# Models for Estimation of Correlation Coefficients Using Conditional Quartiles for Two-dimensional Normal Distributions [1]

**Assoc. Prof. Natalia Nikolova, PhD**
IT Department, Nikola Y. Vaptsarov Naval Academy
e-mail: natalianik@gmail.com

**Assoc. Prof. Daniela Toneva-Zheynova, PhD**
Department of Environmental Protection
Technical University – Varna
e-mail: d_toneva@abv.bg

**Biser Stoyanov, PhD Candidate**
Department of Economics and Management
Technical University-Varna,
e-mail: biserdeev@gmail.com

**Prof. Kiril Tenekedjiev, PhD, DSc, Eng.**
IT Department, Nikola Y. Vaptsarov Naval Academy
e-mail: Kiril.Tenekedjiev@fulbrightmail.org

# Модели за оценка на корелационни коефициенти по условни квартили при двумерни нормални разпределения [1]

**доц. д-р Наталия Николова**
Катедра по информационни технологии
ВВМУ „Никола Й. Вапцаров"
e-mail: natalianik@gmail.com

**доц. д-р Даниела Тонева-Жейнова**
Катедра „Екология и опазване на околната среда"
Технически университет-Варна
e-mail: d_toneva@abv.bg

**докторант Бисер Стоянов**
Катедра „Икономика и мениджмънт"
Технически университет-Варна
e-mail: biserdeev@gmail.com

**проф. д.т.н. инж. Кирил Тенекеджиев**
Катедра по информационни технологии
ВВМУ „Никола Й. Вапцаров"
e-mail: Kiril.Tenekedjiev@fulbrightmail.org

***Abstract**: Constructing the density of a bivariate normal random variable requires finding the covariance matrix. It is then needed to assess the standard deviation of both attributes, as well as the correlation coefficients. Usually, subjective data is used as an input, and it is always in an interval form. The paper presents mathematical models to assess correlation coefficients using conditional subjectively elicited (interval) conditional quartiles for two-dimensional normal distributions. Part of the estimates is found by minimization using weighted least square method.*

***Key words**: bivariate normal variables, conditional density, subjective estimates, correlation coefficients*

***Резюме:*** *Построяването на плътност на двумерна нормална случайна величина изисква намирането на ковариационна матрица. За целта са нужни оценки на стандартното отклонение на двата атрибута, както и оценки на техните корелационни коефициенти. Често като изходни данни за анализа се използват субективни данни, които като правило са в интервална форма. Статията представя математически модели за оценка на корелационни коефициенти по условни субективно оценени (в интервална форма) условни квартили при двумерни нормални разпределения. Част от оценките на коефициента на корелация се намират чрез минимизация по метод на претеглените най-малки квадрати.*

***Ключови думи:*** *двумерни нормални величини, условна плътност, субективни оценки, корелационни коефициенти*

## I. Introduction

The recognition of the condition of a system/process is a subject of many scientific disciplines, among which pattern recognition (Fukunaga, 1990; Duda, et al., 2001), statistics, econometrics (Hanke & Reitsch, 1991), etc. These approaches are related to and support technical diagnostics, regression analysis, financial portfolio optimisation, application of econometric models to test economic theories, etc. Usually, these approaches give a formal description of the object using some of its main parameters, and find the likelihood that the object is at a certain state.

In most cases, the analysed process/object is described in terms of a multivariate random variable. Then it is required to study the correlation between the parameters. As it is usually the case, complete and adequate empirical information in lacking. That is why subjective estimates are employed.

Due to the partial irrationality of people, their subjective estimates come in an interval form. Therefore, in (Nikolova, et al., 2005) real estimators are referred to as fuzzy rational. This paper presents mathematical models that use interval subjective estimates to find correlation coefficients between the attributes of a bivariate normal variable. Initial studies in that respect were presented in (Nikolova et al., 2010a; Nikolova, et al., 2010b).

## I. Въведение

Разпознаване състоянието на дадена система/процес е обект на научни подходи, сред които разпознаване на образи (Fukunaga, 1990; Duda, et al., 2001), статистика, иконометрия (Hanke & Reitsch, 1991) и др. Тези подходи подпомагат техническата диагностика, регресионния анализ, финансовата портфейлна оптимизация, приложението на иконометрични модели към икономически твърдения и др. Обикновено тези подходи дават формално описание на обекта чрез негови основни параметри и определяне правдоподобието обекта да се намира в дадено състояние.
В много случаи, изследвания процес/обект се описва с множество параметри, т.е. описанието е под формата на многомерна случайна величина. В този случай се изследва зависимостта между параметрите. Тъй като обикновено не е налице пълна и достоверна информация, се налага използването на субективна вероятностна информация.
Поради частичната ирационалност на хората, субективните оценки са интервални. Затова в (Nikolova, et al., 2005) реалните оценители се наричат размито рационални. Тук се представят математически зависимости и модел, чрез които по субективни интервални оценки се оценява степента на корелация между атрибути на двумерна нормална случайна величина. Начални разработки са дадени в (Nikolova et al., 2010a; Nikolova, et al., 2010b).

## II. Finding the parameters of a two-dimensional conditional density on subjective estimates

Let's have a random transposed vector $\vec{z} = (x, y)^T$. The transposed vector of mean values is $\vec{\mu} = (\mu_x, \mu_y)^T$ with coordinates – the mean values of the attributes *x* and *y*. Since $\vec{z}$ obeys a two-dimensional normal distribution, then its density is given by:

## II. Намиране на параметри на двумерна нормална условна плътност по субективни оценки

Нека е дадена двумерно-нормална случайна величина $\vec{z} = (x, y)^T$. Транспонирания вектор на средните стойности е $\vec{\mu} = (\mu_x, \mu_y)^T$ с координати – средните стойности на атрибутите *x* и *y*. Тъй като $\vec{z}$ е двумерно-нормално разпределен, то плътността може да се изчисли като:

$$(1)\ f(\vec{z}) = f(x, y) = N(\vec{x}, \vec{\mu}, S) = \frac{1}{2\pi\sqrt{|S|}} exp\left(-\frac{1}{2}(\vec{z} - \vec{\mu})^T S^{-1} (\vec{z} - \vec{\mu})\right).$$

