

Получена: 11 января 2017 / Принята: 2 февраля 2017 / Опубликовано online: 28 февраля 2017

УДК 614.2 + 303.4

## КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ STATISTICA И SPSS

**Андрей М. Гржибовский**<sup>1-4</sup>, <http://orcid.org/0000-0002-5464-0498>,

**Сергей В. Иванов**<sup>5</sup>, <http://orcid.org/0000-0003-0254-3941>

**Мария А. Горбатова**<sup>2</sup>, <http://orcid.org/0000-0002-6363-9595>

<sup>1</sup> Национальный Институт Общественного Здоровья, г. Осло, Норвегия;

<sup>2</sup> Северный Государственный Медицинский Университет, г. Архангельск, Россия;

<sup>3</sup> Международный Казахско-Турецкий Университет им. Х.А. Ясави, г. Туркестан, Казахстан;

<sup>4</sup> Северо-Восточный Федеральный Университет, г. Якутск, Россия;

<sup>5</sup> "Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова", г. Санкт-Петербург, Россия.

### Резюме

В настоящей работе представлены общие сведения о выполнении корреляционного анализа данных с использованием параметрических и непараметрических методов. Приведены алгоритмы расчета коэффициентов корреляции Пирсона, Спирмена и Кендалла с использованием программного обеспечения Statistica 10 и SPSS 20 и интерпретация полученных результатов анализа. Настоящая статья призвана дать общие сведения о корреляционном анализе, и не заменяет прочтения специализированной литературы по статистике и клинической эпидемиологии.

**Ключевые слова:** Statistica, SPSS, корреляционный анализ, коэффициент корреляции Пирсона, коэффициент корреляции Спирмена, коэффициент корреляции Кендалла.

### Abstract

## CORRELATION ANALYSIS OF DATA USING STATISTICA AND SPSS SOFTWARE

**Andrej M. Grjibovski**<sup>1-4</sup>, <http://orcid.org/0000-0002-5464-0498>,

**Sergej V. Ivanov**<sup>5</sup>, <http://orcid.org/0000-0003-0254-3941>

**Maria A. Gorbatova**<sup>2</sup>, <http://orcid.org/0000-0002-6363-9595>

<sup>1</sup> Norwegian Institute of Public Health, Oslo, Norway;

<sup>2</sup> Northern State Medical University, Arkhangelsk, Russia;

<sup>3</sup> International Kazakh-Turkish University, Turkestan, Kazakhstan;

<sup>4</sup> North-Eastern Federal University, Yakutsk, Russia;

<sup>5</sup> Pavlov First Saint Petersburg State Medical University, St. Petersburg, Russia.

In this paper we present the main principles of correlation analysis using parametric and non-parametric methods. Algorithms for calculations of Pearson's, Spearman's and Kendall's coefficients using Statistica 10 and SPSS 20 software are presented. Special emphasis is given to interpretation of results of statistical analysis. The article complements, but does not substitute specialized literature on biostatistics and clinical epidemiology.

**Keywords:** Statistica, SPSS, correlation analysis, Pearson's correlation coefficient, Spearman's correlation coefficient, Kendall's correlation coefficient.

Түйіндеме

## STATISTICA 10 ЖӘНЕ SPSS 20 БАҒДАРЛАМАЛЫҚ ҚАМСЫНДАНДЫРУДЫ ҚОЛДАНУМЕН КОРРЕЛЯЦИЯЛЫҚ МӘЛІМЕТТЕРДІ ТАЛДАУ

**Андрей М. Гржибовский** <sup>1-4</sup>, <http://orcid.org/0000-0002-5464-0498>,

**Сергей В. Иванов** <sup>5</sup>, <http://orcid.org/0000-0003-0254-3941>

**Мария А. Горбатова** <sup>2</sup>, <http://orcid.org/0000-0002-6363-9595>

<sup>1</sup> Қоғамдық Денсаулық сақтау Ұлттық Институты, Осло қ., Норвегия;

<sup>2</sup> Солтүстік Мемлекеттік Медициналық Университеті, Архангельск қ., Ресей;

<sup>3</sup> Х.А. Ясави ат. Халықаралық Қазақ – Түрік Университеті, Туркестан, Қазақстан;

<sup>4</sup> Солтүстік - Шығыс Федералдық Университеті, Якутск қ., Ресей;

<sup>5</sup> Академик И.П. Павлов атынд. бірінші Санкт-Петербург мемлекеттік медициналық университеті, Санкт-Петербург қ., Ресей.

Осы жұмыста параметрлік және параметрлік емес әдістерді қолданумен корреляциялық талдау мәліметтерін орындау туралы жалпы мәліметтер ұсынылды. Statistica 10 және SPSS 20 бағдарламалық қамсындандыруды қолданумен Пирсонның, Спирменның және Кендаллдың корреляциялары коэффициенттерін есептеу алгоритмдері және алынған талдау нәтижелері интерпретациясы келтірілген. Осы мақала корреляциялық мәліметтерді талдау туралы жалпы мәліметтер беруге талап етілген және статистика және клиникалық эпидемиология бойынша мамандандырылған әдебиетті оқудың орнын толтырмайды.

**Негізгі сөздер:** Statistica, SPSS, корреляциялық талдау, Пирсонның корреляциясы коэффициенті, Спирменның корреляциясы коэффициенті, Кендаллдың корреляциясы коэффициенті.

### Библиографическая ссылка:

*Гржибовский А.М., Иванов С.В., Горбатова М.А.* Корреляционный анализ данных с использованием программного обеспечения Statistica и SPSS // Наука и Здравоохранение. 2017. №1. С. 7-36.

*Grijbovski A.M., Ivanov S.V., Gorbatova M.A.* Correlation analysis of data using Statistica and SPSS software. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2017, 1, pp. 7-36.

*Гржибовский А.М., Иванов С.В., Горбатова М.А.* Statistica және SPSS бағдарламалық қамсындандыруды қолданумен корреляциялық мәліметтерді талдау // Ғылым және Денсаулық сақтау. 2017. №1. Б. 7-36.

Настоящая статья продолжает серию публикаций, посвященных статистическому анализу данных биомедицинских исследований [9, 13, 14, 15, 16, 17].

Статистическая обработка данных является завершающим этапом исследования, но никакой статистический анализ не может «исправить» некорректно организованное исследование и неправильно собранные данные. Именно поэтому авторы настоящей статьи настоятельно рекомендуют читателю ознакомиться с литературой по эпидемиологии [32, 23, 35], а также с практическими аспектами организации и анализа результатов

различных типов научных исследований в здравоохранении (одномоментных, когортных, экологических, экспериментальных исследований и «случай-контроль»), которые представлены в статьях, опубликованных в 2015 году в журнале «Наука и Здравоохранение» [10, 11, 12, 18, 19].

Высокое качество статистического анализа является обязательным условием востребованности результатов исследований международным научным сообществом [27, 1], поэтому задачами настоящей серии статей является формирование у начинающего исследователя базисных представлений о

статистической обработке данных, приобретение практического опыта работы с современными статистическими пакетами программ и предупреждение типичных ошибок, возникающих в процессе анализа результатов исследования.

Данная статья посвящена корреляционному анализу данных с использованием программного обеспечения Statistica 10 и SPSS 20.

В процессе изучения различных явлений часто возникает необходимость оценить тесноту (силу) связи между ними. Цель корреляционного анализа – количественная оценка силы и направления взаимосвязи между явлениями.

Термин «корреляция» был впервые введен Ж.Кювье в 1806 году, и только в 1886 г. Ф. Гальтоном он впервые был применен к результатам биомедицинских исследований [24].

По направлению корреляционная связь может быть прямой (положительной), когда увеличение или уменьшение значения одного признака приводит, соответственно, к увеличению или уменьшению значения другого признака, или обратной (отрицательной), когда увеличение значения одного признака приводит к уменьшению значения другого и наоборот.

Например, между степенью бактериального загрязнения воды и количеством кишечных инфекций у проживающего на загрязненной территории может быть обнаружена прямая корреляционная связь, а между уровнем финансирования здравоохранения и смертностью населения от различных заболеваний – обратная корреляционная связь

По характеру связь может быть не только корреляционной, но и функциональной, когда каждому значению одного признака соответствует точное значение другого (например, функциональная связь между массой тела и индексом массы тела).

Исследователь должен ясно понимать, что в результате корреляционного анализа невозможно установить причинно-следственные связи между явлениями, поэтому выводы о влиянии одного явления на другое на основании одного лишь корреляционного анализа данных будут неправомерными. Не допускается подмена понятия причинно-следственной связи корреляционной связью. Юмористическим примером такого заблуждения служит обнаружение сильной положительной корреляционной взаимосвязи между количеством гнезд аистов и количеством новорожденных в Копенгагене в послевоенные годы, как доказательство того, что детей приносят аисты [35].

Для количественной оценки корреляционной связи используются различные коэффициенты корреляции, имеющие разные способы расчета, однако наиболее популярными являются следующие три [7, 32, 5]:

1. Коэффициент корреляции Пирсона (Pearson)  $r$  – параметрический.
2. Коэффициент корреляции Спирмена (Spearman)  $r_s$  – непараметрический.
3. Коэффициенты корреляции Кендалла (Kendall)  $\tau_a$  и  $\tau_b$  – непараметрические.

Все коэффициенты корреляции могут принимать значение от 0 до 1 или от -1 до 0.

Оценка силы корреляционной связи проводится в соответствии с таблицей 1.

Таблица 1.

**Количественные критерии оценки силы и направления корреляционной связи.**

Сила связи	Значения коэффициента корреляции	
	Прямая (+)	Обратная (-)
Отсутствует	0,0	0,0
Слабая	от 0,01 до 0,29	от -0,01 до -0,29
Средняя	от 0,30 до 0,69	от -0,30 до -0,69
Сильная	от 0,70 до 0,99	от -0,70 до -0,99
Полная (функциональная)	1,0	-1,0

Использование коэффициент корреляции Пирсона встречается в биомедицинских исследованиях наиболее широко, и не всегда

обоснованно, поэтому следует указать условия его применения [7, 32]:

1. Обе переменные должны быть количественными и непрерывными.

2. Как минимум один из изучаемых признаков, а лучше оба, должны иметь распределение, близкое к нормальному.

3. Зависимость между переменными должна носить линейный характер.

4. Вариабельность одной переменной не должна зависеть от значения другой переменной, то есть разброс значений одной из переменных должен быть примерно одинаковым для всех значений другой переменной (так называемая «гомоскедастичность»).

5. Наблюдения должны быть независимы друг от друга.

6. Наблюдения должны быть парными (значение обоих анализируемых параметров

регистрируются одновременно у одного и того же объекта исследования).

7. Объем выборки должен составлять не менее 25 наблюдений [39].

Графически зависимость между переменными можно представить в виде скаттерограммы. Из представленных на рисунке 1 скаттерограмм коэффициент корреляции Пирсона можно рассчитать только для первого случая, когда наблюдается линейная зависимость между переменными и скаттерограмма гомоскедастична (для остальных двух случаев коэффициент корреляции Пирсона не может быть рассчитан вследствие несоблюдения условия гомоскедастичности и наличия нелинейной зависимости между переменными).

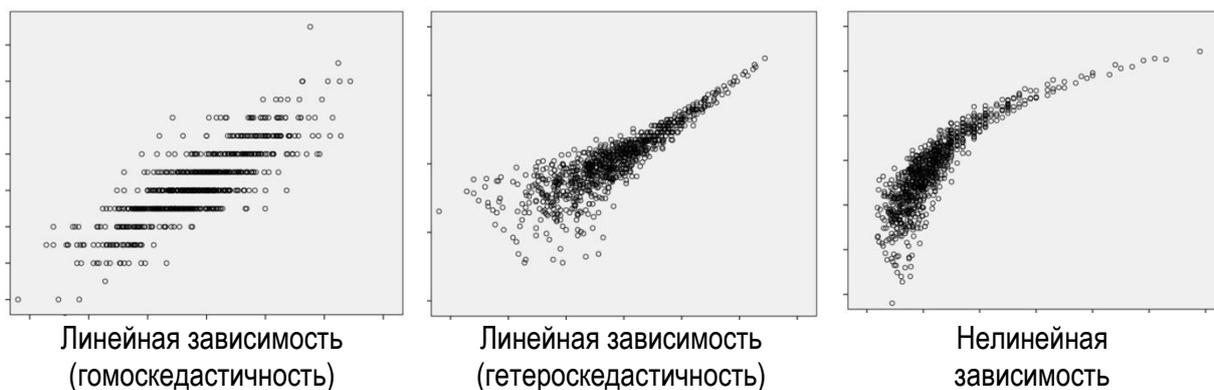


Рис. 1. Примеры скаттерограмм.

Коэффициент корреляции Пирсона для двух переменных (X и Y) рассчитывается следующим образом:

1. Значения переменных X и Y располагают в ряд, в котором каждой величине X соответствует определенная величина Y.

2. Рассчитывают средние арифметические значения для каждой переменной  $X_{cp}$  и  $Y_{cp}$  соответственно.

3. Рассчитывают отклонения каждого значения X и Y от соответствующей средней величины.

4. Отклонения для X и Y перемножают между собой.

5. Рассчитывают стандартные отклонения для X и Y ( $S_x$  и  $S_y$ ) по формулам:

$$S_x = \sqrt{\frac{(X_1 - X_{cp})^2 + (X_2 - X_{cp})^2 + \dots + (X_i - X_{cp})^2}{n - 1}}$$

$$S_y = \sqrt{\frac{(Y_1 - Y_{cp})^2 + (Y_2 - Y_{cp})^2 + \dots + (Y_i - Y_{cp})^2}{n - 1}}$$

6. Рассчитывают коэффициент корреляции Пирсона по формуле (n – количество наблюдений):

$$r = \frac{(X_1 - X_{cp}) \times (Y_1 - Y_{cp}) + (X_2 - X_{cp}) \times (Y_2 - Y_{cp}) + \dots + (X_i - X_{cp}) \times (Y_i - Y_{cp})}{(n - 1) \times S_x \times S_y}$$

7. Сравнивают полученное значение коэффициента Пирсона с критическим значением, взятым из специальных таблиц [30, 24, 5]. В случае, если расчетное значение  $r$  равно или превышает критическое значение для уровня статистической значимости, равного 0,05, то нулевую статистическую гипотезу отвергают и делают вывод о том, что коэффициент корреляции Пирсона статистически значимо отличается от нуля ( $p < 0,05$ ).

