

CLUSTER ANALYSIS OF THE PAP SMEAR SCREENING IN THE STATE OF PARANÁ IN 2010

Nahida Ajala de Carvalho, Clodis Boscaroli, Rosangela Villwock

Western Paraná State University, Cascavel, Paraná, Brazil
nahidacarvalho@yahoo.com.br, clodis.boscaroli@unioeste.br, rosangela.villwock@unioeste.br

Abstract: *The Uterine Cervical Neoplasm is one of the main causes of morbidity and mortality among women in Brazil and all around the world. The objective of this paper was to identify the profile of the results of cytopathological exams (Pap smears) of women aged 25 to 59 years living in 399 municipalities of the State of Paraná, Brazil, during 2010. A database was assembled using data collected in SISCOLO (Information System Cervical Cancer), which provides online information, and the cluster analysis was performed by the K-means algorithm. The analysis made it possible to divide the state into three clusters with similar characteristics between members and define the degree of neoplastic tissue injury that defines the differences. Thus, it was also possible to identify failures in the diagnosis of neoplasms and records request cytopathological exam of cervix incomplete.*

Keywords: *Cervical Cancer, Cluster Analysis, k-means*

1. INTRODUCTION

The global health has been going through countless changes on the health and disease procedures. In the past, such as in the period of the Industrial Revolution, the diseases that most afflicted the population was infectious, which were the cause of numerous deaths, due to the lack of hygiene and basic sanitation. With the epidemiological transition, the chronic-degenerative diseases took the leading cause of death inflicting the population, which is a different picture from today, since a longer life expectancy was established [12].

The cancers are among the main diseases that affect the population in different age groups and in various manifestations. Breast and cervical cancer are the main causes of death among women, both of which, when detected in its initial stages, present a prospect of cure way higher than it normally would if the cancer was detected on a more advanced stage.

The cervical cancer is a malignant neoplasm that affects women mainly on their childbearing age. It is the second most frequent type of cancer among women, with approximately 500 thousand new cases every year worldwide along with, approximately, 230 thousand deaths per year [1]. If diagnosed on its early stages, this type of cancer has a cure, but for several factors involving the diagnosis of the pathology, numerous deaths are caused by it.

In some Brazilian states, the number of cases of cervical cancer reaches higher indices than the risk stipulated by the Brazilian National Cancer Institute (INCA), not considering the non-melanoma skin tumors. The Cervical Cancer is the most frequent in the North region (22/100.000). In the regions South (24/100.000), Central-West (19/100.000) and Northeast (18/100.000), this type of cancer is the second most frequent, and in the Southeast (18/100.000) the fourth most frequent [2].

As contributing factors for the emergence of cervical cancer, are pointed out: “smoking, low intake of vitamins, multiple sexual partners, early sexual initiation and use of

oral contraceptives” [2]. It is possible to notice that even though there are other factors related with the appearance of this type of cancer, the HPV virus is still directly related.

According with [3], they were consolidated in the States and the Federal District of Brazil, with an emphasis to some advances as:

- National standardization of exam reports;
- Dissemination of the techniques of high frequency surgery;
- Expanding the offer of Pap smears;
- Availability of technical and manageable information.

In order to monitor and to manage the activities carried out in the several cities of the country, it was created a software called SISCOLO (Information System of Cervical Cancer), developed by the Information Technology Department of SUS (DATASUS) in partnership with INCA, and the data released from September 2002 [2]. This system is responsible for the data storage of the National Program of Cervical Cancer Control, gathering information about women, as age and education; demographic and epidemiological information, standardized reports and the monitoring of the cases diagnosed with neoplasm precursor lesions [3].

This work aims to study the production profile of the cytopathologic exam (Papanicolaou test or pap smears) of women aged 25 to 59 years, residing in one of the 399 municipalities of the Paraná State - Brazil, in the course of the year of 2010, registered by the DATASUS, considering the coverage and the stage of the disease in the diagnosis, through the data clustering analysis.

The option for such methodology is justified by the fact that the one in particular, when used on the health sector, proves to be widely applicable, since the existence of variables on this area studies is shown as extensive, and the individualized data can provide not very significant information. The data clustering method, along with being the best way to achieve the objectives proposed, it also allows the definition of the clusters with similar characteristics regarding the screening of Pap smears, allowing mapping the State according to the encountered results.