Here, *S* is the covariance matrix

Тук, *S* е ковариационната матрица

$$(2)\ S = \begin{pmatrix} \sigma_x^2 & r\sigma_x\sigma_y \\ r\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix},$$

where $\sigma_x$ is the standard deviation of *x*, $\sigma_y$ is the standard deviation of *y*, and *r* is the correlation coefficient between *x* and *y*. The following applies:

където $\sigma_x$ е стандартното отклонение на *x*, $\sigma_y$ е стандартното отклонение на *y*, а *r* е коефициент на корелация между *x* и *y*. В сила са следните зависимости и условия:

$$(3)\ \sigma_x \ge 0, \sigma_y \ge 0,\ r \in [-1; 1],$$

$$(4)\ |S| = \sigma_x^2\sigma_y^2\left(1 - r^2\right),$$

$$(5)\ \sqrt{|S|} = \sigma_x\sigma_y\sqrt{\left(1 - r^2\right)},$$

$$(6)\ S^{-1} = \frac{1}{|S|}\begin{pmatrix} \sigma_y^2 & -r\sigma_x\sigma_y \\ -r\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix}.$$

Therefore, in order to construct the tow-dimensional density one needs to estimate $\sigma_x$, $\sigma_y$, *r*.

Следователно, за построяване на двумерната нормална плътност е нужно да се оценят $\sigma_x$, $\sigma_y$, *r*.

### *II.1. Estimating standard deviations*

The following is true for the normally distributed *x*:

### *II.1. Оценка на стандартните отклонения*

За нормално разпределения атрибут *x* е в сила:

$$(7)\ f(x) = N(x, \mu_x, \sigma_x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{1}{2}\left(\frac{x - \mu_x}{\sigma_x}\right)^2\right).$$

A normal random variable *z* with mean value of 0 and standard deviation of 1 may be introduced, such that

Може да се въведе нормална величина *z* със средна стойност 0 и стандартно отклонение 1, такава че

$$(8)\ z = \frac{x - \mu_x}{\sigma_x}.$$

$$(9)\ x = z\sigma_x + \mu_x \Rightarrow \sigma_x = \frac{x-\mu_x}{z} .$$

If $x = x_\alpha \in \left[x_\alpha^d, x_\alpha^u\right]$ and $\mu_x = x_{0,5} \in [x_{0.5}^d, x_{0.5}^u]$ are subjectively assessed, then point estimates may be found as:

Ако са налице субективни оценки $x = x_\alpha \in \left[x_\alpha^d, x_\alpha^u\right]$ и $\mu_x = x_{0,5} \in [x_{0.5}^d, x_{0.5}^u]$, то могат да се намерят точкови оценки като средите на тези интервали:

$$(10)\ x_\alpha^{mean} = \frac{x_\alpha^d + x_\alpha^u}{2},\ x_{0.5}^{mean} = \frac{x_{0.5}^d + x_{0.5}^u}{2}.$$

Then $\sigma_x$ may be found as

Тогава $\sigma_x$ може да се намери като

$$(11)\ \sigma_x = \frac{x_\alpha^{mean} - x_{0.5}^{mean}}{z_\alpha} ,$$

$$(12)\ \alpha = \int_{-\infty}^{z_\alpha} N(x,0,1)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} \exp\left(-\frac{x^2}{2}\right) dx .$$

In the previous formula, $\alpha$ should not be 0.5, because it will lead to uncertainty of type 0/0. If $\alpha$ is close to either 0 or 1, the estimates are too imprecise. A good practice is to use 0.25 and 0.75 as values of $\alpha$:

В горната формула, $\alpha$ не трябва да е 0.5, тъй като ще доведе до несигурност от тип 0/0. Ако $\alpha$ е близо или до 1 или до 0, оценките биха били твърде неточни. Подходящо е да се използват стойностите 0.25 и 0.75:

$$(13)\ \sigma_{x,1} = \frac{x_{0.25}^{mean} - x_{0.5}^{mean}}{z_{0.25}} ,$$

$$(14)\ \sigma_{x,2} = \frac{x_{0.75}^{mean} - x_{0.5}^{mean}}{z_{0.75}} .$$

Here, $z_{0.25}$ and $z_{0.75}$ are respectively the 0.25 and 0.75 quantile of a normal distribution $N(0,1)$. If $\sigma_{x,1}$ and $\sigma_{x,2}$ are both estimated, then one obvious possibility is:

Тук, $z_{0.25}$ представлява квантил 0.25 на нормално разпределение $N(0,1)$. Аналогично е и значението на $z_{0.75}$. Ако $\sigma_{x,1}$ и $\sigma_{x,2}$ са оценени, то трета оценка на $\sigma_x$ може да е:

$$(15)\ \sigma_{x,3} = \frac{\sigma_{x,1} + \sigma_{x,2}}{2} .$$

A better approach is to find $\sigma_x$ using weighted least square (WLS) minimisation (Vuchkov & Stoyanov, 1986; Press et al., 1992):

По-добър подход е да се намери $\sigma_x$ чрез минимизация по метод на претеглените най-малки квадрати (WLS) (Vuchkov & Stoyanov, 1986; Press et al., 1992):

$$(16)\ \chi_{\sigma,4}^2 = \left(\frac{\left(x_{0.25}\right)^{model} - x_{0.25}^{mean}}{x_{0.25}^u - x_{0.25}^d}\right)^2 + \left(\frac{\left(x_{0.75}\right)^{model} - x_{0.75}^{mean}}{x_{0.75}^u - x_{0.75}^d}\right)^2 .$$