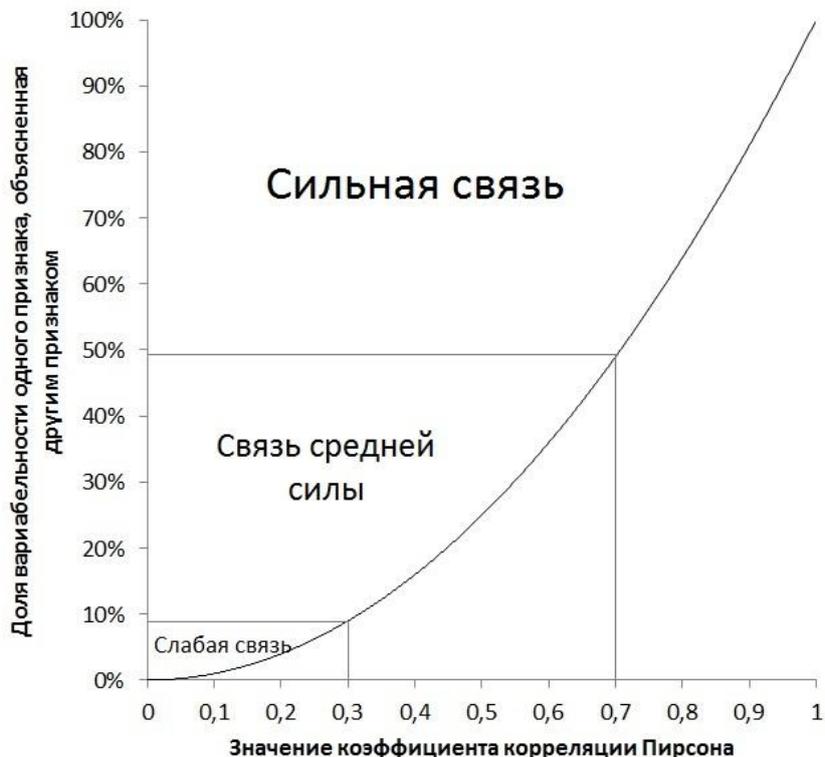
Интерпретация коэффициента корреляции Пирсона включает следующие этапы:

1. Оценка статистической значимости коэффициента корреляции. Если  $p < 0,05$ , то коэффициент корреляции статистически значимо отличается от нуля. Если  $p \geq 0,05$ , то делается вывод об отсутствии корреляционной связи между анализируемыми признаками.

2. Оценка силы и направления корреляционной связи (таблица 1).

3. Оценка степени влияния одного признака на другой. Для понимания степени

«тесноты» связи между признаками используется коэффициент детерминации, который рассчитывается как коэффициент корреляции, возведенный в квадрат ( $r^2$ ). Коэффициент детерминации показывает, какую долю варибельности одного из признаков способно объяснить изменение другого признака. Зависимость значения детерминации коэффициента от коэффициента корреляции представлена на рисунке 2. Из представленного графика видно, что слабая корреляционная связь может объяснить не более 8,4% варибельности признака, обусловленной влиянием другого признака, а связь средней силы – не более 47,6% варибельности. Таким образом, коэффициент детерминации необходимо использовать для адекватного понимания фактической «тесноты» взаимосвязи между признаками, что особенно важно на этапе практических выводов по итогам анализа данных (например, при оценке клинической роли выявленной корреляционной связи).



**Рис. 2. Зависимость значения коэффициента детерминации от значения коэффициента корреляции.**

Существует важное обстоятельство, интерпретации результатов корреляционного анализа: выявленные взаимосвязи между

двумя признаками могут быть обусловлены влиянием некоего третьего фактора. Например, известно, что вероятность рождения ребенка с синдромом Дауна тесно коррелирует с количеством родов у матери до момента настоящей беременности. Но это вовсе не значит, что количество предшествующих родов влияет на риск рождения больного ребенка, так как в данном случае имеется третий фактор, который связан с обеими переменными – возраст женщины. Поэтому между вероятностью рождения ребенка с синдромом Дауна имеется корреляционная, но ни в коем случае не причинно-следственная связь [7].

Следует отметить существенный недостаток коэффициента корреляции Пирсона – высокую чувствительность к наличию «выскакивающих» величин (выбросов). Даже небольшое количество

выбросов может значительно исказить значение коэффициента и буквально «уничтожить» зависимость между величинами, поэтому всегда следует анализировать выбросы и выяснять, не является ли они следствием ошибки регистрации данных.

Приведем для примера гипотетическое исследование, целью которого является оценка корреляционной связи между уровнем IQ респондента и временем, затраченным на решение типовой логической задачи (всего были протестированы 10 респондентов). Следует отметить, что для корреляционного анализа необходимо наличие не менее 25 наблюдений, но в данном примере приведены только 10 наблюдений для уменьшения трудоемкости расчетов вручную.

В таблице 2 представлены результаты наблюдений и расчет промежуточных значений, необходимых для вычисления значения  $r$ .

Таблица 2.

#### Продолжительность решения логической задачи респондентами с различным значением IQ и результаты промежуточных расчетов

№	X (уровень IQ, ед.)	Y (время на решение задачи, сек.)	Для расчета $r$			Для расчета $S_x$ и $S_y$	
			$X_i - X_{cp}$	$Y_i - Y_{cp}$	$(X_i - X_{cp}) \times (Y_i - Y_{cp})$	$(X_i - X_{cp})^2$	$(Y_i - Y_{cp})^2$
1	140	235	16	-43,4	-694,4	256,0	1883,6
2	112	158	-12	33,6	-403,2	144,0	1129,0
3	124	185	0	6,6	0	0,0	43,6
4	130	219	6	-27,4	-164,4	36,0	750,8
5	128	215	4	-23,4	-93,6	16,0	547,6
6	121	176	-3	15,6	-46,8	9,0	243,4
7	115	167	-9	24,6	-221,4	81,0	605,2
8	127	209	3	-17,4	-52,2	9,0	302,8
9	117	165	-7	26,6	-186,2	49,0	707,6
10	126	187	2	4,6	9,2	4,0	21,2
			Сумма		-1853	604	6234,4

Согласно расчетам,  $X_{cp} = 124,0$  ед.,  $Y_{cp} = 191,6$  сек.,  $S_x = 8,2$  ед.,  $S_y = 26,3$  сек.

Соответственно, рассчитываем значение  $r$ :

$$r = \frac{-1853}{(10 - 1) \times 8,2 \times 26,3} = -0,95$$

Из таблицы критических значений критерия корреляции Пирсона для  $n = 10$  и уровня статистической значимости 0,01 критическое значение  $r$  составляет 0,765 [30, 5, 24]. Так как расчетное значение больше критического, выявленная взаимосвязь между уровнем IQ и

длительностью решения типовой логической задачи является статистически значимой ( $p < 0,01$ ).

В данном примере  $r = -0,95$ , что свидетельствует об обратной сильной зависимости: чем выше уровень IQ, тем меньше время, затрачиваемое на решение типовой логической задачи.

Значение коэффициента детерминации ( $r^2 = 0,952 = 0,90$ ) говорит о том, что 90%

вариабельности длительности решения типовой логической задачи обусловлены уровнем IQ (интеллектуального коэффициента) респондента (и 10% вариабельности обусловлены иными факторами).

Также следует рассчитать доверительный интервал (англ. CI – «confidence interval») для коэффициента корреляции Пирсона, так как

$$Z_L = 0,5 \cdot \ln\left(\frac{1+r}{1-r}\right) - \frac{1,96}{\sqrt{n-3}} \quad \text{и} \quad Z_U = 0,5 \cdot \ln\left(\frac{1+r}{1-r}\right) + \frac{1,96}{\sqrt{n-3}},$$

где  $\ln$  обозначает натуральный логарифм, а  $n$  – объем выборки. Само же значение коэффициента корреляции для генеральной совокупности, рассчитанное по данным выборки, будет в 95% случаев находиться в интервале

$$\text{от } r_L = \frac{\exp(2Z_L) - 1}{\exp(2Z_L) + 1} \quad \text{до } r_U = \frac{\exp(2Z_U) - 1}{\exp(2Z_U) + 1}$$

где  $\exp(2z)$  рассчитывается как  $e^{2z}$  (где  $e$  – число Эйлера, или основание натурального логарифма, оно примерно равно 2,72).

В нашем примере коэффициент корреляции Пирсона для взаимосвязи между IQ респондента и продолжительностью решения задачи был равен -0,95 и статистически значимо отличался от нуля ( $p < 0,01$ ). Рассчитаем сначала

$$Z_L \text{ и } Z_U: Z_L = -2,57, Z_U = -1,09.$$

Далее:

$$\exp(2Z_L) = 2,72^{2 \times (-2,57)} = 0,0058$$

$$\exp(2Z_U) = 2,72^{2 \times (-1,09)} = 0,1128$$

Соответственно,

$$r_L = -0,99, r_U = -0,80.$$

Таким образом, мы можем записать результаты корреляционного анализа, учитывающие доверительный интервал:

$$r = -0,95 \text{ (95\% CI } -0,99; -0,80), n = 10, p < 0,01.$$

Доверительный интервал для коэффициента корреляции Пирсона можно рассчитать и с помощью онлайн калькулятора расположенного по адресу: <http://faculty.vassar.edu/lowry/rho.html>. Данный калькулятор использует те же формулы, которые были приведены выше, но требует использовать при внесении коэффициента корреляции в поле «г» не запятую, а точку (например, «-0.95» вместо «-0,95»).

В случае, если условия применения коэффициента корреляции Пирсона не выполняются, для корреляционного анализа следует использовать непараметрические

интервальная оценка любого генерального параметра всегда более информативна, чем точечная. Доверительные интервалы для коэффициента корреляции Пирсона можно рассчитать, с использованием z-преобразование Фишера. Нижняя ( $Z_L$ ) и верхняя ( $Z_U$ ) границы преобразованного 95% доверительного интервала для коэффициента корреляции Пирсона рассчитываются по формулам:

коэффициенты корреляции Спирмена или Кендалла, расчет которых основан на использовании не исходных значений признаков, а их рангов [7, 2, 22]. Если в подобной ситуации применять коэффициент корреляции Пирсона, полученные результаты будут искаженными, а выводы – сомнительными.

Преимуществами непараметрических коэффициентов корреляции по сравнению с коэффициентом корреляции Пирсона является возможность оценки связи между порядковыми и количественными признаками (коэффициент корреляции Спирмена) или только между порядковыми признаками (коэффициент корреляции Кендалла).

Коэффициент корреляции Спирмена для переменных X и Y рассчитывается следующим образом:

1. Два ряда из парных сопоставляемых признаков составляют рядом, первый обозначается как X, второй – как Y.

2. Первый ряд признака ранжируется в убывающем или возрастающем порядке, а числовые значения второго ряда размещаются напротив того значения первого ряда, которым они соответствуют.

3. Значения первой и второй переменных заменяют порядковым номером (рангом) числовым значениям второго признака ранги

должны присваиваться в том же порядке, какой был принят при раздаче их величинам первого признака. При одинаковых величинах признака в ряду ранги следует определять как среднее число из суммы порядковых номеров этих величин.

4. Определяют разности рангов ( $d_i$ ) между ранговыми номерами  $X_i$  и  $Y_i$ .

5. Проводят расчет коэффициента корреляции Спирмена по формуле (где  $n$  – число сравниваемых пар):

$$r_s = 1 - \frac{6 \times (d_1^2 + d_2^2 + d_3^2 + \dots + d_i^2)}{n^3 - n}$$

6. Для оценки статистической значимости выявленной взаимосвязи между переменными расчетное значение коэффициента Спирмена сравнивают с критическим значением, взятым из таблицы [30, 5, 24]. Если расчетное значение  $r_s$  равно или превышает критическое значение, равное 0,05, то нулевая гипотеза отвергается и делается вывод о том, что коэффициент корреляции статистически значимо отличается от нуля ( $p < 0,05$ ).

7. Для коэффициента корреляции Спирмена также можно рассчитать

коэффициент детерминации ( $r_s^2$ ), но он будет означать долю вариабельности рангов одной переменной, которую можно объяснить с помощью рангов другой переменной. Данная интерпретация достаточно громоздка и не совсем понятна, поэтому целесообразность использования коэффициента детерминации  $r_s^2$  в приложении к практической интерпретации результатов корреляционного анализа сомнительна.

8. Приведем для примера гипотетическое исследование, целью которого является оценка корреляционной связи между функциональным классом (ФК) сердечной недостаточности (СН) пациента и количеством приступов стенокардии, возникающих у него в течение месяца. Всего были обследованы 11 пациентов, имеющих I-IV ФК СН.

9. ФК СН является ранговой переменной, а количество приступов стенокардии – дискретной количественной переменной.

10. В таблице 2 представлены результаты наблюдений и расчет промежуточных значений, необходимых для вычисления значения  $r_s$ .

Таблица 3.

### Количество приступов стенокардии у пациентов с различными ФК СН.

№ пациента	ФК СН (переменная X)	Количество приступов стенокардии в месяц (переменная Y)	Ранг X	Ранг Y	Разность рангов (d)	$d^2$
1	3	2,5	2,5	0	0	1
1	3	2,5	2,5	0	0	1
1	4	2,5	8,5	6	36	1
1	4	2,5	8,5	6	36	1
2	4	6	8,5	2,5	6,25	2
2	3	6	2,5	-3,5	12,25	2
2	3	6	2,5	-3,5	12,25	2
3	4	9	8,5	-0,5	0,25	3
3	6	9	14	5	25	3
3	4	9	8,5	-0,5	0,25	3
4	4	12,5	8,5	-4	16	4
4	4	12,5	8,5	-4	16	4
4	5	12,5	13	0,5	0,25	4
Сумма:						176,5

Согласно формуле расчета коэффициента корреляции Спирмена

$$r_s = 1 - \frac{6 \times 176,5}{14^3 - 14} = 0,61$$

Из таблицы критических значений критерия корреляции Спирмена для  $n = 14$  и уровня статистической значимости 0,05 критическое значение  $r_s$  составляет 0,532 [29, 5, 24]. Так как расчетное значение больше критического,

выявленная взаимосвязь между ФК СН и количеством приступов стенокардии в течение месяца является статистически значимой ( $\rho < 0,05$ ).

Рассчитать доверительные интервалы для  $r_s$  можно с использованием уже известного

$$Z_L = 0,5 \cdot \ln\left(\frac{1+\rho}{1-\rho}\right) - \frac{1,96 \cdot \sqrt{1+0,5 \cdot \rho^2}}{\sqrt{n-3}}; \quad Z_U = 0,5 \cdot \ln\left(\frac{1+\rho}{1-\rho}\right) + \frac{1,96 \cdot \sqrt{1+0,5 \cdot \rho^2}}{\sqrt{n-3}},$$

где  $\rho$  – рассчитанное значение коэффициента корреляции Спирмена.

Далее полученные значения  $Z_L$  и  $Z_U$  следует подставить в уже упоминавшуюся

$$\text{от } r_L = \frac{\exp(2Z_L) - 1}{\exp(2Z_L) + 1}$$

Приведенная формула, по мнению D. Bonnett и T. Wright [36], является наиболее адекватной для расчета доверительного интервала для коэффициента корреляции Спирмена.