This article is organized as following: Section 2 covers the concepts about Clustering Analysis, describing the K-means algorithm, utilized in this study. Section 3 describes the methodology adopted for the experimental evaluation. Section 4 presents the results and discussions and, finally, Section 5 provides the final considerations as well as the perspectives of the research.

2. CLUSTERING ANALYSIS AND K-MEANS ALGORITHM

Cluster Analysis is the process of data clustering such that the objects inside a group have a high similarity when compared to the other objects of said group, and high dissimilarity to the objects of other clusters. To [6], the task of data clustering is based in two basic ideas: the internal cohesion of the objects (homogeneity) and the external isolation (separation) among the clusters.

According to [7], the cluster analysis is a generic denomination for a wide-scale of numerical methods utilized to examine multivariate data, aiming to find the homogenous sets of observations. Given a sample with an n amount of data (or individuals), each one of them measured by p variables, the objective is to look for a scheme that cluster them into g clusters. With this cluster, it's possible to identify useful relations between the data, like similarities and differences that were not previously revealed.

It is an unsupervised learning process, since there are no predefined classes or examples which demonstrate that some sort of relation should be valid among the data, or even the presence of tutors of the field to supervise the learning process. Both the optimal number of clusters as the particular characteristics that reveal similarities (or differences) should be determined by the process itself.

The cluster analysis is a difficult problem to solve because many critical factors are not included in the given problem, such as proximity measures, definition of the criteria functions, proper algorithms and its initial conditions. Furthermore, it is known that no clustering method can properly handle all types of group structures (with different shapes, sizes and density) [11].

The many existing approaches for cluster analysis can produce different clusters for the same kind of data. And if, for the same algorithm, there are any changes in the parameters, or even changes in the order of the presentation of input patterns, the final results can be affected.

The algorithm used here is the K-means, that is one of the most popular clustering algorithms [8], based in partitioning, so that giving a database with n objects and k clusters to form, it organizes the objects in k parts ($k \leq n$), where each part represents a group.

It is based on centroids, using the geometric center of each group to represent it, and the clusters are formed so as to optimize a criterion of objective partitioning, commonly called as the similarity function, as a distance measurement. Thus, the objects allocated within a group are similar and the objects allocated within a different group as dissimilar, regarding the attributes that make up the database. The algorithm can be described as follows:

Input: Data set and a value for k .

Output: Clustered data set.

Select k points, randomly, as initial centroids.

REPEAT

 Assign each point to the centroid closest to it;

 Recalculate the centroid for each group;

UNTIL {Stabilize}

3. METHODOLOGY

The database was compiled from extracted data from the Information System in the National Program of Cervical Cancer Control [5] selecting the statistical information from Information System SISCOLO 4.00, considering the chosen year for the analysis to have been the one of 2010 and that the previous versions only have data from the years prior to 2006. Afterwards, the conducted selection was of "Cervicovaginal and microflora cytopathological exam - procedure 12.011.01-0", choosing the State of Paraná.

The online system offers several options of variables for the obtaining of data. For the first table obtained, the Line contained the town of residence, the Column was non active, the Content was the quantity of exams, the sampling interval selected all the categories as well as the outcome intervals, time of exam, ethnicity/race, city of residence, city of the sampling unit, service provider city, time of the last preventive and suitability of the material.

In terms of age group, the chosen period was of 25 to 59 years old, since this is the age group in which the occurrence of cervical cancer neoplasm is higher. The period was from

January to December of 2010. To obtain the other data only the field Content was altered, being selected the low (NIC I), high level (NIC II), micro-invasion (NIC III) and invasive adenocarcinoma (invasive Carcinoma).

To better analyze the outcomes of the study, the values were standardized by performing the division of the founded values for each variable by the total of women in the chosen age group in each town [9]. The values provided by the IBGE (Brazilian Institute of Geography and Statistics) are arranged every 4 years, what made necessary the manual addition of the number of women in all age groups, between 25 and 59 years.

The similarity measurement adopted in the K-means algorithm was the Euclidian distance. The software used in the application of the technique and data analysis was the Statgraphics Centurion XVI.I demo version [13]. The choice was determined by the extent of methods of clustering analysis provided by the software.

To set the number of clusters to be formed, simulations were made of 1 to 5 clusters. The choice to form three clusters was made due the higher heterogeneity of the clusters, thereby making it easier to differentiate the regions. Thus, the formation of three clusters was made from the 399 cities within the four variables: carcinoma, NIC I, NIC II, NIC III and total of conducted exams.

