The use of statistical tools for the demonstration of effects of polluted environment on congenital defects and spontaneous abortions in the Czech Republic

Loster Tomas[1]

[1] University of Economics, Faculty of Informatics and Statistics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, losterto@vse.cz

Abstract. The environment is being polluted as a result of developed industrial production and energy industry. Every region of the Czech Republic has been affected by industrial activities on different level and therefore the quality of the environment is different in each of 14 regions of the Czech Republic. The objective of this article is to prove that pollution of the environment certainly affects, among others, the spontaneous abortions and congenital defects. One way ANOVA has been used to analyze the spontaneous abortion rate. Statistical system STATGRAPHICS Plus has been used for analysis of variance. Development of the spontaneous abortions in all Czech regions in the period 2000 – 2005 is part of this analysis.

Key words: congenital defects, spontaneous abortion, environment, one-way analysis of variance

1 Introduction

Congenital defects and spontaneous abortions can be consequences of adverse environmental effects, among others of those of the polluted atmosphere. In attempts to demonstrate effects of the environment condition on spontaneous abortions and congenital defects, is suitable to use dividing of the Czech Republic territory into particular territorial areas – regions. Depending on the condition of the environment in particular regions, given *inter alia* by the general level of the industrial production and power engineering industry, it is possible to investigate effects on quantities of interest. For the analysis of effects of environmental conditions on these quantities, it is suitable to determine the measure of spontaneous abortions as a ratio of the number of spontaneous abortions to the number of pregnant women, where the number of pregnant women can be established as a sum of numbers of spontaneous abortions and numbers of living children delivered. The measure of spontaneous abortions is henceforth referred to as *measure of SA.* The quality of the atmosphere in particular regions is examined based on levels of different substances present in the atmosphere. These are e.g. the following substances: CO, O_3, SO_2, etc.

The target of the present contribution is to demonstrate that the polluted atmosphere affects both the congenital defects and spontaneous abortions and thus, regions having larger capacity of the industrial production and of the power engineering industry (as e.g. the Region Ústí nad Labem and the Moravskoslezský kraj) exert an enhanced measure of spontaneous abortions and congenital defects.

2 Congenital defects

It is to expect that congenital defects (CD) will be adversely involved by the polluted environment. For the quantification of consequences of effects of the polluted atmosphere on congenital defects, different extents of the congenital defects will be considered, which are summarized in Table 1 in a structure by regions, and appropriate tests will be carried out, which will verify the assumption of environmental effects on the congenital defects. The processing of data by these tests is summarized in Table 2.

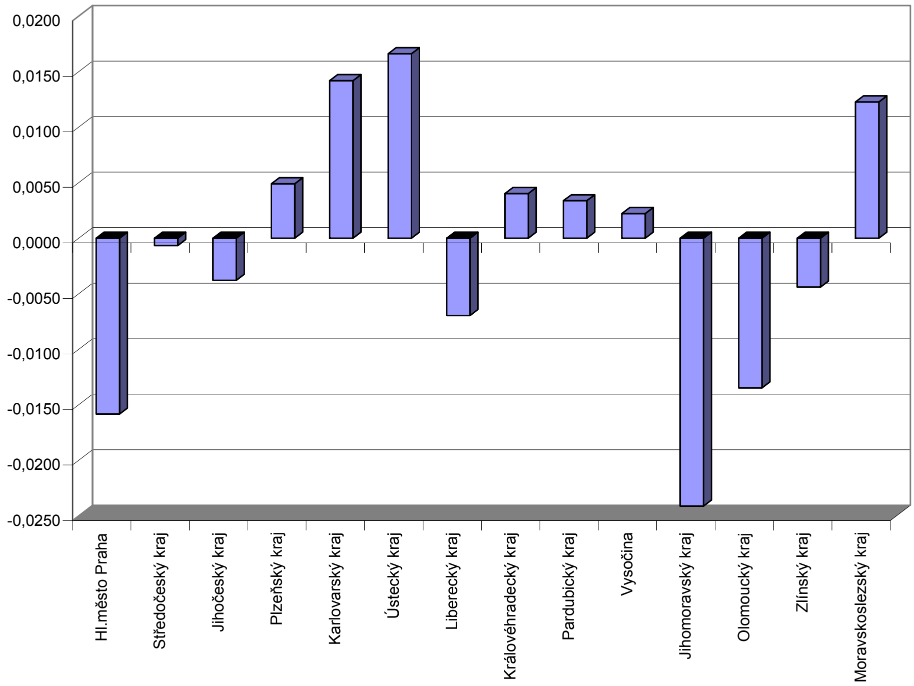
The first column of Table 1 comprises particular territorial areas of the Czech Republic – regions. The second column presents ratios of particular regions to the total numbers of selected congenital defects (as e.g. lip cleft, cystic renal disease, congenital diaphragmatic hernia). In the third column, there are contributions of particular regions to the total number of all the congenital defects. The fourth column presents the contribution of each particular region to the total number of living children delivered (DA). The fifth column contains differences between the contribution of each region to the total number in terms of all the congenital defects and delivered living children (i.e. differences between third and fourth columns). In the fifth column (and subsequently in Graph 1), it is possible to find, what regions exert higher contributions to congenital defects compared with their contribution to children delivered alive. For example, it is obvious that the Region Praha brings a lower contribution to the congenital defects compared with its contribution to the total number of children delivered. In contrast, the Region Ústí nad Labem exerts a larger contribution to the congenital defects compared with the numbers of children delivered alive. This suggests that the congenital defects depend on the territorial area (region). Given the fact that different regions have different levels of the substances examined in the atmosphere, it is obvious that the atmosphere (environment) affects the measure, and of course also number, of the congenital defects.

