Professor Ionel SINESCU PhD[1]
"Carol Davila" University of Medicine
and Pharmacy", Bucharest
Primary Doctor Calin CHIBELEAN, PhD[2]
Eng. Cristian SURCEL[3]
Professor Maria Viorica STEFANESCU PhD
Department of Mathematics
The Bucharest Academy of Economic Studies
Professor Massimiliano FERRARA, PhD
University of Reggio, Calabria

PRINCIPAL COMPONENT ANALYSIS AND CLASSIFICATION WITH APPLICATIONS IN MEDICINE

Abstract. *PCA and a classification algorithm are summarized. These are then used for processing data collected by GfK Romania on urinary incontinence.*

For urology, the study conducted by GfK Romania is a premiere, it is a beginning helping doctors to familiarize with the condition of patients in Romania.

The study may also be useful to a drug company for an optimal distribution upon regions of products needed for treatment. Results obtained by GFK Romania were processed with data analysis techniques and presented further

Key words: *classification, urology, urinary incontinence.*

JEL Classification C02, C61, C62

Principal Component Analysis (PCA) is one of the most common methods of factorial analysis of multidimensional data.

PCA analyzes quantitative numerical data in order to form homogeneous groups of statistical units and investigate interdependencies between variables. Being a descriptive method, it highlights fundamental properties of data, using numerical parameters and graphic plots.

[1] Head of the Center for Urological Surgery, Dialysis and Renal Transplantation, Fundeni Clinical Institute;

[2] Centre for urological Surgery, Dialysis and Renal Transplantation, Fundeni Clinical Institute;

[3] Centre for urological Surgery, Dialysis and Renal Transplantation, Fundeni Clinical Institute.

The initial data is represented by different valued observations also known as variables, denoted by $X_1, \dots, X_j, \dots, X_p$ on a set of statistical units numbered from $i = 1$ to $i = n$. Frequently these data are presented as a table for which rows correspond to statistical units and columns represent the observed variables.

Variables

Statistical Units	X_1	X_2	$\dots$	X_j	$\dots$	X_p
1	$x_1(1)$	$x_2(1)$	$\dots$	$x_j(1)$	$\dots$	$x_p(1)$
2	$x_1(2)$	$x_2(2)$	$\dots$	$x_j(2)$	$\dots$	$x_p(2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$⋰$	$\vdots$
i	$x_1(i)$	$x_2(i)$	$\dots$	$x_j(i)$	$\dots$	$x_p(i)$
$\vdots$	$\vdots$	$\vdots$	$⋰$	$\vdots$	$\ddots$	$\vdots$
n	$x_1(n)$	$x_2(n)$	$\dots$	$x_j(n)$	$\dots$	$x_p(n)$

The generic element of the table, situated at the crossing between row i and column j, $x_j(i)$, is the observation of variable X_j for the statistical unit i.

Data used for PCA must be quantitative, i.e. the notion of *average* must have meaning. PCA can be performed for data defined by a preference order between the p variables, but it is often recommended to apply Correspondence Analysis to this data.

Quantitative variables can be homogeneous (same units of measure, dispersion of same magnitude as the data) or heterogeneous. The variables can be discrete (can only take a finite number of values) or continuous (can take any value inside an interval). These do not affect the PCA method since the fundamental property of data is still that of being quantitative.

PCA will provide relevant results for sufficiently large data tables. The number of statistical units should be greater than 15 and the number of rows superior to 4. Obviously, this is only a suggestion, since often we can perform PCA on a smaller data set. Most commonly in practice, tables have hundreds of rows (statistical units) and tens of columns (variables).

To make easier the interpretation of results, we can insert in the table *supplementary data*. Supplementary statistical units are those statistical units for which we have observations upon the variables, but we do not wish to take them into account when computing the statistical parameters. Similarly, we can also introduce *supplementary variables*.

Using the supplementary data we can characterize groups of statistical units on graphic plots, or highlight bonds between initial variables and various other variables.

In the beginning, we must define (by measuring) the *distance* or *similitude* between two statistical units. Two statistical units are similar if the observed variables take similar values.

Our objective is to quantify the distance between two statistical units, reflecting as much as possible reality – we must take into account all the variables (except supplementary ones) in order to define the function which expresses the distance between two statistical units. A first definition for the distance function would be the sum of the squares of the distances between the variable observations. This definition is not satisfactory because it would depend on the measurement units of the statistical units.

To stabilize the distance, we must *center* and *reduce* the data, obtaining the formula:

$$d^2(i,k) = \sum_{j=1}^{p} \left(x_j(i) - x_j(k)\right)^2 / \sigma_j^2$$

The distance no longer depends on the units of measure in which the variables are expressed. Using the formula above, we can calculate all distances between statistical units, that is $n(n-1)/2$ distances for n statistical units.

PCA best describes the data, providing a system of orthonormal axes conserving as good as possible the distances between data. The axes have additional properties: they are the straight lines that best fit the cloud of points corresponding to observations according to the least squares criterion and they are called the *factor axes*. Their directing vectors of the axes are called *eigenvectors* and are denoted by u_i.

