

# ОПТИМИЗАЦИЯ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ КЛИНИКО- ЛАБОРАТОРНЫХ ДАННЫХ ПРИ ОСТРЫХ ОТРАВЛЕНИЯХ

А.Н. Ельков<sup>1,2</sup>, Ю.С. Гольдфарб<sup>1,2</sup>,  
М.М. Поцхверия<sup>1,2</sup>, А.В. Бадалян<sup>1,2</sup>

<sup>1</sup>ГБУЗ «НИИ СП им. Н.В. Склифосовского  
ДЗМ», 129090, г. Москва, Российская  
Федерация

<sup>2</sup>ФГБОУ ДПО РМАНПО Минздрава России,  
125993, г. Москва, Российская Федерация

**А**вторами разработано специализированное программное приложение, обобщающее длительный опыт анализа клинико-лабораторных данных в клинической токсикологии с использованием средств объектно-ориентированного программирования Visual Basic for Applications фирмы Microsoft в среде Excel. Основным назначением приложения является существенное (в 5–10 раз) ускорение выполнения комплекса наиболее часто применяемых статистических алгоритмов. Приложение вычисляет описательные выборочные характеристики, выявляет отклонения выборочных распределений от нормального закона по нескольким общепринятым критериям, производит статистическое сравнение матриц данных по нескольким критериям и автоматически формирует ряд готовых для размещения в научных текстах таблиц различного назначения в формате Microsoft Word. Представление выходных данных позволяет при необходимости легко верифицировать результаты расчетов путем сравнения их с выходными формами пакета прикладных статистических программ Statistica. Ускорение получения конечного результата достигается: 1 – за счет отсутствия необходимости перемещения данных из Excel в пакет прикладных статистических программ; 2 – в результате автоматизации определения параметров задачи посредством специальной цветовой разметки; 3 – путем объединения ограниченного числа наиболее употребительных статистических методов в один неизменяемый комплекс; 4 – в результате программной реализации статистического сравнения всех возможных сочетаний участвующих в исследовании матриц данных по совокупности показателей как единой операции, не требующей вмешательства пользователя; 5 – благодаря использованию приложением синтаксических преобразований выходных данных и автоматического формирования таблиц посредством заполнения готовых шаблонов результатами расчетов.

**Ключевые слова:** острые отравления, вычислительная статистика, пакеты прикладных статистических программ, Microsoft Excel, Visual Basic for Applications.

Цит.: А.Н. Ельков, Ю.С. Гольдфарб, М.М. Поцхверия, А.В. Бадалян. Оптимизация статистической обработки клинико-лабораторных данных при острых отравлениях. Токсикологический вестник. 2021; 2:14-22.

**Введение.** Высокой эффективностью при работе с клинико-лабораторными данными обладают пакеты прикладных статистических программ (ППСП), содержащие значительное количество

надежных алгоритмов и имеющие удобный интерфейс [1, 2, 3]. Однако, при использовании этих пакетов в клинических исследованиях обнаружались недостатки. На наш взгляд, среди них – не-

**Ельков Александр Никанорович (El'kov Aleksandr Nikanorovich)**, кандидат физико-математических наук, старший научный сотрудник отделения лечения острых отравлений Государственного бюджетного учреждения здравоохранения города Москвы «Научно-исследовательский институт скорой помощи им. Н.В. Склифосовского Департамента здравоохранения города Москвы», сотрудник кафедры клинической токсикологии Государственного бюджетного образовательного учреждения дополнительного профессионального образования «Российская медицинская академия последиplomного образования», 2bobika@mail.ru;

**Гольдфарб Юрий Семенович (Gol'dfarb Yuriy Semenovich)**, доктор медицинских наук, профессор, заведующий отделом внешних научных связей Государственного бюджетного учреждения здравоохранения города Москвы «Научно-исследовательский институт скорой помощи им. Н.В. Склифосовского Департамента здравоохранения города Москвы», заведующий кафедрой клинической токсикологии Государственного бюджетного образовательного учреждения дополнительного профессионального образования «Российская медицинская академия последиplomного образования», goldfarb@mail.ru;