Table 1: Proportions of congenital defects (CD) and children delivered alive to total numbers of children delivered alive (DA)

Regions	Ratio of Selected CD	Ratio of All CD	Ratio of DA	Differences between all CD and DA
Hl.město Praha	0,0858	0,0948	0,1106	-0,0158
Středočeský kraj	0,1049	0,1129	0,1136	-0,0006
Jihočeský kraj	0,0650	0,0567	0,0604	-0,0038
Plzeňský kraj	0,0527	0,0581	0,0532	0,0049
Karlovarský kraj	0,0422	0,0453	0,0311	0,0142
Ústecký kraj	0,0915	0,1044	0,0877	0,0166
Liberecký kraj	0,0451	0,0366	0,0436	-0,0070
Královéhradecký kraj	0,0536	0,0577	0,0537	0,0040
Pardubický kraj	0,0533	0,0527	0,0494	0,0034
Vysočina	0,0606	0,0525	0,0503	0,0022
Jihomoravský kraj	0,0966	0,0838	0,1079	-0,0241
Olomoucký kraj	0,0569	0,0474	0,0609	-0,0135
Zlínský kraj	0,0496	0,0514	0,0557	-0,0044
Moravskoslezský kraj	0,1356	0,1343	0,1220	0,0123

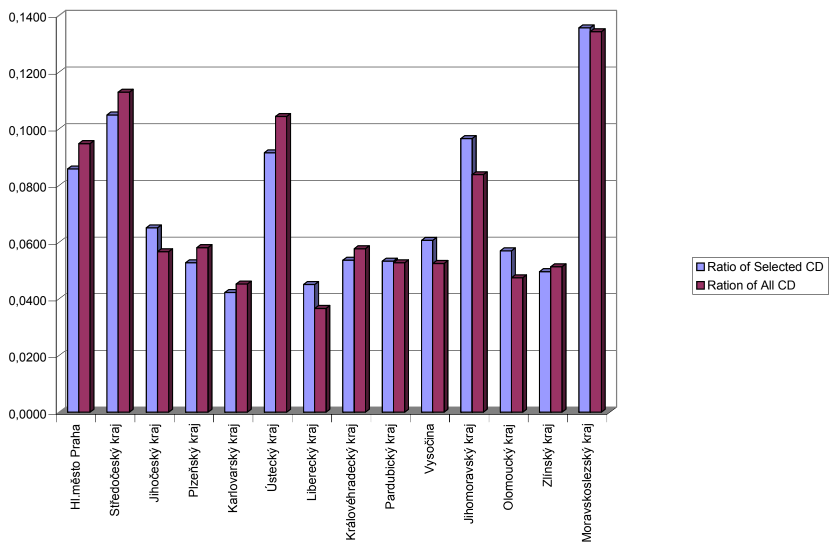
Graph 1 shows a representation of the last column of Table 1. From this graph it is obvious at the first sight, what regions of the Czech Republic have larger proportions of congenital defects compared with the proportion of the children delivered alive to the total numbers. These are the following regions: Plzeň**, Karlovy Vary, Ústí nad Labem,** Hradec Králové, Vysočina and **Moravskoslezský kraj.** The regions printed in bold exceed +1% in the difference of proportions of congenital defects and children delivered alive to the total numbers. The effect of the polluted atmosphere on congenital defects is obvious when considering the list of the regions.

