

Получена: 11 декабря 2015 / Принята: 18 января 2016 / Опубликовано online: 25 Марта 2016

УДК 614.2 + 303.4

ОПИСАТЕЛЬНАЯ СТАТИСТИКА С ИСПОЛЬЗОВАНИЕМ ПАКЕТОВ СТАТИСТИЧЕСКИХ ПРОГРАММ STATISTICA И SPSS

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

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

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

¹ Национальный Институт Общественного Здравоохранения, г. Осло, Норвегия

² Северный Государственный Медицинский Университет, г. Архангельск, Россия

³ Международный Казахско-Турецкий Университет им. Х.А. Ясави, г. Туркестан, Казахстан

⁴ Северо-Восточный Федеральный Университет, г. Якутск, Россия

⁵ Северо-Западный Государственный Медицинский Университет им. И.И. Мечникова, г. Санкт-Петербург, Россия

В настоящей работе представлены основные методы предварительной статистической обработки данных с использованием программного обеспечения Statistica и SPSS для начинающих исследователей. Описаны различные типы данных, встречающихся в биомедицинских исследованиях, параметры нормального распределения количественных признаков, способы графического представления количественных результатов и методы проверки типа распределения. Настоящая статья призвана дать общие сведения об описательной статистике и проверке распределения данных, и не заменяет прочтения специализированной литературы по статистике и клинической эпидемиологии.

Ключевые слова: Statistica, SPSS, описательная статистика, распределение, Казахстан.

DESCRIPTIVE STATISTICS USING STATISTICA AND SPSS SOFTWARE

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

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

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

¹Norwegian Institute of Public Health, Oslo, Norway

²Northern State Medical University, Arkhangelsk, Russia

³North-Eastern Federal University, Yakutsk, Russia

⁴International Kazakh-Turkish University, Turkestan, Kazakhstan

⁵North-Western State Medical University n.a. I.I. Mechnikov, St. Petersburg, Russia

This is the first paper of a series of articles where we describe basic principles of statistical data analysis using Statistica and SPSS software for beginners. Step-by-step algorithms for presentation of quantitative and qualitative data including normality testing are presented. Measures of central tendency and variability are explained and practical examples using the data from earlier studies are given. The main aim of this paper is to provide basic knowledge on descriptive statistics with practical examples using popular software. The article complements, but does not substitute specialized literature on biostatistics and clinical epidemiology.

Keywords: Statistica, SPSS, descriptive statistics, distribution, Kazakhstan.

STATISTICA ЖӘНЕ SPSS СТАТИСТИКАЛЫҚ БАҒДАРЛАМАЛАРЫ ПАКЕТТЕРІН ҚОЛДАНУМЕН СИПАТТАУШЫ СТАТИСТИКА

Андрей М. Гржибовский¹⁻⁴, <http://orcid.org/0000-0002-5464-0498>,

Сергей В. Иванов⁵, <http://orcid.org/0000-0003-0254-3941>

Мария А. Горбатова², <http://orcid.org/0000-0002-6363-9595>

¹ Қоғамдық Денсаулық сақтау Ұлттық Институты, Осло қ., Норвегия;

² Солтүстік Мемлекеттік Медициналық Университеті, Архангельск қ., Ресей;

³ Х.А. Ясави ат. Халықаралық Қазақ – Түрік Университеті, Туркестан, Қазақстан;

⁴ Солтүстік - Шығыс Федералдық Университеті, Якутск қ., Ресей;

⁵ И. И. Мечников атынд. Солтүстік – Батыс мемлекеттік медициналық университеті, Санкт-Петербург қ., Ресей.

Осы жұмыста бастаушы зерттеушілер үшін Statistica және SPSS бағдарламалық қамсыздандыруды қолданумен мәліметтерді алдын ала статистикалық өңдеудің негізгі әдістері көрсетілген. Биомедициналық зерттеулерде кездесетін әртүрлі мәліметтер түрлері суреттелген, сандық белгілерді қалыпты бөлу параметрлері, сандық нәтижелердің графикалық көріністері әдістері және бөлу сияқты тексеру әдістері. Осы мақала сипаттаушы статистика және мәліметтерді таратуды тексеру туралы жалпы мәліметтер беруге бейімделген және статистика және клиникалық эпидемиология бойынша мамандандырылған әдебиетті оқуды ауыстырмайды.

Негізгі сөздер: Statistica, SPSS, суреттеме статистика, тарату, Қазақстан.

Библиографическая ссылка:

Гржибовский А. М., Иванов С. В., Горбатова М. А. Описательная статистика с использованием пакетов статистических программ Statistica и SPSS // Наука и Здравоохранение. 2016. №1. С. 7-23.

Grijbovski A. M., Ivanov S. V., Gorbatoва M. A. Descriptive statistics using Statistica and SPSS software. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2016, 1, pp. 7-23.

Гржибовский А. М., Иванов С. В., Горбатова М. А. Statistica және SPSS статистикалық бағдарламалары пакеттерін қолданумен сипаттаушы статистика // Ғылым және Денсаулық сақтау. 2016. №1. Б. 7-23.

Настоящая статья открывает серию публикаций, посвященных статистическому анализу данных биомедицинских исследований. В последующих статьях будут последовательно представлены базисные статистические методы, наиболее часто используемые для анализа результатов биомедицинских исследований, а также особенности их применения и интерпретация полученных результатов.

По мнению авторов, настоящая серия статей направлена на решение трех задач:

1. Формирование у начинающего исследователя базисных представлений о статистическом анализе данных.

2. Обучение начинающего исследователя основным практическим навыкам статистической обработки результатов собственных исследований с использованием современного программного обеспечения.

3. Предупреждение типичных ошибок, возникающих в процессе статистической обработки данных и приводящих к некорректной интерпретации результатов анализа.

Для наиболее полного понимания представленного материала, авторы настоящей статьи настоятельно рекомендуют читателю предварительно ознакомиться с литературой по эпидемиологии и получить

достаточные базисные представления о планировании, организации и проведении эпидемиологических исследований [14, 22, 24]. Практические аспекты организации и анализа результатов различных типов научных исследований в здравоохранении (одномоментных, когортных, экологических, экспериментальных исследований и «случай-контроль») представлены в серии статей, опубликованных в журнале «Наука и Здравоохранение» в 2015 году [7, 8, 9, 10, 11].

Настоящая статья посвящена предварительной статистической обработке данных, представлению различных типов данных, описательной статистике и проверке типа распределений изучаемых признаков.

В результате любого исследования ученый получает некий набор сведений, которые необходимо соответствующим образом классифицировать и систематизировать для последующего анализа. Определение типов собственных данных является необходимым условием их корректного представления и правильного выбора методов статистической обработки [1, 18].

Все признаки, характеризующие какое-либо явление или процесс, прежде всего, делятся на количественные (числовые) и качественные (категориальные) [6, 15, 17, 21]. Количественные признаки названы так потому, что их количественная мера точно определена, и они, в свою очередь, делятся на непрерывные, которые могут принимать любое значение (например, масса тела, артериальное давление, уровень гемоглобина, концентрация вещества и др.) и дискретные, которые могут выражаться только целыми числами (например, количество рецидивов заболевания, количество беременностей, число заболевших и др.). Качественные признаки не имеют непосредственной количественной меры и характеризуют принадлежность явления к какой-либо определенной категории. Качественные признаки подразделяются на номинальные, то есть просто имеющие определенное наименование (например, номер палаты, семейное положение, основная или контрольная группа наблюдения), и порядковые (ранговые, ординальные), которые отличаются от номинальных тем, что могут быть размещены в порядке возрастания или

убывания (например, состояние пациента, стадия сердечной недостаточности, степень угнетения сознания). Хотя порядковые переменные, как и количественные, могут быть представлены в числовом выражении, но они принципиально отличаются от количественных тем, что не могут быть подвергнуты арифметическим операциям. В случае порядковых данных невозможно оценить степень различий между находящимися рядом уровнями переменной. Например, если между значениями количества гемоглобина 105 г/л, 106 г/л и 107 г/л разница, в любом случае, составляет 1 г/л, то в случае оценки тяжести состояния пациента невозможно утверждать, что состояние средней тяжести в той же степени хуже удовлетворительного состояния, в которой состояние средней тяжести хуже тяжелого состояния.