**Поцхверия Михаил Михайлович (Potskhveriya Mikhail Mikhailovich)**, кандидат медицинских наук, заведующий научным отделением острых отравлений и соматопсихиатрических расстройств Государственного бюджетного учреждения здравоохранения города Москвы «Научно-исследовательский институт скорой помощи им. Н.В. Склифосовского Департамента здравоохранения города Москвы», доцент кафедры клинической токсикологии Государственного бюджетного образовательного учреждения дополнительного профессионального образования «Российская медицинская академия последиplomного образования», potskhveriya@mail.ru;

**Бадалян Амаяк Вазгенович (Badalyan Amayak Vazgenovich)**, кандидат медицинских наук, заведующий отделением острых отравлений Государственного бюджетного учреждения здравоохранения города Москвы «Научно-исследовательский институт скорой помощи им. Н.В. Склифосовского Департамента здравоохранения города Москвы», доцент кафедры клинической токсикологии Государственного бюджетного образовательного учреждения дополнительного профессионального образования «Российская медицинская академия последиplomного образования», drbadalian@mail.ru.

высокая исходная степень механизации требуемых от пользователя повторяющихся действий, приводящая к весьма значительным потерям времени; неполное соответствие предлагаемой универсальными пакетами формы представления результатов расчетов принятым de facto стандартам, используемым при оформлении научной продукции и, наконец, недостаточное внимание к некоторым критериям сравнения выборок, считающимся не вполне эффективными, например, к медианному критерию. Устранение этих недостатков нам представляется весьма интересным при острых отравлениях, научные исследования при которых сопровождаются использованием большого числа количественных данных, характеризующих изменение состояния больных в процессе лечебных мероприятий.

*Целью исследования* явилось создание независимого специализированного программного приложения для ускорения статистической обработки клинико-лабораторных данных при острых отравлениях.

**Материалы и методы исследования.** Для анализа были использованы массивы клинико-лабораторных данных, описывающие состояние 350 больных с острыми отравлениями различной этиологии, находившихся в отделении лечения острых отравлений и соматопсихиатрических расстройств НИИ СП им. Н.В. Склифосовского, у которых в процессе лечения проводили интенсивную терапию и реабилитационные мероприятия. Из них 144 пациента с острыми отравлениями психофармакологическими средствами и 206 – веществами прижигающего действия (ВПД). При этом данные группы из 25 больных с острыми отравлениями ВПД при их динамическом наблюдении были использованы в иллюстративной части работы. В качестве математических методов применялись алгоритмы вычисления выборочных показателей, статистические критерии согласия и проверки гипотез, а также модульное и объектно-ориентированное программирование. Инструментами исследования послужили входящие в Microsoft Office программы Excel и Word, а также среда программирования Visual Basic for Applications (VBA). Для верификации результатов расчетов применяли ППСР STATISTICA.

Как известно, после того как задача сформулирована, процесс статистической обработки в клинических исследованиях включает следующие этапы: 1 – сбор данных; 2 – вычисления и 3 – подготовка научной продукции, неотъемлемой частью которой являются таблицы и другой иллюстративный материал [4].

На первом этапе формируются матрицы данных в виде электронных таблиц – в настоящее время наиболее широко с этой целью применя-

ется программа Excel, которая входит в состав Microsoft Office. Если ограничиваться готовыми средствами обработки данных, то на втором этапе возможны два варианта. Один из них состоит в продолжении работы в Excel с использованием встроенного в него набора статистических функций. Другой вариант – перенос полученных данных в один из статистических пакетов, среди которых по некоторым сведениям [3, 5] наибольшее распространение получил ППСР Statistica. Что же касается третьего этапа обработки данных – оформления результатов, то обычно для этой цели применяется приложение Microsoft Office Word различных версий.

Не вникая подробно в содержание первого этапа, поскольку это выходит за рамки темы данной работы, рассмотрим второй и третий этапы.