Graph l: The difference between proportions of all the congenital defects and numbers of children delivered alive



Graph 2 depicts proportions of particular regions to the total number of congenital defects (considered based on the choice and all congenital defects). Thus, it concerns second and third columns of Table 1. It also supports the conclusion resulting from Graph 1 and it indicates that e.g. the Moravskoslezský kraj takes one of fore positions in terms of the proportion (i.e. also of the number) of congenital defects. This fact results, among other factors, from the environmental pollution. It is obvious from the graph that the contribution of the Moravskoslezský kraj to the total number of the congenital defects is of about 14%.

Graph 2: Proportions of congenital defects occurring in particular regions



The statement that the environment affects the measure of the occurrence of congenital defects in particular regions due to different levels of harmful substances in the atmosphere should be verified with the help of a statistical test. A test is used, which verifies an agreement between two basic proportions. The first set is formed by congenital defects and the second one is formed by the children delivered alive. The tested hypothesis alleges that the proportion (measure) of the total number of the congenital defects occurring in a certain region to the whole number is the same as the proportion of the number of children delivered alive in that region, i.e. that the region considered participates in the total measure of the congenital defects to the same extent compared with the proportion of the number of children delivered alive; in other words, this hypothesis means that the measure of the congenital effects is not affected by the environment. The values of the test criteria for particular regions, alternative hypothesis and refusal or non-refusal of the hypothesis tested are summarized in Table 2.

Table 2: Test criteria, hypotheses and conclusions of tests of the independence of the measure of the CD of the environment

Regions	Ratio of all CD (p_1)	Ratio of DA (p_2)	Value of TC	Tested hypothesis (H_0)	Alternative hypothesis	Conclusion
Hl.město Praha	0,0948	0,1106	-35,92187	refused	$\Pi_1 < \Pi_2$	better
Středočeský kraj	0,1129	0,1136	-1,415835	non-refused	$\Pi_1 < \Pi_2$	better
Jihočeský kraj	0,0567	0,0604	-14,79212	refused	$\Pi_1 < \Pi_2$	better
Plzeňský kraj	0,0581	0,0532	21,39797	refused	$\Pi_1 > \Pi_2$	worse
Karlovarský kraj	0,0453	0,0311	100,4198	refused	$\Pi_1 > \Pi_2$	worse
Ústecký kraj	0,1044	0,0877	45,1903	refused	$\Pi_1 > \Pi_2$	worse
Liberecký kraj	0,0366	0,0436	-37,34595	refused	$\Pi_1 < \Pi_2$	better
Královéhradecký kraj	0,0577	0,0537	17,41522	refused	$\Pi_1 > \Pi_2$	worse
Pardubický kraj	0,0527	0,0494	15,83255	refused	$\Pi_1 > \Pi_2$	worse
Vysočina	0,0525	0,0503	10,23728	refused	$\Pi_1 > \Pi_2$	worse
Jihomoravský kraj	0,0838	0,1079	-56,23856	refused	$\Pi_1 < \Pi_2$	better
Olomoucký kraj	0,0474	0,0609	-52,96507	refused	$\Pi_1 < \Pi_2$	better
Zlínský kraj	0,0514	0,0557	-18,54714	refused	$\Pi_1 < \Pi_2$	better
Moravskoslezský kraj	0,1343	0,1220	25,1224	refused	$\Pi_1 > \Pi_2$	worse

Table 2 *inter alia* indicates that there are distinct differences between contributions of particular regions to the numbers of congenital defects and to the numbers of children delivered alive (the second and third columns of the table) and presents an alternative hypothesis (the sixth column of the table). If the alternative hypothesis is defined as the inequality $\Pi_1 > \Pi_2$, then this statement tells that the region participates in the measure of the congenital defects to a larger extent compared to the measure of children delivered alive, which means that the environment is harmful and that it adversely affects the occurrence of the congenital defects. With the use of values of test criteria in particular tests (column 4) and with the help of critical values, the individual tests are evaluated either by refusal or by non-refusal of hypothesis H_0 (column 5). The results demonstrate that except for the Region Středočeský kraj, all the regions and their measures of congenital defects are affected by their environment. The last column of the table shows whether the condition of the environment affects the measure of the congenital defects negatively or positively. If the statement “better” is specified in the column, then this means that in the region, the proportion of the number of congenital effects is smaller than the proportion of the number of children delivered alive.