Each eigenvector u_i has the components:

$$u_i = \left(u_i^1, u_i^2, \dots, u_i^j, \dots, u_i^p\right).$$

The axes origin characterizes the statistical unit defined by the average of the initial variables. This property has fundamental consequences in interpreting the results.

Next, the axes are taken in the descending order of closeness to the statistical units. The plane 1×2 will be the closest to the statistical units. On each projection plane the distances between points are inferior to the distances between the statistical units.

A *principal component* denoted by c_i is the list of coordinates of statistical units for the factor axis generated by u_i. Each principal component defines a new variable, because for each statistical unit there is a corresponding coordinate on the factor axis.

The principal components are centered and each pair is uncorrelated. They have great importance in interpreting the results as they explain the relationships between initial variables and justify the formation of homogeneous groups of statistical units. In doing this, we use correlation coefficient between principal components and initial variables.

The dispersion of a principal component is called the *eigenvalue* or *inertia* corresponding to factor axis of same rank. The eigenvalues λ_i are sorted in descending order and their number is equal to the number of initial variables, p. Each factor axis corresponds to an eigenvalue. We usually take into account only the first k nonzero eigenvalues.

The eigenvectors which generate the factor axes are the eigenvectors of the correlation matrix associated to the eigenvalues. These vectors

$u_l = \left(u_l^1, u_l^2, \dots, u_l^j, \dots, u_l^p\right)$ for $(\forall)\, l = 1, \dots, p$

are unitary (the sum of the squares of their components is equal to 1) and orthogonal (the sum of products of components of same rank for any pair of different vectors is null).

The coordinates $c_l(i)$ of the statistical units on the axis generated by u_l is given by

$$c_l(i) = \sum_{j=1}^{p} u_l^j x_j'(i).$$

The computation of coordinates of supplementary statistical units is performed using the same formula but without modifying the average and dispersion used in determining the reduced centered value of a variable. We use the formula for data reconstruction which expresses the reduced centered initial variables x' as a function of the principal components:

$$(\forall)\, i = 1, \dots, n \quad (\forall)\, j = 1, \dots, p \quad x_j'(i) = \sum_{l=1}^{k} u_l^j c_l(i).$$

Graphic representations are obtained using the above results.

The statistical units are in the planes whose axes are the factor axes, which are orthonormal. The coordinate of the statistical unit i on the axis l is equal to the $c_l(i)$ value of the principal component c_l referring to the statistical unit i. The origin of the axes characterizes the statistical unit whose values are equal to the averages of the initial variables. These planes are called *factor planes.*

Variables are represented using correlation circles: the coordinates of a variable are the correlation coefficients of this variable with respect to principal components which define the circle.

Classification with Algorithm Based on an Ultrametric Distance

Proving that a total hierarchy determines and is determined by an ultrametric structure on *X*, S.C. Johnson ([3]) has firstly proposed a general scheme for constructing a classification based on ultrametric distance. Essentially, this scheme determines a chain of partitions which contains classes with growing diameters.

Let δ be an ultrametric distance on *X*.

STAGE 0:

Let **P** $^0 = \{P_1^0, \dots, P_n^0\}$ be the discrete partition, with $P_i^0 = \{x^i,\ i = \overline{1, n}\}$.

Define $\delta\left(P_i^0, P_j^0\right) = \delta\left(x^i, x^j\right),\ v_0 = 0$, **L** $^0 = \{1,2,\dots, n\}$.

STAGE T (T ≥ 1):

1^0. Determine $v_t = \min\{\delta(P_i^{t-1}, P_j^{t-1}) \mid P_i^{t-1}, P_j^{t-1} \in \mathbf{P}^{t-1}\}$

2^0. $\mathbf{C}^t = \{(i,j) \mid i,j \in \mathbf{L}^{t-1}, \delta(P_i^{t-1}, P_j^{t-1}) = v_t\}$,

$\mathbf{I}^t = \{i \in \mathbf{L}^{t-1} \mid \exists j, (i,j) \in \mathbf{C}^t\}$.

3^0. For each $i, j \in \mathbf{L}^{t-1}$, with $\{i \mid (i,j) \in \mathbf{C}^t\} \neq \emptyset$, put

$P_i^t = P_i^{t-1} \cup \{P_j^{t-1} \mid j \in \mathbf{L}^{t-1}, (i,j) \in \mathbf{C}^t\}$, $P_i^t = P_i^{t-1}$, if $i \in \mathbf{L}^{t-1} \setminus \mathbf{I}^t$.

4^0. $\mathbf{P}^t = \{P_i^t \mid \text{distinct } P_i^t \text{ obtained at } 3^0\} \cup \{P_i^t \mid i \in \mathbf{L}^{t-1} \setminus \mathbf{I}^t$,

$\mathbf{I}^t$ = indices of elements of $\mathbf{P}^t\}$.

5^0. If $|\mathbf{P}^T| = 1$, write $\mathbf{L}^T = \{\mathbf{P}^0, \mathbf{P}^1, \dots, \mathbf{P}^T\}$. STOP.

If $|\mathbf{P}^T| > 1$, go to 6^0.

6^0. Define $\delta(P_i^t, P_j^t) = \delta(P_i^{t-1}, P_j^{t-1})$, $i, j \in \mathbf{L}^t$.

Repeat the cycle for $t = t + 1$.