It is possible to prove that

Тогава може да се докаже, че

$$(17)\ \sigma_{x,4} = \arg\left\{\min_{\sigma_x}\left(\chi^2_{\sigma,4}\right)\right\} = \frac{\dfrac{z_{0.25}\left(x^{mean}_{0.25} - x^{mean}_{0.5}\right)}{\left(x^u_{0.25} - x^d_{0.25}\right)^2} + \dfrac{z_{0.75}\left(x^{mean}_{0.75} - x^{mean}_{0.5}\right)}{\left(x^u_{0.75} - x^d_{0.75}\right)^2}}{\left(\dfrac{z_{0.25}}{x^u_{0.25} - x^d_{0.25}}\right)^2 + \left(\dfrac{z_{0.75}}{x^u_{0.75} - x^d_{0.75}}\right)^2}.$$

### *II.2. Estimation of the correlation coefficient*

It is possible to prove that the distribution $f(\vec{z})$ is:

### *II.2. Оценка на корелационния коефициент*

Може да се докаже, че нормалната плътност $f(\vec{z})$ е:

$$f(\vec{z}) = f(x,y) = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-r^2}} exp\left(-\frac{1}{2}\left(\frac{x - (\mu_x + ry\sigma_x/\sigma_y - r\mu_y\sigma_x/\sigma_y)}{\sigma_x\sqrt{1-r^2}}\right)^2\right).$$

$$.\frac{1}{\sqrt{2\pi}\sigma_y} exp\left(-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right) =$$

$$= N\left(x, \mu_x + ry\sigma_x/\sigma_y - r\mu_y\sigma_x/\sigma_y, \sigma_x\sqrt{1-r^2}\right)N\left(y,\mu_y,\sigma_y\right) = f(x \mid y).f(y)$$

Then the conditional density $f(x \mid y)$ may be presented as

Оттук условната плътност $f(x \mid y)$ може да се представи като

$$(18)\quad \begin{aligned} f(x \mid y) &= \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-r^2}} exp\left(-\frac{1}{2}\left(\frac{x - \left(\mu_x + ry\sigma_x/\sigma_y - r\mu_y\sigma_x/\sigma_y\right)}{\sigma_x\sqrt{1-r^2}}\right)^2\right) = \\ &= N\left(x, \mu_x + ry\sigma_x/\sigma_y - r\mu_y\sigma_x/\sigma_y, \sigma_x\sqrt{1-r^2}\right) = N\left(x, \mu_{x|y}, \sigma_{x|y}\right), \end{aligned}$$

$$(19)\ \mu_{x|y} = \mu_x - r\mu_y\sigma_x/\sigma_y + yr\sigma_x/\sigma_y \qquad \sigma_{x|y} = \sigma_x\sqrt{1-r^2}$$

Then the correlation coefficient $r$ may be found as:

Тогава корелационния коефициент $r$ може да се намери като:

$$(20)\ r = \frac{\mu_{x|y} - \mu_x}{y\sigma_x/\sigma_y - \mu_y\sigma_x/\sigma_y} = \left(\frac{\mu_{x|y} - \mu_x}{\sigma_x}\right) \Big/ \left(\frac{y - \mu_y}{\sigma_y}\right).$$

If $y = y_\alpha \in \left[y^d_\alpha, y^u_\alpha\right]$ and $\mu_{x|y} = \left(x_{0.5} \mid y_\alpha\right) \in \left[\left(x_{0.5} \mid y_\alpha\right)^d, \left(x_{0.5} \mid y_\alpha\right)^u\right]$ are subjectively elicited, then:

Ако са налице субективни оценки $y = y_\alpha \in \left[y^d_\alpha, y^u_\alpha\right]$ и $\mu_{x|y} = \left(x_{0.5} \mid y_\alpha\right) \in \left[\left(x_{0.5} \mid y_\alpha\right)^d, \left(x_{0.5} \mid y_\alpha\right)^u\right]$, то:

$$(21)\quad y^{mean}_\alpha = \frac{y^d_\alpha + y^u_\alpha}{2},$$

$$(22)\ \left(x_{0.5} \mid y_\alpha\right)^{mean} = \frac{\left(x_{0.5} \mid y_\alpha\right)^d + \left(x_{0.5} \mid y_\alpha\right)^u}{2},$$

$$(23)\ r = \left(\frac{\left(x_{0.5} \mid y_\alpha\right)^{mean} - \mu_x}{\sigma_x}\right) \Big/ \left(\frac{y^{mean}_\alpha - \mu_y}{\sigma_y}\right).$$

In the previous formula, $\alpha$ should not be 0.5, because it will lead to uncertainty of type 0/0. If $\alpha$ is close to either 0 or 1, the estimates are too imprecise. A good practice is to use 0.25 and 0.75 as values of $\alpha$. Then two possible estimates of *r* follow from (23):

В горната формула, $\alpha$ не трябва да е 0.5, тъй като ще доведе до несигурност от тип 0/0. Ако $\alpha$ е близо до 0 или 1, оценките ще бъдат твърде неточни. Препоръчителните стойности за $\alpha$ са 0.25 и 0.75. Тогава от (23) следват две възможни оценки на *r*:

$$(24)\; r_1 = \left( \frac{(x_{0.5} \mid y_{0.25})^{mean} - \mu_x}{\sigma_x} \right) \Bigg/ \left( \frac{y_{0.25}^{mean} - \mu_y}{\sigma_y} \right),$$

$$(25)\; r_2 = \left( \frac{(x_{0.5} \mid y_{0.75})^{mean} - \mu_x}{\sigma_x} \right) \Bigg/ \left( \frac{y_{0.75}^{mean} - \mu_y}{\sigma_y} \right).$$

Then if both $r_1$ and $r_2$ are estimated, one obvious possibility is

Ако $r_1$ и $r_2$ са оценени, то трета оценка на *r* може да се намери като:

$$(26) \quad r_3 = \frac{r_1 + r_2}{2}.$$

A better approach is to find *r* using weighted least square (WLS) minimisation of:

По-добър подход е да се намери *r* чрез WLS минимизация:

$$(27) \quad \chi_4^2 = \left( \frac{(x_{0.5} \mid y_{0.25})^{model} - (x_{0.5} \mid y_{0.25})^{mean}}{(x_{0.5} \mid y_{0.25})^{u} - (x_{0.5} \mid y_{0.25})^{d}} \right)^2 + \left( \frac{(x_{0.5} \mid y_{0.75})^{model} - (x_{0.5} \mid y_{0.75})^{mean}}{(x_{0.5} \mid y_{0.75})^{u} - (x_{0.5} \mid y_{0.75})^{d}} \right)^2$$

