ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 162, 2002

Małgorzata Kobylińska[*], *Wiesław Wagner*[**]

NUMERICAL ASPECTS OF DETERMINING MEASURES AND CONTOURS IN DEPTH FOR DATA IN R^2

ABSTRACT. Measures and contours in depth are new statistical techniques applied in the analysis of observations. They are particularly applied in the visualisation of 2-dimentional samples in R^2 space. The theory of measures and contours in depth for the case of R^2 has been presented in numerous scientific papers by Donoho and Gasko (1992), He and Wang (1997), Rousseeuw and Ruts (1996, 1999), Ruts and Rousseeuw (1996). The papers by the above authors are mainly theoretical. They have put less emphasis on applications. Such situation could be explained by the lack of adequate software in this field in such common statistical packages as SAS, SPSS, or STATISTICA.

This paper focuses on the numerical aspects of construction of the contour for samples in space R^2. Certain numerical aspects with their direct implementation in the TURBO-PASCAL programming language were presented. The prepared program did numerical calculations. It allowed us to focus attention on the basic features of contours in depth being the graphical visualisation of 2-dimentional samples.

The theoretical basis, as regards measures in depth and contours in depth, are included in the above-mentioned papers and in the article by Wagner and Kobylińska (2000).

I. THEORETICAL AND NUMERICAL BASIS

The basic numerical denotations and numerical aspects referring to a 2-dimentional sample were specified in the following points. The ❖ *TP* symbol denotes the implementation of these issues in the TURBO-PASCAL programming language.

[*]Dr, University of Warmia and Mazury in Olsztyn,
[**]Prof., The Academy of Physical Education in Poznań.

(a) Assumptions:

- n – element set size,
- (X,Y) – a pair of observable random variables,
- (x_i, y_i) – two-dimensional observation of the i–th element,
- $X = \{(x_i, y_i): i = 1, 2, ..., n\}$ – two-dimensional sample (*TDS*),
- $\theta = (\theta_1, \theta_2)' \in R^2$ – a given point for determining depth measure in *TDS*;

(b) Samples in a non-decreasing order:

❖ TP: $SORT(n, X)$ procedure,

$$X: x_{(1)} \le x_{(2)} \le ... \le x_{(n)}, \quad Y: y_{(1)} \le y_{(2)} \le ... \le y_{(n)},$$

where (.) is a rank of observation in a disordered sample,

❖ *TP*: $\{no., rank, x\}$,

(c) A rectangular of the dispersion (RD) of TDS

$$A(x_{(1)}, y_{(1)}), \quad B(x_{(n)}, y_{(1)}), \quad C(x_{(n)}, y_{(n)}), \quad D(x_{(1)} y_{(n)}),$$

$$RD = \langle x_{(1)}, x_{(n)} \rangle \times \langle y_{(1)}, y_{(n)} \rangle,$$

❖ TP:

$$A(x_{\min}, y_{\min}) = (x_{\min}, y_{\min}), \quad B(x_{\max}, y_{\min}) = (x_{\max}, y_{\min}),$$

$$C(x_{\max}, y_{\max}) = (x_{\max}, y_{\max}), \quad D(x_{\min}, y_{\max}) = (x_{\min}, y_{\max}),$$

Point $\Theta = (\theta_1, \theta_2)$ belongs to *RD*, i.e. is its internal or peripheral point, if at the same time its coordinates belong to the variation ranges of the features X and Y, i.e. when $\theta_1 \in \langle x_{(1)}, x_{(2)} \rangle$ and $\theta_2 \in \langle y_{(1)}, y_{(2)} \rangle$. If $\Theta \in RD$, then the distance from the *RD* sides could be calculated and illustrated by fig. 1.