Proposition 1. For every $i, j \in \mathbf{L}^t$ and $x, u \in P_i^t$, $y, v \in P_j^t$, it follows that $\delta(x, v) = \delta(u, v)$.

Proof. Inductively, in respect to *t*.

Let t = 1, $|P_i^1| \geq 2$ be.

According to 1^0 and 2^0, $\delta(x, u) = v_1 = \min\{\delta(a, b) \mid a, b \in X\}$.

Let $y \in P_j^1$ and assume that $\delta(x, y) > \delta(u, y)$.

But, $\delta(x, u) \leq \max\{\delta(x, u), \delta(u, y)\} = \delta(u, y)$, contradicting the previous inequality.

Hence, $\delta(x, y) = \delta(u, y)$.

In the same way, it results δ(x, v) = δ(u, v). On the other hand, if y ≠ v, then δ (y, v) ≠ v_1. Interchanging the two pairs, we obtain another two inequalities, which complete the proof for t = 1.

Further, let we presume that the proposition is verified until the stage t -1, (t ≥ 2) and let $P_i^t, P_j^t \in \mathbf{P}^t$. If $P_i^t, P_j^t \in \mathbf{P}^{t-1}$, the induction assumption ensures the truth of proposition.

1) Suppose that $P_i^t \notin \mathbf{P}^{t-1}$, $P_j^t \in \mathbf{P}^{t-1}$. Then there exist $i_1, i_2 \in \mathbf{L}^{t-1}$ such that $x, u \in P_i^{t-1} \cup P_{i_1}^{t-1} \cup P_{i_2}^{t-1}$, $u \in P_{i_2}^{t-1}$ and $\delta(P_i^{t-1}, P_{i_1}^{t-1}) =$ $= \delta(P_i^{t-1}, P_{i_2}^{t-1}) = v_t = \min\{\delta(P_k^{t-1}, P_l^{t-1}) \mid k, l \in \mathbf{L}^{t-1}, k \neq l\}$.

Consider three situations:

a) *x*, *u* belongs to the same set from the above union. Then, the desired equality results from induction assumption.

b) $x \in P_i^{t-1}$, $u \in P_{i_1}^{t-1}$. Then, $\delta(x, u) = \delta(P_i^{t-1}, P_{i_1}^{t-1}) = v_t =$ $\min\{\delta(a, b) \mid a \in P_k^t, b \in P_l^{t-1} \mid k, l \in \mathbf{L}^{t-1}, k \neq l\}$. As in the first step, we conclude that $\delta(x,y) = \delta(u,y)$ and $\delta(x,v) = \delta(u,v)$. But, from the induction assumption, $\delta(x,y) = \delta(x,v)$ and, hence, $\delta(x,y) = \delta(u,v)$.

c) $x \in P_{i_1}^{t-1}$, $u \in P_{i_2}^{t-1}$. It is sufficient to show that $\delta(x,u) = v_t$ and argue as in case *b)*.

Consider $z \in P_i^{t-1}$. Then we may write $\delta(x,u) \leq \max\{\delta(x,z), \delta(z,u)\} = v_t$.
In addition, if $\delta(x,u) = (P_{i_1}^{t-1}, P_{i_2}^{t-1})$, the inequality fails (otherwise one contradicts the choice of v_t).

2) $P_i^t, P_j^t \notin \mathbf{P}^{t-1}$. We may find j_1, $j_2 \in \mathbf{L}^t$ such that $y, v \in P_j^{t-1} \cup P_{j_1}^{t-1} \cup P_{j_2}^{t-1}$ and $\delta(P_j^{t-1}, P_{j_1}^{t-1}) = \delta(P_j^{t-1}, P_{j_2}^{t-1}) = v_t$.
From the first part of the proof it results that $\delta(x,y) = \delta(u,y)$ and $\delta(x,v) = \delta(u,v)$. Interchanging P_i^t and P_j^t we obtain again from 1) that $\delta(y,x) = \delta(v,x)$ and $\delta(y,u) = \delta(v,u)$.

Corollary. If $P_j^{t-1}, P_k^{t-1} \subseteq P_i^t$ for same $t \geq 1$ and j, $k \in \mathbf{L}^{t-1}$, then $\delta(P_k^{t-1}, P_l^{t-1}) = \delta(P_j^{t-1}, P_l^{t-1})$ for every $l \in \mathbf{L}^{t-1}$.

Proof. Indeed, the algorithm assures that $\delta(P_j^{t-1}, P_k^{t-1}) \leq v_t$. Then, we follow as in stage $t = 1$ of the previous proof.

Remarks. 1) Proposition 1 justifies the instructions of the algorithm. Since $\delta(P_j^t, P_j^t)$ is the same with the distance between any two points of the two sets, we can take $\delta(P_i^t, P_j^t) = \delta(P_i^{t-1}, P_j^{t-1})$ if, at the step t, P_i^t has been obtained as a union which includes P_i^{t-1}.