Выделяют также дихотомические (бинарные) переменные, которые относятся к качественным переменным, и могут принимать только два взаимоисключающих значения (например, мужской или женский пол, больной или здоровый участник исследования, благоприятный или неблагоприятный исход заболевания, наличие или отсутствие осложнений). Типы признаков (переменных) в биомедицинских исследованиях представлены на рисунке 1.

Помимо исходных данных, полученных непосредственно в результате исследования, для статистического анализа используются также и производные данные (например, индекс массы тела, рассчитываемый на основании измеренных роста и массы тела).

Качественные данные, как правило, могут быть представлены в виде частот или процентов (например, доля мужчин в изучаемой выборке пациентов, распространенность инфекционного заболевания в городе, частота встречаемости различных степеней ожирения и др.). Следует учесть, что если исследование является выборочным, то есть не включает в себя всю популяцию, на которую будут транслированы результаты выборочного исследования, корректно представлять данные о частотах в виде 95% доверительных интервалов, алгоритмы расчета которых представлены в соответствующей литературе [5, 25].

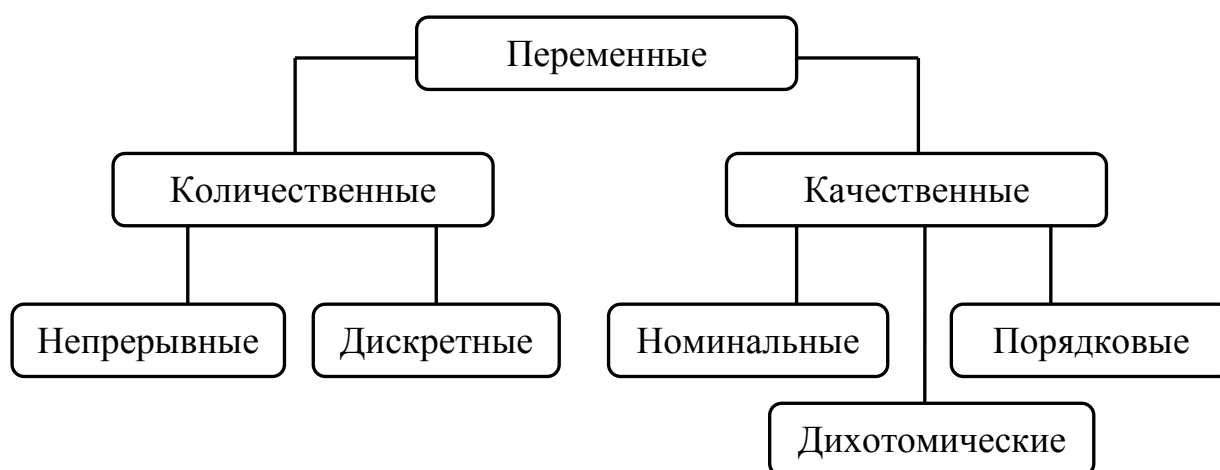


Рис. 1. Типы признаков (переменных) в биомедицинских исследованиях.

Представление количественных данных производится с использованием определенного набора показателей. При этом для того, чтобы выбрать, какие именно показатели использовать для корректного представления количественных данных, и какие статистические критерии использовать, важно предварительно оценить тип распределения изучаемого признака.

Существуют различные виды распределений (биномиальное распределение, распределение Пуассона и др.), с которыми можно ознакомиться в специальной литературе по статистике [17]. В настоящей статье будет

рассмотрено нормальное (гауссовское) распределение количественных данных.

Нормальное распределение характерно для многих биологических явлений и процессов. Например, нормальному распределению подчиняется рост людей в генеральной совокупности: у наибольшего количества людей рост средний, меньшее количество людей обладает ростом ниже и выше среднего, и совсем немного людей имеют очень низкий или очень высокий рост.

Примеры нормального и некоторые виды отличного от нормального распределений представлены на рисунке 2.

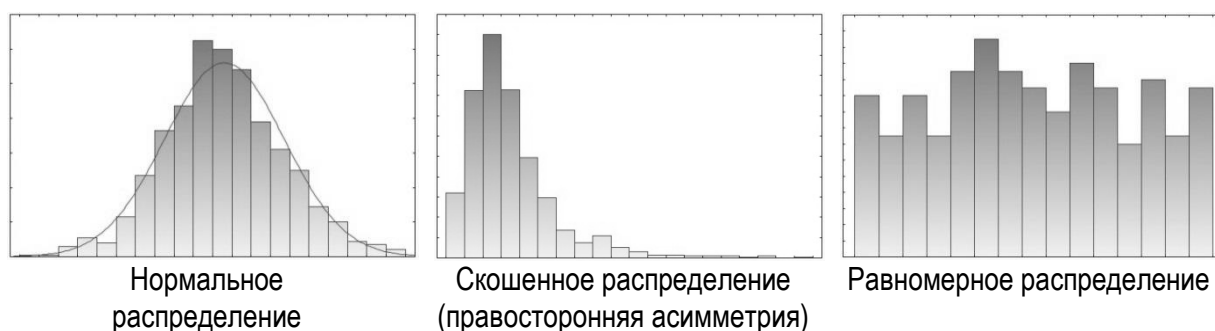


Рис. 2. Некоторые виды распределения количественных признаков.

Нормальное распределение может быть математически описано с использованием двух параметров – среднего арифметического значения и стандартного отклонения.

Среднее арифметическое значение рассчитывается по формуле:

$$M = \frac{X_1 + X_2 + X_3 + \dots + X_i}{n}$$

где

$X_1 \dots X_i$ – значения количественного признака,
 n – количество наблюдений.

Стандартное отклонение рассчитывается путем извлечения квадратного корня из значения дисперсии (σ^2), которая является

суммой квадратов отклонений значений признака от среднего арифметического значения:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\left(\frac{(X_1 - M)^2 + (X_2 - M)^2 + (X_3 - M)^2 + \dots + (X_i - M)^2}{n} \right)}$$

Стандартное отклонение является мерой разброса значений вокруг среднего арифметического: в область от $M - \sigma$ до $M + \sigma$ попадает 68% всех значений количественного показателя, в область от $M - 1,96\sigma$ до $M + 1,96\sigma$ попадает 95% значений, а область от $M - 2,58\sigma$ до $M + 2,58\sigma$ включает в себя 99% всех значений признака.

Для графического представления значений количественного признака, имеющего нормальное распределение, можно использовать столбчатую диаграмму, с указанием границ значения стандартного отклонения (рис. 3, слева). В то же время, во многих журналах рекомендуется представлять не одно, а два стандартных отклонения на столбчатой диаграмме, то есть 95% доверительный интервал.

Если же распределение отличается от нормального, то для описания данных

целесообразно использовать медиану и квартили. Медиана – значение признака, которое делит распределение пополам – половина значений признака располагается слева, а другая половина – справа от медианы. 25-й процентиль (или первый квартиль) отсекает четверть самых низких значений признака, а 75-й процентиль (или третий квартиль) – четверть самых высоких значений (медиана является 50-м процентилем, или вторым квартилем). Таким образом, между 1-м и 3-м квартилями располагается 50% значений признака.

Для графического представления значений признака, имеющего отличное от нормального распределение, можно использовать квартильную диаграмму, представленную на рисунке 3 (диаграмма типа «ящик с усами», англ. «box-whisker plot»).