Then it can be proven that

Тогава може да се докаже, че

$$(28) \quad r_4 = \arg\left\{ \min_r \left( \chi_4^2 \right) \right\} =$$

$$= \frac{\left( \frac{\left( (x_{0.5} \mid y_{0.25})^{mean} - \mu_x \right)\left( y_{0.25}^{mean} \sigma_x / \sigma_y - \mu_y \sigma_x / \sigma_y \right)}{\left( (x_{0.5} \mid y_{0.25})^{u} - (x_{0.5} \mid y_{0.25})^{d} \right)^2} + \frac{\left( (x_{0.5} \mid y_{0.75})^{mean} - \mu_x \right)\left( y_{0.75}^{mean} \sigma_x / \sigma_y - \mu_y \sigma_x / \sigma_y \right)}{\left( (x_{0.5} \mid y_{0.75})^{u} - (x_{0.5} \mid y_{0.75})^{d} \right)^2} \right)}{\left( \frac{y_{0.25}^{mean} \sigma_x / \sigma_y - \mu_y \sigma_x / \sigma_y}{(x_{0.5} \mid y_{0.25})^{u} - (x_{0.5} \mid y_{0.25})^{d}} \right)^2 + \left( \frac{y_{0.75}^{mean} \sigma_x / \sigma_y - \mu_y \sigma_x / \sigma_y}{(x_{0.5} \mid y_{0.75})^{u} - (x_{0.5} \mid y_{0.75})^{d}} \right)^2}$$

Similar formulae may be derived for *y* and *x*:

Аналогични формули могат да се изведат и за условното разпределение на *y* при дадено *x*:

$$(29) \quad \begin{aligned} & f(\vec{z}) = f(x, y) = N(\vec{x}, \vec{\mu}, S) = \\ & = N\left( y, \mu_y + r x \sigma_y / \sigma_x - r \mu_x \sigma_y / \sigma_x, \sigma_y \sqrt{1 - r^2} \right) N(x, \mu_x, \sigma_x) = f(y \mid x) f(x), \end{aligned}$$

$$\Rightarrow f(y\,|\,x)=\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-r^2}}\exp\left(-\frac{1}{2}\left(\frac{y-(\mu_y+rx\sigma_y/\sigma_x-r\mu_x\sigma_y/\sigma_x)}{\sigma_y\sqrt{1-r^2}}\right)^2\right)=$$

$$=N\left(y,\mu_y+rx\sigma_y/\sigma_x-r\mu_x\sigma_y/\sigma_x,\sigma_y\sqrt{1-r^2}\right)=N\left(y,\mu_{y|x},\sigma_{y|x}\right)$$

$$(30)\quad \begin{aligned}\mu_{y|x}&=\mu_y-r\mu_x\sigma_y/\sigma_x+rx\sigma_y/\sigma_x,\\ \sigma_{y|x}&=\sigma_y\sqrt{1-r^2}\;.\end{aligned}$$

Then the correlation coefficient $r$ is

Тогава корелационния коефициент $r$ може да се намери като:

$$(31)\; r=\frac{\mu_{y|x}-\mu_y}{x\sigma_y/\sigma_x-\mu_x\sigma_y/\sigma_x}=\left(\frac{\mu_{y|x}-\mu_y}{\sigma_y}\right)\Big/\left(\frac{x-\mu_x}{\sigma_x}\right).$$

If $x=x_\alpha\in\left[x_\alpha^d,x_\alpha^u\right]$ and $\mu_{y|x}=\left(y_{0.5}\,|\,x_\alpha\right)\in\left[\left(y_{0.5}\,|\,x_\alpha\right)^d,\left(y_{0.5}\,|\,x_\alpha\right)^u\right]$ are subjectively elicited, then:

Ако са налице субективни оценки $x=x_\alpha\in\left[x_\alpha^d,x_\alpha^u\right]$ и $\mu_{y|x}=\left(y_{0.5}\,|\,x_\alpha\right)\in\left[\left(y_{0.5}\,|\,x_\alpha\right)^d,\left(y_{0.5}\,|\,x_\alpha\right)^u\right]$ то:

$$(32)\quad x_\alpha^{mean}=\frac{x_\alpha^d+x_\alpha^u}{2}\;,$$

$$(33)\;\left(y_{0.5}\,|\,x_\alpha\right)^{mean}=\frac{\left(y_{0.5}\,|\,x_\alpha\right)^d+\left(y_{0.5}\,|\,x_\alpha\right)^u}{2}\,,$$

$$(34)\; r=\left(\frac{\left(y_{0.5}\,|\,x_\alpha\right)^{mean}-\mu_y}{\sigma_y}\right)\Big/\left(\frac{x_\alpha^{mean}-\mu_x}{\sigma_x}\right).$$

If the values 0.25 and 0.75 are assigned to $\alpha$, then:

При стойности 0.25 и 0.75 за $\alpha$ следват две възможни оценки на $r$:

$$(35)\; r_5=\left(\frac{\left(y_{0.5}\,|\,x_{0.25}\right)^{mean}-\mu_y}{\sigma_y}\right)\Big/\left(\frac{x_{0.25}^{mean}-\mu_x}{\sigma_x}\right),$$

$$(36)\; r_6=\left(\frac{\left(y_{0.5}\,|\,x_{0.75}\right)^{mean}-\mu_y}{\sigma_y}\right)\Big/\left(\frac{x_{0.75}^{mean}-\mu_x}{\sigma_x}\right).$$

If both $r_5$ and $r_6$ are assessed, then:

Ако $r_5$ и $r_6$ са оценени, тогава:

$$(37)\quad r_7=\frac{r_5+r_6}{2}\,,$$

$$(38)\;\chi_8^2=\left(\frac{\left(y_{0.5}\,|\,x_{0.25}\right)^{model}-\left(y_{0.5}\,|\,x_{0.25}\right)^{mean}}{\left(y_{0.5}\,|\,x_{0.25}\right)^{u}-\left(y_{0.5}\,|\,x_{0.25}\right)^{d}}\right)^2+\left(\frac{\left(y_{0.5}\,|\,x_{0.75}\right)^{model}-\left(y_{0.5}\,|\,x_{0.75}\right)^{mean}}{\left(y_{0.5}\,|\,x_{0.75}\right)^{u}-\left(y_{0.5}\,|\,x_{0.75}\right)^{d}}\right)^2,$$

$$r_8 = \arg\left\{\min_r\left(\chi_8^2\right)\right\} =$$

(39)
$$= \frac{\left(\frac{\left(\left(y_{0.5} \mid x_{0.25}\right)^{mean} - \mu_y\right)\left(x_{0.25}^{mean}\sigma_y/\sigma_x - \mu_x\sigma_y/\sigma_x\right)}{\left(\left(y_{0.5} \mid x_{0.25}\right)^{u} - \left(y_{0.5} \mid x_{0.25}\right)^{d}\right)^2} + \frac{\left(\left(y_{0.5} \mid x_{0.75}\right)^{mean} - \mu_y\right)\left(x_{0.75}^{mean}\sigma_y/\sigma_x - \mu_x\sigma_y/\sigma_x\right)}{\left(\left(y_{0.5} \mid x_{0.75}\right)^{u} - \left(y_{0.5} \mid x_{0.75}\right)^{d}\right)^2}\right)}{\left(\frac{x_{0.25}^{mean}\sigma_y/\sigma_x - \mu_x\sigma_y/\sigma_x}{\left(y_{0.5} \mid x_{0.25}\right)^{u} - \left(y_{0.5} \mid x_{0.25}\right)^{d}}\right)^2 + \left(\frac{x_{0.75}^{mean}\sigma_y/\sigma_x - \mu_x\sigma_y/\sigma_x}{\left(y_{0.5} \mid x_{0.75}\right)^{u} - \left(y_{0.5} \mid x_{0.75}\right)^{d}}\right)^2}$$

If both $r_4$ and $r_8$ are estimated, then a better approach is to find $r$ using WLS minimisation of:

Ако са намерени стойности за $r_4$ и $r_8$, то е подходящо да се намери $r$ чрез WLS минимизация на:

$$\chi_9^2 = \chi_4^2 + \chi_8^2 . \qquad (40)$$

$$r_9 = \arg\left\{\min_r\left(\chi_9^2\right)\right\} =$$

(41)
$$= \frac{\left(\begin{aligned}&\frac{\left(\left(x_{0.5} \mid y_{0.25}\right)^{mean} - \mu_x\right)\left(y_{0.25}^{mean}\sigma_x/\sigma_y - \mu_y\sigma_x/\sigma_y\right)}{\left(\left(x_{0.5} \mid y_{0.25}\right)^{u} - \left(x_{0.5} \mid y_{0.25}\right)^{d}\right)^2} + \frac{\left(\left(x_{0.5} \mid y_{0.75}\right)^{mean} - \mu_x\right)\left(y_{0.75}^{mean}\sigma_x/\sigma_y - \mu_y\sigma_x/\sigma_y\right)}{\left(\left(x_{0.5} \mid y_{0.75}\right)^{u} - \left(x_{0.5} \mid y_{0.75}\right)^{d}\right)^2} +\\ &\frac{\left(\left(y_{0.5} \mid x_{0.25}\right)^{mean} - \mu_y\right)\left(x_{0.25}^{mean}\sigma_y/\sigma_x - \mu_x\sigma_y/\sigma_x\right)}{\left(\left(y_{0.5} \mid x_{0.25}\right)^{u} - \left(y_{0.5} \mid x_{0.25}\right)^{d}\right)^2} + \frac{\left(\left(y_{0.5} \mid x_{0.75}\right)^{mean} - \mu_y\right)\left(x_{0.75}^{mean}\sigma_y/\sigma_x - \mu_x\sigma_y/\sigma_x\right)}{\left(\left(y_{0.5} \mid x_{0.75}\right)^{u} - \left(y_{0.5} \mid x_{0.75}\right)^{d}\right)^2}\end{aligned}\right)}{\left(\begin{aligned}&\left(\frac{y_{0.25}^{mean}\sigma_x/\sigma_y - \mu_y\sigma_x/\sigma_y}{\left(x_{0.5} \mid y_{0.25}\right)^{u} - \left(x_{0.5} \mid y_{0.25}\right)^{d}}\right)^2 + \left(\frac{y_{0.75}^{mean}\sigma_x/\sigma_y - \mu_y\sigma_x/\sigma_y}{\left(x_{0.5} \mid y_{0.75}\right)^{u} - \left(x_{0.5} \mid y_{0.75}\right)^{d}}\right)^2 +\\ &+\left(\frac{x_{0.25}^{mean}\sigma_y/\sigma_x - \mu_x\sigma_y/\sigma_x}{\left(y_{0.5} \mid x_{0.25}\right)^{u} - \left(y_{0.5} \mid x_{0.25}\right)^{d}}\right)^2 + \left(\frac{x_{0.75}^{mean}\sigma_y/\sigma_x - \mu_x\sigma_y/\sigma_x}{\left(y_{0.5} \mid x_{0.75}\right)^{u} - \left(y_{0.5} \mid x_{0.75}\right)^{d}}\right)^2\end{aligned}\right)}$$

In that way, all necessary parameters for the creation of the covariance matrix are estimated.

Така са пресметнати всички необходими елементи за създаването на ковариационната матрица.

## III. Numerical example

### *III.1. Input data*

Let's analyse a bivariate vector $\vec{z} = (x, y)$, whose parameters are related with Bulgarian men: $x$ – height (in cm); $y$ – weight (in kg). Using algorithms for elicitation of subjective quantiles (Nikolova et al., 2004), the following subjective estimates of the quartiles for $x$ and $y$ are found: $x_{0.25} \in [171; 174]$, $x_{0.5} \in [178; 182]$, $x_{0.75} \in [184; 190]$, $y_{0.25} \in [69; 73]$, $y_{0.5} \in [79; 84]$, $y_{0.75} \in [90; 92]$.