2) Moreover, the adopted notation has the usual signification:

$$\delta\left(P_i^t, P_j^t\right) = \min_{x \in P_i^t,\ y \in P_j^t} \delta(x, y) .$$

Theorem 1. The family $\mathbf{L}^k = \{\mathbf{P}^0, \mathbf{P}^1, \ldots, \mathbf{P}^T\}$ obtained with algorithm ($|\mathbf{P}^T| = 1$) is a complete chain of partitions of X. Moreover,

$$v_{t-1} = \max_{P \in P^{t-1}} \rho(P) < \max_{P \in P^t} \rho(P) = v_t,\ t \geq 1.\ \left(\rho(P) = \max_{x, y \in P} \delta(x, y)\right).$$

Proof. The first part of theorem is easily verified by induction. For the second affirmation it is sufficient to verify that $v_t = \max_{P \in P^t} \rho(P)$ and $v_{t-1} < v_t$.

For $t = 1$ both properties are trivial. Suppose that they are true for t -1. Assume that $P_i^t = P_i^{t-1} \cup P_j^{t-1}, P_i^{t-1}, P_j^{t-1} \in \mathbf{P}^{t-1}$. Then,

$$\rho\left(P_i^t\right) = \max\left\{\rho\left(P_i^{t-1}\right), \rho\left(P_j^{t-1}\right), \delta(x,y) \,\middle|\, x \in P_i^{t-1}, y \in P_j^{t-1}\right\} = \max\left\{\rho\left(P_i^{t-1}\right), \rho\left(P_j^{t-1}\right), v_t\right\}$$

Let $x, z \in P_i^{t-1}$, $y \in P_j^{t-1}$. Then, $\delta(x,z) \le \max\{\delta(x,y), \delta(y,z)\} = \delta\left(P_i^{t-1}, P_j^{t-1}\right) = \nu_t$. Hence, $\rho\left(P_i^{t-1}\right) \le \nu_t$. Consequently, $\rho\left(P_i^t\right) = \nu_t$ and $\max\limits_{P \in P^t} \rho(P) = \nu_t$.

Finally, we notice that $\nu_{t-1} < \nu_t$. Indeed, in the case of equality the set P_i^t would be formed at the step t -1.

Theorem 2. The set A of distinct classes of partitions $\mathbf{P}^0, \mathbf{P}^1, \ldots, \mathbf{P}^T$ is a total hierarchy of X, indexed by the mapping $\nu : \mathbf{A} \to R^+$, $\nu(A) = \nu_\tau$, if $\tau = \min\{t \mid \mathbf{A} \in \mathbf{P}^t\}$.

Proof. Evidently, from the inequalities $\nu_0 < \nu_1 < \ldots < \nu_T$, above proved.

Corollary. The algorithm constructs a total ascending hierarchy.

Remark. The algorithm is well defined, that is, for each ultrametric distance constructs a unique total classification.

Application in medicine of PCA and Classification

Urinary incontinence is not so much a disease as a symptom that can be attributed to various physical or mental disorders: diabetes, stroke, Parkinson's disease, etc.

Urinary incontinence may occur at any age, but shows an increased incidence in women over 60 years. The incidence in case of women over 65 years old is 25% and for men is approximately 15%. Moreover, 10% of children suffer from enuresis (involuntary or unconscious emission of urine at night). Some forms of urinary incontinence may be temporary and disappear after treatment (for a urinary infection, for example). Other forms are due to long-term problems, but in most cases, there are solutions to eliminate or keep them under control - especially if treatment is instituted promptly. In the absence of treatment, incontinence can be a disability that limits physical activities, social, sexual or professional life, without considering the increased risk of depression and anxiety.

Treatment depends on the cause and severity of urinary incontinence. In many cases, urinary incontinence can be corrected with drugs. Often, medication is used in combination with behavioral techniques. Medications commonly used to treat urinary incontinence are anticholinergics, antispasmodics, antibiotics. Also, among commonly used medical devices are: urethral inserts - small devices inserted into the urethra, pessary - rubber rings to support bladder, catheter - flexible tube that is inserted into the urethra several times a day for drainage.

For urology, the study conducted by GfK Romania is a premiere. A beginning which helps doctors to familiarize with the condition of patients in Romania. The study may also help a drug company for an optimal distribution upon regions of products needed for treatment. Results obtained by GFK Romania were processed with data analysis techniques and presented further.

Presentation of data

Subject:	Urinary Incontinence
Field:	03 – 22 August 2008
Population:	Men and women aged 40 years old and over
Region:	Nationally representative
Interviewing technique:	Face-to-face, “paper and pencil”, in respondents’ home
Sampling method	Multi stratified sample on regions and locality size; Random distribution of sampling areas and statistic step for household selection; Birthday rule for selecting the respondent inside the household.
Total number of interviews:	874 interviews
Weighting:	On gender, age, region and residence area
Questionnaire:	Developed by MMD and adapted by GfK Romania Approved by the client
Research management:	GfK Romania

OBJECTIVES

The study covers the following objectives:

- Incidence of urinary incontinence among the population over 40;
- How common urinary leakage is, their amount and when they occur;
- Finding the distributions of patients by geographical area.

METHODOLOGY

- The study was developed as part of GfK*Omnibus, unfolding between 3 and 22 August 2008.
- The interviews were conducted face to face at the respondents’ residence.
- Sampling method was based on a probabilistic sample, stratified by region and size of locality. Selection of sampling areas and households were made randomly.
- In the household, respondents were selected by the "first birthday".
- Respondents were men and women from urban and rural area, aged at least 40.
- For 874 respondents, the maximum sampling error is + / - 3.3% at a confidence of 95%.