4. RESULTS AND DISCUSSIONS

When performing the analysis of the general data of Paraná while analyzing the exam screening, the formed 3 clusters were separated as shown on Table 1, according to the number and the percentage of members in each cluster.

Table 1. Number and percentage of members by clusters formed according to the screening of Pap smears in the 399 cities in Paraná State, 2010

CLUSTERS	NUMBER OF MEMBERS	PERCENTAGE OF MEMBERS
1	256	64,16%
2	24	6,02%
3	119	29,82%

Through this initial analysis it is possible to notice the great difference in the number of cities among the clusters, which shows the existing differences among the municipalities from the Paraná State in the screening of Pap smears. The main disparity is verified in Cluster 2, in which only 6.02% of the cities have similar characteristics. Clusters 1 and 3 present differences though in a more elevated number. The characteristics of the member of the clusters (from the municipalities) will be detailed below, as well as the similarity factors that left them in the same cluster (similarities in the screening of exams made in the municipalities).

According to the Macroregion to which the municipalities belong, the Cluster 1 has, among the 256 municipalities, approximately 13.3% in the Eastern Region, 15.2% in the Western Region, 29% in the North Region, 29.3% in the Northeast Region, 6.25% in the

Region of Campos Gerais and 7% in the Center-South Region; mostly North and Northwest regions.

Cluster 2 presents among the 24 municipalities, approximately 16.7% in the Eastern Region, 20.8% in the Western Region, 12.5% in the North Region, 33.3% in the Northwest Region, 4.2% in the Region of Campos Gerais and 12.5% in the Center-South Region; mostly in the West and Northwest regions.

As for Cluster 3, with 119 member cities, has approximately 6.7% of them in the Eastern Region, 30.2% in the Western Region, 16.8% in the North Region, 25.2% in the Northwest Region, 8.4% in the Region of Campos Gerais and 12.6% in the Center-South Region; mostly West and Northwest regions.

This must not be a decisive (or influent) factor in the cluster division, for the Northwest Region presented a large number of municipalities in all the clusters and the Region of Campos Gerais presented a small number of municipalities in all the clusters.

In a study conducted in the State of Minas Gerais in the year of 2002 [8] five clusters were formed, of which the Cluster 1, with 70% of the municipalities located in six mesoregions, it was concentrated from the center to the south of the State; the Cluster 2, with 70% of the municipalities scattered in four mesoregions, it was concentrated 26% in the South/Southwest, with tendency to locations further North than the Cluster 1; the Cluster 3, with 70% of the municipalities in five mesoregions, it was concentrated in the South/Southwest (20.6%), extending to the Center and the North of the State; the Cluster 4, with 70% of the municipalities concentrated in five mesoregions, reduced the proportion of municipalities on the South and expanded it from the Center to concentrate in the North mesoregion (23%); Cluster 5 had 80% of the municipalities located in four mesoregions, predominantly to the North (mesoregion North with 27.8% and Jequitinhonha with 23%) and Eastern of the State.

By comparing the conducted studies with the work of [8], it is possible to notice that in the Minas Gerais State as well as in the Paraná State, there was a similar concentration of the municipalities in certain clusters, in both studies the regions with increased number of municipalities stayed with about 20% of them.

In the conducted clustering analysis, for the cluster settling and the appearance of the noted differences in Table 1, the points that represent the centroids of these clusters and each observation/municipality, belong to the cluster of which the centroid/average is closer (according to the Euclidian distance metric), as shown in Table 2:

Table 2. Centroids of each cluster by variable

CLUSTER	CARCINOMA	NIC I	NIC II	NIC III	SAMPLIN
S					G
1	-	0,000435431	0,00039298	0,0000275843	0,195133
2	0,0	0,00310371	0,00239062	0,000135394	0,342708
3	0,0	0,00133298	0,000523959	0,0000304253	0,319993

Through the analysis of Table 2, it is possible to verify that the clusters divide themselves by the variables of NIC I, NIC II, NIC III and the number of samples by municipality, since that the centroid for the variable "carcinoma" stays in zero, because in Paraná there were 16 cases of this type of cervical cancer, what makes the variable not much significant.

The following municipalities reported cases of invasive carcinoma: Cambé, Capitão Leônidas Marques, Catanduvas, Curitiba, Guarapuava, Londrina, Paranaguá, Quatiguá, São José dos Pinhais, São Mateus do Sul and Telêmaco Borba. It is possible to notice the presence of small, medium and large sized cities among them, and, as they are from the same cluster, one should take into account that the number of detected cases in one city is considered high in relation to the size of its population.