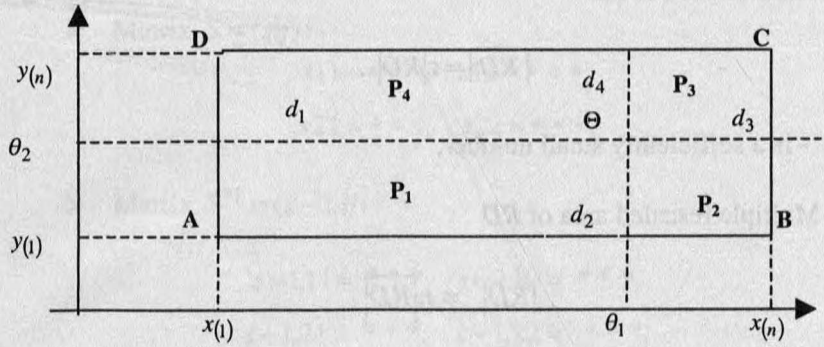

Fig. 1. The distance from the RD

Distances d_i for $i=1,2,3,4$ are expressed by the following formulas:

$$d_1=\theta_1-x_{(1)},\quad d_2=\theta_2-y_{(1)},\ d_3=x_{(n)}-\theta_1,\quad d_4=y_{(n)}-\theta_2.$$

The area of RD is divided into four disjoint areas P_i, $i=1,2,3,4$, described as follows:

$$P_1=\left\{(x,y)\in RD;x\in\left\langle x_{(1)},\theta_1\right\rangle,y\in\left\langle y_{(1)},\theta_2\right\rangle\right\},$$

$$P_2=\left\{(x,y)\in RD;x\in\left\langle \theta_1,x_{(n)}\right\rangle,y\in\left\langle y_{(1)},\theta_2\right\rangle\right\},$$

$$P_3=\left\{(x,y)\in RD;x\in\left\langle \theta_1,x_{(n)}\right\rangle,y\in\left\langle \theta_2,y_{(n)}\right\rangle\right\},$$

$$P_4=\left\{(x,y)\in RD;x\in\left\langle x_{(1)},\theta_1\right\rangle,y\in\left\langle \theta_2,y_{(n)}\right\rangle\right\},$$

in such a manner that $RD=P_1+P_2+P_3+P_4$. Then the size of these areas is calculated $n_i=\#\{P_i\}$, $i=1,2,3,4$, for which the condition $n=n_1+n_2+n_3+n_4$ is met.

(d) Area of RD

$$|RD|=(x_{(n)}-x_{(1)})(y_{(n)}-y_{(1)}),$$

➢ Rescaled area of *RD*

$$\overline{|RD|} = \varepsilon |RD| ,$$

where ε – is a sufficiently small number,

➢ Multiple rescaled area of *RD*

$$|RD|^{*} = n\left|\overline{RD}\right|,$$

(e) Classic numerical characteristics

➢ Ranges: – R_x, R_y, ➢ Standard deviation: – s_x, s_y,

➢ Arithmetic means: – $\bar{x}, \bar{y}$, ➢ Variation coefficients: – v_x, v_y,

❖ *TP*:

Ranges: $x - * * *$, $y - * * *$,

Arithmetic means: $x - * * *$, $y - * * *$,

Standard deviation: $x - * * *$, $y - * * *$,

Variation coefficients: $x - * * *$, $y - * * *$,

where the $* * *$ symbol is a numerical value with respectively given formant,

(f) A sample matrix of covariance S and its inverse matrix S^{-1}

$$S = \begin{bmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{bmatrix},$$

where s_x^2, s_y^2 are variances and $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ is a covariance between variables X and Y, and

$$S^{-1} = \begin{bmatrix} s^{11} & s^{12} \\ s^{21} & s^{22} \end{bmatrix} = \frac{1}{|S|}\begin{bmatrix} s_y^2 & -s_{xy} \\ -s_{xy} & s_x^2 \end{bmatrix},$$

where $s^{11} = s_y^2 / |S|$, $s^{12} = s^{21} = -s_{xy} / |S|$, $s^{22} = s_x^2 / |S|$ and $|S| = s_x^2 s_y^2 - s_{xy}^2$,

❖ *TP*:

➢ Matrix $S=(sij)$:

$$s11=***,\quad s12=***,$$

$$s21=***,\quad s22=***,$$

➢ Matrix $S^{-1}=(s-1,ij)$

$$s-1,11=***,\quad s-1,12=***,$$

$$s-1,21=***,\quad s-1,22=***;$$

(g) Classic typical dispersion areas (CTDA) of TDS

➢ $TOR_k=\langle\bar{x}-ks_x;\bar{x}+ks_x\rangle\times\langle\bar{y}-ks_y,\bar{y}+ks_y\rangle$ for $k=1,2,3,$

➢ The area of $|CTDA_k|=4k^2s_xs_y$,

➢ The percentage ratio of the TDA_k and the dispersion rectangle area in %

$$\tau_k=\frac{100|TDA_k|}{|RD|},\quad k=1,2,3\ ,$$

❖ *TP*: for $k=1,2,3$ are given coordinates of the vertex points A,B,C,D for the *TDA*;

(h) The distance of the Euklidean d_{ij} and the diameter δ of the RD set

➢ $d_{ij}=\{(x_i-x_j)'(x_i-x_j)\}^{1/2},\ x_i,x_j\in R^2$ for $1\le i<j\le n$,

❖ *TP*: $(i,j,d_{ij})_m,\ m=1,2,...,\binom{n}{2}$,

➢ $\delta=d_{pq}=\max\limits_{1\le i<j\le n}\{d_{ij}\}$,

❖ *TP*: (p,q,d_{pq});

(i) The distance of the Euclid ean vector observations *RD* from a given vector depth θ

$$\tilde{d}_i=\{(x_i-\theta)'(x_i-\theta)\}^{1/2},\quad i=1,2,...,n$$

❖ TP: $(nr,\tilde{d}_i)$;

(j) Collinearity of three points from within the *TDS*

➢ For each three points $x_i, x_j, x_k \in TDS$ a triangle $\Delta_{ijk} = \Delta(x_i, x_j, x_k)$ can be constructed and its area calculated

$$S_{ijk} = \frac{1}{2} abs \begin{vmatrix} 1 & 1 & 1 \\ x_i & x_j & x_k \end{vmatrix} = \frac{1}{2} abs \begin{vmatrix} 1 & 1 & 1 \\ x_{i1} & x_{j1} & x_{k1} \\ x_{i2} & x_{j2} & x_{k2} \end{vmatrix},$$

➢ The total number of all triangles

$$q = \binom{n}{3} = \frac{1}{6} n(n-1)(n-2),$$

➢ The total area of all triangles Δ_{ijk}

$$P = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} S_{ijk}.$$

TDS includes all collinear observations if $P \le |RD|^*$, where $|RD|^*$ was given in point (d). Three points $x_i, x_j, x_k \in R^2$ from the *TDS* sample are collinear when, $S_{ijk} \le \varepsilon$, when the Δ_{ijk} triangle area is not larger than the given sufficiently small number of ε,

❖ TP: (nr, i_1, j_1, k_1), where $i_1, j_1, k_1 \in \{1,2,...,n\}$;

(k) Determination of observation multiple occurrence in *PD*.

Conditional samples are determined $Y|X = x$ and $X|Y = y$, which corresponds to the projection of observation in *TDS*, respectively onto *OX* and *OY* axis. The algorithm for determining the conditional samples includes the following steps:

(α) *TDS* is projected onto the *OX* axis:

(i) pairs (x_j, y_j) are arranged, $j = 1, 2, ..., n$ according to the x_j value, obtaining a non-decreasingly arranged array $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$,

(ii) m of different values is determined $x_1', x_2', ..., x_m'$ in the $x_{(1)}, x_{(2)}, ..., x_{(n)}$ array,

(iii) a conditional *TDS* for a given x_i' is created

$$CTDS(X = x_i') = \{(x_j, y_j) : x_j = x_i'; j = 1, 2, ..., n\},$$

as well as their sizes $n_{i\cdot} = \#\{CTDS(X = x_i')\}$ for $i = 1, 2, ..., m$,