The results obtained from the survey

The Moment of Leaking Urine – analysis on region

Base: N=874 respondents (all respondents)

%	Total (base=874)	Region						
		Ardeal (base=153)	Banat/Crisana/ Maramures (base=104)	Muntenia (base=191)	Oltenia (base=94)	Dobrogea (base=39)	Moldova (base=214)	Bucharest (base=78)
Never - urine does not leak	82	89	85	81	74	80	77	88
Leaks before you can get to the toilet	6	4	3	7	6	7	5	9
Leaks when you cough or sneeze	7	3	9	6	12	4	7	4
Leaks when you are asleep	2	1	1	4	1	0	1	0
Leaks when you are physically active/ exercising	4	3	4	3	5	8	5	2
Leaks when you have finished urinating and are dressed	2	1	1	4	5	0	2	3
Leaks for no obvious reason	4	3	2	4	6	2	7	2
Leaks all the time	0	0	0	0	0	0	1	1

The Quantity of Urine Leaked – analysis on region Base: N=874 respondents (all respondents)

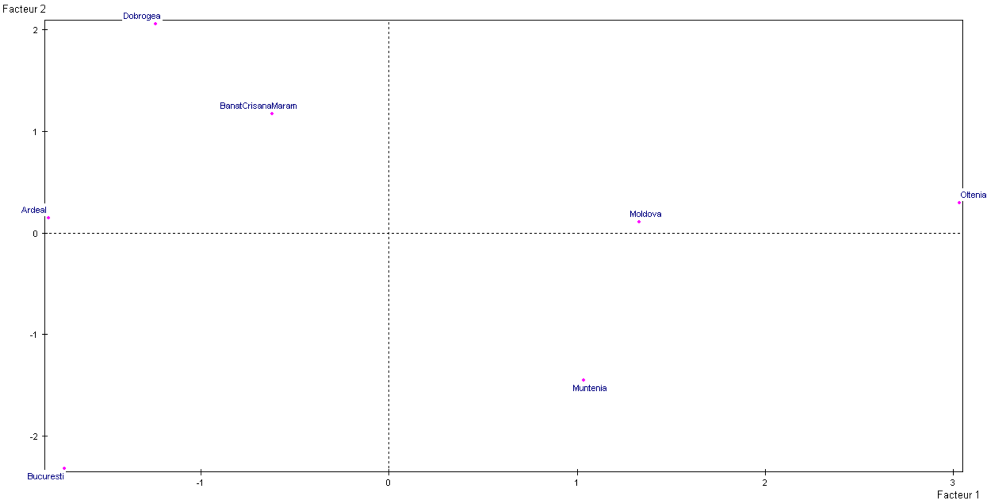
%	Total (base=874)	Region						
		Ardeal (base=153)	Banat/Crisana/ Maramures (base=104)	Muntenia (base=191)	Oltenia (base=94)	Dobrogea (base=39)	Moldova (base=214)	Bucharest (base=78)
None	82	89	85	81	74	79	77	88
A small amount	14	8	14	15	24	14	15	9
A moderate amount	3	2	2	3	2	5	6	2
A large amount	1	1	0	1	0	2	2	1

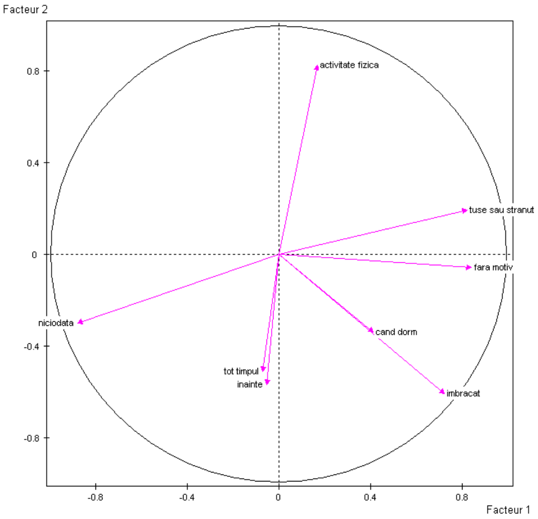
The Frequency of Urine Leaking – analysis on region Base: N=874 respondents (all respondents)

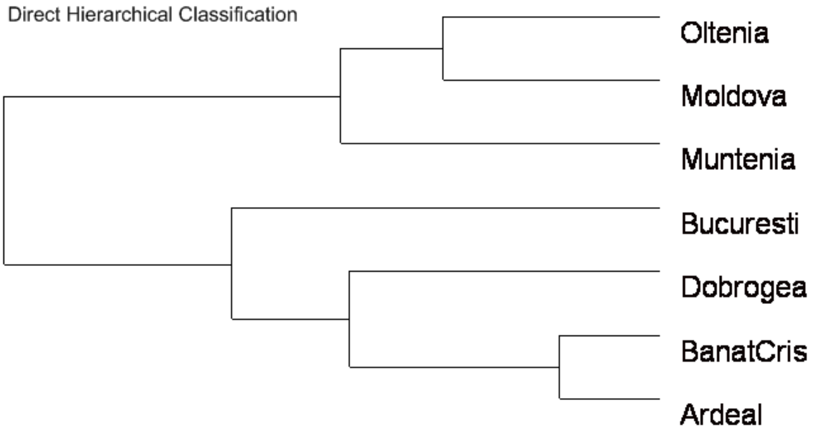
%	Total (base=874)	Region						
		Ardeal (base=153)	Banat/Crisana/ Maramures (base=104)	Muntenia (base=191)	Oltenia (base=94)	Dobrogea (base=39)	Moldova (base=214)	Bucharest (base=78)
Never	82	89	85	81	74	79	77	88
About once a week or less often	9	4	10	8	14	20	12	6
Two of three times a week	3	4	1	4	5	2	2	3
About once a day	1	1	1	2	3	0	2	0
Several times a day	4	2	4	5	4	0	5	3
All the time	1	1	0	0	0	0	3	0