Below, it is presented the population of each municipality, the number of women between 25 and 59 years and the number of cases of invasive carcinoma detected in 2010, according to [4, 9].

- Cambé – 96.733 inhabitants in which 30.625 are women and 1 case;
- Capitão Leônidas Marques –14.970 inhabitants in which 3.527 are women and 1 detected case;
- Catanduvas - 10.202 inhabitants in which 2.223 are women and 1 case;
- Curitiba - 1.751.907 inhabitants in which 478.350 are women and 3 cases;
- Guarapuava - 167.328 inhabitants in which 40.852 are women and 1 case;
- Londrina - 506.701 inhabitants in which 131.925 are women and 2 cases;
- Paranaguá - 140.469 inhabitants in which 33.530 are women and 2 diagnosed cases;
- Quatiguá - 7.045 inhabitants in which 1.812 are women and 1 case;
- São José dos Pinhais - 264.210 inhabitants in which 66.512 are women and 1 case;
- São Mateus do Sul - 41.257 inhabitants in which 9.665 are women and 1 case;
- Telêmaco Borba - 69.872 inhabitants in which 16.864 are women and 2 cases.

Comparing the female population in the selected age group with the number of cases, one can find that despite Curitiba and Londrina being the cities with the most cases of invasive carcinoma, they do not present a high number of cases proportional to the population.

The proportionality of cases/female population in the selected age group had a variation of approximately 0.00001 in São José dos Pinhais to 0.00055 in Quatiguá.

It is important to note that the carcinoma is found when the neoplasm is already in an advanced state, which afflicts most internal body structures, and in the case of the cervical cancer, women. It is thus important to analyze the interval of time between the Papanicolaou sampling, what influences the late diagnosis of the disease.

In the year of 2010, in the cities where it was detected cases of invasive carcinoma, it was obtained the following number of women who collected their last Pap smear in more than 5 years between the ages from 25 to 59 years: in Cambé 277, Capitão Leônidas Marques 27, Catanduvas 14, Curitiba 2.081, Guarapuava 575, Londrina 1237, Paranaguá 44, Quatiguá 16,

São José dos Pinhais 656, São Mateus do Sul 54 and Telêmaco Borba with 22 women, while in the whole State there was 25250 women,

Despite the campaign, the offer of exam in the public health network and the wide dissemination by communication networks, many women fail to collect their Pap smear. As pointed out, about 0.9% of women, with an age in risk of developing the cervical cancer in the year 2010, they didn't seek any service for the collection of material in the last 5 years.

The data allowed the visualization that the centroids of NIC I and the Total of Exams characterizes the formation of Cluster 1, which is the largest cluster formed among the three of them. As for the formation of Cluster 2, the smallest cluster and the one that possibly has most differentiated characteristics from the others, differs as to NIC II and III.

Cluster 2 is composed of 24 cities, as follows: Adrianópolis, Ampére, Bela Vista da Caroba, Boa Vista da Aparecida, Bocaiúva do Sul, Imbaú, Itaguajé, Itaipulândia, Itapejara d'Oeste, Miraselva, Munhoz de Melo, Nova Olímpia, Peabiru, Pérola, Piên, Quedas do Iguaçu, Santa Inês, Santana do Itararé, Saudade do Iguaçu, Tamarana, Tijucas do Sul, Ubiratã, Vitorino, Xambrê. According to the geographical distribution of these cities, 8 of them are located in the Northwest Macroregion of the State, 5 in the West, 4 in the Eastern Region, 3 in the North, 3 in the Center-South and 1 in the Campos Gerais. Therefore, the Northwest Region of Paraná presented the highest number of cities that are part of the cluster that was differentiated by a more elevated centroid in the variables, NIC II and NIC III.

Among the cities listed in Cluster 2, that had a diagnose of NIC II, it is possible to verify that the schooling of these women are given as absent, for that in some municipalities all the reported neoplasm cases had the educational level overlooked, meaning the several times the questioning is not carried out, what makes it harder to perform a more effective analysis in this aspect. As for the municipalities in which the collection of data regarding schooling was made, most women showed to have not completed their elementary education. Regarding the results of NIC III, the educational level followed the same tendency, identifying once again a high number of absent (ignored) values.

