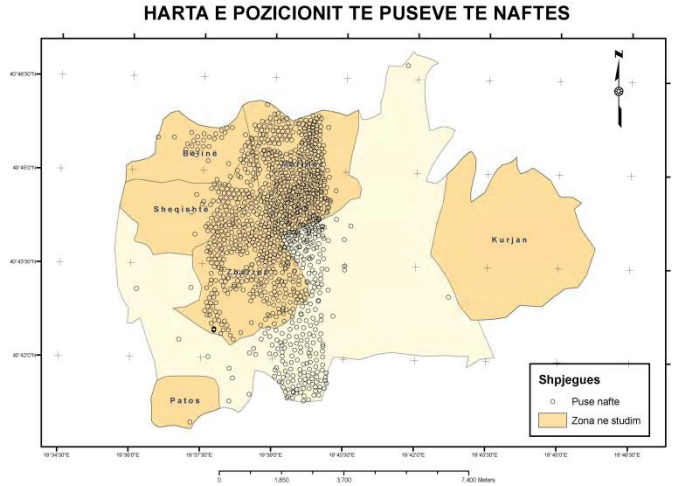

Figure 2. Oil-wells posicions

The **factors** able to cause **respiratory and dermatologic problems** among habitants are listed by **Health Public Institution** and regarding that, the questionnaire is formed.

A brief medical history was taken for each of 171 people of each site. This history consisted of age, sex, length of residence at current and previous address, history of respiratory problems (5 questions), dermatological problems (5 questions) and general health questions (3 questions).

- The general health questions regard *dizziness, vomiting, headache.*
- Questions for dermatological problems regard

skin nodule, itchy skin, rash,
skin redness and skin inflammation.

- While questions on respiratory problems regard

watery eyes, itchy throat, dry throat, phlegm and cough.
Thirteen variables are defined by coding of samples' responses, which, as seen, are dependent variables.

STATISTICAL ANALYSIS

We used two methods to study the effect of 'villages' on the 13 medical health questions.

Method of **Principal Component analysis** (PAC) followed by varimax rotation is performed, that converts the set of observations of possibly correlated variables into two sets of values of linearly uncorrelated variables.

This method of data reduction reduces our 13 measures to a few principal components. As a result, we took two sets of eigenvalues, whose corresponding eigenvectors were at right angles to each other.

Root No.	*Eigenvalue*	*Pct.*	*Cum. Pct.*	*Canon Cor.*
1	*1.25375*	*100.00000*	*100.00000*	*.74585*
2	*2.3E-5*	*.00000*	*100.00000*	*.00021*

- The first set of eigenvalues, gives us the total variances of 13 variables, so that the first component explains $1.25375 0.74585^2 = 0.556\%$ of the total variance.

'super_impact=-.20647* *watery eyes* +.00410* *itchy throat* -.5260* *dry throat*

+ .54939* *phlegm* – .10457* *cough* + .90458* *skin nodule*

+ .00185* *itchy skin* + .19729* *rash*

+ 1.58608* *skin redness* +.28266* *skin inflammation*

+ .64448* *dizziness* +.16127* *vomiting* + 1.20061* *headache*

- The second set of eigenvalues, gives us a second eigenvector, but the total variances of 13 variables of this second supervector explain a small part of the total variance, because the sum of eigenvalues is only 2.3E-5

Note that the result of this method is only for a simple presentation, since all values of 13 variables must be measured in order to compute the value of principal components.

After that we use ANOVA to see the effect of 'village' on 'super_impact'.

The means of industrial villages are not significantly different, but they are different from the means of other non industrial villages, Kurjan.

The result of ANOVA method gives indication, that 54.7% of variance of variable 'super_impact' is explained by the variable 'village'(F is 101.28 and p-value <.001)

FA estimates how much the factors affect the existing variables. For this large datasets, *Promax Rotation*, an oblique rotation which allows factors to be correlated, is used.

- The results show that 13 variables seems to measure three specific factors. This happens because, the values of their eigenvalues are greater than 1 only for the 3-4 first components (Figure 3).