Results obtained using data analysis
Time of urine leakage
The results obtained from the surve

```
Correlation  matrix
     |  nici   ina    tust   dorm   acfi   imbr   fara   tot
-----+------------------------------------------------------
nici |   1.00
ina  |  -0.04   1.00
tust |  -0.68  -0.30   1.00
dorm |  -0.15  -0.05   0.14   1.00
acfi |  -0.59  -0.07   0.14  -0.33   1.00
imbr |  -0.43   0.38   0.55   0.46  -0.40   1.00
fara |  -0.75  -0.13   0.53   0.26   0.11   0.52   1.00
tot  |   0.06   0.38  -0.20  -0.33  -0.27   0.08   0.26   1.00
-----+------------------------------------------------------
```


Facteur 2
Dobrogea
BanatCrisanaMaram
Oltenia
Ardeal
Moldova
Muntenia
Bucuresti
2
1
0
-1
-2
-1
0
1
2
3
Facteur 1

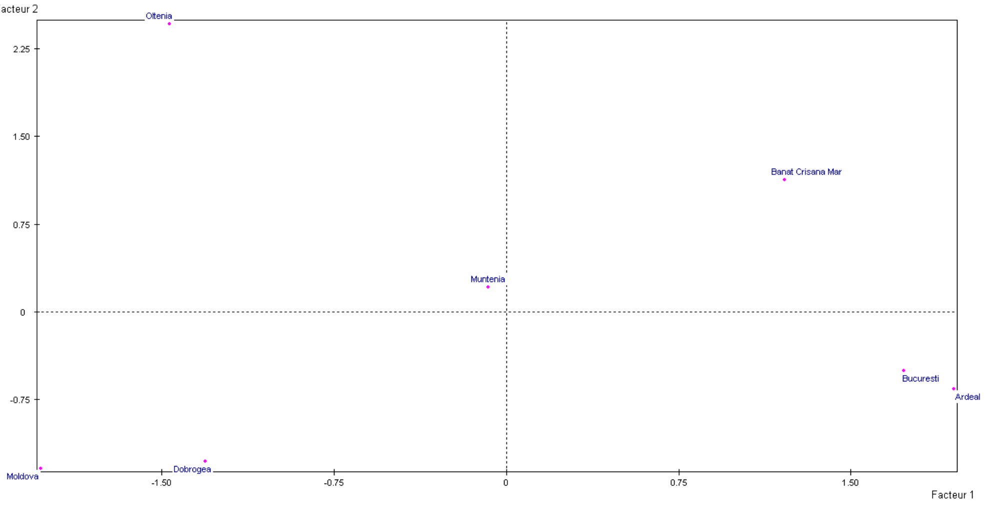

Facteur 2
activitate fizica
tuse sau stranut
fara motiv
niciodata
cand dorm
tot timpul
inainte
imbracat
0.8
0.4
0
-0.4
-0.8
-0.8
-0.4
0
0.4
0.8
Facteur 1