Рассчитаем значения для нижней и верхней границ 95% доверительного интервала коэффициента корреляции Спирмена, равного 0,61:

$$Z_L = 0,069, \quad Z_U = 1,349.$$

Далее:

$$\text{Exp}(2Z_L) = 2,72^{2 \times (0,069)} = 1,15$$

$$\text{Exp}(2Z_U) = 2,72^{2 \times (1,349)} = 14,86$$

$$\text{Соответственно, } r_{SL} = 0,07, \quad r_{SU} = 0,87.$$

Таким образом, мы можем записать результаты корреляционного анализа, учитывающие доверительный интервал:

$$r_s = 0,61 \text{ (95\% CI } 0,07; 0,87), \quad n = 14, \quad p < 0,05.$$

Следует обратить внимание на значительную ширину 95% доверительного интервала – от 0,07 (практически полное отсутствие связи) до 0,87 (сильная связь). Данный факт связан с небольшим количеством наблюдений в выборке. Например, если бы данное значение коэффициента корреляции Спирмена было бы получено на основании 28 наблюдений, то 95% доверительный интервал располагался бы в пределах от 0,28 до 0,81, а если бы расчеты проводились на основании 140 наблюдений, то границы сузились бы до значений от 0,48 до 0,71. В нашем примере большая широта доверительного интервала не позволяет делать каких-либо клинически значимых выводов о наличии корреляционной

преобразования [37, 7]. Формулы для расчета  $Z_L$  и  $Z_U$  для 95% доверительного интервала для коэффициента корреляции Спирмена будут следующими:

ранее формулу для расчета верхней и нижней границ 95% доверительного интервала:

$$\text{до } r_U = \frac{\exp(2Z_U) - 1}{\exp(2Z_U) + 1}$$

связи между ФК СН и количеством приступов стенокардии, и требует увеличения количества наблюдений в исследовании для того, чтобы выводы приобрели достаточную степень определенности.

Таким образом, приведенный пример наглядно продемонстрировал, как важно использовать интервальную, а не точечную оценку коэффициентов корреляции, так как точечная оценка «скрадывает» информацию, которая может оказаться крайне важной с клинической точки зрения.

Третьим рассматриваемым в настоящей статье коэффициентом корреляции является непараметрический коэффициент корреляции Кендалла. Существуют 3 его разновидности –  $\tau_a$ ,  $\tau_b$  и  $\tau_c$

Рассмотрим наиболее простой вариант коэффициента корреляции Кендалла –  $\tau_a$ . Допустим, речь идет о двух участниках исследования  $i$  и  $j$ , у которых в ходе исследования изучаются признаки  $X$  и  $Y$ . Изучаемыми признаками могут, например, быть рост и масса тела, индекс массы тела и артериальное давление, и др. Пару наблюдений можно обозначить как  $X_i, Y_i$  и  $X_j, Y_j$ .

Если разности  $X_j - X_i$  и  $Y_j - Y_i$  будут одинаковы по знаку (либо  $X_j > X_i$  и  $Y_j > Y_i$ , либо  $X_j < X_i$  и  $Y_j < Y_i$ ), то пару называют конкордантной (например, и рост, и вес участника  $i$  больше, чем рост и вес участника  $j$ ). Количество конкордантных пар (или проверсий) обозначается как  $S$ .

Если разности  $X_j - X_i$  и  $Y_j - Y_i$  различаются по знаку (либо  $X_j > X_i$  и  $Y_j < Y_i$ , либо  $X_j < X_i$  и  $Y_j > Y_i$ ), то пару называют дискордантной.

$> Y_i$ ), то такая пара называется дискордантной (например, рост участника  $i$  больше роста участника  $j$ , а вес участника  $i$  меньше роста участника  $j$ ). Количество дискордантных пар (или инверсий) обозначается как  $D$ . Если выборка состоит из  $n$  участников исследования, то возможно формирование  $n \times (n - 1) / 2$  пар, для которых  $1 \leq i < j \leq n$ .

Коэффициент корреляции Кендалла  $\tau_a$  рассчитывается по формуле [7, 40, 41]:

$$\tau = \frac{2 \cdot (C - D)}{n \cdot (n - 1)}$$

Недостатком  $\tau_a$  является то, что он не учитывает одинаковых (связанных, равных) рангов (англ. «ties»), которые возникают в тех случаях, когда у нескольких участников исследования изучаемый признак имеет одно и то же значение (например, одинаковый рост, или одинаковая стадия заболевания). Из формулы видно, что максимально возможное значение  $\tau_a = 1$  достигается только в том случае, если все пары являются конкордантными. Соответственно, если все пары являются дискордантными,  $\tau_a$  принимает минимально возможное значение  $-1$ . Если количество конкордантных и дискордантных пар равно, то  $\tau_a = 0$ , что говорит об отсутствии взаимосвязи между изучаемыми признаками.

Если  $C$  представляет собой количество конкордантных пар из возможных в выборочной совокупности  $n \times (n - 1) / 2$  пар, то оценить вероятность того, что пара наблюдений будет конкордантной ( $\pi_c$ ), можно с помощью формулы:

$$\pi_c = \frac{2 \cdot C}{n \cdot (n - 1)}$$

Аналогично, вероятность того, что пара наблюдений будет дискордантной ( $\pi_d$ ), можно оценить с помощью формулы:

$$\pi_d = \frac{2 \cdot D}{n \cdot (n - 1)}$$

Таким образом, для любой пары наблюдений, отобранных случайно, коэффициент корреляции Кендалла  $\tau_a$  может интерпретироваться как разность между вероятностью того, что пара окажется конкордантной, и того, что она окажется дискордантной, то есть

$$\tau_a = \pi_c - \pi_d$$

Отрицательное значение  $\tau_a$  будет говорить о том, что вероятность того, что любая случайно отобранная пара наблюдений с характеристиками  $(X_i, Y_i$  и  $X_j, Y_j)$  будет скорее дискордантной, чем конкордантной, и наоборот.

Кроме того, в генеральной совокупности, для которой коэффициент корреляции Кендалла равен  $\tau_a$ , вероятность того, что любая случайно отобранная пара наблюдений с характеристиками  $(X_i, Y_i$  и  $X_j, Y_j)$  окажется конкордантной, будет в  $(1 + \tau_a) / (1 - \tau_a)$  раза выше, чем вероятность того, что эта пара будет дискордантной. Таким образом, если в исследовании с использованием случайно отобранной репрезентативной выборки был получен коэффициент корреляции Кендалла  $\tau_a = 0,5$ , это означает, что вероятность того, что любая случайно отобранная из генеральной совокупности пара окажется конкордантной, в среднем в  $(1 + 0,5) / (1 - 0,5) = 3$  раза выше, чем вероятность того, что эта пара будет дискордантной.

Коэффициент корреляции Кендалла  $\tau_a$  отличается от  $\tau_b$  тем, что учитывает связанные ранги. Появление связанных рангов, то есть, когда два или более наблюдений по любой из переменных имеют одинаковые ранги, неизбежно при изучении порядковых признаков, таких как, например, стадия заболевания, степень тяжести, уровень образования и других, имеющее ограниченное количество значений (например, использование 3-х степеней тяжести заболевания).

В целом, использование коэффициентов корреляции Кендалла предпочтительно при анализе корреляционной связи между порядковыми (ординальными) переменными. Например, коэффициент корреляции  $\tau_b$  рекомендован для анализа связи между порядковыми признаками, которые проще всего представить в виде многопольных таблиц, у которых число рядов равно числу столбцов.

Коэффициент корреляции Кендалла  $\tau_c$  используется при расчете связи между порядковыми переменными, формирующими таблицу, в которой количество рядов и количество столбцов не равны (например, оценка корреляционной связи между порядковой переменной  $X$ , имеющей 5 наименований, и другой порядковой переменной  $Y$ , имеющей 3 наименования).

В данной статье подробный алгоритм расчета коэффициентов корреляции Кендалла на примерах не приводится по причине трудоемкости их ручного вычисления.

Более подробно с различными коэффициентами корреляции Кендалла можно познакомиться в [7, 6].

В статистическом программном обеспечении в рамках корреляционного анализа как правило производится расчет коэффициента корреляции Кендалла  $\tau_b$ .

$$Z_L = 0,5 \cdot \ln\left(\frac{1+\tau}{1-\tau}\right) - \frac{1,96 \cdot \sqrt{0,437}}{\sqrt{n-4}};$$

Значения  $Z_L$  и  $Z_U$  которые затем подставляют в уже известную формулу для расчета 95% доверительного интервала:

$$\text{от } r_L = \frac{\exp(2Z_L) - 1}{\exp(2Z_L) + 1}$$

В завершении теоретической части настоящей статьи следует отметить, что корреляционный анализ встречается в русскоязычной научной периодике очень часто, поэтому целесообразно остановиться на основных ошибках его использования:

- Применение параметрического коэффициента корреляции Пирсона при несоблюдении необходимых условий его использования.

- Подмена понятия корреляционной связи понятием причинно-следственной связи.

- Беспорядочный расчет коэффициентов корреляции для всех пар переменных по принципу «сравнить все со всем».

- Смешивание понятий корреляционного и регрессионного анализа.

- Неполное представление результатов корреляционного анализа (значения коэффициента корреляции, объема выборки, значения уровня статистической значимости).

- Представление только точечной оценки (игнорирование доверительных интервалов).

- Отождествление статистически значимых коэффициентов корреляции с клинически важными.

- Отсутствие обсуждения, почему были получены те или иные коэффициенты корреляции (выяснение вопросов истинной или ложной является выявленная зависимость, возможность присутствия неких

Большой практический интерес представляет расчет доверительного интервала для  $\tau_b$ , который также можно рассчитать с помощью преобразования Фишера, которое дает достаточно адекватную интервальную оценку коэффициента корреляции Кендалла для генеральной совокупности при объеме выборки не менее 10 наблюдений и значении  $\tau$  не более 0,8. Отличие будет заключаться в расчете вспомогательных значений  $Z_L$  и  $Z_U$ :

$$Z_U = 0,5 \cdot \ln\left(\frac{1+\tau}{1-\tau}\right) + \frac{1,96 \cdot \sqrt{0,437}}{\sqrt{n-4}},$$

переменных, тесно коррелирующих с обеими изучаемыми переменными).

- Однозначное заключение о полном отсутствии взаимосвязи между признаками при значении коэффициента корреляции близком к нулю (возможно, что взаимосвязь между переменными носит нелинейный характер, но исследователь этого не учел).

- Редкое применение скаттерграмм для графического представления зависимостей.

Для того, чтобы читатель приобрел практические навыки проведения корреляционного анализа, будет рассмотрен фрагмент данных, которые были собраны в ходе исследования, направленного на изучение метаболического синдрома и его детерминант в условиях неблагополучной социально-экологической ситуации в Южном Казахстане [20, 21, 25, 29].

В ходе данного исследования у 277 пациентов получены значения индекса массы тела (ИМТ), окружности талии, уровне креатинина и мочевины в крови (все четыре анализируемых признака являются непрерывными количественными переменными). Также были собраны данные об уровне образования пациентов (4 градации – высшее, незаконченное высшее, среднее и начальное).

На предварительном этапе обработки данных количественная шкала значений ИМТ была переведена в номинальную: были выделены 3 «рамки» значений ИМТ:

нормальная масса тела, избыточная масса тела и ожирение. Подобная группировка значений позволяет проводить сравнения между различными категориями пациентов и часто используется при анализе данных. Отметим, что получившаяся номинальная переменная, имеющая значения «нормальная масса тела», «избыточная масса тела» и «ожирение», является как номинальной, так и ординальной, поскольку все три значения могут быть ранжированы по возрастанию или убыванию.

Корреляционный анализ будет проведен с использованием программного обеспечения Statistica 10 [28, 3] и SPSS 20 [4], демонстрационные версии которого можно загрузить с официальных сайтов разработчиков ([www.stastsoft.com](http://www.stastsoft.com) и [www.ibm.com](http://www.ibm.com) соответственно).

Представленные ниже алгоритмы действий являются не более чем инструментом анализа данных, в то время как корректная интерпретация полученных результатов требует наличия базисных знаний в области биомедицинской статистики, которые могут быть получены только путем изучения специализированной литературы [5, 26, 24, 34, 38].

### Корреляционный анализ с использованием программы Statistica 10.

Для начала работы необходимо открыть файл 7\_Correlation\_STAT.sta, который потребуется загрузить с сайта журнала «Наука и Здоровье». В данном файле представлены следующие вариационные ряды:

1. ИМТ (переменная «BMI»): непрерывная количественная переменная.
2. Окружность талии (переменная «Waist\_circum»): непрерывная количественная переменная.
3. Категория ИМТ (переменная «Category\_BMI»): номинальная (или ординальная) переменная.
4. Уровень образования (переменная «Education»): номинальная (или ординальная) переменная.
5. Уровень креатинина крови (переменная «Creatinin»): непрерывная количественная переменная.

6. Уровень мочевины крови (переменная «Carbamide»): непрерывная количественная переменная.

В результате статистического анализа данных будет проведена оценка корреляционной связи между следующими переменными:

- ИМТ и окружностью талии.
- Уровнем креатинина и уровнем мочевины крови.
- Уровнем образования пациента и категорией ИМТ.

На начальном этапе обработки данных требуется построить скаттерограмму, чтобы визуально оценить степень связи между переменными.

Для этого войдем в меню «Graphs» в верхней части экрана и выберем раздел «Scatterplots...». В появившемся окне «2D Scatterplots» (рисунок 3) нажмем на кнопку «Variables» и выберем переменные, значения которых будут отложены по осям абсцисс и ординат, как это показано на рисунке 4. Выбор подтвердим нажатием на кнопку «OK» и в снова открывшемся окне «2D Scatterplots» снова нажимаем на кнопку «OK».

В результате наших действий программа сформирует скаттерограмму зависимости между переменными «BMI» и «Waist\_circum» (рисунок 5). Видно, что скаттерограмма отражает линейную зависимость и в достаточной мере гомоскедастична: разброс значений одной переменной практически не зависит от разброса другой переменной, и разброс точек вокруг линии тренда примерно одинаков.

На следующем этапе обработки данных требуется определить тип распределения, чтобы понять, можно ли использовать параметрический метод (коэффициент корреляции Пирсона), или потребуется сравнивать группы с помощью непараметрического критерия Спирмена. Пошаговый алгоритм проверки распределения переменных на «нормальность» нескольких групп подробно описан в [13, 8, 31].

Проверка на «нормальность» распределения изучаемых количественных переменных показала, что обе переменные имеют близкое к нормальному распределение (читатель может

самостоятельно убедиться в этом, выполнив проверку распределения переменных с использованием разделов меню «Statistics»

– «Basic Statistics/Tables» – «Descriptive statistics» – «Normality» – кнопка «Histograms» программы Statistica 10).