❖ TP: $(i, n_{i\cdot}, x_i' ** y_k,\ k \in \{j_1, j_2, ..., j_{n_{i\cdot}}\})$, $i = 1, 2, ..., m$;

(β) *TDS* is projected onto the *OY* axis:

(i) pairs (x_j, y_j) are arranged, $j = 1, 2, ..., n$ according to the y_j value, obtaining the following array $y_{(1)} \le y_{(2)} \le ... \le y_{(n)}$,

(ii) r of different values is determined $y_1', y_2', ..., y_r'$ in the $y_{(1)}, y_{(2)}, ..., y_{(n)}$ array,

(iii) a conditional *TDS* for a given y_i' is created

$$CTDS(Y = y_i') = \{(x_j, y_j) : y_j = y_i'; j = 1,2,...,n\},$$

and their sizes $n_{i\cdot} = \#\{CTDS(Y = y_i')\}$ for $i = 1, 2, ..., r$,

❖ TP: $(i, n_{i\cdot}, y_i' ** x_k, k \in \{j_1, j_2, ..., j_{n_{i\cdot}}\})$, $i = 1,2,...,r$;

(l) the $y = a + bx$ lines include the given two points $x_i, x_j \in TDS$

➢ The determinant form

$$L_{ij} : \begin{vmatrix} 1 & 1 & 1 \\ x & x_{i1} & x_{j1} \\ y & x_{i2} & x_{j2} \end{vmatrix} = 0,$$

➢ The extended form

$$L_{ij} : (x_{j1} - x_{i1})(y - x_{i2}) - (x_{j2} - x_{i1})(x - x_{i1}) = 0.$$

The total number of lines $\binom{n}{2} = \frac{1}{2}n(n-1)$. The $y = a + bx$ line in the directional form including points x_i and x_j:

$$a = \frac{\begin{vmatrix} x_{j1} & x_{j2} \\ x_{i1} & x_{i2} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ x_{i1} & x_{j1} \end{vmatrix}}, \quad b = \frac{\begin{vmatrix} 1 & 1 \\ x_{i2} & x_{j2} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ x_{i1} & x_{j1} \end{vmatrix}}.$$

Let set Γ of the $\#\{\Gamma\} = \binom{n}{2}$ size express the set of all lines. For a given L_{ij} line, the (x_p, x_q) point, lies either on:

(i) a line, then $L_{ij}(x_p, x_q) = 0$,

(ii) on the left side, when $L_{ij}(x_p, x_q) < 0$,

(iii) on the right side, when $L_{ij}(x_p, x_q) > 0$.

Case (i) is a π_0 points set lying on the L_{ij} line, case (ii) is a half-plane π_L and case (iii) determines the half-plane π_P. The L_{ij} line is a limiting line (dividing) for $\leq n-2$ points within the *TDS* in R^2. The numbers of points lying on the half-planes π_L and π_P are determined by:

$$j_a = \#\{\pi_L\} \text{ and } j_b = \#\{\pi_P\},$$

in such a way that $j_a + j_b = n-2$, and their minimum is expressed by $j_z = \min\{j_a, j_b\}$. In the set of numbers j_z the maximum number is calculated, for example $j_{\max}$, which allows to determine the number of possible contours for the analysed *TDS*, which is $k_{\max} = [\frac{j_{\max}}{2}] + 1$. where [.] is an integer part of the integer function argument.

❖ TP: $(nr, i, j, a, b, \mathrm{tg}(a), j_a, j_b, j_z)$, for *nr* going from *1* to $\binom{n}{2}$;

(m) Contour TDS.