Второй этап – вычисления. Следует прежде всего отметить, что, несмотря на широкое распространение Excel, имеющиеся в нем готовые механизмы статистических расчетов, то есть включенные в него статистические функции, не объединены в единый комплекс и, согласно некоторым оценкам [6], характеризуются недостаточными точностью и полнотой, что может привести к некорректным результатам на больших массивах данных. В то же время, несмотря на то, что средства, которыми располагают ППСР, содержат чрезвычайно развитый аппарат статистического анализа, они, как это ни парадоксально, характеризуются довольно низкой степенью механизации. В процессе однотипных вычислений для подготовки каждого из них приходится всякий раз в ручном режиме выполнять практически одну и ту же последовательность действий, как-то: загрузка или выбор матрицы данных, определение или коррекция параметров, задание метода обработки и его опций. Более того, для некоторых важных статистических критериев – например, для точного теста Фишера в программе Statistica, обработка матрицы данных как единого целого просто отсутствует.

Это привело к появлению в некоторых из ППСР (Statistica, Stata) определенных средств, позволяющих записывать и затем выполнять сценарии обработки данных [7, 3], а также к созданию специального языка программирования статистических вычислений R, не получившему, однако, достаточно широкого распространения среди неспециалистов [7]. В то же время, в решении задачи повышения уровня автоматизации расчетов можно пойти по другому пути – применив встроенный в программу Excel инструмент Visual Basic for Applications (VBA), который представляет собой полноценную версию объектно-ориентированного языка программирования высокого уровня. Несмотря на то, что статистический аппарат в Excel в сравнении с ППСР гораздо менее

развит, средства алгоритмизации, включенного в него VBA существенно лучше тех, что реализованы в специализированных статистических пакетах [8].

Одним из свойств набора статистических методов, применяемых в клинических исследованиях, является то, что в нем можно выделить довольно небольшое устоявшееся подмножество, которое является практически обязательной частью любого процесса статистической обработки данных. Если не рассматривать связи между признаками, то в это подмножество входят: вычисление описательных статистик, проверка выборочных распределений признаков на соответствие нормальному закону и статистическое сравнение распределений в различных группах пациентов. Независимая программная реализация этих методов не представляет больших сложностей, позволяя в более узком, по сравнению с ППСП, диапазоне статистических методов достичь более высокой интеграции указанных выше алгоритмов и степени автоматизации вычислений.

**Результаты и обсуждение.** Исходя из изложенного и ориентируясь в плане выбора статистических процедур на ППСП StatSoft Statistica (далее Statistica) как наиболее распространенный и признанный в медицинских исследованиях пакет статистических программ [5], нами была создано приложение VBA Excel, работающее со сформированными на первом этапе клинических исследований матрицами данных. Эти матрицы должны находиться на рабочем листе Excel и располагаться одна под другой в одних и тех же столбцах. Как это принято в клинических исследованиях [3, 5], каждая строка такой матрицы описывает состояние пациента посредством набора признаков, последовательность которых одинакова для всех строк во всех матрицах. Имена признаков помещаются в ячейки соответствующих столбцов над самой верхней матрицей. При этом матрицы данных и строки с именами признаков мы выделяли цветом, что дает приложению возможность непосредственно после запуска самому автоматически определить все необходимые параметры задачи и сразу перейти к выполнению фиксированного набора алгоритмов.

Сначала приложение проводит проверку данных на валидность по принципу «цифровой/нецифровой». Нецифровые данные отмечаются и исключаются из дальнейших расчетов. Далее по всем признакам для каждой из матриц данных вычисляются описательные статистики.

Затем, с целью оценки статистической значимости результатов проводимого статистического анализа и границ применимости соответствующих процедур, приложением выполняется контроль выборочных распределений на соответствие нормальному закону с использовани-

ем критериев и таблиц, включенных в ГОСТ Р ИСО 5479-2002 [9, 10]. Для этого в приложении реализованы: критерии проверки на симметричность и на эксцесс, критерий Шапиро–Уилка [11] и критерий Эпса–Палли [12]. При решении вопроса о статистической значимости отклонения от нормальности приложение либо обращается к приведенным в Стандарте таблицам, определяя по ним (с использованием интерполяции) критические величины (первые два критерия), либо непосредственно вычисляет р-значения (критерии Шапиро–Уилка и Эпса–Палли).