In the field of the congenital defects it is to conclude that based on appropriate calculations of measures and tests as shown in Tables 1, 2 and Graphs 1 and 2, it was possible to demonstrate a dependence of the congenital defects on particular regions. Given the differences in particular regions in terms of the atmosphere and environment quality, it is possible to conclude that the congenital defects are affected by the quality of the atmosphere.

3 Spontaneous abortions

Similarly as in congenital defects, in spontaneous abortions, it is possible to expect their involvement by adverse environment (polluted atmosphere). The analysis of variance can be used for the analysis of this assumption with respect to the nature of data. For purposes of this analysis, a quantity was established named the *measure of SA*, which is a proportion of the number of spontaneous abortions to the number of pregnant women (number of children delivered alive plus number of spontaneous abortions).

The purpose of the work was to demonstrate that larger capacity of the industrial production (as e.g. in the Region Ústí nad Labem and Moravskoslezský kraj) affects the measure of spontaneous abortions. For the analysis of the measure of spontaneous abortions on particular regions, the single-factor analysis of variance was employed, solved in the STATGRAPGHICS Plus system. Assumptions for the use of the analysis of variance were verified with the help of the Bartlett test (test of agreement of variances). Graph 3 shows that the assumption of the equality of group variances may be considered as satisfied.

Graph 3: Output of the analysis of variance from the STRATIGRAPHICS Plus system

```
Variance Check

Cochran's C test: 0,225241   P-Value = 0,0640827
Bartlett's test: 1,36287   P-Value = 0,0897009
Hartley's test: 18,0919
```

From the calculated values in the output from the STRATIGRAPHICS Plus system in Graph 4, where the dependence of the *measure of SA* depending on particular regions is verified, it is obvious that at all the commonly chosen significance levels (5% or 1%) the test hypothesis (about the equality of mean values) is refused and thus, there is at last one pair of regions, which are statistically significantly different one form another in the *measure of SA.* Thus, based on this calculation, the dependence between the measure of spontaneous abortions and atmosphere quality is demonstrated.

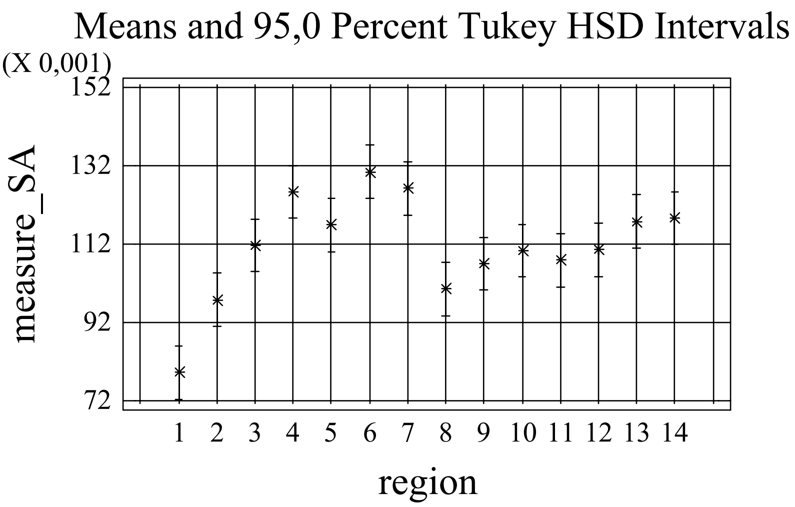
Graph 4: Output of the analysis of variance from the STRATIGRAPHICS Plus system

```
ANOVA Table for mira_SP by kraj

                            Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Between groups          0,0136653     13   0,00105117      22,99       0,0000
Within groups          0,00320073     70 0,0000457247
-----------------------------------------------------------------------------
Total (Corr.)            0,016866     83
```

Given the demonstration of the dependence of the *measure of SA* on particular regions, it is suitable to furthermore investigate what pairs of regions are statistically significantly different in terms of their *measure of SA* (at a 5% significance level) and what is the difference between them in percentage. For this comparison, the method by Tukey was employed, which was calculated again with the help of the STRATIGRAPHICS Plus system. It is possible to determine that the statistically significantly different pair is that, in which the difference in the *measure of SA* is larger than ±1.35%. The mean level of the *measure of SA* over the whole period studied (2000 to 2005) is shown in Graph 5.