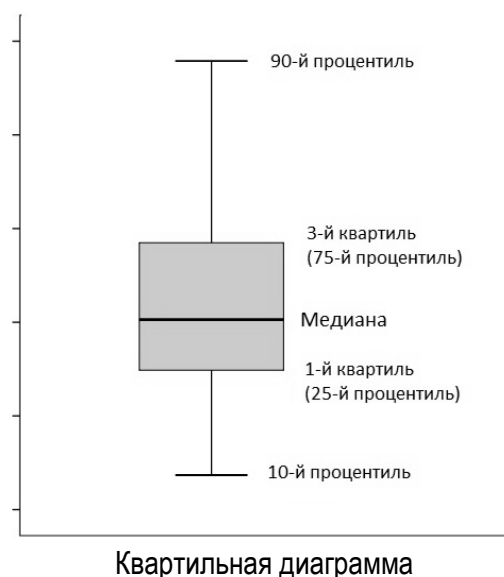
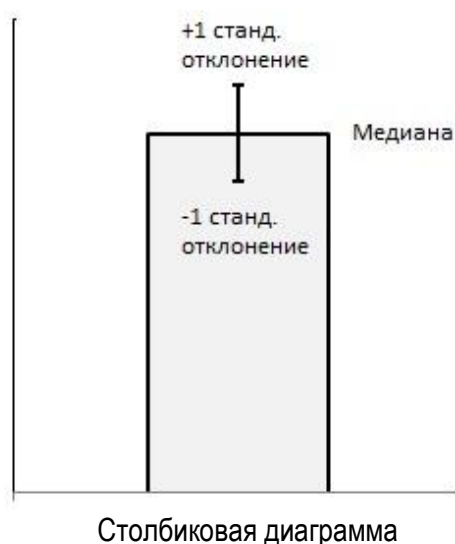


Рис. 3. Графическое представление количественных данных: при нормальном распределении (столбчатая диаграмма) и при распределении, отличном от нормального (квартильная диаграмма).

Более подробно о распределении признаков и математическом описании данных можно узнать из литературы по статистике [4, 15, 17].

Наличие нормального распределения изучаемого признака расширяет возможности исследователя в отношении использования методов статистического анализа данных и

повышает чувствительность статистических критериев. Поэтому, если фактическое распределение похоже на скошенное влево или вправо нормальное распределение, во многих случаях рекомендуется приблизить распределение к нормальному с помощью математического преобразования (разумеется, если это возможно). Например, при правосторонней асимметрии (рисунок 2) извлекают квадратный корень из значений, проводят логарифмическое преобразование с использованием натурального или десятичного логарифма или гармоническое преобразование по формуле $x' = -1/x$. При левосторонней симметрии данные возводят во вторую или третью степень [6]. Подбор того или иного способа «нормализации» распределения, как правило, проводят эмпирически.

Сопоставление фактического распределения количественных данных и нормального распределения проводится следующими взаимодополняющими методами:

1. Построение гистограммы распределения: колоколообразная форма гистограммы свидетельствует в пользу нормального распределения.

2. Построение квантильной диаграммы: расположение точек на прямой линии свидетельствует в пользу нормального распределения.

3. Проверка соответствия имеющегося распределения нормальному с помощью критерия Колмогорова-Смирнова с поправкой Lilliefors (для больших выборок) и критерия Шапиро-Уилка (для небольших выборок). Если достигнутое в результате расчета значения критериев значение статистической значимости превышает 0,05 ($p > 0,05$) то фактическое распределение можно формально считать не отличающимся от нормального.

С целью приобретения читателем практических навыков представления описательных статистических данных и проверки распределения, в качестве практического примера будет рассмотрен фрагмент данных, которые были собраны в процессе крупного исследования, направленного на изучение метаболического синдрома и его детерминант в условиях

неблагополучной социально-экологической ситуации в Южном Казахстане [12, 13, 16]. В ходе данного исследования получены значения массы тела и уровня холестерина крови 298 пациентов. Масса тела и уровень холестерина крови являются непрерывными количественными признаками, вариационные ряды которых будут в проанализированы с использованием программного обеспечения Statistica 10 [19, 2] и SPSS 20 [3]. Выбор данных пакетов статистических программ обусловлен тем, что они являются одними из наиболее удобных в использовании начинающими исследователями инструментами анализа данных, не требуют специального образования и приобрели заслуженную популярность среди исследователей стран СНГ, работающих в области медицины. Важной причиной популярности данного программного обеспечения является, в том числе и, наличие русифицированных версий обеих программ. Демо-версии программ Statistica и SPSS можно загрузить с официальных сайтов разработчиков (www.statsoft.com и <http://www-01.ibm.com/software/analytics/spss/> соответственно).

Следует отметить, что представленные ниже алгоритмы действий являются только инструментом анализа данных, так как корректная интерпретация получаемых результатов требует наличия базисных знаний в области биомедицинской статистики, которые можно приобрести путем изучения соответствующей литературы [4, 15, 17, 23, 25].

Анализ вариационного ряда с использованием программы Statistica 10.

Для начала работы необходимо открыть файл 1_Bodymass.sta, который необходимо загрузить с сайта журнала «Наука и Здравоохранение». В данном файле представлен один вариационный ряд значений массы тела пациентов.

Далее следует выполнить ряд действий, целью которых является настройка вывода программой необходимых данных.

Выберем меню «Statistics» и войдем в раздел «Basic Statistics/Tables» (рисунок 4), где выберем раздел «Descriptive Statistics», подтвердив выбор нажатием на кнопку ОК.

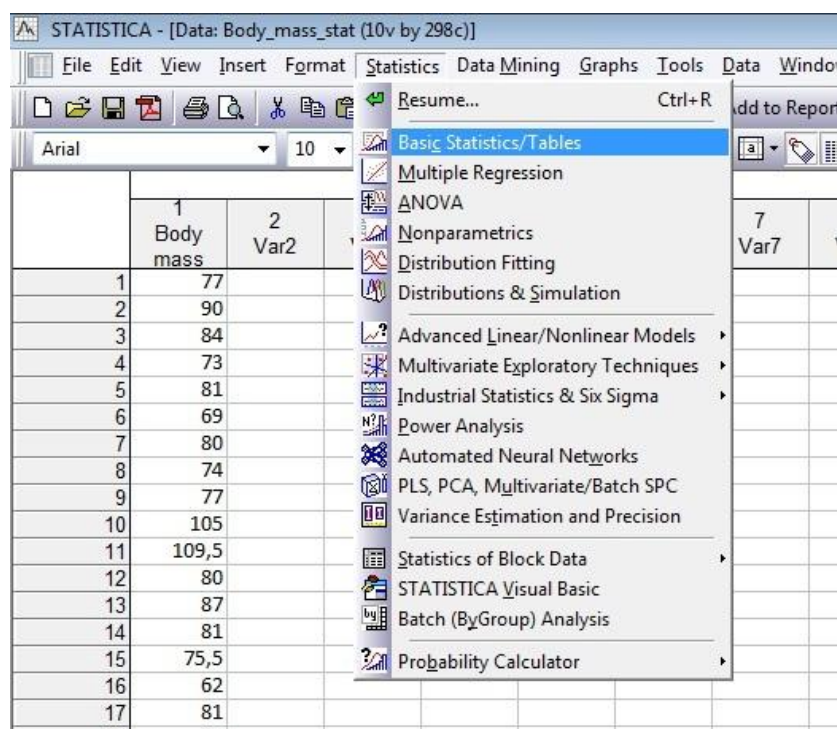


Рис. 4. Меню «Statistics» программы Statistica 10.

Далее в появившемся окне «Descriptive Statistics» нажмем на кнопку «Variables» (в левом верхнем углу окна) и выберем для

анализа вариационный ряд «Body mass», как показано на рисунке 5, после чего подтвердим выбор нажатием на кнопку OK.

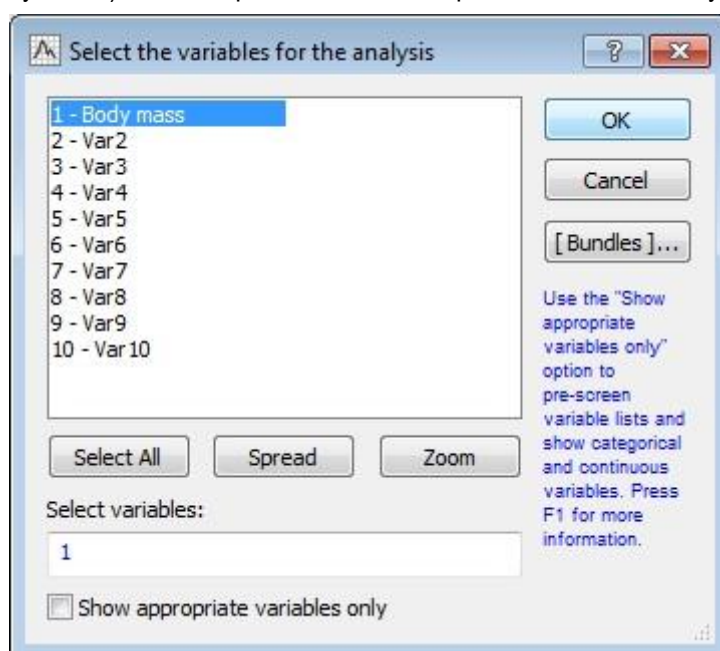


Рис. 5. Выбор вариационного ряда для последующего анализа в программе Statistica 10.