## III. Числов експеримент

### *III.1. Начални данни*

Нека се изследва двумерната случайна величина $\vec{z} = (x, y)$, свързана с два физически параметъра на мъжете в България: $x$ – ръст (в cm); $y$ – тегло (в kg). Чрез използване на алгоритми за субективна оценка (Nikolova et al., 2004) са намерени субективни оценки на квартилите на разпределенията на $x$ и $y$: $x_{0.25} \in [171; 174]$, $x_{0.5} \in [178; 182]$, $x_{0.75} \in [184; 190]$, $y_{0.25} \in [69; 73]$, $y_{0.5} \in [79; 84]$, $y_{0.75} \in [90; 92]$.

The medians of the height at the two outer quartiles of weight are also subjectively elicited: $(x_{0.5} \mid y_{0.25}) \in [172, 173]$, $(x_{0.5} \mid y_{0.75}) \in [187, 189]$, as well as the medians of weight at the two outer quartiles of height: $(y_{0.5} \mid x_{0.25}) \in [70, 71]$, $(y_{0.5} \mid x_{0.75}) \in [89, 90]$.

### *III.2. Finding estimates for* $\sigma_x$ *and* $\sigma_y$

It is easy to find the 0.25 and 0.75 quantiles of the normal distribution $N(0,1)$, as $z_{0.75}$=0.6745, $z_{0.25}$= −0.6745 (Hanke & Reitsch, 1991). According to (13) and (14), the first two estimates of $\sigma_x$ are $\sigma_{x,1} = 11.1195$, $\sigma_{x,2} = 10.3782$.

Then following (15), $\sigma_{x,3}$ = 10.7489. Using (16) and (17), it is possible to find $\sigma_{x,4} = 10.9713$. This estimate is used to construct the distribution of $x$ on the elicited quartiles (fig. 1). In the same fashion, the first two estimates of $\sigma_y$ are $\sigma_{y,1} = 15.5673$, $\sigma_{y,2} = 14.0847$. Then according to (15), $\sigma_{y,3}$ = 14.8260. Using (16) and (17), it is possible to find $\sigma_{y,4} = 15.1111$. This estimate is used to construct the distribution of $y$ on the elicited quartiles (fig. 2). According to (10), the point estimates of the quartiles of $x$ are: $x_{0.25}^{mean} = \frac{171+174}{2} = 172.5$, $x_{0.5}^{mean} = \frac{178+182}{2} = 180$, $x_{0.75}^{mean} = \frac{184+190}{2} = 187$. According to (21), the point estimates of the quartiles of $y$ are: $y_{0.25}^{mean} = \frac{69+73}{2} = 71$, $y_{0.5}^{mean} = \frac{79+84}{2} = 81.5$, $y_{0.75}^{mean} = \frac{90+92}{2} = 91$.

Субективно са оценени медианите на ръста при двата квартила на теглото: $(x_{0.5} \mid y_{0.25}) \in [172, 173]$, $(x_{0.5} \mid y_{0.75}) \in [187, 189]$, както и медианите на теглото при двата квартила на ръст: $(y_{0.5} \mid x_{0.25}) \in [70, 71]$, $(y_{0.5} \mid x_{0.75}) \in [89, 90]$.

### *III.2. Намиране на оценки на* $\sigma_x$ *и* $\sigma_y$

За нормално разпределение $N(0,1)$, могат лесно да се намерят $z_{0.75}$=0.6745, $z_{0.25}$= −0.6745 (Hanke & Reitsch, 1991). Съгласно (13) и (14), първите две оценки на $\sigma_x$ ще са $\sigma_{x,1}$ = 11.1195, $\sigma_{x,2} = 10.3782$.

Тогава, съгласно (15), $\sigma_{x,3}$ = 10.7489. По (16) и (17), може да се намери и $\sigma_{x,4}$ = 10.9713. Тази оценка е използвана при построяване на разпределението на $x$ по оценените квартили (фиг. 1). Аналогично, първите две оценки на $\sigma_y$ ще са $\sigma_{y,1} = 15.5673$, $\sigma_{y,2}$ = 14.0847. Тогава, съгласно (15), $\sigma_{y,3} = 14.8260$. По (16) и (17), може да се намери $\sigma_{y,4} = 15.1111$. Тази оценка е използвана при построяване на разпределението на $y$ по оценените квартили (фиг. 2). Съгласно (10), точковите оценки на квартилите на $x$ са: $x_{0.25}^{mean} = \frac{171+174}{2} = 172.5$, $x_{0.5}^{mean} = \frac{178+182}{2} = 180$, $x_{0.75}^{mean} = \frac{184+190}{2} = 187$. Съгласно (21), точковите оценки на квартилите на $y$ са: $y_{0.25}^{mean} = \frac{69+73}{2} = 71$, $y_{0.5}^{mean} = \frac{79+84}{2} = 81.5$, $y_{0.75}^{mean} = \frac{90+92}{2} = 91$.

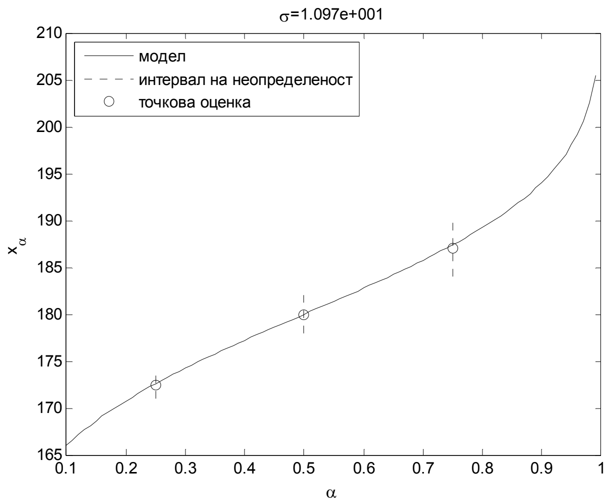


Fig. 1. Distribution law of $x$ (height) on the elicited interval quartiles, for $\sigma_{x,4} = 10.9713$