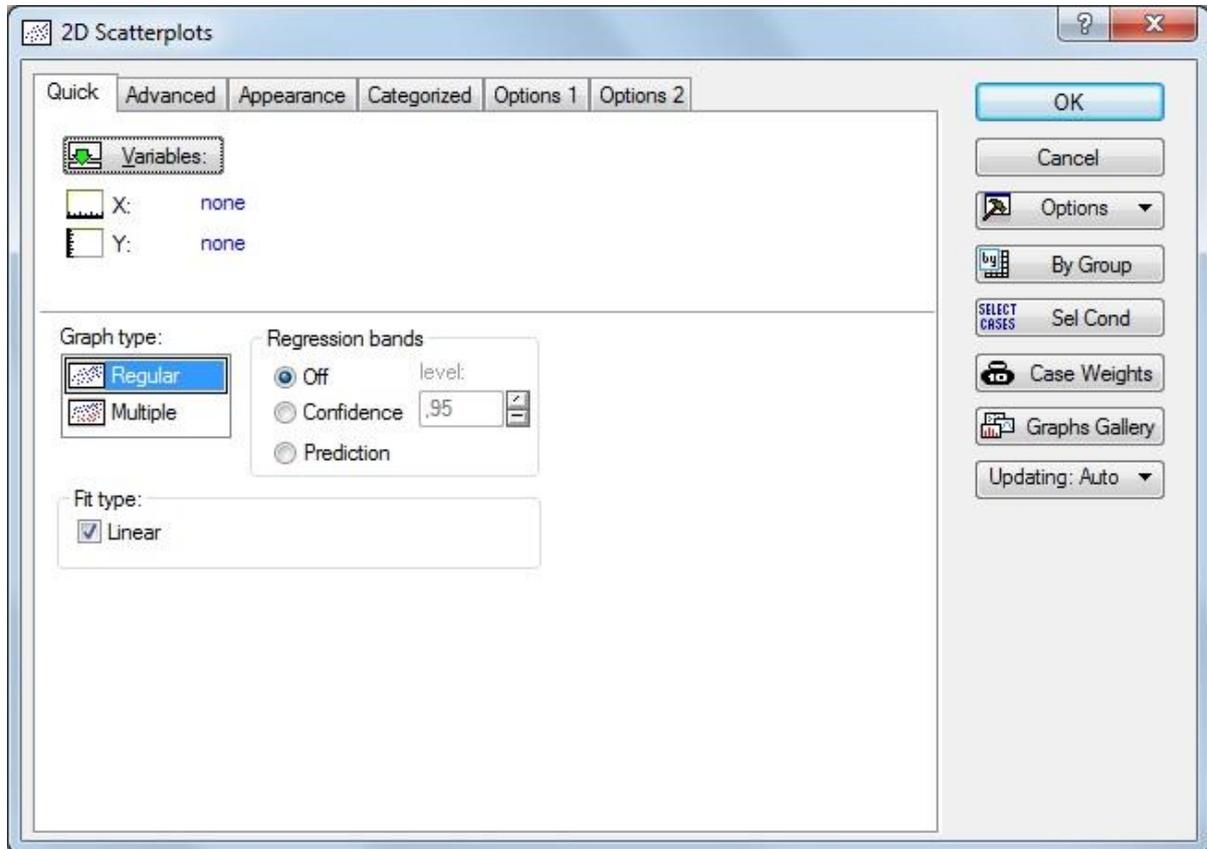


Рис. 3. Окно «2D Scatterplots» программы Statistica 10.

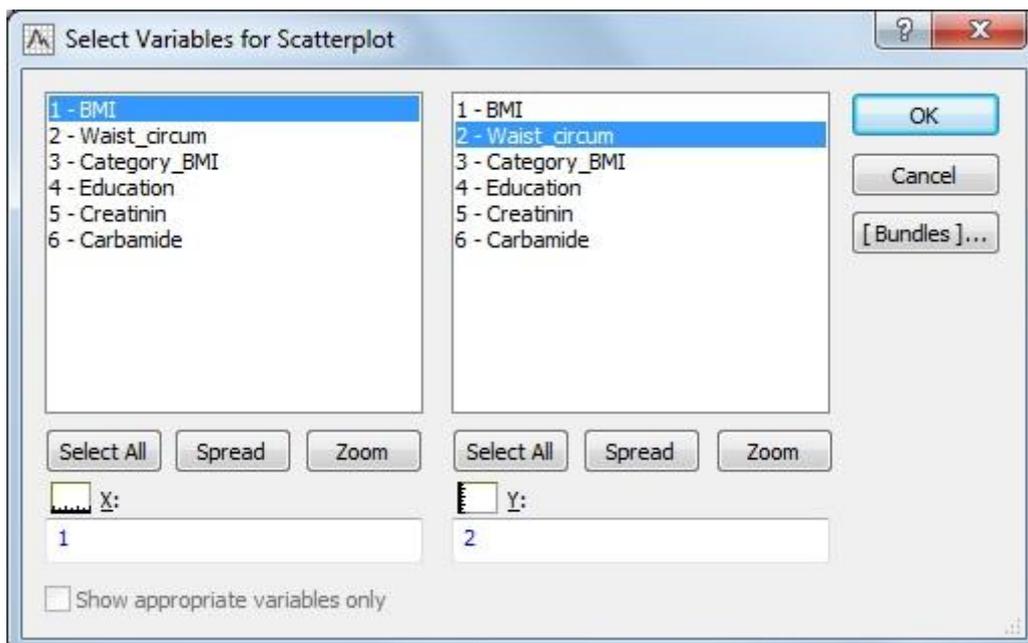
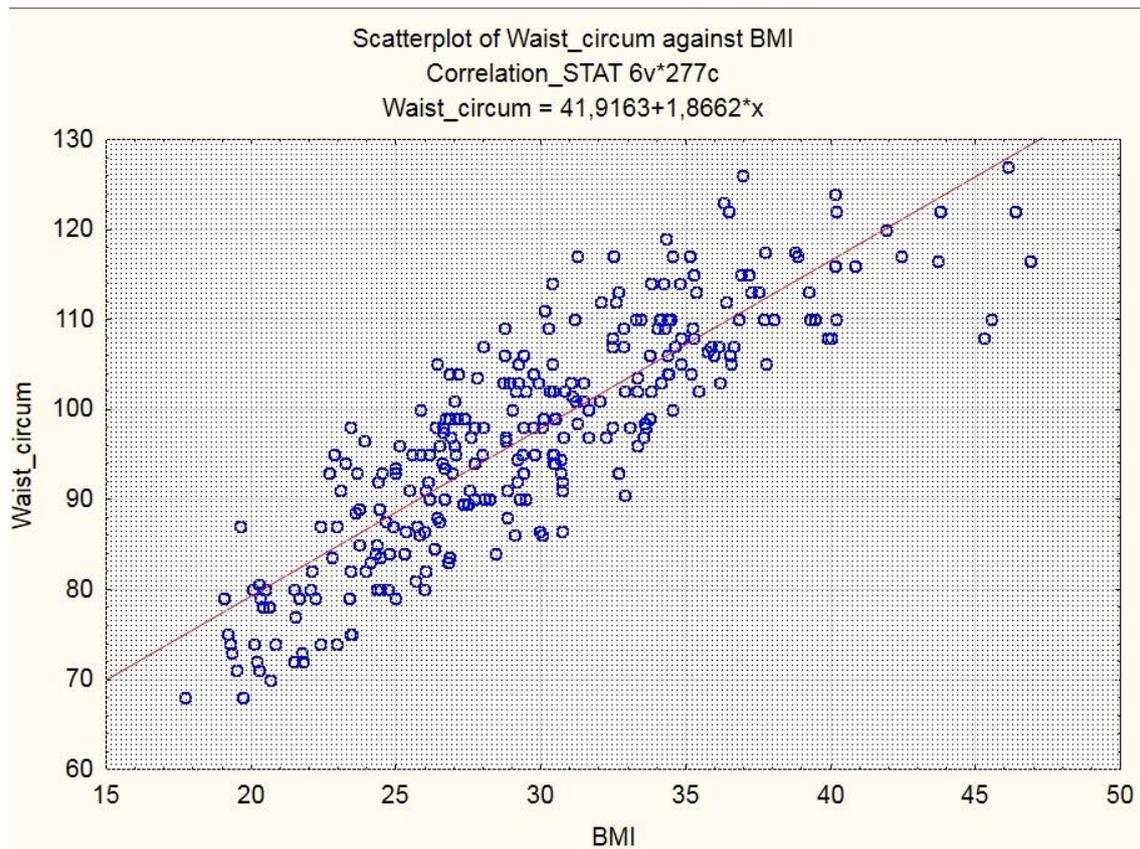


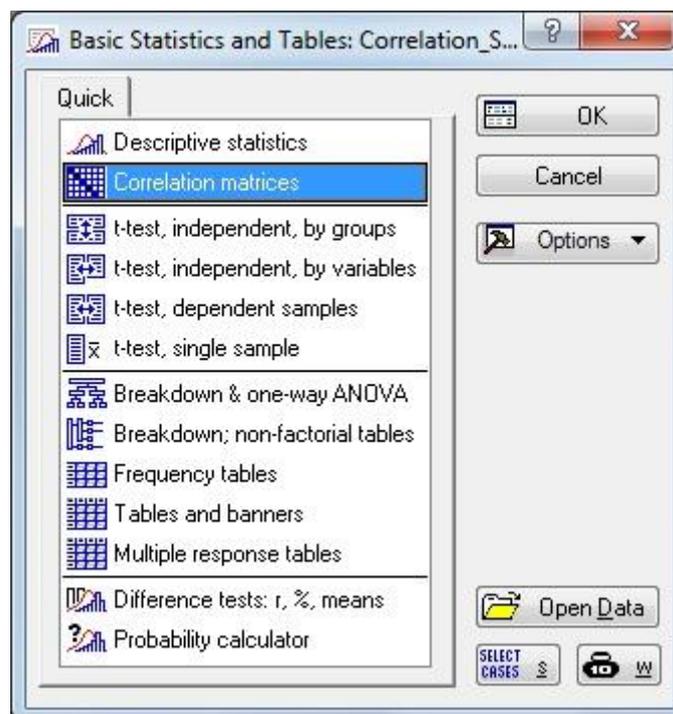
Рис. 4. Окно «Select Variables for Scatterplot» программы Statistica 10.



**Рис. 5. Скаттерграмма зависимости между переменными «BMI» и «Waist\_circum» (программа Statistica 10).**

Для проведения корреляционного анализа с использованием коэффициента корреляции Пирсона выбираем меню «Statistics» (в верхней части рабочего пространства программы) и

входим в раздел «Basic Statistics/Tables». В появившемся окне выбираем раздел «Correlation matrices», и подтверждаем выбор нажатием на кнопку «ОК» (рисунок 6).



**Рис. 6. Окно «Basic Statistics and Tables» программы Statistica 10.**

В открывшемся окне нажимаем на кнопку «One variable list» (рисунок 7) и в окне «Select the variables for the analysis» выбираем переменные «BMI» и «Waist\_circum», как показано на рисунке 8, и подтверждаем выбор

кнопкой «OK» (для выбора обеих переменных необходимо левой кнопкой мыши выбрать одну переменную, после чего удерживая кнопку «Ctrl» также левой кнопкой мыши выбрать вторую переменную).

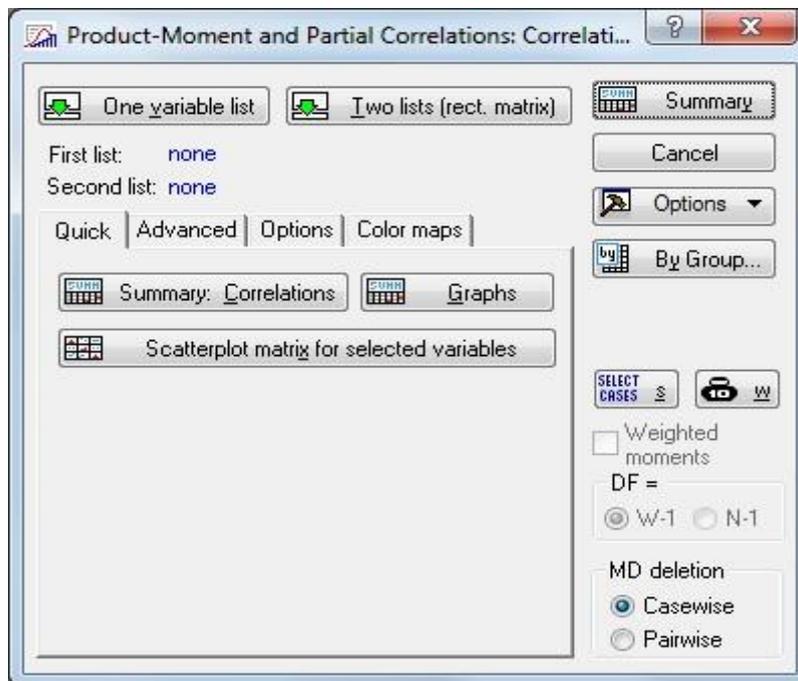


Рис. 7. Окно «Product-Moment and Partial Correlations...» программы Statistica 10.

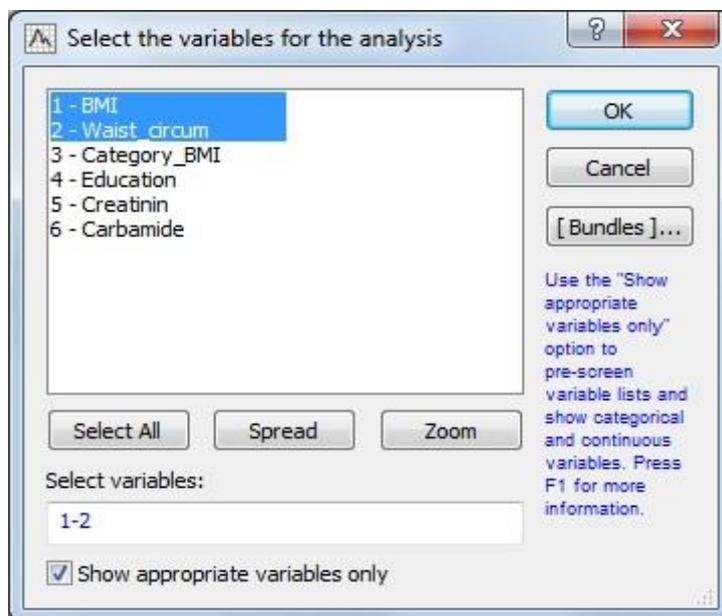


Рис. 8. Окно «Select the variables for the analysis» программы Statistica 10 (выбор переменных «BMI» и «Waist\_circum»).

Программа вернется к окну «Product-Moment and Partial Correlations...», в котором нажмем на кнопку «Graphs» для формирования скаттерограммы (рисунок 9). Сформированная программой скаттерограмма соответствует представленной на рисунке 5, также по осям представляется гистограмма

распределения, для визуальной оценки соответствия распределения включенных в анализ переменных нормальному распределению. Выше графика в строке «Correlation:  $r = ,85705$ » представлено значение коэффициента корреляции Пирсона ( $r = 0,86$ ).