For order to create contours of *TDS* the lines given in point (l) are used. The contours are built from the edges of calculated lines and their intersection points. As mentioned before, for n points are defined $g = n(n-1)/2$ straight lines of L_{ij} made of pairs (i, j) which meet the condition: $1 \leq i_1 < j_1 \leq n$. We consider pairs of straight lines $L_{i_1 j_1}$ and $L_{i_2 j_2}$ when $(i_1, j_1) \in \Gamma$ and $(i_2, j_2) \in \Gamma$ for which

index pairs include the following ranges: $1 \le i_1 < j_1 \le n$; $(i_2, j_2) \in \{i_1+1, i_2+2,...,n\} - \{j_1\}$. This way we can exclude the pairs of straight lines which had a common observation point from a sample (e.g. pairs of the straight lines (2, 4), (4, 5) have a common point 4, which is also their intersecting point). It has been recorded that the total number of possible pair of straight lines with repeated point numbers to be created could be as large as: $\binom{g}{2} = \frac{1}{8} n(n-2)(n^2-1)$. The total number of intersection points of two lines is expressed by the formula:

$$q_n = 3\sum_{k=1}^{n-3} \binom{n-k}{3} = \frac{1}{2}\sum_{k=1}^{n-3} (n-k)(n-k-1)(n-k-2), \; n = 4, 5, 6,....$$

In particular for n = 4, 5, 6, 7 the following equation is true $q_n = 2n^3 - 21n^2 + 79n - 105$, whereas for n = 8, 9, 10, 11 the equation is the following form $q_n = 4n^3 - 66n^2 + 422n - 990$.

Table 1 includes the illustration of the above formula for a sample of exemplary n = 6 elements. The set I = {1, 2, 3, 4, 5, 6} was divided into two sub-sets $\{i, j\}$ and $\{i+1,...,6\} - \{j\}$, and all possible pairs that could be created from the elements in set $\{i+1, ..., 6\} - \{j\}$ were given. By summing all the obtained pairs (i, j) and (i', j'), is

$$q_6 = 3\sum_{k=1}^{3} \binom{6-k}{3} = 3\left[\binom{5}{3} + \binom{4}{3} + \binom{3}{3}\right] = 45$$

or

$$q_6 = \frac{1}{2}\sum_{k=1}^{3} (6-k)(5-k)(4-k) = \frac{1}{2}\{5 \cdot 4 \cdot 3 + 4 \cdot 3 \cdot 2 + 3 \cdot 2 \cdot 1\} = 45$$

➢ The contour convex hull. It is a convex closed polygon built on the vertexes of certain observations from *TDS* and its each side is determined by the lines whose one of the two separating planes is empty. It means that, according to point (*l*), the limiting lines $y = a + bx$ of this polygon are determined the areas π_L and π_P such as, either $\#\{\pi_L\} = 0$ or $\#\{\pi_P\} = 0$ (Fig. 2).

Table 1

The division of the 6-element set into two sub-sets

$\{i, j\}$	$\{i+1,...,6\}-\{j\}$	Pairs (i', j')
(1,2)	{3, 4, 5, 6}	(3,4), (3,5), (3,6), (4,5), (4,6), (5,6)
(1,3)	{2, 4, 5, 6}	(2,4), (2,5), (2,6), (4,5), (4,6), (5,6)
(1,4)	{2, 3, 5, 6}	(2,3), (2,5), (2,6), (3,5), (3,6), (5,6)
(1,5)	{2, 3, 4, 6}	(2,3), (2,4), (2,6), (3,4), (3,6), (4,6)
(1,6)	{2, 3, 4, 5}	(2,3), (2,4), (2,5), (3,4), (3,5), (4,5)
(2,3)	{4, 5, 6}	(4,5), (4,6), (5,6)
(2,4)	{3, 5, 6}	(3,5), (3,6), (5,6)
(2,5)	{2, 4, 6}	(2,4), (2,6), (4,6)
(2,6)	{3, 4, 5}	(3,4), (3,5), (4,5)
(3,4)	{4, 5}	(4,5)
(3,5)	{4, 6}	(4,6)
(3,6)	{4, 5}	(4,5)