In the data of the entire population of these municipalities, it is found the following rates of poverty and illiteracy among women older than 15 years old, respectively [10]: Adrianópolis – 43.44% and 25.9%; Ampére – 24.82% and 10.6%; Bela Vista da Caroba – 37.92% and 14.1%; Boa Vista da Aparecida – 39.99% and 18.2%; Bocaiúva do Sul – 29.02% and 13.4%; Imbaú – 43.95% and 23.3%; Itaguajé – 34.91% and 21.3%; Itaipulândia – 25.02% and 9.6%; Itapejara d'Oeste – 25.62% and 10.9%; Miraselva – 16.42% and 17.8%; Munhoz de Melo – 23.57% and 15.8%; Nova Olímpia – 30.02% and 20.3%; Peabiru – 31.90% and 15.3%; Pérola – 27.21% and 14.8%; Piên – 27.95% and 6.5%; Quedas do Iguaçu – 35.27% and 13.9%; Santa Inês – 29.57% and 17.8%; Santana do Itararé – 40.59% and 21.2%; Saudade do Iguaçu – 32.54% and 13.6%; Tamarana – 39.26% and 22.7%; Tijucas do Sul – 32.21% and 13.3%; Ubiratã – 33.16% and 15.1%; Vitorino – 25.84% and 12.5%; and, Xambrê – 32.76% and 19%.

In Brazil the illiteracy rate is of 13.3% and in the South Region, it is of 7.8%, and poverty rate is of 27.6% in the whole country and of 22.2% in the South Region [9]. Among the 24 cities belonging to Cluster 2 7.5% of them presented an illiteracy rate higher than the national one, and approximately 95.8% of them higher the illiteracy rate in the South Region. Only Piên had illiteracy rates inferior from the national and the South Region ones. As for the poverty rate, 75% of the cities had a higher rate than the national one and 87.5% higher than the rate from the South Region. Ampére and Miraselva had lower rates for both.

Based on this information, one can say that the municipalities belonging to Cluster 2 presented illiteracy and poverty rate higher than expected in a national level and regarding the South Region of the country.

Concerning the largest cluster formed, Cluster 1 with 256 members, the variables that characterized it the most were the NIC I and the total of collected exams. Despite having the lower median in the total of exams (0.1955133), NIC I and NIC III, it was slightly more elevated for NIC II.

Related to the time of performance in the last preventive, 25.250 of the 2.798.459 women aged from 25 to 59 years collected the last preventive 5 years back or longer, 15.077 collected 4 years back, 37.216 collected 3 years back, 99.997 collected 2 years back, 182.592 collected 1 year back and 11.585 collected it in the same year two Pap smears.

As for the exams collected in the same year, i.e. the last collecting made by the woman was in the same year, there are cases in which it is necessary to confirm the diagnosis and, therefore, the analysis laboratory requests a new collection.

5. FINAL CONSIDERATIONS

The conducted study showed that even with the available media campaigns and the offers to conduct the cytopathologic cervical exams, a lot of women remain to not make it for several reasons, such as rural areas where there is no health unit and thus making it difficult for users to have access to the service.

The collected data also allowed checking that among the 2.798.459 women aged from 25 to 59 years, only 492.328 conducted a collection of preventives in the year of 2010, what is equivalent to approximately 17,6% of the women of Paraná in this age group.

Knowing that the poor adherence to the women health programs favors the late diagnose of pathologies like cervical cancer, it was verified that approximately 7.6/10.000 women from Paraná had a detection of NIC I, 6.2/10.000 of NIC II, 0.3/10.000 of NIC III and 0.05/10.000 of invasive type carcinoma.

In separate analysis of the Macroregions of the State it was verified that there is a possibility of occurring errors in the diagnosis, for that when it is detected a more advanced stage of cancer, the woman's last Papanicolaou's test does not always have been conducted in long time. For example, there are only a few invasive carcinoma cases that were detected in women that do not conduct a preventive for 5 years of more.

Although the Regional Health Services are divided according to the sanitary and epidemiological characteristics of the municipalities, it was not found any relationship between the cities of the same Regional and the production of cytopathological cervical exams, i.e. the formed clustering were presenting municipalities of the various Regionals with characteristics that distinguished them from the others.

Unfortunately, the requisition forms are not completely filled, hindering the data gathering alongside SISCOLO. Important information for the research (already noted the relation between these variables with cervical cancer) is omitted, such as education, race/ethnicity and time from the last preventive.