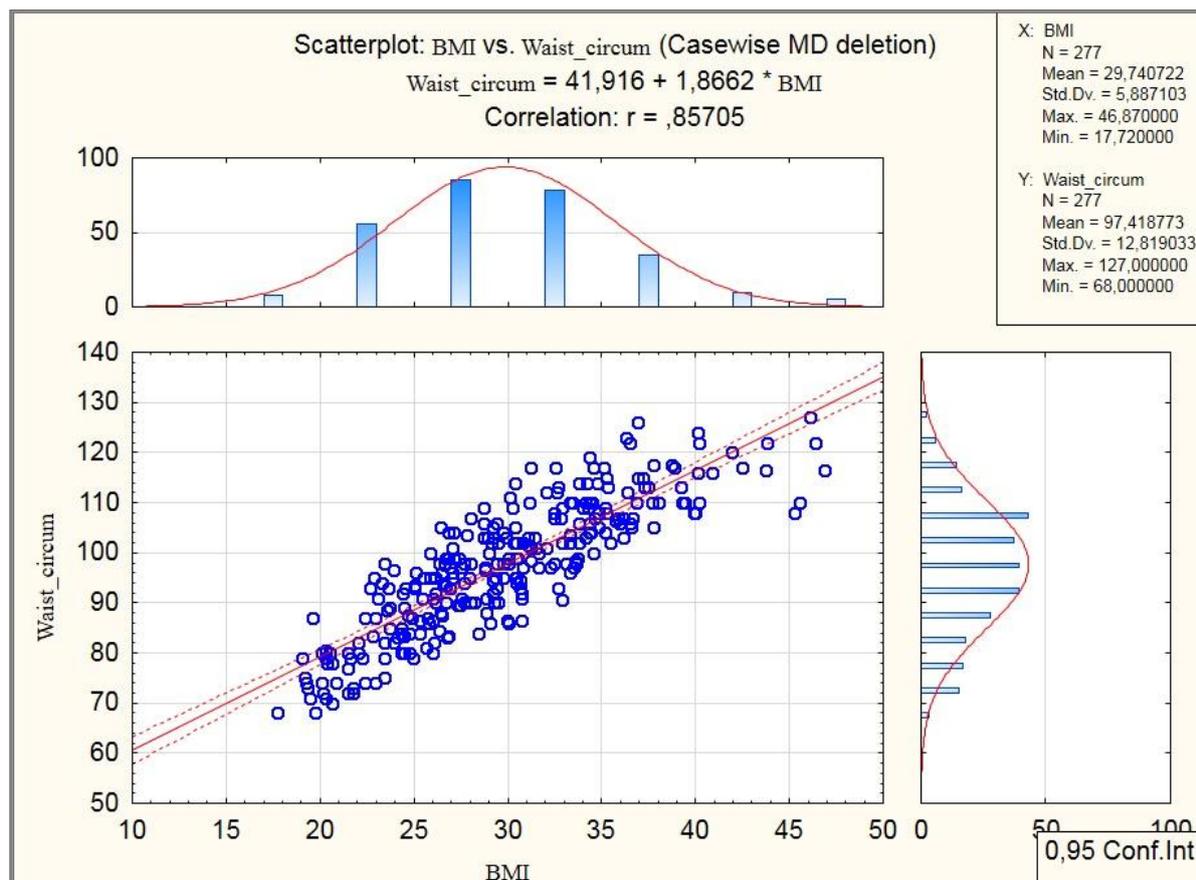


Рис. 9. Результаты корреляционного анализа с использованием коэффициента корреляции Пирсона (скаттерограмма).

Далее нажмем на вкладку «Product-Moment...» в нижнем левом углу рабочего поля программы, чтобы вернуться к окну Product-Moment and Partial Correlations..., и нажмем на кнопку «Summary: Correlations». Программа представит таблицу (рисунок 10), в которой также приведено значение

коэффициента корреляции Пирсона, а выше таблицы программа указывает, что значение  $r$  является статистически значимым («Marked correlations are significant at  $p < ,05000$ »). В данном случае значение коэффициента статистически значим, поэтому в таблице программа выделяет его красным цветом.

Correlations (Correlation_STAT)				
Marked correlations are significant at $p < ,05000$				
N=277 (Casewise deletion of missing data)				
Variable	Means	Std.Dev.	BMI	Waist_circum
BMI	29,74072	5,88710	1,000000	<b>0,857052</b>
Waist_circum	97,41877	12,81903	<b>0,857052</b>	1,000000

Рис. 10. Результаты корреляционного анализа с использованием коэффициента корреляции Пирсона (таблица).

Таким образом, на основании проведенного корреляционного анализа установлено, что между ИМТ и окружностью живота существует сильная положительная корреляционная связь ( $r = 0,86$ ,  $n = 277$ ,  $p < 0,05$ ). Соответственно, коэффициент детерминации равен  $r^2 = 0,86^2 = 0,74$ , то есть варибельность одной переменной способно объяснить 74% варибельности второй переменной.

Далее рассмотрим корреляционный анализ переменных «Creatinin» и «Carbamide», имеющих распределение, отличное от нормального.

Для проведения корреляционного анализа с использованием коэффициента корреляции Спирмена выбираем меню «Statistics» (в верхней части рабочего пространства программы) и входим в раздел

«Nonparametrics». В появившемся окне выбираем раздел «Correlations (Spearman, Kendall tau, gamma)» и подтверждаем выбор нажатием на кнопку «OK» (рисунок 11).

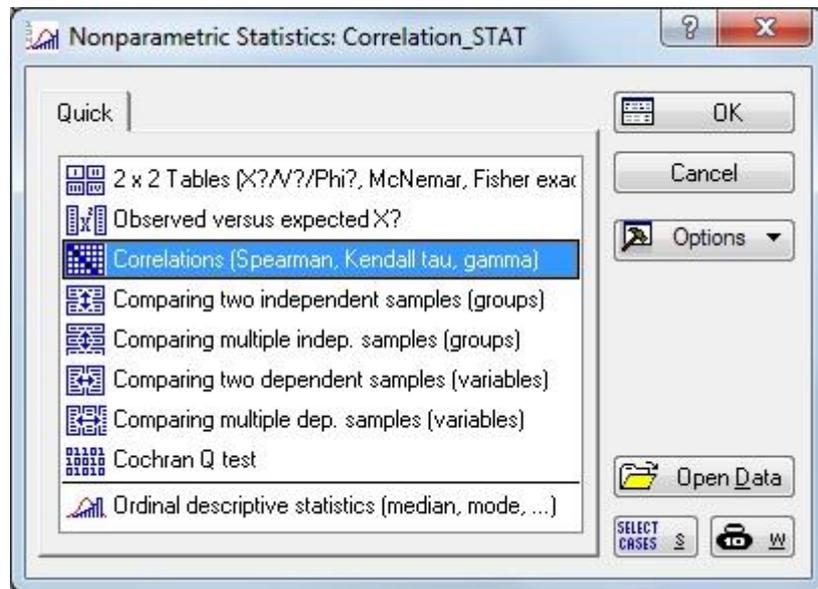


Рис. 11. Окно «Nonparametric Statistics» программы Statistica 10.

Программа откроет окно «Nonparametric Correlation» (рисунок 12), в котором нажмем на кнопку «Variables» для выбора переменных для анализа. В открывшемся окне выберем

переменные «Creatinin» и «Carbamide», после чего подтвердим выбор кнопкой «OK» (рисунок 13).

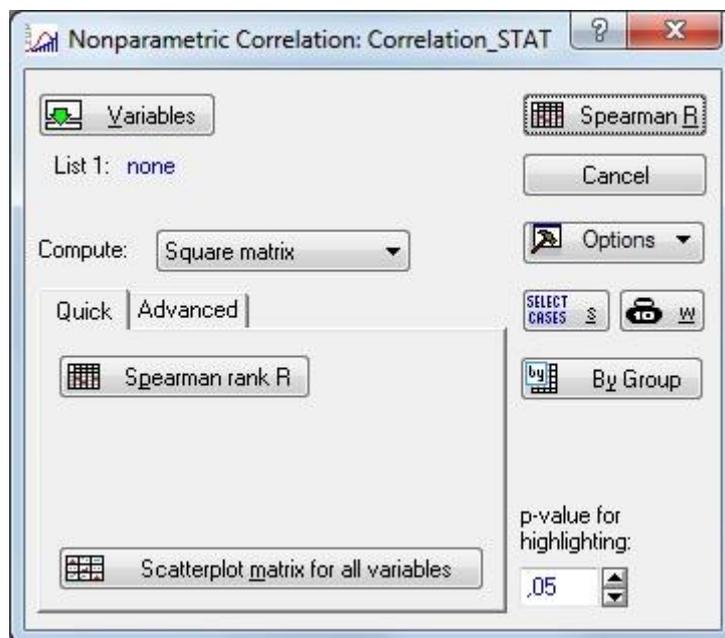


Рис. 12. Окно «Nonparametric Correlation» программы Statistica 10.

Программа вернется к окну «Nonparametric Correlation», в котором нажмем кнопку «Scatterplot matrix for all variables», которая сформирует представленную на рисунке 14 скаттерограмму и гистограммы распределения

переменных «Creatinin» и «Carbamide» (очевидно, что обе переменные имеют отличное от нормального, скошенное вправо распределение).

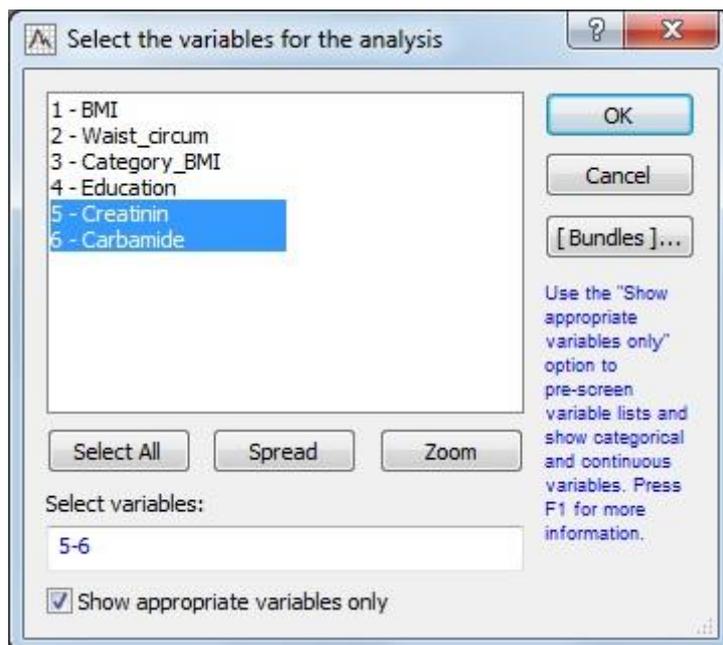


Рис. 13. Окно «Select the variables for the analysis» программы Statistica 10 (выбор переменных «Creatinin» и «Carbamide»).

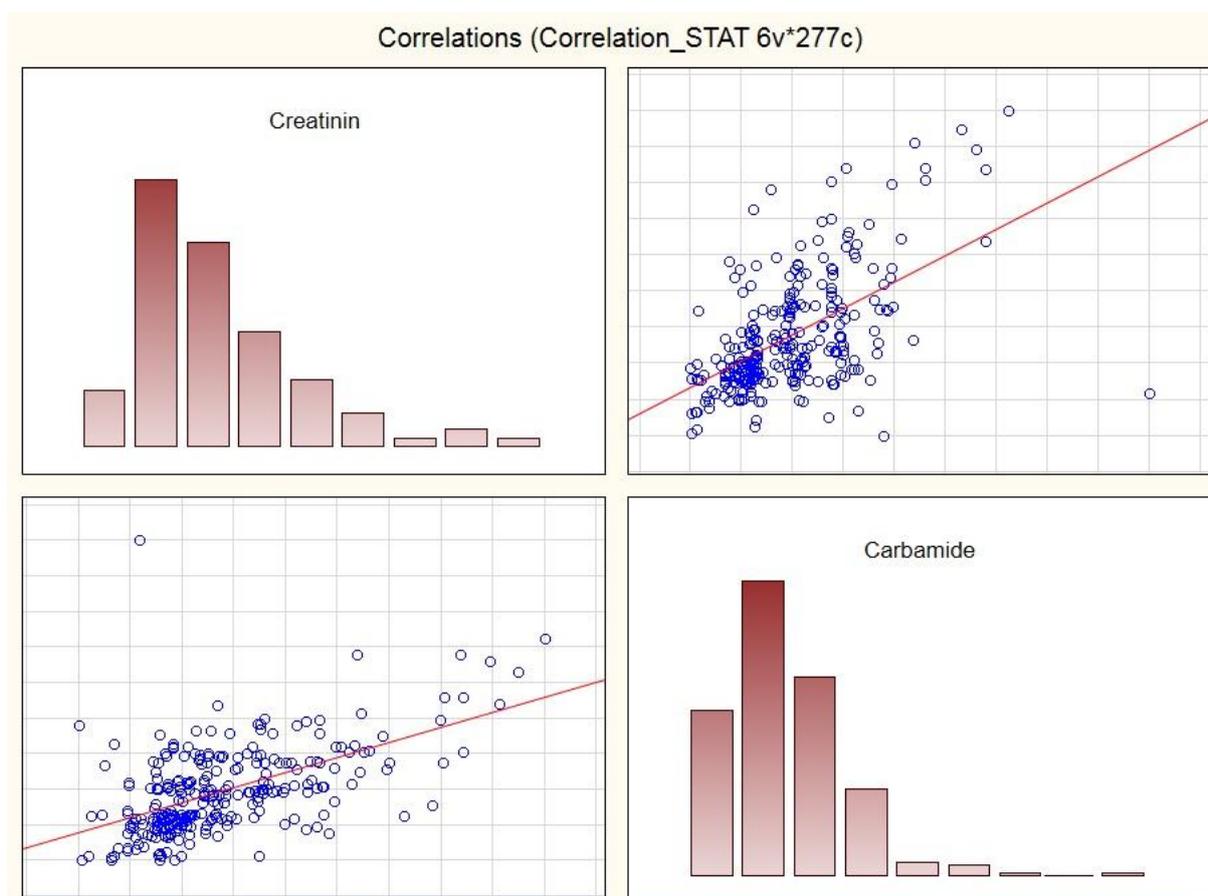


Рис. 14. Результаты корреляционного анализа с использованием коэффициента корреляции Спирмена (скаттерграмма).

Вернемся к окну анализа с помощью нажатия на вкладку «Nonparametric correlations» в нижнем левом углу рабочего

поля программы, и нажмем на кнопку «Spearman rank R» для запуска анализа (рисунок 12).

В таблице, представленной на рисунке 15, программа демонстрирует значение коэффициента корреляции Спирмена (0,534125) и сообщает, что значение коэффициента статистически значимо

(«Marked correlations are significant at  $p < ,05000$ »). Как и при расчете коэффициента корреляции Пирсона, программа выделяет красным цветом значение статистически значимого коэффициента.

Spearman Rank Order Correlations (Correlation_STAT)	
MD pairwise deleted	
Marked correlations are significant at $p < ,05000$	
Variable	
	Creatinin Carbamide
Creatinin	1,000000 0,534125
Carbamide	0,534125 1,000000

Рис. 15. Результаты корреляционного анализа с использованием коэффициента корреляции Спирмена (таблица).

Таким образом, между уровнем креатинина и мочевины сыворотки крови обнаружена положительная корреляционная связь средней силы:  $r_s = 0,53$ ,  $n = 277$ ,  $p < 0,05$ .

Далее на примере ординальных переменных «Category\_BMI» и «Education» рассмотрим алгоритм расчета коэффициента корреляции Кендалла  $\tau_b$ .

Для этого снова входим в раздел «Nonparametrics». В появившемся окне выберем раздел «Correlations (Spearman, Kendall tau, gamma)» и подтвердим выбор нажатием на кнопку «ОК» (рисунок 11).

В окне «Nonparametric Correlation» (рисунок 12), нажмем на кнопку «Variables» и выберем переменные «Category\_BMI» и «Education» после чего подтвердим выбор кнопкой «ОК» (рисунок 16).