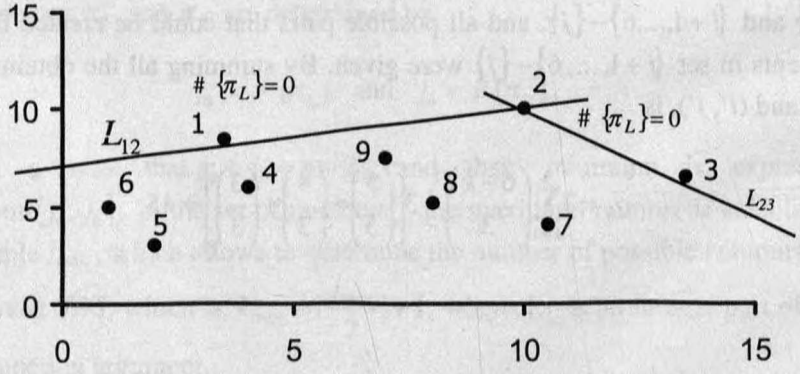

Fig. 2. Illustration of the separating lines

❖ *TP*: ((lp), i, j, a, b), where (i, j) point numbers from set Γ determining edges of the convex hull and a and b are the coefficients of the line crossing the observations $x_i, x_j \in R^2$.

If all observations from *TDS* are included in the convex hull, the analysis of *TDS* is finished. It occurs when $\#\{M_0\}=n$, that is when set M_0 includes all the observations from *TDS* included in the convex hull.

➢ k-th, k = 1, 2,... degree closed convex contours. In order to determine contour Kon_k of k-th degree, the arrangement of lines from set Γ defined in point (*l*) is used, for which one of the halfplanes either π_L or π_P includes k points from *TDS* (Fig.3).

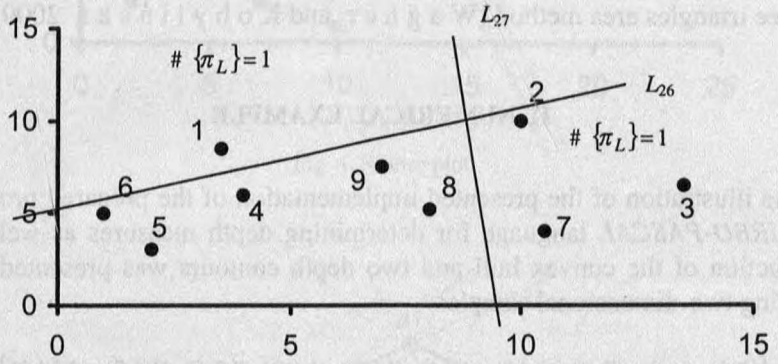

Fig. 3. The illustration of separating half-planes of the k = 1 size

Let set Ψ_k be the set of the lines, of the $m_k = \#\{\Psi_k\}$ size. Actual selection of lines to set Ψ_k is done through reviewing j_z value, for which $j_z = k$, what is done in point (*l*). For m_k lines from set Ψ_k a set of contour vertexes Kon_k is determined, from intersection of two lines $(i_1, j_1, a_{i_1}, b_{j_1})$ and $(i_2, j_2, a_{i_1}, b_{j_2})$, when $(i_1, j_1), (i_2, j_2) \in \Psi_k$ and r_k $i_1 \neq i_2$, $j_1, \neq j_2$. Let set Φ_k express such established set of vertexes about sizes $r_k = \#\{\Phi_k\}$. The size reduction of r_k in the set Φ_k is completed by eliminating the following intersecting points: $x_0 \equiv x_0(i_1, j_1, i_2, j_2)$, $y_0 \equiv y_0(i_1, j_1, i_2, j_2)$ of straight lines $L_{i_1 j_1}, L_{i_2 j_2} \in \Psi_k$ for which:

a) the conditions $i_1 = i_2$ or $j_1 = j_2$, are met and the number of such cases is determined by number f_k,

b) $(x_0, y_0) \notin RD$, i.e. a pair of co-ordinates (x_0, y_0) does not belong to the area of a scattering rectangle *(RD)*, and the number of such pairs is determined by number g_k.