Nowadays, the woman is included in the job market and very often she ends up being deprived of some of her legal rights as her own health, for that a work in the contracting companies could raise awareness that the prevention brings a lot less damage to the companies than the distancing of their workers from the job for a neoplasm treatment. In this situation, there are also women who start to take care of some relative or acquaintance and the

local health care teams are interested only in the one that is being taken care of, thus forgetting about the caretaker.

To prevent that these problems repeat themselves in a foreseeable future, it would be necessary that the instructions regarding the prevention programs and the health promotion were taught and encouraged from an early childhood, even in elementary schools, so that it became a habit and a routine among the Brazilian citizens.

There are several possibilities that this study will be continued, investigating furthermore the concepts here exposed and providing future studies to clarify even more the differences of healthcare within the same State, as in the case of Paraná State. Possibilities are the study made accordingly with the State Regionals Health Service and an analysis of the current situation of the not complete fulfilling of the forms before the exams collection.

Other experiments also may be conducted by expanding the database with new registers, or contrasting the data of the Paraná State with other Brazilian states. Also, other algorithms may be applied, in a comparative way, to the approach that was here described for the clusters formation.

REFERENCES

- [1] Brasil, Ministério da Saúde, Instituto Nacional do Câncer. **Estimativa 2010: Incidência de Câncer no Brasil**. Rio de Janeiro; 2009. URL: <http://www.inca.gov.br/estimativa/2010/estimativa20091201.pdf>. Last access ago 2011.
- [2] Brasil, Ministério da Saúde, Instituto Nacional do Câncer. **Estimativa 2008: Incidência de Câncer no Brasil**. Rio de Janeiro; 2007. URL: http://bvsms.saude.gov.br/bvs/publicacoes/estimativa_incidencia_cancer_2008.pdf. Last access ago 2011.
- [3] Brasil, Ministério da Saúde, Instituto Nacional do Câncer. **Viva Mulher. Câncer do colo do útero: ações técnicos-gerenciais e ações desenvolvidas**. Rio de Janeiro, 2002.
- [4] Brasil, Ministério da Saúde, **DATASUS**. Brasília; 2008. URL: <http://www2.datasus.gov.br/DATASUS/index.php>. Last access nov 2011.
- [5] Brasil, Ministério da Saúde. **SisCam - Sistema de Informação do Câncer da Mulher**. Brasília; 2008. URL: <http://w3.datasus.gov.br/siscam/index.php?area=01>. Last access sep 2011.
- [6] Cormack, R. M. **A Review of Classifications**. In: Journal of the Royal Statistical Society. v. Series A, n. 134, p. 321-353 (1971).
- [7] Everitt, B. S.; Landau, S.; Morven, L. **Cluster Analysis**. Hodder Arnold Publishers: 4th Ed., London (2001).
- [8] Feitosa, T. M. P; Almeida, R. T. **Perfil de produção do exame citopatológico para controle do câncer do colo do útero em Minas Gerais, Brasil, em 2002**. In: Cadernos de Saúde Pública, Rio de Janeiro, v. 23, n. 4, abr., 2007. Disponível em: <http://www.scielo.br/pdf/csp/v23n4/17.pdf>. Last access mai 2011.

- [9] IBGE. **Sinopse do Censo Demográfico 2010 – Paraná.** URL: <http://www.censo2010.ibge.gov.br/sinopse/index.php?dados=26&uf=41>. Last access oct 2011.
- [10] IPARDES. **Perfil dos Municípios,** 2010. URL: <http://www.ipardes.pr.gov.br/modules/conteudo/conteudo.php?conteudo=6>. Last access nov. 2011.
- [11] Kotsiantis, S. B.; Pintelas, P. E. **Recent Advances in Clustering: A Brief Survey.** In: Transactions on Information Science and Applications, v. 1, n. 1, p. 73-81 (2004).
- [12] Prata P. R. **A Transição Epidemiológica no Brasil.** Caderno de Saúde Pública, v. 8, n. 2, p. 168-175. Rio de Janeiro; 1992. URL: <http://www.scielosp.org/pdf/csp/v8n2/v8n2a08.pdf>. Last access sep 2011.
- [13] Statigraphics, **Statigraphics XVI.** StatPoint Technologies, 2009. URL: <http://www.statgraphics.com/downloads.htm>. Last access oct 2011.