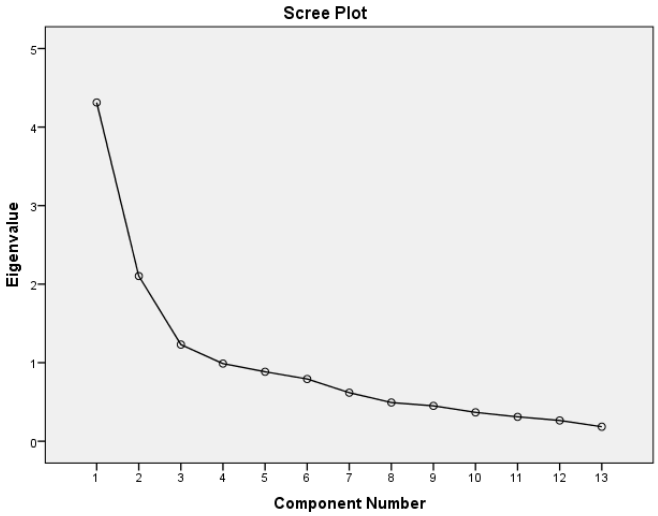

Figure 3. Plot of Eigenvalues

These three specific factors explain 60% of the total variance.

- A careful analysis of the given values of the coefficient before the 13 variables for the three new components shows that:
  - The records for the first are higher near the dermatological variables, that is why we call it *dermatological_component.*

Dermatological_component = .324* *watery eyes* +.460* *itchy throat* +.312* *dry throat* +.142* *phlegm* +.062* *cough* +.834* *skin nodule* +.773* *itchy skin* +.872* *rash* -.020* *skin redness* + + .692* *skin inflammation* -.168* *dizziness* (1)

+.161* *vomiting* + .256* *headache*

This variable justifies 33% of the total variance.

- The second factor involves common condition variables. So we called it *general_component*

General_component = .542* *watery eyes* +.273* *itchy throat* +.614* *dry throat*
+.098* *phlegm* +.063* *cough* +.026* *skin nodule* +.165* *itchy skin* (2)
+.104* *rash* +.437* *skin redness* + .108* *skin inflammation*
+.830* *dizziness* +.576* *vomiting* + .770* *headache*

These specific factors explain 50% of the total variance.

- While the third factor includes variables of pneumological conditions.

Respiratory_component=.122 watery eyes +.178* itchy throat*
+.405 dry throat +.730* phlegm +.783* cough*
-.084 skin nodule +.034* itchy skin*
-.068 rash +.600* skin redness + .310* skin inflammation*
-.104 dizziness +.275* vomiting + .167* headache*

Because of this we named it *respiratory_component*.

Using ANOVA method, we see that the effect of *village* variable is significant at all these three components. But the change of this variable justifies only 44% of the variance of *general_component,* 35% of the change of *respiratory_component,* and 22% of *dermatological_component***.**

CONCLUSIONS

Two methods of rotation Varimax and Promax have the ability to recover the underlying factor.

But oblique method is better able to identify the presence of a "simple structure"

Oblique method suggests how much factors are maesured by a large datebase and which questions are associated with which factors.

Factor Analysis , Promax helps us to conclude:

- This study documented significant associations between exposure to air pollution and some general health problems, based on samples taken in the human population living near a polluted zone.
- So, statistically significant changes are seen in respiratory parameters, from industrial sites to nonindustrial ones.
- A (statistically significant) effect of the air pollution was noted in dermatological problems.

Principal Component Method is a useful method

- That suggests the use of one single variable, 'super_impact'.
- The method of Principal Component, tells us that the pollution affects 55% of the changes of general conditions, represented by supervariable -'super_impact"

Two methods have the ability to find the correctly link items to factors.

Finally:

- The presence of these associations with SO_2 draws attention to the fact that environmental protection concerns should be considered a top priority for primary prevention of the diseases, too.

REFERENCES

1. Chen JC, Schwartz J. Metabolic syndrome and inflammatory responses to long-term particulate air pollutants. Environ Health Perspect.(2008);116:612
2. Steinvil A, Kordova-Biezuner L, Shapira I, Berliner S, Rogowski O. Short-term exposure to air pollution and inflammation-sensitive biomarkers. Environ Res.(2008); 106:51-61
3. Forbes LJ, Patel MD, Rudnicka AR, Cook DG, Bush T, Stedman JR, et al. Chronic exposure to outdoor air pollution and markers of systemic inflammation. Epidemiology.(2009);20;245-253
4. VMR Muggeo (march 2004) Statistics in Medicine**,** Volume 23, Issue 7