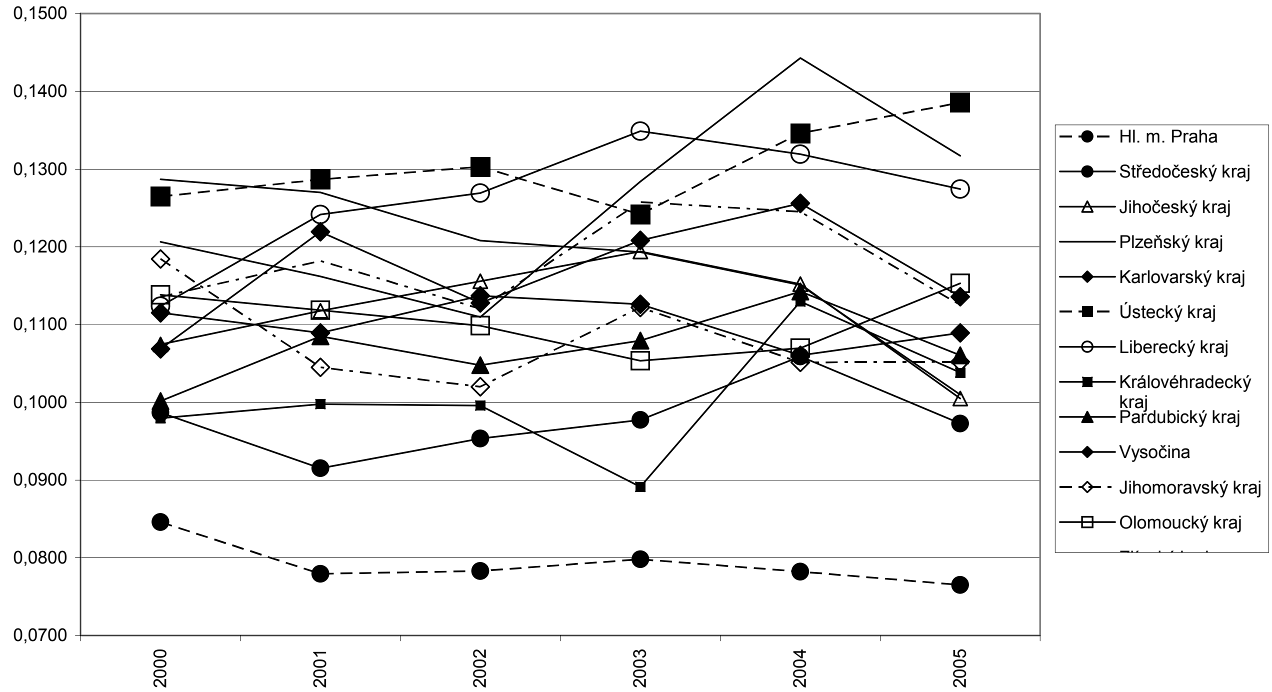
Graph 5: Mean value of the measure of SA over the whole period in the structure by particular regions



Graph 5 demonstrates that the highest measure of SA over the period of interest was in region 6 (Region Ústí nad Labem). Region Moravskoslezský kraj, denoted by number 14, is above the average in the measure of spontaneous abortions (similarly as in the case of congenital defects).

Graph 6 shows the development of the measure of spontaneous abortions in particular years in the structure by the regions. The course over 2000 to 2005 is obvious when considering the movement along the connecting line. For example, in the Region Středočeský kraj, there is a decrease in the measure of spontaneous abortions over the period considered.

Graph 6: Measures of spontaneous abortions in particular regions (development over the years)



4 Conclusion

It is possible to conclude that, based on results and graphic outputs obtained, the measure of spontaneous abortions and congenital defects (and thus also their numbers) are affected by the level of the atmospheric pollution. Particular regions are different in the level of the atmospheric pollution and this was *inter alia* manifested by the measure of spontaneous abortions and congenital defects. As mentioned above, in terms of spontaneous abortions, the Region Ústí nad Labem belongs to the worst regions within the Czech Republic.

References

HEBÁK, P. and collective (2004): Multivariate stastistical methods (part 1) (in Czech), Prague, 155-179.

JAROŠOVÁ, E., PECÁKOVÁ, I. (2000): Examples to course of Statistics B (in Czech), Prague, 25-48.

LÖSTER, T. (2007): Dependence of spontaneous abortion on environment (in Czech), Prague, 48-53.

MAREK, L. and collective (2005): Statistics for economists application (in Czech), Prague, 204-216.