Далее в окне «Descriptive Statistics» настроим процесс проверки распределения: для этого нажмем на вкладку «Normality» и проставим галочки, напротив позиций, соответствующих расчету значений критериев Колмогорова-Смирнова (Kolmogorov-Smirnov & Lilliefors test for normality) и Шапиро-Уилка

(Shapiro-Wilk's W test), как это показано на рисунке 6. Здесь же можно указать количество интервалов (столбиков) гистограммы, на которые будет разбита выборка (Number of intervals), которое по умолчанию задано равным 10-ти. Количество интервалов, на которые необходимо разбить выборку для

построения гистограммы распределения, можно рассчитать с помощью формулы Стерджеса: $m = 1 + 3,32 \times \lg n$, где n – количество наблюдений в выборке [23]. В нашем случае количество наблюдений равно

298, и рассчитанное значение количества интервалов равно 9, то есть оно всего лишь на единицу отличается от количества интервалов, указанному программой по умолчанию.

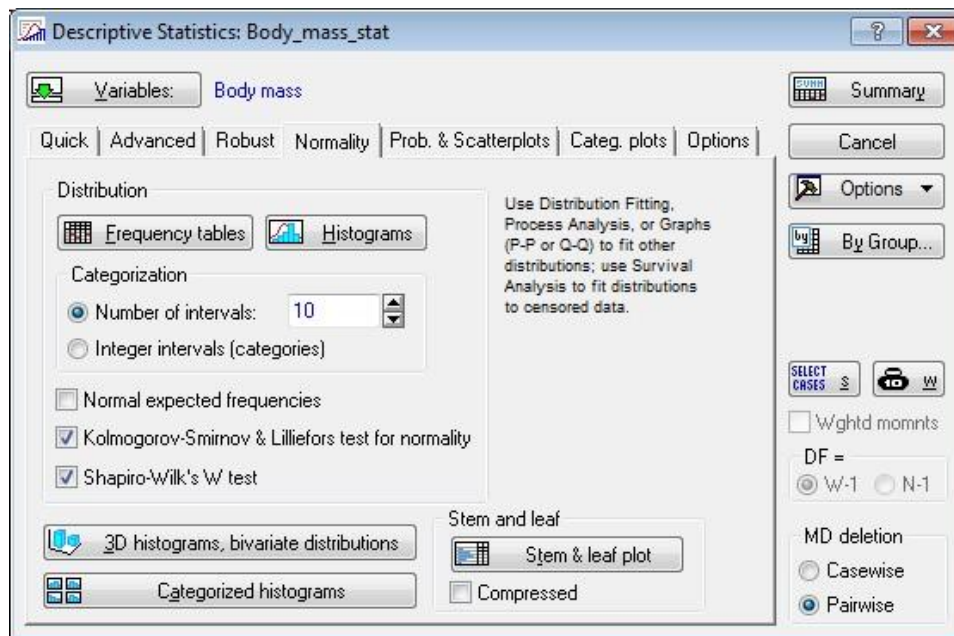


Рис. 6. Вкладка «Normality» окна «Descriptive Statistics» программы Statistica 10.

Далее нажмем на вкладку «Advanced» и, как указано на рисунке 7, проставим галочки, соответствующие выводу показателей описательной статистики: среднего

арифметического значения (Mean), медианы (Median), стандартного отклонения (Standard Deviation), степени асимметрии (Skewness) и эксцесса (Kurtosis).

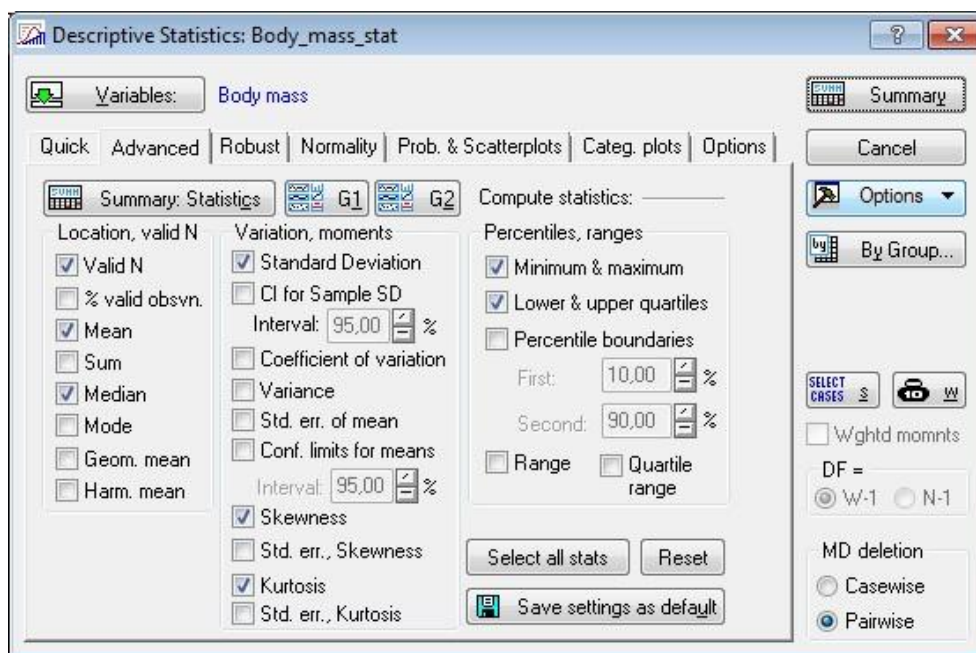


Рис. 7. Вкладка «Advanced» окна «Descriptive Statistics» программы Statistica 10.

Также целесообразно указать минимум и максимум (Minimum & maximum) и верхний и нижний квартили (Lower & upper quartiles). При необходимости можно выбрать и другие показатели описательной статистики (моду, дисперсию, коэффициент вариации, стандартную ошибку среднего арифметического, межквартильный интервал и другие).

Для запуска анализа в той же вкладке «Advanced» нажмем на кнопку G1, которая активизирует вывод графиков, результатов расчетов показателей и значений статистических критериев.

Результаты анализа представлены на рисунке 8. В левом верхнем углу окна вывода данных расположена гистограмма распределения, имеющая колоколообразную форму. Выше гистограммы размещены результаты расчета значений и статистической значимости критериев проверки на соответствие фактического распределения нормальному: уровень статистической значимости (p) для критерия

Колмогорова-Смирнова (K-S), который составил $> 0,20$. Для критерия Колмогорова-Смирнова с поправкой Lilliefors он также составил $> 0,20$, для критерия Шапиро-Уилка (Shapiro-Wilk W) – 0,053. В правом верхнем углу окна вывода данных представлена квантильная диаграмма (Normal P-plot), на которой видно, что почти все точки располагаются на одной линии. В левом нижнем углу окна представлены значения показателей описательной статистики, которые были выбраны на этапе настройки анализа данных. В правом нижнем углу окна представлена диаграмма по типу «box plot», которую не следует путать с квартильной диаграммой, так в представленной диаграмме центральное значение занимает среднее арифметическое, а не медиана, а верхним и нижним значениями «ящика» являются соответственно, значения $M - \sigma$ и $M + \sigma$ (выбор различных вариантов вывода диаграмм «box plot» можно настроить во вкладке «Options» окна «Descriptive Statistics»).