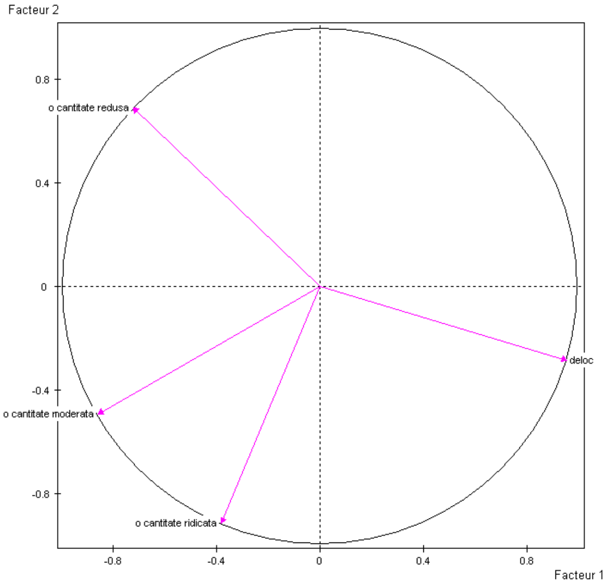


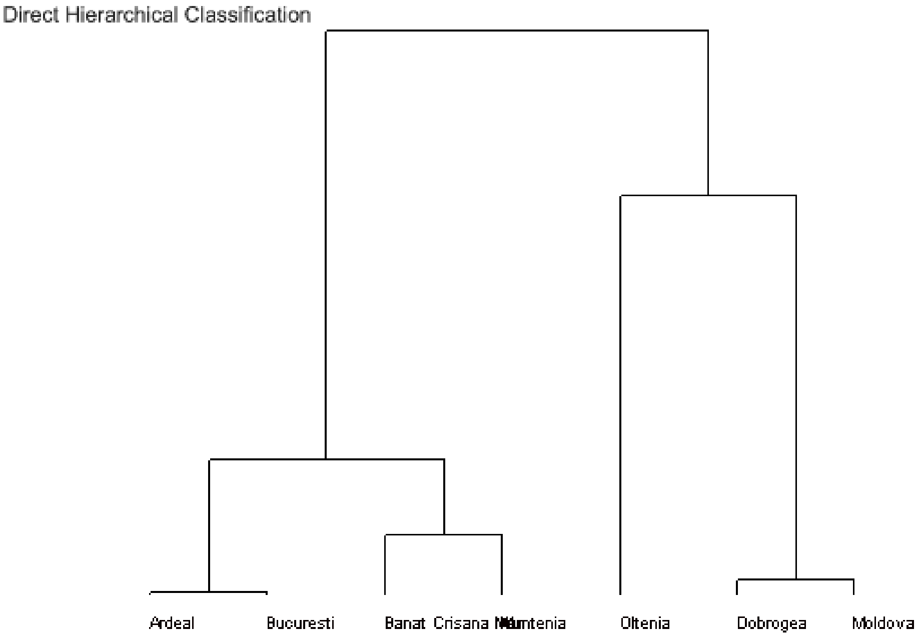
The amount of urine leaked

Correlation matrix

```
     |  delc   redu   mode   ridi
-----+----------------------------
delc |   1.00
redu |  -0.89   1.00
mode |  -0.67   0.27   1.00
ridi |  -0.11  -0.35   0.76   1.00
-----+----------------------------
```



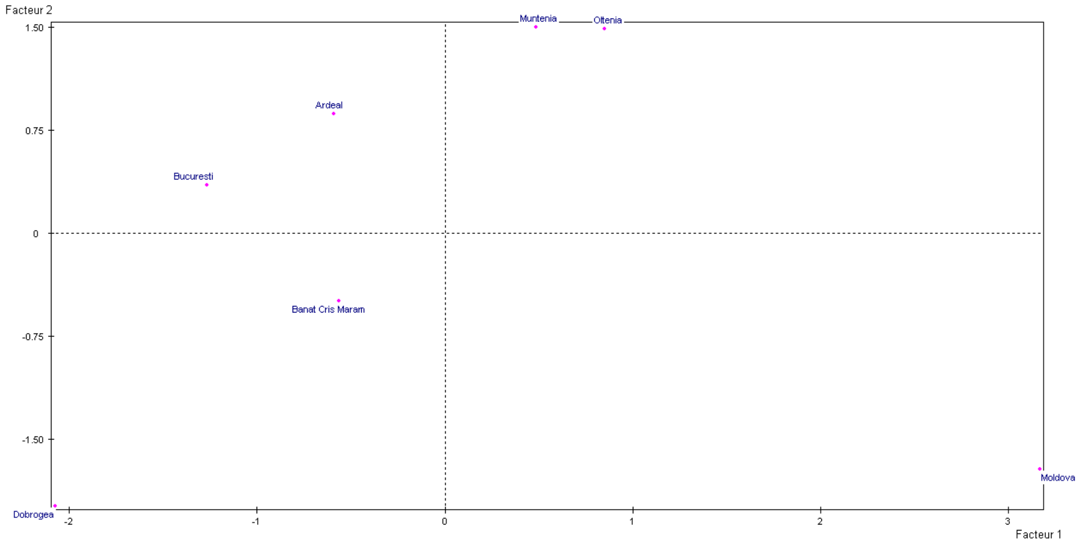

Facteur 2
0.8
0.4
0
-0.4
-0.8
-0.8
-0.4
0
0.4
0.8
Facteur 1
o cantitate redusa
deloc
o cantitate moderata
o cantitate ridicata

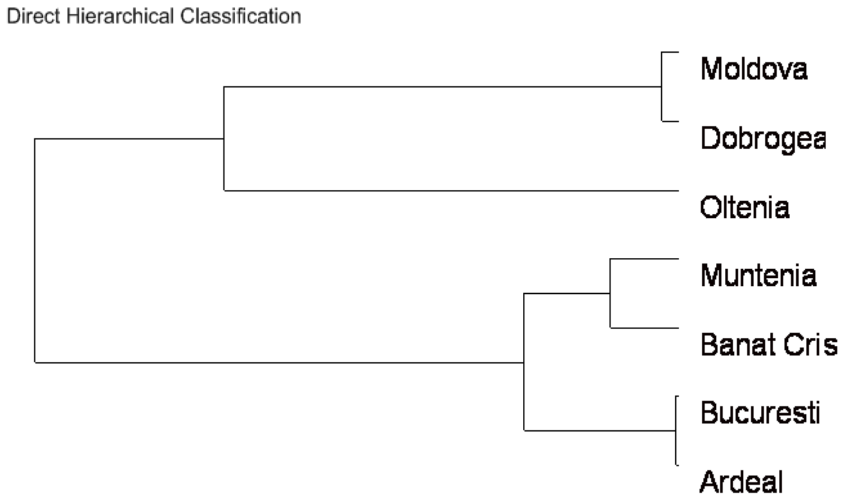

Direct Hierarchical Classification
Ardeal
Bucuresti
Banat
Crisana
Muntenia
Oltenia
Dobrogea
Moldova