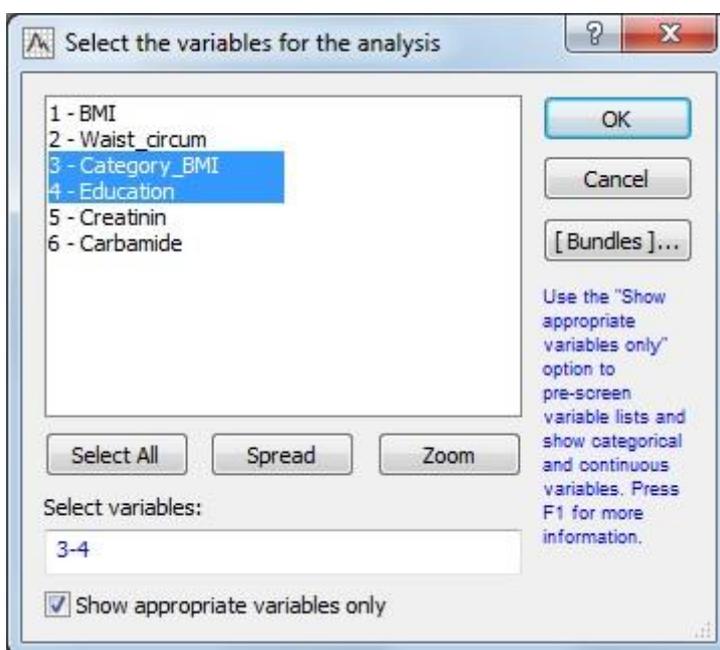


Рис. 16. Окно «Select the variables for the analysis» программы Statistica 10 (выбор переменных «Category\_BMI» и «Education»).

Программа вернется к окну «Nonparametric Correlation» (рисунок 12), где нажмем на вкладку «Advanced», которая открывает

возможность расчета критерия Кендалла  $\tau_b$ . Для запуска анализа нажмем на кнопку «Kendall Tau» (рисунок 17).

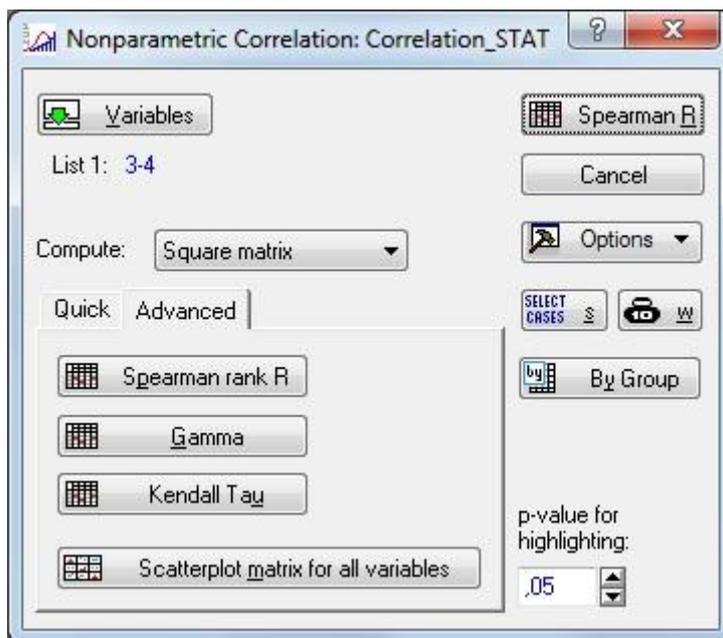


Рис. 17. Вкладка «Advanced» окна «Nonparametric Correlation» программы Statistica 10.

Результаты анализа программа представляет в виде таблицы (рисунок 18) где указано и выделено красным цветом значение коэффициента корреляции (0,17778). Строка

«Marked correlations are significant at  $p < ,05000$ » сообщает о том, что значение коэффициента корреляции Кендалла  $\tau_b$  статистически значимо.

Kendall Tau Correlations (Correlation_STAT)	
MD pairwise deleted	
Marked correlations are significant at $p < ,05000$	
Variable	
Category_BMI	1,000000
Education	0,177778

Рис. 18. Результаты корреляционного анализа с использованием коэффициента корреляции Кендалла  $\tau_b$ .

Таким образом, между уровнем образования и категорией ИМТ пациентов выявлена слабая

положительная корреляционная связь:  $\tau_b = 0,18$ ,  $n = 277$ ,  $p < 0,05$ ).

### Корреляционный анализ с использованием программы SPSS 20.

Для начала работы необходимо открыть файл 7\_Correlation\_SPSS.sav, который потребуется загрузить с сайта журнала «Наука и Здоровоохранение». В файле представлены те же вариационные ряды, что и в файле данных программы Statistica: ИМТ («BMI»), окружность талии («Waist\_circum») категория ИМТ («Category\_BMI»), уровень образования («Education»), уровень креатинина

(«Creatinin») и мочевины («Carbamide») крови пациентов.

На начальном этапе корреляционного анализа с использованием коэффициента корреляции Пирсона необходимо построить скаттерограмму.

Для этого войдем в меню «Graphs», раздел «Legacy Dialogs», подраздел «Scatter/Dot» (рисунок 14).

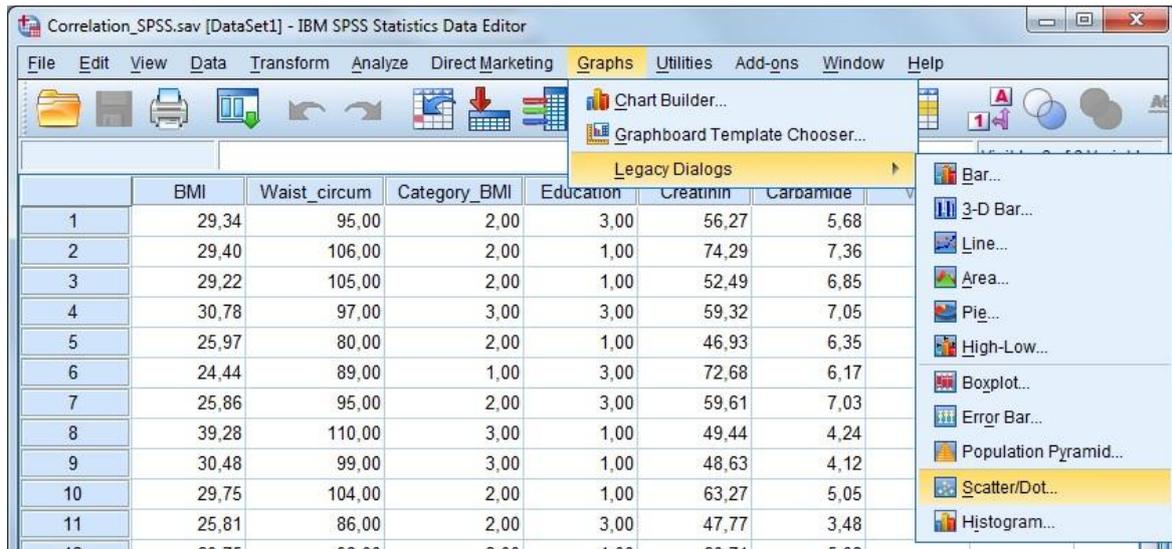


Рис. 14. Выбор подраздела «Scatter/Dot» меню «Graphs» программы SPSS 20.

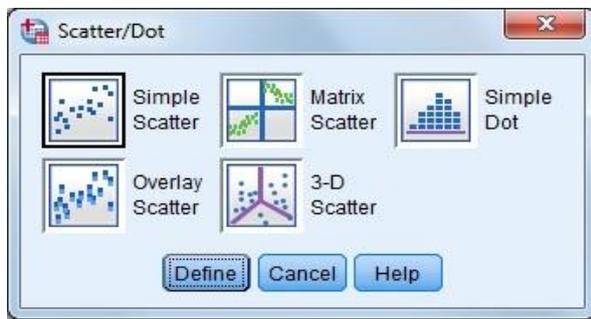


Рис. 15. Окно «Scatter/Dot» программы SPSS 20.

В открывшемся окне «Scatter/Dot» выберем простую скаттерограмму «Simple Scatter» и нажмем на кнопку «Define» (рисунок 15).

В открывшемся окне «Simple Scatterplot» с помощью стрелок между полями перенесем переменную «BMI» в поле «X Axis», а переменную «Waist\_circum» – в поле «Y Axis» и нажмем на кнопку «OK» (рисунок 16).

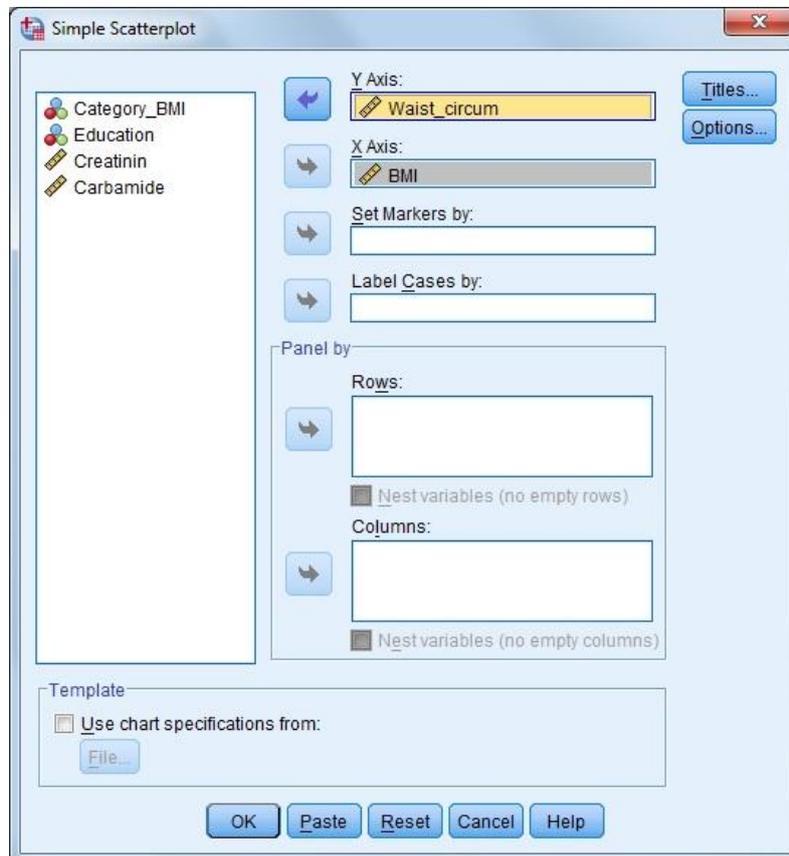
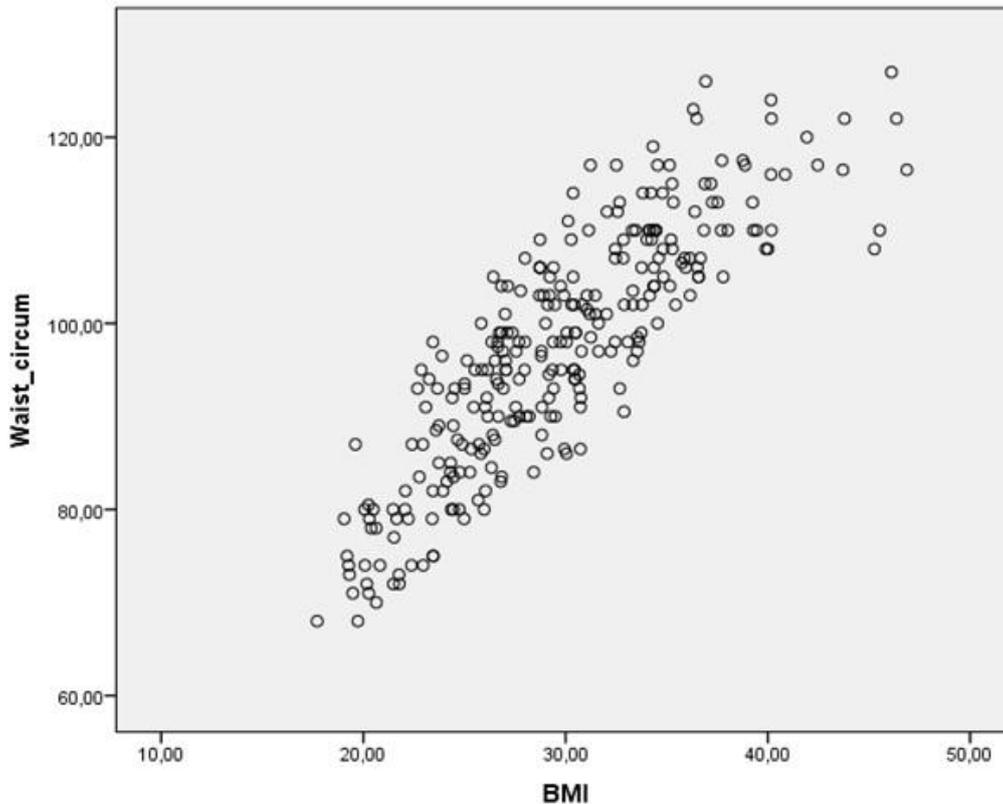


Рис. 16. Окно «Simple Scatterplot» программы SPSS 20.

В результате программа сформирует  
 скаттерограмму, представленную на рисунке

17. Данный график полностью соответствует  
 представленному на рисунке 5.

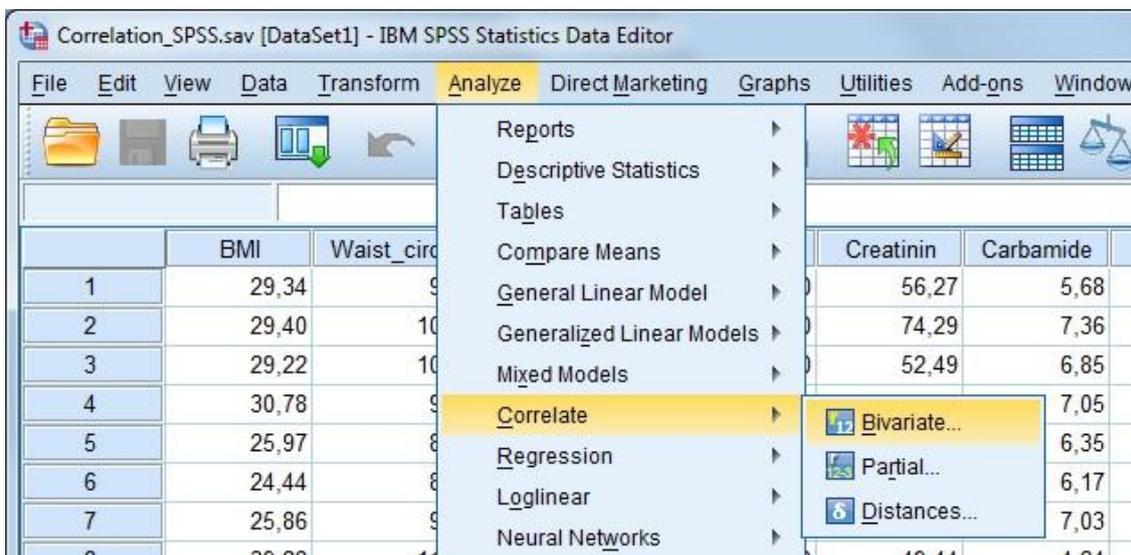


**Рис. 17. Скаттерограмма зависимости между переменными «BMI» и «Waist\_circum» (программа SPSS 20).**

Для расчета коэффициента корреляции Пирсона в первую очередь необходимо проверить соответствие распределения имеющихся переменных закону нормального распределения. Алгоритм проверки

распределения с помощью программы SPSS подробно описан в [13, 8].