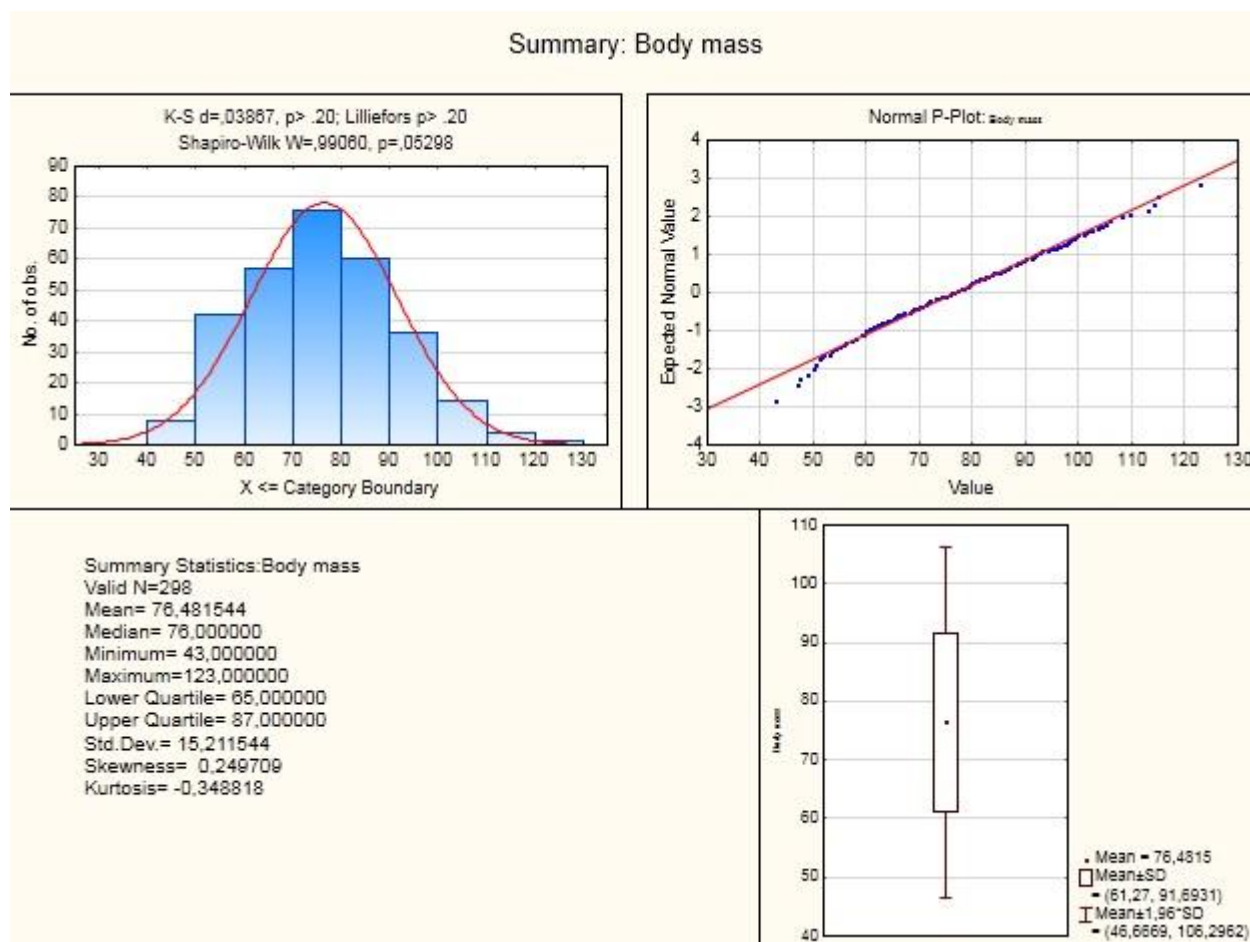


Рис 8. Результаты анализа переменной «Body mass» в программе Statistica 10.

Таким образом, по результатам представленных программой графиков и расчетов следует заключить, что распределение имеет форму, близкую к колоколообразной, статистическая значимость критериев проверки распределения на нормальности имеет значение, превышающее 0,05, почти все точки на квантильной диаграмме располагаются на прямой линии, значение среднего арифметического и медианы близки друг к другу, а значения асимметрии и эксцесса не превышают значение 1 (либо не ниже значения - 1), что в совокупности свидетельствует в пользу нормальности распределения массы тела пациентов – участников исследования.

Значение массы тела пациентов можно представить в виде среднего арифметического и стандартного отклонения: $M=76,5$, ($SD=15,2$) кг. Данное математическое выражение имеет

следующий практический смысл: в границы от 61,3 до 91,7 кг ($M \pm SD$) попадает масса тела 68% пациентов, а в границы от 46,1 до 106,9 ($M \pm 1,96*SD$) попадает масса тела 95% пациентов.

Анализ вариационного ряда с использованием программы SPSS 20.

Для начала работы необходимо открыть файл 1_Cholesterol.sav, который потребуется загрузить с сайта журнала «Наука и здравоохранение». В данном файле представлен один вариационный ряд значений уровня холестерина крови пациентов.

Как и в процессе работы с программой Statistica, прежде всего потребуется подготовить параметры вывода данных. Для этого войдем в меню «Analyze», раздел «Descriptive Statistics», и выберем подраздел «Explore», как показано на рисунке 9.

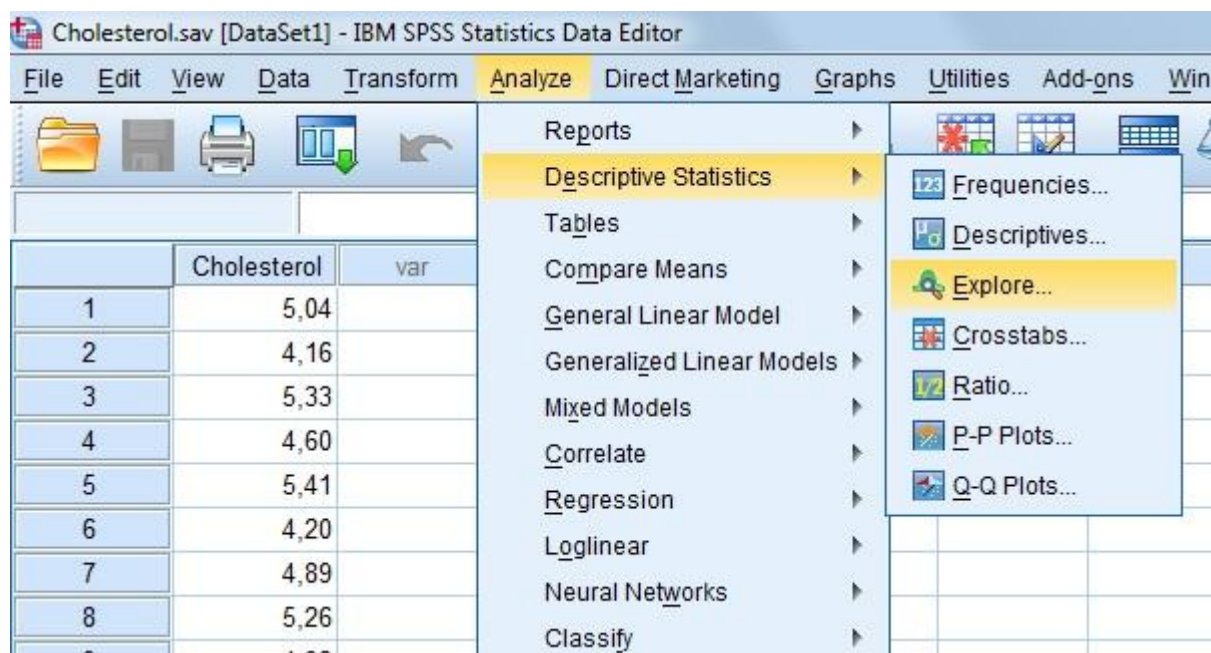


Рис. 9. Меню «Analyze» программы SPSS 20.