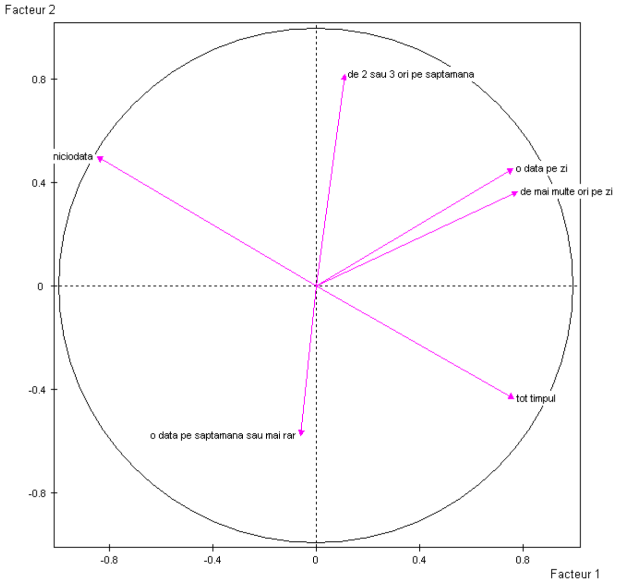
Frequency of urine leaking

Correlation matrix

```
     |  nic    1sap   23sp   1zi    muzi   tot
-----+-----------------------------------------
nic  |   1.00
1sap |  -0.30   1.00
23sp |   0.23  -0.26   1.00
1zi  |  -0.42   0.02   0.53   1.00
muzi |  -0.43  -0.36   0.13   0.70   1.00
tot  |  -0.87  -0.06  -0.21   0.25   0.32   1.00
-----+-----------------------------------------
```







Conclusions

82% of respondents are not suffering of urinary incontinence. Among those who have urinary leakage (18% of the total number of respondents), half reported that they happen once a week or less, while 21% said they had urinary leakage several times a day.

Compared with the overall results, in Transylvania there are less people suffering from urinary incontinence, while in Dobrogea 20% of respondents said they had urinary leakage once a week or less (higher percentage than in the entire sample).

Most (76%) who had urinary leakage consider that they are losing a small amount of urine. 61% of those who lose a small amount of urine said this happens once a week or less often. Several respondents in Oltenia lose smaller amounts of urine, compared with the total sample.

Moments when urinary leakage occurs differ: 35% lose urine when coughing or sneezing, 30% before reaching the toilet, 25% do not have a precise time or reason, and for 20% it is caused by physical activity. The number of women over 70 in Oltenia suffering from urine leakage when coughing or sneezing is greater than the entire sample.We noticed that from the point of view of time when it causes loss of bladder, Banat, Crisana and Maramures are very similar to Ardeal. Bucharest is then added. Oltenia is like Moldova and Muntenia.

The size of urinary leakage Ardeal with Bucharest joins first. Then, we add the class consisting of Banat, Crisana, Maramures and Muntenia. Class consisting of Dobrogea and Oltenia, Moldova is also added. The same classification is obtained and the frequency urinary loss. These are observations that may help later in the prevention of urinary incontinence. Also the results obtained may be useful for a drug company for an optimal distribution upon regions of products needed for treatment.

REFERENCES

[1] **Jardine, N., Sibson, R. (1968),** ***The Construction of Hierarchic Classifications***, *Computer Journal*, 11, 177-184;

[2] **Ştefãnescu, M.V, Ferrara, M., Dedu, S. (2008),** ***Algorithms for Hierarchical Classification with Applications in Portfolio Management***. Economic Computation and Economic Cybernetics Studies and Research, 3-4, 42, 109-122, ASE Publishing House, Bucharest;

[3] **Saporta, G. (2006),** ***Probabilités, Analyse des Donées et Statistique***, *Editions Techip, Paris;*

[4] **Saporta, G., Ştefanescu , M.V. (1996),** ***Analiza Datelor şi Informatică*** , *Economică*, *Publishing House,* Bucharest;

[5] **Ştefanescu M.V. (1985),** ***Fundamental Algorithms for Hierarchical Classification***, *Economic Computation and Economic Cybernetics Studies and Research*, 20, 55-60, 37-48, ASE Publishing House, Bucharest;

[6] **Ştefanescu, M.V. (1984),** ***The Scatter Function and Hierarchical Classification,*** *Proceedings of the Fifth Symposium "Cybernetic Modelling of Production Processes"*, 336-340;

[7] **Ştefanescu , M.V. (1984),** ***Mathematical Basis of Total Hierarchical Clustering***, *Economic Computation and Economic Cybernetics Studies and Research*, 19, 55-6;

[8] **Thierry F. (1997),** ***L'analyse des données, mode d'employ***, Press Universitaires de Rennes.