Важнейшее место в исследованиях занимают операции статистического сравнения выборочных данных. С этой целью в приложении реализованы критерии Стьюдента, Манна–Уитни и медианный критерий для сравнения двух независимых выборок, а также критерий Уилкоксона и критерий знаков для зависимых выборок. При этом наименьшие трудности вызвало написание кода для критерия Стьюдента. Разработка программных реализаций непараметрических критериев сравнения двух выборок выполнялась согласно их описаниям в открытых источниках [11, 13, 14].

Недостаточная степень автоматизации вычислений в стандартных статистических пакетах особенно хорошо видна при статистическом сравнении данных по совокупности показателей матриц. Сложность расчета здесь состоит в том, что в клинических исследованиях количество таких матриц всегда больше двух, и требуется осуществить не одно, а целую серию статистических сравнений матриц данных. Количество таких сравнений по одному критерию равно числу элементов в верхнем треугольнике корреляционной матрицы за вычетом главной диагонали, т.е. для трех матриц оно равно трем, для четырех матриц – шести и т.д. Однако число еще нужно умножить на количество используемых критериев. В итоге операция статистического сравнения матриц данных не сводится только к собственно сравнениям – важное место занимает также организация самого процесса, которая не является оптимальной в стандартных статистических пакетах, где в обычном режиме сравнение каждой пары матриц рассматривается как независимая операция, включающая в каждом случае задание или коррекцию параметров в диалоговом окне и другие процедуры, выполняемые пользователем, что требует значительного объема трудозатрат. Чтобы избежать этого, в разработанном нами приложении реализован цикл верхнего уровня для попарного статистического сравнения всех возможных сочетаний матриц данных по совокупности показателей по нескольким критериям, который выполняется в автоматическом режиме, не требуя от пользователя каких-либо действий,

что способствует резкому уменьшению времени, расходуемому на подготовку к вычислениям.

Третий этап – подготовка таблиц. Для описания этого этапа рассмотрим результаты работы приложения на примере статистической обработки массива информации клинико-лабораторных данных, состоящего из трех матриц, которые содержат динамику изменения 8 показателей клинического анализа крови, характеризующих состояние группы 25 больных с отравлениями ВПД, на 1-е, 2–3-и и 4–5-е сутки от момента поступления в клинику.

По результатам вычислений приложение генерирует несколько выходных форм, каждая из которых представляет собой таблицу в формате Word и создается приложением посредством загрузки в Word заранее подготовленных шаблонов и заполнения их результатами расчетов.

Создаваемая приложением базовая выходная форма по содержанию почти полностью соответствует выходной форме, получающейся при запуске вычислительной процедуры вкладки Advanced диалогового окна Descriptive Statistics пакета Statistica, что позволяет легко верифицировать результаты расчетов с его помощью.

Следующая, выходная форма (табл.1) дает возможность проследить динамику изменения выборочных медиан исследуемой задачи.

Из таблицы видно, что приложение выполняет синтаксическое преобразование выборочных квартилей к общепринятому табличному формату, рассчитывает относительные изменения

показателей ( $\Delta\%$ ) по отношению исходному значению и расставляет маркеры статистической значимости изменений показателей относительно исходного значения. Эти возможности отсутствуют в пакете Statistica. Приложением формируется также полностью аналогичная таблица, содержащая динамику средних величин.

Также приложение генерирует таблицу, в которой содержится информация, позволяющая выявить наличие отклонений от нормального закона выборочных распределений исследуемых клинико-лабораторных показателей. Вместе с медианами и средними, которые даны для удобства визуальной оценки степени асимметричности выборочных распределений, она содержит выборочные значения асимметрии и эксцесса, а также критические значения для отклонения нулевой гипотезы о симметричности и кривизне распределений. Кроме этого, в таблице приводятся вычисленные программой р-значения для критериев Шапиро–Уилка и Эпса–Палли.

Как уже говорилось выше, приложение осуществляет в автоматическом режиме по нескольким критериям и по совокупности показателей полный цикл статистического сравнения всех представленных в исследовании матриц данных по принципу каждой с каждой. В рассматриваемом примере таких сравнений должно быть три, однако для экономии места далее представлено сравнение только первых двух матриц.