В открывшемся окне «Исследовать» с помощью стрелочки между полями перенесем переменную «Cholesterol» из левого поля в поле «Dependent List» (рисунок 10), после чего настроим ряд параметров вывода данных. Для этого нажмем на кнопку «Statistics», и в открывшемся окне отметим позиции «Descriptives» и «Percentiles», подтвердив выбор нажатием кнопки «Continue» (рисунок

11), после чего программа снова вернется к окну «Explore».

Далее нажмем на кнопку «Plots», и в открывшемся окне «Explore: Plots» отметим позиции «Factor levels together» (поле «Boxplots»), «Histogram» (поле «Descriptive») и «Normality plots with tests», подтвердив выбор нажатием кнопки «Continue» (рисунок 12).

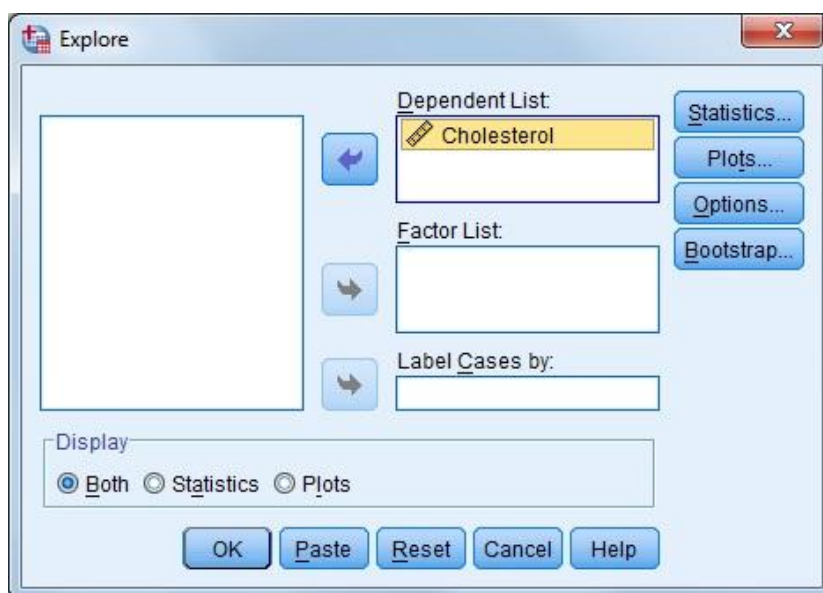


Рис. 10. Раздел «Explore» программы SPSS 20.

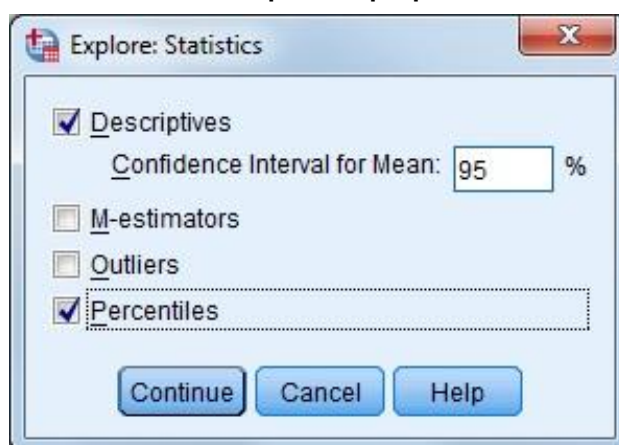


Рис. 11. Окно «Statistics» раздела «Explore» программы SPSS 20.

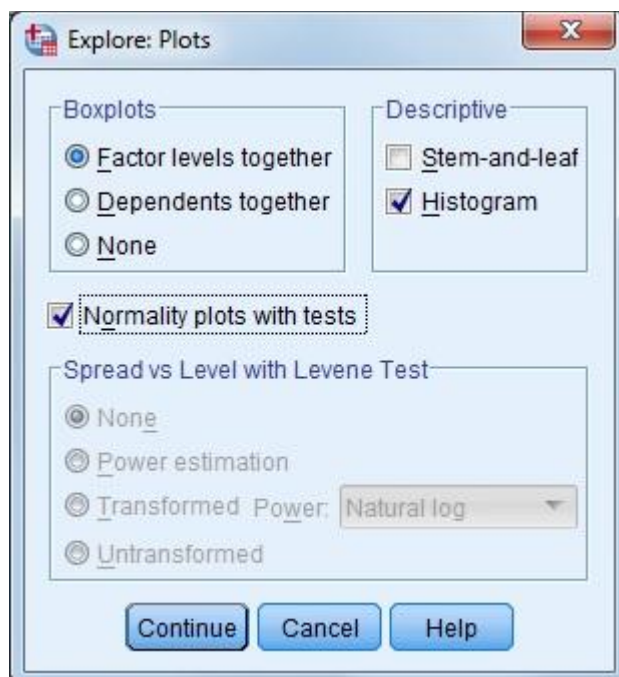


Рис. 12. Окно «Explore: Plots» раздела «Explore» программы SPSS 20.

Программа снова вернется к окну «Analyze» (Рисунок 10), в котором нажатием на кнопку «ОК» запустим анализ данных.

Рассчитанные параметры описательной статистики для уровня холестерина крови пациентов программа представляет в виде таблиц и графиков.

В таблице 1 представлены среднее арифметическое и его 95% доверительный интервал, медиана, дисперсия, стандартное отклонение, асимметрия и эксцесс, а также ряд других параметров переменной «Cholesterol».

В таблице 2 представлены процентиля (квартили) для переменной «Cholesterol». В

данной таблице следует обратить внимание на 25-й процентиль, соответствующий 1-му квартилю, и 75-й процентиль, соответствующий 3-му квартилю (50-й процентиль соответствует медиане).

Результаты проверки распределения изучаемого признака с помощью статистических критериев представлены в таблице 3. Уровень статистической значимости («Sig.») оказался меньше 0,001 как для рассчитанного критерия Колмогорова-Смирнова с поправкой Lilliefors, так и для значения критерия Шапиро-Уилка.

Таблица 1.

Описательная статистика для переменной «Cholesterol»

			Statistic	Std. Error
Cholesterol	Mean		5,0548	,06155
	95% Confidence Interval for Mean	Lower Bound	4,9337	
		Upper Bound	5,1759	
	5% Trimmed Mean		4,9614	
	Median		4,8600	
	Variance		1,129	
	Std. Deviation		1,06258	
	Minimum		3,18	
	Maximum		10,79	
	Range		7,61	
	Interquartile Range		1,19	
	Skewness		1,735	,141
	Kurtosis		4,760	,281

Таблица 2.

Процентиля (квартили) для переменной «Cholesterol»

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	Cholesterol	3,8195	4,0870	4,3000	4,8600	5,4925	6,4240	6,9615

Таблица 3.

Результаты проверки распределения переменной «Cholesterol» с помощью статистических критериев

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Cholesterol	,127	298	,000	,874	298	,000

a. Lilliefors Significance Correction

Гистограмма распределения, сформированная программой SPSS, представлена на рисунке 13. На рисунке видно, что данная гистограмма не имеет равносостроенной

колоколообразной формы, характерной для нормального распределения.

Квантильная диаграмма распределения представлена на Рисунке 14: видно, что точки отклоняются от прямой линии.

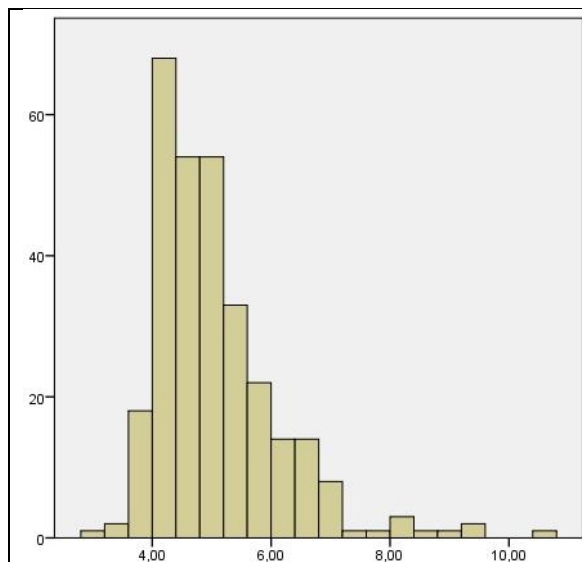


Рис. 13. Гистограмма распределения переменной «Cholesterol», сформированная программой SPSS 20.