Фиг. 1. Закон на разпределение на величината $x$ (ръст) по оценени интервални квартили, при $\sigma_{x,4} = 10.9713$

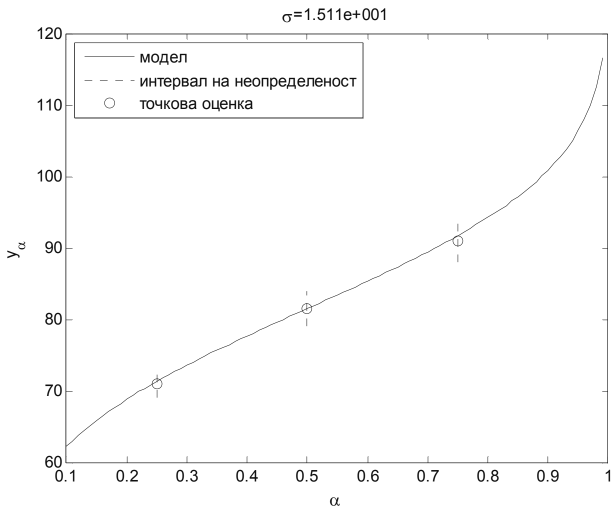


Fig. 2. Distribution law of $y$ (weight) on the elicited interval quartiles, for $\sigma_{y,4} = 15.1111$

Фиг. 2. Закон на разпределение на величината $y$ (тегло) по оценени интервални квартили, при $\sigma_{y,4} = 15.1111$

According to (22), the point estimates of the median of $x$ at quartiles $y_{0.25}$ and $y_{0.75}$ are:

Съгласно (22), точкови оценки на медианата на $x$ при квартили $y_{0.25}$ и $y_{0.75}$ са:

$$\left(x_{0.5} \mid y_{0.25}\right)^{mean} = \frac{172+173}{2} = 172.5,$$

$$\left(x_{0.5} \mid y_{0.75}\right)^{mean} = \frac{187+189}{2} = 188$$

Then, according to (24) and (25), $r_1$=0.5247, $r_2$=0.5074. Since those estimates are available, then $r_3 = \frac{r_1 + r_2}{2} = \frac{0.5247 + 0.5074}{2} = 0.5161$ .

The fourth estimate of $r$ may be found using WLS minimisation of (27), according to (28), as $r_4$=0.5145. The linear dependence between $y$ and $x$ at $r_4$=0.5145 is given in fig. 3.

Тогава, съгласно (24) и (25), $r_1$=0.5247, $r_2$=0.5074. При наличие на тези две оценки, може да се намери $r_3 = \frac{r_1 + r_2}{2} = \frac{0.5247 + 0.5074}{2} = 0.5161$ .

Чрез WLS минимизация на (27), съгласно (28), може да се намери и четвърта оценка $r_4$=0.5145. Линейната връзка между $y$ и $x$ при $r_4$=0.5145 е дадена на фиг. 3.

In the same fashion, according to (33), the point estimates of the median of the weight $y$ at the quartiles $x_{0.25}$ and $x_{0.75}$ are: $\left(y_{0.5} \mid x_{0.25}\right)^{mean} = \frac{70+71}{2} = 70.5$ , $\left(y_{0.5} \mid x_{0.25}\right)^{mean} = \frac{89+90}{2} = 89.5$.

Аналогично, съгласно (33), точкови оценки на медианата на $y$ при квартили $x_{0.25}$ и $x_{0.75}$ са: $\left(y_{0.5} \mid x_{0.25}\right)^{mean} = \frac{70+71}{2} = 70.5$ , $\left(y_{0.5} \mid x_{0.25}\right)^{mean} = \frac{89+90}{2} = 89.5$.

Then, according to (35) and (36), $r_5$=0.5808, $r_6$=0.5186. Since those estimates are available, then $r_7 = \frac{r_5 + r_6}{2} = \frac{0.5808 + 0.5186}{2} = 0.5497$ . The fourth estimate of $r$ may be found using WLS minimisation of (38), according to (39), as $r_8$=0.5519. The linear dependence between $x$ and $y$ at $r_8$=0.0.5519 is given in fig. 4. If both $r_4$ and $r_8$ are available, then using (40) and (41), $r_9$=0.5411. This is employed to construct the linear dependencies on fig. 5.

Тогава, съгласно (35) и (36), $r_5$=0.5808, $r_6$=0.5186. При наличие на тези две оценки, може да се намери $r_7 = \frac{r_5 + r_6}{2} = \frac{0.5808 + 0.5186}{2} = 0.5497$ . Чрез WLS минимизация на (38), съгласно (39), се намира $r_8$=0.5519. Линейната връзка между $x$ и $y$ при $r_8$=0.0.5519 е дадена на фиг. 4. При наличие на стойности за $r_4$ и $r_8$, то чрез (40) и (41) се намира, че $r_9$=0.5411. Тази оценка е използвана за построяване на двете линейни зависимости на фиг. 5.

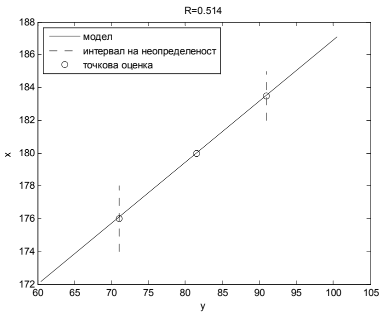


Fig. 3. Linear dependence of $x$ at given values of $y$, constructed on conditional subjectively elicited medians of $x$ at $y_{0.25}$ and $y_{0.75}$, for $r_4$=0.5145

Фиг. 3. Линейна зависимост на $x$ при дадени стойности на $y$, построена по условни субективно оценени медиани на $x$ при $y_{0.25}$ и $y_{0.75}$, при коефициент на корелация $r_4$=0.5145