Finally we obtain set Φ_k^* of the size $r_k^* = r_k - f_k - q_h \cdot h_k \geq 1$ points from 2-dimensional sample (*TDS*) are included in the contour Kon_k. It means that each contour includes at least one observation from *TDS*. The set of these points is expressed by set M_k. Conditions $i_1 \neq i_2$ and $j_1 \neq j_2$ aim at eliminating such line intersecting points that may overlap with these observations in *TDS*, that are included in the convex hull and in the previously determined contours $Kon_1, Kon_2, ..., Kon_k$.

Depth measures for elements of set Φ_k^*, $k = 1, 2, ...$ were determined using the three triangles area method (W a g n e r and K o b y l i ń s k a 2000).

II. NUMERICAL EXAMPLE

The illustration of the presented implementation of the prepared program in the *TURBO-PASCAL* language for determining depth measures as well as the construction of the convex hull and two depth contours was presented for the following two-dimensional sample:

$$\{(2,3), (4,9), (7,3), (9,12), (10,1), (11,9), (14,9), (13,6), (17,5), (20,10)\}.$$

The numerical data was listed in the correlation chart (Fig. 4). The main numerical statistics of the given set:

- dispersion rectangle determined by the vertexes: *A*(2, 1), *B*(20, 1), *C*(20, 16) and *D*(2, 16),
- diameter of set 18.68 between points (2, 3) and (20, 8),
- means: $\bar{x} = 10.7$, $\bar{y} = 7.7$,
- medians: $Med_x = 10.5$, $Med_y = 9$.
- Standard deviation: $s_x = 5.58$, $s_y = 4.64$,
- Variation coefficients (%): $v_x = 51.1$, $v_y = 60.3$,
- Skewness coefficients: 0.07 and 0.20,
- Linear correlation coefficients $r = 0.322$,

The observations in *TDS* are included in the following (Fig. 5).

- Convex hull = $\{1, 2, 5, 8, 9, 10\}$,
- Contours: $Kon_1 = \{3, 4\}$, $Kon_2 = \{7\}$ and $Kon_3 = \{6\}$.

The point of co-ordinates (11, 9), is Tukey's median for the analysed sample.