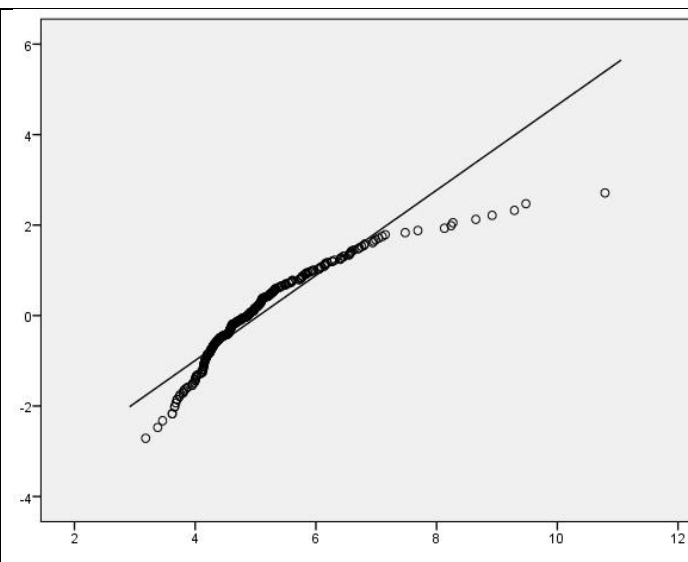


Рис. 14. Квантильная диаграмма переменной «Cholesterol», сформированная программой SPSS 20.

Квартильная диаграмма уровня холестерина представлена на рисунке 15. Точки, расположенные выше верхней границы

диаграммы, представляют собой значения, превышающие уровень 90-го перцентиля.

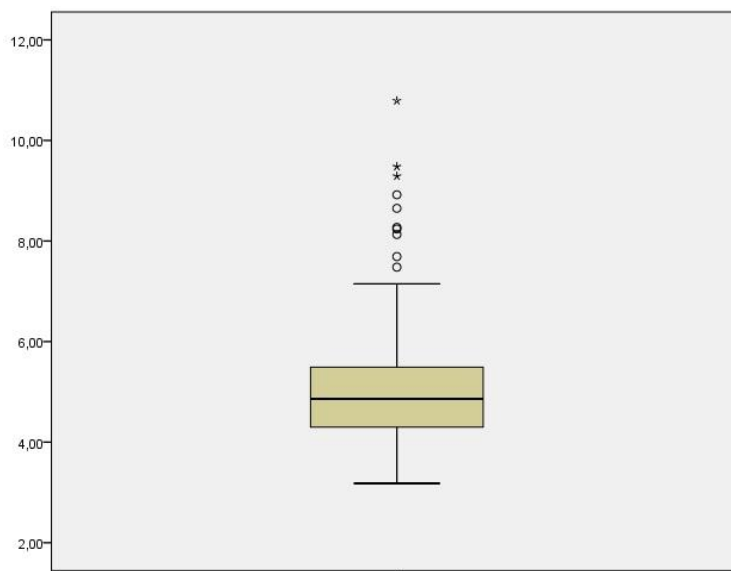


Рис. 15. Квартильная диаграмма переменной «Cholesterol», сформированная программой SPSS 20.

Таким образом, по результатам представленных программой SPSS данных видно, что распределение имеет форму,

отличную от формы нормального распределения, статистическая значимость критериев проверки распределения на

нормальность имеет значение меньше 0,001, точки на квантильной диаграмме отклоняются от прямой линии, что в совокупности свидетельствует в пользу того, что фактическое распределение значений уровня холестерина существенно отличается от нормального распределения.

В связи с отличием распределения уровня холестерина от нормального, использовать для представления данных среднее арифметическое значение и стандартное отклонение будет некорректным. Значение уровня холестерина пациентов следует представить в виде медианы и квартилей (Me (Q_1 ; Q_3)): 4,86 (4,30; 5,49) ммоль/л. Данное математическое выражение имеет следующий практический смысл: в границы уровня холестерина от 4,30 до 5,49 ммоль/л попадают 50% пациентов, у 25% уровень холестерина будет выше 5,49 ммоль/л, у 25% – ниже 4,30 ммоль/л.

Результаты проверки распределения на нормальность следует всегда учитывать в совокупности. Так, результаты проверки с помощью одних только статистических критериев Колмогорова-Смирнова и Шапиро-Уилка следует интерпретировать с осторожностью, так как они чувствительны к объему выборок: вероятность получить статистически значимое отличие фактического распределения от нормального ($p < 0,05$) при количестве наблюдений $n = 1000$ значительно выше, чем, например, при $n = 30$. Существуют следующие практические рекомендации: при количестве наблюдений от 30 до 100, если критерии проверки на нормальность покажут отличие распределения от нормального следует считать его отличным от нормального только в том случае, если гистограмма распределения, квантильная диаграмма и значения асимметрии и эксцесса не свидетельствуют об обратном. Для условного соответствия фактического распределения нормальному допускается нахождение показателей асимметрии и эксцесса в пределах от -1 до 1 (некоторые исследователи сужают допустимые границы до значений от -0,5 до 0,5). Если количество наблюдений превышает 100, и статистическая значимость критериев проверки распределения на нормальность превышает 0,05, то

распределение считают нормальным, если графики и значения асимметрии и эксцесса не говорят об обратном.

Но возможны ситуации, когда для представления данных используются параметры нормального распределения, хотя фактические данные распределены иначе. Примером такой ситуации являются результаты исследования, посвященного проблеме кариеса у детей и подростков, выполненные по методике ВОЗ [26, 27, 28]. В вышеупомянутых работах описаны значения индекса КПУ («кариес-пломбы-удаленные»), который рассчитывается как сумма кариозных зубов, пломбированных зубов и зубов, удаленных по поводу кариеса или его осложнений. В данных публикациях для представления описательной статистики индекса КПУ для различных категорий детского и подросткового населения было использовано среднее арифметическое значение и его 95% доверительный интервал, а для статистического сравнения различных категорий детского и подросткового населения – непараметрический критерий Манна-Уитни. В данном случае использование непараметрического критерия обусловлено обнаруженным отличием фактического распределения значений индекса КПУ от нормального, а представление данных в виде среднего арифметического значения обусловлено требованиями ВОЗ описывать индекс КПУ как признак, имеющий нормальное распределение. Согласно центральной предельной теореме, при достаточно большом размере выборки средние значения имеют нормальное распределение, поэтому при больших выборках допускается использование параметрических методов статистики даже при отклонениях фактического распределения от нормального. Кроме того, если известно, что в генеральной совокупности распределение признака является нормальным, то данные можно представлять с помощью параметров нормального распределения и анализировать с помощью параметрических методов статистики независимо от распределения признака в выборке.

В следующем выпуске будет представлен алгоритм статистического сравнения

количественных данных двух независимых выборок с использованием программ Statistica и SPSS.

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. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: для профессионалов. СПб. : Питер, 2003. 688 с.
3. Бююль А., Цефель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. СПб. : ООО «ДиаСофтЮП», 2005. 608 с.
4. Гланц С. Медико-биологическая статистика. М. : Практика, 1998. 459 с.
5. Гржибовский А.М. Доверительные интервалы для частот и долей // Экология человека. 2008. №5. С. 57-60.
6. Субботина А.В., Гржибовский А.М. Описательная статистика и проверка нормальности распределения количественных данных // Экология человека. 2014. №2. С. 51-57.
7. Гржибовский А.М., Иванов С.В. Исследования типа «случай-контроль» в здравоохранении // Наука и Здравоохранение. 2015. № 4. С. 5-17.
8. Гржибовский А.М., Иванов С.В. Когортные исследования в здравоохранении // Наука и Здравоохранение. 2015. № 3. С. 5-16.
9. Гржибовский А.М., Иванов С.В. Поперечные (одномоментные) исследования в здравоохранении // Наука и Здравоохранение. 2015. № 2. С. 5-18.
10. Гржибовский А.М., Иванов С.В. Экологические (корреляционные) исследования в здравоохранении // Наука и Здравоохранение. 2015. № 5. С. 5-18.
11. Гржибовский А.М., Иванов С.В. Экспериментальные исследования в здравоохранении // Наука и Здравоохранение. 2015. № 6. С. 5-17.
12. Жунисова М.Б., Шалхарова Ж.С., Шалхарова Ж.Н., Гржибовский А.М. Типы

пищевого поведения и абдоминальное ожирение // Медицина. 2015. №4. С. 92-95.