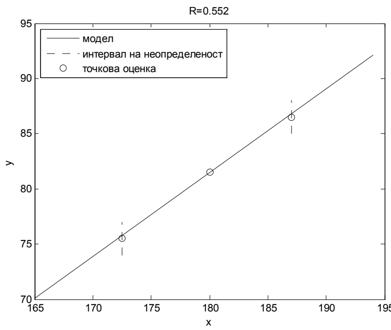


Fig. 4. Linear dependence of $y$ at given values of $x$, constructed on conditional subjectively elicited medians of $y$ at $x_{0.25}$ and $x_{0.75}$, for $r_8$=0.5519

Фиг. 4. Линейна зависимост на $y$ при дадени стойности на $x$, построена по условни субективно оценени медиани на $y$ при $x_{0.25}$ и $x_{0.75}$, при коефициент на корелация $r_8$=0.5519

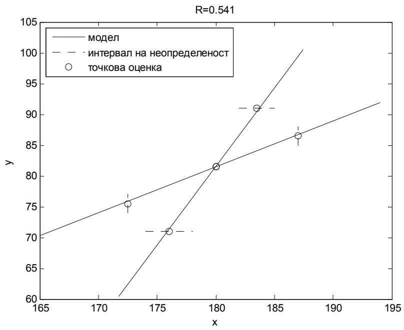


Fig. 5. Linear dependence of $x$ from $y$ and of $y$ from $x$, for $r_9$=0.5411

Фиг. 5. Линейна зависимост на $x$ от $y$ и на $y$ от $x$ при коефициент на корелация $r_9$=0.5411

## IV. Conclusions

The paper presented the task to estimate the correlation coefficients for a bivariate normal random variable. Subjective interval probability estimates were used to find those estimates. Four estimates for the standard deviation, as well as nine estimates of $r$ were proposed. Once those estimates were available, it was possible to find the covariance matrix and construct the density of the variable.

## IV. Заключение

Статията разгледа задача за оценка на корелационни коефициенти при двумерна нормална случайна величина. За целта бяха използвани субективни интервални вероятности. Предложени бяха четири оценки на стандартни отклонения и девет на корелационния коефициент. Така може да се намери ковариационната матрица и да се построи плътността.

The elaborated procedures are realised in original software in Matlab R2010a. The results from the paper may be applied in different analyses in economic and financial processes, technical diagnostics, etc. The formal proof of the mathematical dependencies is the subject of future publications.
All algorithms for subjective elicitation of the elements of the covariance matrix often generate negative estimates, which does not comply with its classical structure. Such a matrix is referred to as fictitious (Tenekedjiev et al., 2000) and cannot be further used, especially in pattern recognition tasks. That is why (Tenekedjiev et al., 2000; Nikolova, et al., 2010a) present algorithms to transform fictitious covariance matrices into classical ones.

Създадените процедури са реализирани в оригинален софтуер в Matlab R2010a. Резултатите в статията имат приложение в анализа на икономически и финансови процеси, техническа диагностика и др. Формалното извеждане на математическите зависимости и модели е обект на бъдещи публикации.

В алгоритмите за субективна оценка на елементите на ковариационната матрица често се получават отрицателни стойности, което не е в съгласие с нейната класическа структура. Такава матрица се нарича фиктивна (Tenekedjiev et al., 2000) и не може да се използва, особено в задачите за статистическо разпознаване на образи. Затова в (Tenekedjiev et al., 2000; Nikolova, et al., 2010a) се представят алгоритми за преобразуване на фиктивни в класически ковариационни матрици.

## Reference/ Литература

[1]. **Clemen, R**. (1996) Making Hard Decisions: an Introduction to Decision Analysis. Second Edition. Duxbury Press, Wadsworth Publishing Company, 1996.

[2]. **Duda, R., Hart, P., Stork, D**. (2001) Pattern Classification. A Wiley Interscience Publication.

[3]. **Fukunaga, K**. (1990) Introduction to the Statistical Pattern Recognition. Second Edition. Academic Press.

[4]. **Hanke, J. E., Reitsch, A. G**. (1991) Understanding Business Statistics. Irwin.

[5]. **Nikolova, N.D, Stoyanov, B., Tenekedjiev, K**. (2010a), Expert Estimate of the correlation Coefficient at Multi-Dimensional Normal Distribution with Conditional Quantiles, Automatics and Informatics'2010, 3-6.10, Sofia, pp. II-403 – II.406 (in Bulgarian).

[6]. **Niolova, N.D, Toneva, D., Stoyanov, B., Tenekedjiev K**. (2010b) Subjective estimation of correlation coefficients, Annual Proceedings of the Ruse University (in Bulgrian).

[7]. **Nikolova, N.D., Dimitrakiev, D., Tenekedjiev, K**. (2004) Fuzzy Rationality in the Elicitation of Subjective Quantiles, Proc. Second International IEEE Conference on Intelligent Systems IS'2004, Volume III, pp. 32-34, June, Varna, Bulgaria.

[8]. **Nikolova, N. D., A. Shulus, D. Toneva, K. Tenekedjiev** (2005) Fuzzy Rationality in Quantitative Decision Analysis. – Journal of Advanced Computational Intelligence and Intelligent Informatics, 9(1), 65-69.

[9]. **Press, W. H., Teukolski S. A., Vetterling, W. T., Flannery, B. P**. (1992) Numerical Recipes – The Art of Scientific Computing, Cambridge University Press.

[10]. **Tenekedjiev, K., Karakatsanis, N., Bekiaris, A**. (2000) Fictitious Covariance Matrices, Proc. Forth International Conference, Adaptive Computing in Design and Manufacture ACDM'2000, pp. 23-26, Plymouth, UK.

[11]. **Vuchkov, I., Stoyanov, S**. (1986) Mathematical Modeling and Optimization of Technological Objects, Second Edition, Sofia, Technika, 1986 (in Bulgarian).

[1] This paper is developed and financially supported by project DVU01/0031 "INPORT" and by project TK01/045: „Intelligent systems for diagnostics and decision making in technological processes" of the National Science Fund.

[1] Трудът е разработен и финансово подпомогнат в проект DVU01/0031 "ИНПОРТ" и в проект TK01/045: „Интелигентни системи за диагностика и вземане на решения в технологични процеси" на Фонд „Научни изследвания".