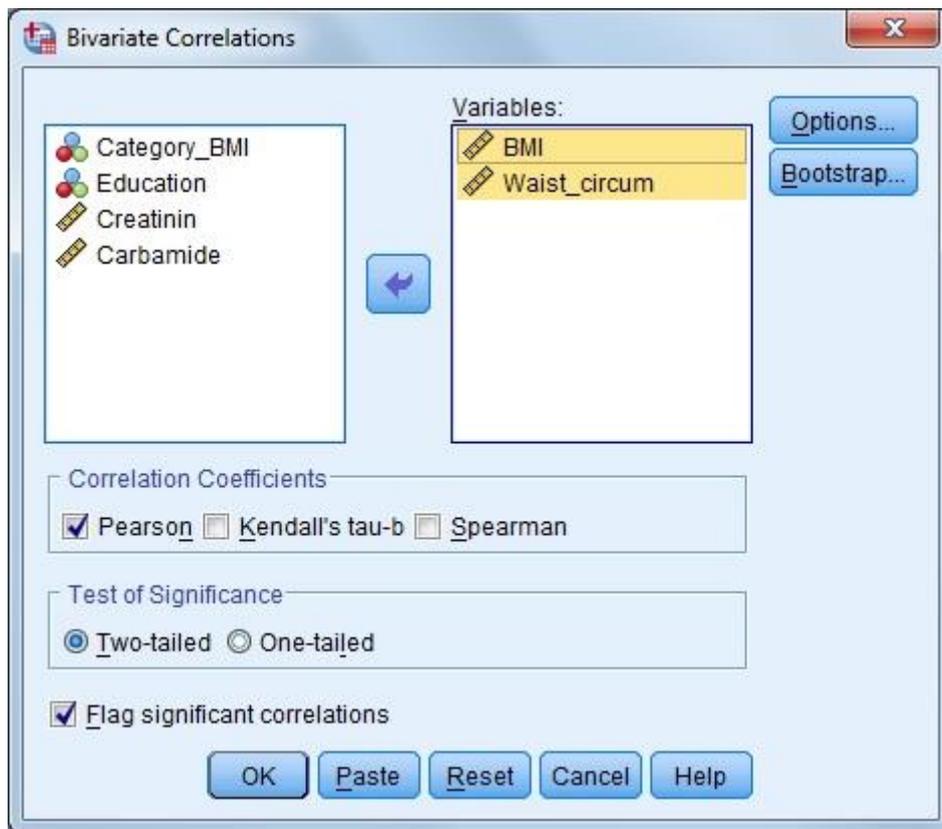
Непосредственно для проведения корреляционного анализа войдем в меню «Analyze», раздел «Correlate», подраздел «Bivariate» (рисунок 18).



**Рис. 18. Выбор подраздела «Correlate» – «Bivariate» меню «Analyze» программы SPSS 20.**

Откроемся окно «Bivariate Correlations», в котором с помощью стрелки необходимо перенести переменные «BMI» и «Waist\_circum»

в правое поле (рисунок 19). Оставим отмеченной галочкой позиции «Pearson» и «Two-tailed» и нажмем на кнопку «OK» для запуска анализа.



**Рис. 19. Окно «Bivariate Correlations» программы SPSS 20 (выбор переменных «BMI» и «Waist\_circum»).**

Результаты расчета коэффициента корреляции Пирсона для переменных «BMI» и «Waist\_circum» представлены в таблице 4. В таблице указано, что значение коэффициента

корреляции Пирсона составляет 0,857 (строка «Pearson Correlation»), а уровень статистической значимости менее 0,001 (строка «Sig. (2-tailed)»).

*Таблица 4.*

**Результаты расчета коэффициента корреляции Пирсона для переменных «BMI» и «Waist\_circum»**

		BMI	Waist_circum
BMI	Pearson Correlation	1	,857**
	Sig. (2-tailed)		,000
	N	277	277
Waist_circum	Pearson Correlation	,857**	1
	Sig. (2-tailed)	,000	
	N	277	277

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Для расчета коэффициента корреляции Спирмена для переменных «Creatinin» и «Carbamide» снова войдем в меню «Analyze», раздел «Correlate», подраздел «Bivariate» (рисунок 18). В окне «Bivariate Correlations»

перенесем в правое поле переменные «Creatinin» и «Carbamide» и отметим галочкой позицию «Spearman», после чего запустим анализ кнопкой «OK» (рисунок 20). Результаты расчетов представлены в таблице 5.

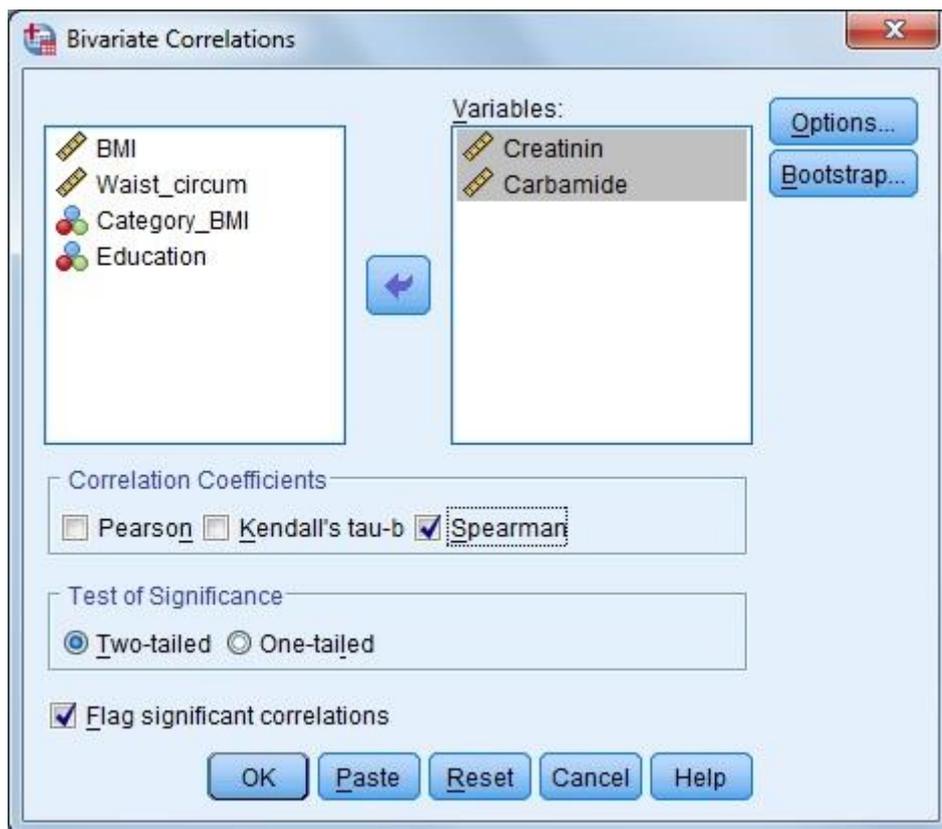


Рис. 20. Окно «Bivariate Correlations» программы SPSS 20 (выбор переменных «Creatinin» и «Carbamide»).

Таблица 5.

Результаты расчета коэффициента корреляции Спирмена для переменных «Creatinin» и «Carbamide».

			Creatinin	Carbamide
Spearman's rho	Creatinin	Correlation Coefficient	1,000	,534**
		Sig. (2-tailed)	.	,000
		N	277	277
	Carbamide	Correlation Coefficient	,534**	1,000
		Sig. (2-tailed)	,000	.
		N	277	277

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Для расчета коэффициента корреляции Кендалла  $\tau_b$  для переменных «Creatinin» и «Carbamide» снова войдем в меню «Analyze», раздел «Correlate», подраздел «Bivariate» (рисунок 18), и в окне «Bivariate Correlations»

перенесем в правое поле переменные «Category\_BMI» и «Education» и отметим галочкой позицию «Kendall's tau-b», после чего запустим анализ кнопкой «OK» (рисунок 21). Результаты расчетов представлены в таблице 6.

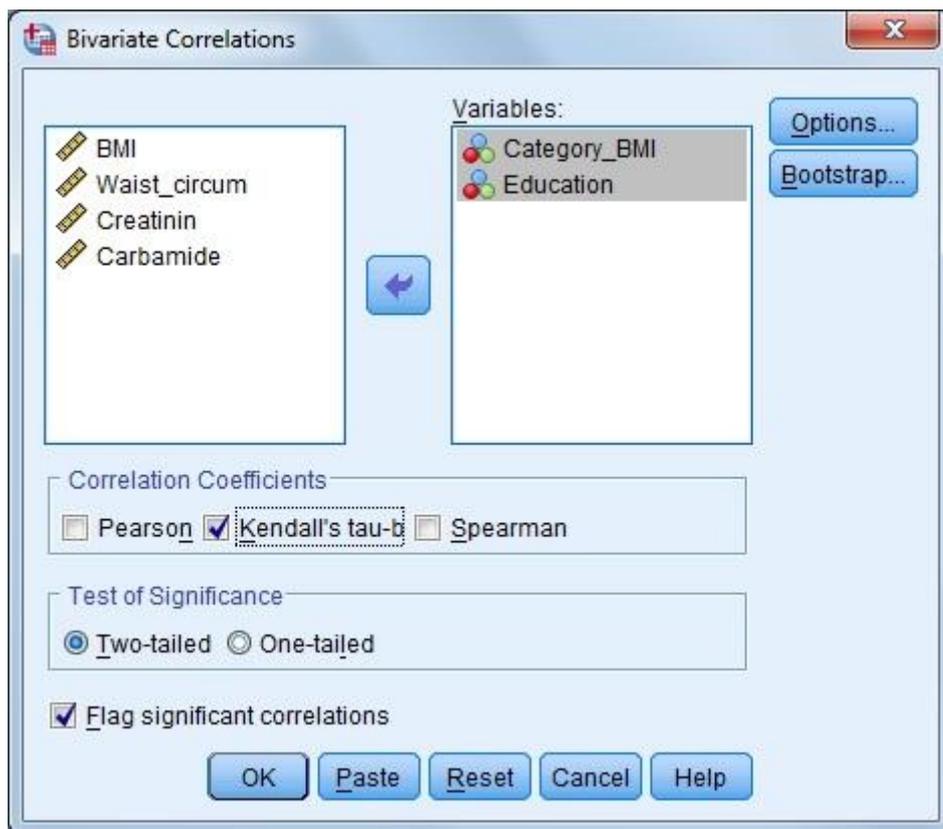


Рис. 21. Окно «Bivariate Correlations» программы SPSS 20 (выбор переменных «Category\_BMI» и «Education»).

Таблица 6.

Результаты расчета коэффициента корреляции Кендалла  $\tau_b$  для переменных «Category\_BMI» и «Education».

		Category_BMI	Education
Kendall's tau_b	Category_BMI	Correlation Coefficient	1,000
		Sig. (2-tailed)	,178**
		N	,002
	Education	Correlation Coefficient	277
		Sig. (2-tailed)	,178**
		N	,002
			277

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Следует отметить, что программа SPSS показывает точные значения достигнутого уровня статистической значимости для коэффициентов корреляции в отличие от Statistica 10, которая только указывает, выше или ниже критического уровня находится значение коэффициента корреляции.

В остальном результаты расчета критериев корреляции Пирсона, Спирмена и Кендалла с помощью программы SPSS 20 полностью соответствуют полученным при использовании программы Statistica 10 и не требуют дополнительных комментариев.

**Литература:**

1. Аканов А.А., Турдалиева Б.С., Изекенова А.К., Рамазанова М.А., Абдраимова Э.Т., Гржибовский А.М. Оценка использования статистических методов в научных статьях медицинских журналов Казахстана // Экология человека. 2013. №5. С. 61-64.
2. Банержи А. Медицинская статистика понятным языком: вводный курс. М. : Практическая медицина, 2007. 287 с.
3. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: для профессионалов. СПб. : Питер, 2003. 688 с.

4. Бююль А., Цефель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. СПб. : ООО «ДиаСофтЮП», 2005. 608 с.
5. Гланц С. Медико-биологическая статистика. М. : Практика, 1998. 459 с.
6. Гржибовский А.М. Анализ порядковых данных // Экология человека. 2008. №8. С. 56-62.
7. Гржибовский А.М. Корреляционный анализ // Экология человека. 2008. №9. С. 50-60.
8. Гржибовский А.М. Типы данных, проверка распределения и описательная статистика // Экология человека. 2008. №1. С. 52-58.
9. Гржибовский А.М., Иванов С.В. Анализ номинальных и ранговых переменных данных с использованием программного обеспечения Statistica и SPSS // Наука и Здравоохранение. 2016. № 6. С. 5-39.
10. Гржибовский А.М., Иванов С.В. Исследования типа «случай-контроль» в здравоохранении // Наука и Здравоохранение. 2015. № 4. С. 5-17
11. Гржибовский А.М., Иванов С.В. Когортные исследования в здравоохранении // Наука и Здравоохранение. 2015. № 3. С. 5-16.
12. Гржибовский А.М., Иванов С.В. Поперечные (одномоментные) исследования в здравоохранении // Наука и Здравоохранение. 2015. № 2. С. 5-18.
13. Гржибовский А.М., Иванов С.В., Горбатова М.А. Описательная статистика с использованием пакетов статистических программ Statistica и SPSS: и проверка распределения // Наука и Здравоохранение. 2016. № 1. С. 7-23.
14. Гржибовский А.М., Иванов С.В., Горбатова М.А. Сравнение количественных данных двух независимых выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и Здравоохранение. 2016. № 2. С. 5-28
15. Гржибовский А.М., Иванов С.В., Горбатова М.А. Сравнение количественных данных двух парных выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и Здравоохранение. 2016. № 3. С. 5-25.
16. Гржибовский А.М., Иванов С.В., Горбатова М.А. Сравнение количественных данных трех и более независимых выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и Здравоохранение. 2016. № 4. С. 5-37.
17. Гржибовский А.М., Иванов С.В., Горбатова М.А. Сравнение количественных данных трех и более парных выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и Здравоохранение. 2016. № 5. С. 5-29.
18. Гржибовский А.М., Иванов С.В. Экологические (корреляционные) исследования в здравоохранении // Наука и Здравоохранение. 2015. № 5. С. 5-18.
19. Гржибовский А.М., Иванов С.В. Экспериментальные исследования в здравоохранении // Наука и Здравоохранение. 2015. № 6. С. 5-17.
20. Жунисова М.Б., Шалхарова Ж.С., Шалхарова Ж.Н., Гржибовский А.М. Типы пищевого поведения и абдоминальное ожирение // Журн. Медицина. 2015. №4. С. 92-95.
21. Жунисова М.Б., Шалхарова Ж.С., Шалхарова Ж.Н., Нускабаева Г.О., Садыкова К.Ж., Маденбай К.М., Гржибовский А.М. Психоземotionalный стресс как предиктор типа пищевого поведения в Казахстане // Экология человека. 2015. №5. С. 36-45.
22. Зайцев В.М., Лифляндский В.Г., Маринкин В.И. Прикладная медицинская статистика. СПб. : Фолиант, 2003. 428 с.
23. Зуева Л.П., Яфаев Р.Х. Эпидемиология : учебник. СПб : ООО «Издательство Фолиант», 2008. 752 с.
24. Лакин Г.Ф. Биометрия. М. : Высшая школа, 1990. 351 с.
25. Маденбай К.М., Шалхарова Ж.С., Шалхарова Ж.Н., Жунисова М.Б., Садыкова К.Ж., Нускабаева Г.О., Гржибовский А.М. Оценка связи между площадью подкожной жировой ткани и показателями электронейромиографии // Экология человека. 2015. №7. С. 58-64.