13. Жунисова М.Б., Шалхарова Ж.С., Шалхарова Ж.Н., Нускабаева Г.О., Садыкова К.Ж., Маденбай К.М., Гржибовский А.М. Психозомоциональный стресс как предиктор типа пищевого поведения в Казахстане // Экология человека. 2015. №5. С. 36-45.
14. Зуева Л.П., Яфаев Р.Х. Эпидемиология : учебник. СПб. : ООО «Издательство Фолиант», 2008. 752 с.
15. Лакин Г.Ф. Биометрия. М. : Высшая школа, 1990. 351 с.
16. 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 электрокардиографии // Экология человека. 2015. №7. С. 58-64.
17. Петри А., Сэбин К. Наглядная статистика в медицине. М. : ГЭОТАР-Мед, 2003. 140 с.
18. Рахымбеков Т.К., Гржибовский А.М. К вопросу о необходимости повышения качества казахстанских научных публикаций для успешной интеграции в международное научное сообщество // Наука и Здравоохранение. 2015. №1. С. 5-11.
19. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. М. : МедиаСфера, 2002. 312 с.
20. 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.
21. Субботина А.В., Гржибовский А.М. Описательная статистика и проверка нормальности распределения количественных данных // Экология человека. 2014. №2. С. 51-57.
22. Флетчер Р. Клиническая эпидемиология. Основы доказательной медицины: пер. с англ. / Р. Флетчер, С. Флетчер, Э. Вагнер. М. : Медиа Сфера, 1998. 352 с.

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

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

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

26. Gorbatoва M.A., Gorbatoва L.N., Grijbovski A.M. Dental caries experience among 15-year-old adolescents in north-west Russia // Int. J. Circumpolar Health. 2011. Vol. 3. P. 232-235.

27. Gorbatoва M.A., Gorbatoва L.N., Pastbin M.U., Grijbovski A.M. Urban-rural differences in dental caries experience among 6-year-old children in the Russian north // Rural Remote Health. 2012. Vol. 12. P. 1999.

28. Gorbatoва M.A., Grijbovski A.M., Gorbatoва L.N., Honkala E. Dental caries experience among 12-year-old children in Northwest Russia // Community Dent. Health. 2012. Vol. 1. P. 20-24.

References:

1. Akanov A.A., Turdalieva B.S., Izenkova A.K., Ramazanova M.A., Abdramova, Grijbovski 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. 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.

3. 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]

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

5. Grijbovski A.M. Doveritel'ny einterval dlya chastot i dolei [Confidence intervals for

proportions]. *Ekologiya cheloveka* [Human Ecology]. 2008. No.5. PP. 57-60. [in Russian].

6. Subbotina A.V., Grijbovski A.M. Opisatel'naya statistika i proverka normalnosti kolichestvennykh dannyykh [Descriptive statistics and normality testing for quantitative data]. *Ekologiya cheloveka* [Human Ecology]. 2014. No.2. PP. 51-57. [in Russian].

7. Grijbovski A.M., Ivanov S.V. Issledovaniya tipa sluchay-kontrol v zdavoohranenii [Case-control studies in health sciences]. *Nauka i Zdavoohranenie* [Science & Healthcare]. 2015, 2, pp. 5-17. [in Kazakhstan].

8. Grijbovski A.M., Ivanov S.V. Poperechnie (odnomomentnie) issledovaniya v zdavoohranenii [Cross-sectional studies in health sciences]. *Nauka i Zdavoohranenie* [Science & Healthcare]. 2015, No2, PP. 5-18.

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

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

11. Grijbovski A.M., Ivanov S.V. Eksperimentalnie issledovaniya v zdavoohranenii [Experimental studies in health sciences]. *Nauka i Zdavoohranenie* [Science & Healthcare]. 2015, 6. pp. 5-17. [in Kazakhstan].

12. 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]. *Meditsina* [Medicine]. 2015. No.4. PP. 92-95.

13. 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]

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

15. Lakin G.F. Biometria [Biometrics]. M. : Viscshaya shkola. 1990. P. 351. [in Russian]
16. Madenbay K.M., Shalkarova Zh.S., Shalkarova Zh. N., Zhunisova M.B., Sadykova K. Zh., Nuskabayeva G.O., Grijbovski A.M. Otsenka svyazi mezhdu ploshadyu podkojnoj jirovoy tkani i pokazatelyami electroneyromiografii [Assesment of the relationship between subcutaneous fat tissue an results of electromyoneurography]. *Ekologiya cheloveka* [Human Ecology]. 2015. 7. PP. 58-64. [in Russian]
17. Petri A., Sebin K. Naglyadnaya statistika v meditsine [Demonstrative statistics in medicine]. M. : GEAOTAR-Med, 2003. p. 140. [in Russian]
18. Rakhypbekov T.K., Grijbovski A.M. K voprosu o nedhodimosti povisheniya kachestva kazakhstanskikh nauchih publicatsiy dlya uspeshnoy 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 Zdravoohranenie* [Science & Healthcare]. 2015. No.1. PP. 5-11.
19. Rebrova O. Yu. Statisticheskiy analiz meditsinskih daanikh. Primenenie paketa prikladnikh program STATISTICA [Statistical analysis of medical data. Using of STATISTICA software]. M. : MediaSphera, 2002. P. 312.
20. 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 - demograficheskkiye 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. P. 58-64. [in Russian]
21. Subbotina A.V., Grijbovski A.M. Opisatel'naya statistika i proverka normal'nosti raspredeleniya kolichestvennih dannikh [Descriptive statistics and normality testing for quantitative data]. *Ekologiya cheloveka* [Human Ecology]. 2014. No.2. PP. 51-57. [in Russian]
22. Fletcher R. et al. Klinicheskaya ehpidemiologiya. 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].
23. Yunkerov V.I., Grigoryev S.G. Matematiko-statisticheskaya obrabotka daanikh meditsinskih issledovaniy [Mathematical and statistical analysis of medical research data]. SPb : VMedA, 2002. P. 266. [in Russian]
24. Beaglehole R., Bonita R. Basic epidemiology. World Health Organization, Geneva, 1993.
25. Cleopas T.J. et al. Statistics Applied to Clinical Trials. 4th ed. Springer, 2009.
26. Gorbatoва M.A., Gorbatoва L.N., Grijbovski A.M. Dental caries experience among 15-year-old adolescents in north-west Russia. *Int. J. Circumpolar Health*. 2011. Vol. 3. P. 232-235.
27. Gorbatoва M.A., Gorbatoва L.N., Pastbin M.U., Grijbovski A.M. Urban-rural differences in dental caries experience among 6-year-old children in the Russian north. *Rural Remote Health*. 2012. Vol. 12. P. 1999.
28. Gorbatoва M.A., Grijbovski A.M., Gorbatoва L.N., Honkala E. Dental caries experience among 12-year-old children in Northwest Russia. *Community Dent. Health*. 2012. Vol. 1. P. 20-24.

Контактная информация:

Грибовский Андрей Мечиславович – доктор медицины, Старший советник Национального Института Общественного Здравоохранения, г. Осло, Норвегия; Руководитель отдела международных программ и инновационного развития ЦНИЛ СГМУ, г. Архангельск, Россия; Профессор кафедры общественного здоровья и здравоохранения Северо-Восточного Федерального Университета, г. Якутск, Россия; Профессор, почетный доктор Международного Казахско-Турецкого Университета г. Туркестан, Казахстан; Почетный профессор ГМУ г. Семей, Казахстан.

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

E-mail: Andrej.Grijbovski@gmail.com

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