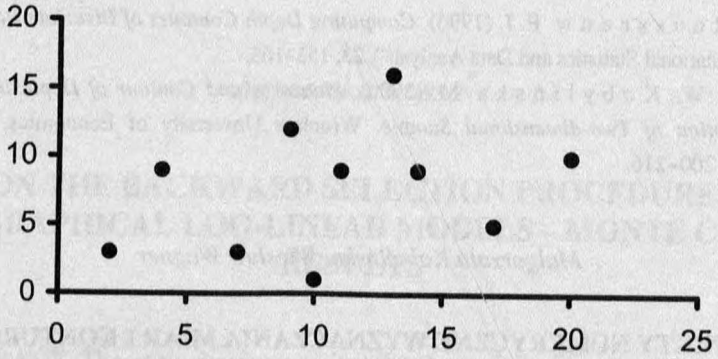

Fig. 4. Scatter plot

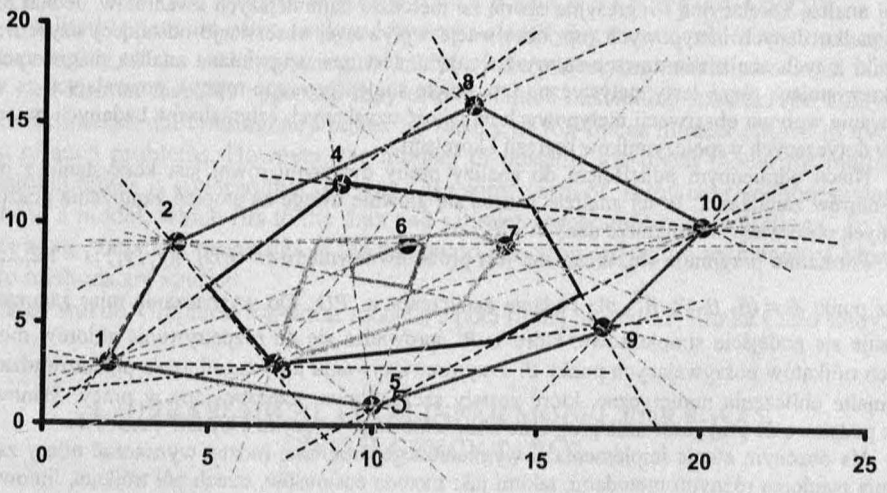

Fig. 5. Convex hull and contours

REFERENCES

Donoho D. L., Gasko M. (1992), *Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness*, „The Annals of Statistics", 20, 1803–1827.

He X., Wang G. (1997), *Convergence of Depth Contours for Multivariate Datasets*, „The Annals of Statistics", **25**, 495–504.

Rousseeuw P. J., Ruts I. (1999), *The Depth Function of a Population Distribution*, „Metrika", **49**, 213–244.

Ruts I., Rousseeuw P. J. (1996), *Computing Depth Contours of Bivariate Point Clouds*, „Computational Statistics and Data Analysis", **23**, 153–168.

Wagner W., Kobylińska M. (2000), *Measures and Contour of Depth in Statistical Description of Two-dimensional Sample*, Wrocław University of Economics, Publishing House, 200–216.

Małgorzata Kobylińska, Wiesław Wagner

ASPEKTY NUMERYCZNE WYZNACZANIA MIAR I KONTURÓW ZANURZENIA DLA DANYCH W R^2

W statystycznej analizie mierzalnych ciągłych danych liczbowych w R^2 stosuje się najczęściej analizę korelacyjną i regresyjną opartą na metodzie najmniejszych kwadratów. Jednakże w przypadku danych nietypowych (np. obserwacje wpływowe, obserwacje odstające) uzyskiwane wyniki z tych analiz nie zawsze są wystarczające. Jest ona uzupełniana analizą diagnostyczną wykorzystującą różne testy statystyczne (np. ucięte studentyzowane reszty), pozwalającą na wykrywanie wpływu obserwacji nietypowych na jakość uzyskanych estymatorów badanych parametrów dotyczących współczynników regresji i korelacji.

Nieco odmiennym podejściem do analizy próby dwuwymiarowej jest korzystanie z miar i konturów zanurzenia. W tej analizie zwraca się głównie uwagę na stopień zanurzenia poszczególnych obserwacji w strukturze danych z R^2.

Formalnie przyjmuje się, iż zadana jest próba dwuwymiarowa (PD) $\{(x_i, y_i) : i = 1,2,...,n\}$, oraz punkt $\theta = (\theta_1, \theta_2)' \in R^2$ dla badania zanurzenia w PD. Do wyznaczania miar zanurzenia stosuje się podejście simpleksowe, które w R^2 sprowadza się do rozpatrywania zbiorów możliwych trójkątów pokrywających punkt θ. Przy rozwiązywaniu tego zagadnienia przeprowadza się rozmaite obliczenia numeryczne, które zostały szczegółowo przedstawione w pracy. Stanowiły one podstawę do przygotowania programu obliczeniowego w języku TURBO-PASCAL.

Na obecnym etapie implementacji wymienionego programu można wyznaczać miary zanurzenia pięcioma różnymi metodami, takimi jak: metoda cosinusów, trzech pól trójkąta, liniowych kombinacji wypukłych, trzech półpłaszczyzn rozdzielających, przekształcenia kątowego oraz metodą odległości Mahalanobisa.

Także nadmienionym programem wyznacza się kontury zanurzenia, w tym ich punkty wierzchołkowe oraz przynależność punktów z PD do poszczególnych stopni konturów, wraz ze wskazaniem punktu medianowego w PD.