26. Петри А., Сэбин К. Наглядная статистика в медицине. М. : ГЭОТАР-Мед, 2003. 140 с.

27. Рахымбеков Т.К., Гржибовский А.М. К вопросу о необходимости повышения качества казахстанских научных публикаций для успешной интеграции в международное научное сообщество // Наука и Здоровоохранение. 2015. №1. С. 5-11.

28. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. М. : МедиаСфера, 2002. 312 с.

29. Садыкова К.Ж., Шалхарова Ж.С., Нускабаева Г.О., Садыкова А.Д., Жунисова М.Б., Маденбай К.М., Гржибовский А.М. Распространенность анемии, ее социально-демографические детерминанты и возможная связь с метаболическим синдромом в г. Туркестан, Южный Казахстан // Экология человека. 2015. №8. С. 58-64.

30. Статистический анализ эмпирических исследований [электронный ресурс]. URL: [www.stateexpert.org/articles/таблицы\\_критических\\_значений\\_статистических\\_критериев](http://www.stateexpert.org/articles/таблицы_критических_значений_статистических_критериев) (дата обращения 08.09.2015).

31. Субботина А.В., Гржибовский А.М. Описательная статистика и проверка нормальности распределения количественных данных // Экология человека. 2014. №2. С. 51-57.

32. Унгурияну Т.Н., Гржибовский А.М. Корреляционный анализ с использованием пакета статистических программ STATA // Экология человека. 2014. №9. С. 60-64.

33. Флетчер Р. Клиническая эпидемиология. Основы доказательной медицины: пер. с англ. / Р. Флетчер, С. Флетчер, Э. Вагнер. М. : Медиа Сфера, 1998. 352 с.

34. Юнкеров В.И., Григорьев С.Г. Математико-статистическая обработка данных медицинских исследований. СПб : ВМедА, 2002. 266 с.

35. Anderson M. RSM simplified: optimizing processes using response surface methods for design of experiments / M. Anderson P., Whitcomb. – London : Taylor & Francis, 2005. P. 39-42.

36. Beaglehole R., Bonita R. Basic epidemiology. World Health Organization, Geneva, 1993.

37. Bonett D. Wright T. Sample size requirements for estimating Pearson, Kendall and Spearman correlations // Psychometrica. 2000. Vol. 65. P. 23-28.

38. Cleopas T.J. et al. Statistics Applied to Clinical Trials. 4th ed. Springer, 2009.

39. David F. Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples. Cambridge : Cambridge University Press, 1938.

40. Kendall M. A new method of rank correlation // Biometrika. 1938. Vol. 30. P. 91-93.

41. Kruskal W. Ordinal measures of association // Journal of the American Statistical Association. 1958. Vol. 53. P. 814-861.

#### References:

1. Akanov A.A., Turdalieva B.S., Izenkova A.K., Ramazanova M.A., Abdraimova, Grjibovski A.M. Otsenka ispolzovania statisticheskikh metodov v nauchnih statyakh Kazakhstana [Assesment of use of statistical methods in scientific articles of the Kazakhstan's medical journals]. *Ekologiya cheloveka* [Human Ecology]. 2013. No.5. PP. 61-64. [in Russian]

2. Banerzhi A. *Meditinskaya statistika ponyatnym yazykom : vvodnyy kurs* [Medical statistics in plain language : an introductory course]. М. : Prakticheskaya meditsina, 2007. P. 287. [in Russian].

3. Borovikov V. *STATISTICA. Iskusstvo analiza dannikh na kompyutere: dlya professionalov* [STATISTICA. The art of data analysis using computer: for professionals]. SPb. : Piter, 2003. P. 688.

4. Buhl A., Zofel P. *SPSS: isskustvo obrabotki informatsii. Analiz statisticheskikh daanikh i vosstanovlenie skritikh zakonomernostey* [SPSS: the art of information analysis. Statistical data analysis and hidden regularities identification]. SPb. : ООО «DiaSoftUP», 2005. P. 608. [in Russian]

5. Glants S. *Mediko-biologicheskaya statistika* [The biomedical statistics]. М. : Praktika, 1998. PP. 459. [in Russian]

6. Grjibovski A.M. Analiz poryadkovikh dannikh [Analysis of ordinal data]. *Ekologiya*

*cheloveka* [Human Ecology]. 2008. No.1. PP. 56-62. [in Russian].

7. Grijbovski A.M. Korrelatsionniy analiz [Correlation analysis]. *Ekologiya cheloveka* [Human Ecology]. 2008. No.9. PP. 50-60. [in Russian].

8. Grijbovski A.M. Tipy dannikh, proverka raspredeleniya I opisatel'naya statistika [Types of data, distribution estimation and descriptive statistics]. *Ekologiya cheloveka* [Human Ecology]. 2008. No.1. PP. 52-58. [in Russian].

9. Grijbovski A.M., Ivanov S.V. Analiz nominalnykh I rangovykh peremennykh dannikh s ispol'zovaniyem programmogo obespecheniya Statistica i SPSS [Analysis of nominal and ordinal data using Statistica and SPSS software]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2016. № 6. pp. 5-39. [in Russian].

10. Grijbovski A.M., Ivanov S.V. Issledovaniya tipa sluchay-kontrol v zdravookhraneni [Case-control studies in health sciences]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2015, 4, pp. 5-17 [in Russian].

11. Grijbovski A.M., Ivanov S.V. Kogortnie issledovaniya v zdravookhraneni [Cohort studies in health sciences]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2015, 3, pp. 5-16. [in Russian]

12. Grijbovski A.M., Ivanov S.V. Poperechnye (odnomomentnye) issledovaniya v zdravookhraneni [Cross-sectional studies in health sciences]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2015, No2, pp. 5-18. [in Russian]

13. Grijbovski A.M., Ivanov S.V., Gorbatova M.A. Opisatel'naya statistika s ispol'zovaniyem paketov statisticheskikh programm Statistica i SPSS I proverka raspredeleniya [Descriptive statistics using Statistica and SPSS software]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2016, 1, pp. 7-23 [in Russian].

14. Grijbovski A.M., Ivanov S.V., Gorbatova M.A. Sravneniye kolichestvennykh dannikh dvukh nezavisimyykh vyborok s ispol'zovaniyem programmogo obespecheniya Statistica i SPSS : parametricheskiye i neparametricheskiye kriterii [Comparing the quantitative data of two independent groups using the software Statistica and SPSS: parametric and nonparametric tests]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2016, 2, pp.5-28 [in Russian].

15. Grijbovski A.M., Ivanov S.V., Gorbatova M.A. Sravneniye kolichestvennykh dannikh dvukh parnykh vyborok s ispol'zovaniyem programmogo obespecheniya Statistica i SPSS : parametricheskiye i neparametricheskiye kriterii [Comparing the quantitative data of two dependent variations using the software Statistica and SPSS: parametric and nonparametric tests]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2016, 3, pp. 5-25. [in Russian].

16. Grijbovski A.M., Ivanov S.V., Gorbatova M.A. Sravneniye kolichestvennykh dannikh trekh i boleye nezavisimyykh vyborok s ispol'zovaniyem programmogo obespecheniya Statistica i SPSS : parametricheskiye i neparametricheskiye kriterii [Comparing of the quantitative data of three or more independent samples using Statistica and SPSS software: parametric and nonparametric methods]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2016. 4. pp. 5-37. [in Russian].

17. Grijbovski A.M., Ivanov S.V., Gorbatova M.A. Sravneniye kolichestvennykh dannikh trekh i boleye parnykh vyborok s ispol'zovaniyem programmogo obespecheniya Statistica i SPSS : parametricheskiye i neparametricheskiye kriterii [Comparing of the quantitative data of three or more dependent samples using Statistica and SPSS software: parametric and nonparametric methods]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2016. 5. C. 5-29 [in Russian].

18. Grijbovski A.M., Ivanov S.V. Ekologicheskiye (korrelyatsionnye) issledovaniya v zdravookhraneni. [Ecological (correlational) studies in health sciences]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2015, 5, pp. 5-18. [in Russian].

19. Grijbovski A.M., Ivanov S.V. Eksperimentalnye issledovaniya v zdravookhraneni [Experimental studies in health sciences]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2015, 6, pp. 5-17. [in Russian].

20. Zhunisova M.B., Shalkarova Zh.S., Shalkarova Zh. N., Nuskabayeva G.O., Sadykova K.Zh., Madenbay K.M., Grijbovski A.M. Tipy pischevogo povedeniya i abdominalnoe ozhirenie [Eating behavior types and abdominal obesity]. *Meditcina* [Medicine]. 2015. No.4. pp. 92-95. [in Russian].

21. Zhunisova M.B., Shalkarova Zh.S., Shalkarova Zh. N., Nuskabayeva G.O., Sadykova K. Zh., Madenbay K.M., Grijbovski A.M.

Psykhoemotsionalniy stress kak predictor tipa pischevogo povedeniya v Kazakhstane [Psychoemotional stress and eating behavior in Kazakhstan]. *Ekologiya cheloveka* [Human Ecology]. 2015. No.5. pp. 36-45. [in Russian]

22. Zaytsev V.M., Lifyandskiy V.G., Marinkin V.I. *Prikladnaya meditsinskaya statistika* [Applied medical statistics]. SPb.: Foliant, 2003. P. 428. [in Russian]

23. Zueva L.P., Yafaev R.H. *Epidemiologiya: uchebnik* [Epidemiology: the textbook]. SPb.: OOO «Izdatelstvo Foliant», 2008. P. 752. [in Russian].

24. Lakin G.F. *Biometria* [Biometrics]. M.: Vyschshaya shkola. 1990. P. 351. [in Russian]

25. Madenbay K.M., Shalkarova Zh.S., Shalkarova Zh. N., Zhunisova M.B., Sadykova K.Zh., Nuskabayeva G.O., Grijbovski A.M. Otsenka svyazi mezhdru ploshadyu podkojnoj jirovoy tkani i pokazatelyami elektroneymyografi [Assesment of the relationship between subcutaneous fat tissue and results of electromyoneurography]. *Ekologiya cheloveka* [Human Ecology]. 2015. No.7. pp. 58-64. [in Russian]

26. Petri A., Sebin K. *Naglyadnaya statistika v meditsine* [Demonstrative statistics in medicine]. M.: GEAOTAR-Med, 2003. p. 140. [in Russian]

27. Rakhypbekov T.K., Grijbovski A.M. K voprosu o neobhodimosti povysheniya kachestva kazakhstanskih nauchnykh publikatsii dlya uspeshnoi integratsii v mezhdunarodnoe nauchnoe soobshchestvo [The need for improvement of the quality of Kazakhstani publications for successful integration in the international research community]. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2015. No.1. pp. 5-11. [in Russian]

28. Rebrova O. Yu. *Statisticheskii analiz meditsinskikh danykh. Primenenie paketa prikladnykh program STATISTICA* [Statistical analysis of medical data. Using of STATISTICA software]. M.: MediaSphera, 2002. P. 312.

29. Sadykova K.Zh., Shalkharova Zh.S., Shalkharova Zh.N. Nuskabaeva G.O., Sadykova A.D., Zhunisova M.B., Madenbay K.M., Grijbovski A.M. Rasprostranennost' anemii, yeye sotsial'no - demograficheskiye determinanty i vozmozhnaya svyaz' s metabolicheskim sindromom v g. Turkestan, Yuzhniy Kazakhstan [Prevalence of anemia, its socio-demographic

determinants and potential association with metabolic syndrome in residents of Turkestan, Southern Kazakhstan]. *Ekologiya cheloveka* [Human Ecology]. 2015. No.8. pp. 58-64. [in Russian]

30. Statistical analysis of empirical research [website] Available at: [www.stateexpert.org/articles/таблицы\\_критически\\_значений\\_статистических\\_критериев](http://www.stateexpert.org/articles/таблицы_критически_значений_статистических_критериев) (Accessed 08 September 2015).

31. Subbotina A.V., Grijbovski A.M. Opisatel'naya statistika i proverka normal'nosti raspredeleniya kolichestvennykh danykh [Descriptive statistics and normality testing for quantitative data]. *Ekologiya cheloveka* [Human Ecology]. 2014. No.2. pp. 51-57. [in Russian].

32. Unguryanu T.N., Grijbovski A.M. Korrelatsionnyi analiz s ispol'zovaniyem paketa statisticheskikh programm STATA [Correlation analysis using STATA] *Ekologiya cheloveka* [Human Ecology]. 2014. No.9. PP. 60-64. [in Russian].

33. Fletcher R. et al. *Klinicheskaya epidemiologiya. Osnovy dokazatel'noi meditsiny* [Clinical epidemiology. Basics of the evidence-based medicine] / R. Fletcher, C. Fletcher, E. Vagner. M.: Media Sphere, 1998. 352 p. [in Russian].

34. Yunkerov V.I., Grigoryev S.G. *Matematiko-statisticheskaya obrabotka danykh meditsinskikh issledovaniy* [Mathematical and statistical analysis of medical research data]. SPb.: VMedA, 2002. P. 266. [in Russian]

35. Anderson M. RSM simplified: optimizing processes using response surface methods for design of experiments / M. Anderson P., Whitcomb. London: Taylor & Francis, 2005. P. 39-42.

36. Beaglehole R., Bonita R. *Basic epidemiology*. World Health Organization, Geneva, 1993.

37. Bonett D. Wright T. Sample size requirements for estimating Pearson, Kendall and Spearman correlations // *Psychometrica*. 2000. Vol. 65. P. 23-28.

38. Cleopas T.J. et al. *Statistics Applied to Clinical Trials*. 4th ed. Springer, 2009.

39. David F. *Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples*. Cambridge: Cambridge University Press, 1938.

40. Kendall M. A new method of rank correlation // *Biometrika*. 1938. Vol. 30. P. 91-93.

41. Kruskal W. Ordinal measures of association // *Journal of the American Statistical Association*. 1958. Vol. 53. P. 814-861.

**Контактная информация:**

**Гржибовский Андрей Мечиславович** – доктор медицины, магистр международного общественного здравоохранения, Старший советник Национального Института Общественного Здравоохранения, г. Осло, Норвегия; Заведующий ЦНИЛ СГМУ, г. Архангельск, Россия; Профессор Северо-Восточного Федерального Университета, г. Якутск, Россия; Профессор, Почетный доктор Международного Казахско-Турецкого Университета им. Х.А. Ясяви, г. Туркестан, Казахстан; Почетный профессор ГМУ г. Семей, Казахстан.

**Почтовый адрес:** INFA, Nasjonalt folkehelseinstitutt, Postboks 4404 Nydalen, 0403 Oslo, Norway.

**Email:** Andrej.Grijibovski@gmail.com

**Телефон:** +4745268913 (Норвегия), +79214717053 (Россия), +77471262965 (Казахстан).