Итоговая таблица результатов вычислений, произведенных приложением при выполнении

Таблица 1

**Динамика изменений выборочных квартилей показателей анализа крови при отравлениях веществами прижигающего действия**

	Показатель	n	1-е сутки	n	2-3-и сутки	$\Delta\%$	n	4-5-е сутки	$\Delta\%$
1	Гемоглобин	25	136,0 (116,0;149,0)	21	123,0 <sup>1</sup> (109,0;135,0)	-9,6	19	131,0 (110,0;149,0)	-3,7
2	Гематокрит	25	42,4 (35,2;44,2)	21	35,9 <sup>1</sup> (33,5;41,0)	-15,3	19	38,3 (33,9;45,2)	-9,7
3	Эритроциты	25	4,47 (4,08;4,76)	21	3,96 <sup>1</sup> (3,68;4,40)	-11,4	18	4,22 (3,65;4,71)	-5,6
4	Лейкоциты	25	10,5 (8,80;13,1)	21	8,30 <sup>1</sup> (6,00;10,0)	-20,9	18	9,50 (6,40;11,5)	-9,5
5	Тромбоциты	25	179,0 (158,0;233,0)	21	174,0 (136,0;208,0)	-2,8	18	212,5 <sup>1,2</sup> (172,0;274,0)	18,7
6	MCV	25	92,0 (89,0;96,0)	21	93,0 (89,0;96,0)	1,1	18	91,0 (87,0;97,0)	-1,1
7	MCH	25	30,3 (29,0;31,7)	21	30,7 (29,2;32,2)	1,3	18	29,9 (28,4;31,4)	-1,3
8	Средняя концентрация	25	328,0 (321,0;333,0)	21	331,0 (323,0;337,0)	0,91	18	325,0 (319,0;334,0)	-0,9

Примечания: <sup>1</sup> – статистически значимое различие показателя по отношению к 1-м суткам ( $p < 0,05$ ); <sup>2</sup> – статистически значимое различие показателя по отношению к 2–3-м суткам ( $p < 0,05$ );  $\Delta\%$  – по отношению к 1-м суткам. Обозначения: MCV – средний объем эритроцита; MCH – среднее содержание гемоглобина.

сравнений показателей по критерию Стьюдента, здесь не приводится, так как она полностью аналогична форме, создаваемой пакетом Statistica. Частично эти результаты используются в заполняемой приложением следующей выходной форме (табл. 2).

Ведущим критерием в данной таблице является ранговый непараметрический критерий Манна–Уитни для оценки различий между двумя независимыми выборками. Основой для разработки этой таблицы в части критерия Манна–Уитни послужила соответствующая форма ППСР Statistica и, как сказано выше, описание алгоритма вычисления рангов, приведенное в литературе [13]. Однако, поскольку этот критерий отличается повышенной чувствительностью к совпадающим значениям в выборочных данных [15], не лишено интереса сравнение результатов его применения с результатами, с одной стороны критерия Стьюдента, обладающего большей мощностью там, где его применение оправдано и, с другой стороны, менее мощного, но более устойчивого к совпадающим значениям медианного критерия. Такое сравнение, как нам кажется, позволяет повысить надежность статистических выводов.

При вычислении критерия Манна–Уитни описываемое приложение, в сравнении с пакетом Statistica, дает некоторую дополнительную информацию, как-то: число серий, т.е. совпадающих рангов, их размеры и, кроме того, при небольшом объеме выборки, – критическое значение для статистики критерия [13].

Правую часть таблицы занимает медианный критерий (отметим, что результаты расчетов по этому критерию в пакете Statistica представлены в гораздо менее удобной форме). Базовым показателем этого критерия является медиана объединенного вариационного ряда сравниваемых выборок (Медиана). В каждой клетке следующего за медианой столбца располагается таблица сопряженности  $2 \times 2$ , позволяющая оценить сколько элементов каждой из двух выборок меньше общей медианы (A, C) или больше нее, или равны ей (B, D). Далее отражены результаты обработки таблиц сопряженности точным тестом Фишера, а также по критерию  $\chi^2$  без поправки Йейтса, ориентированной на предотвращение завышения статистической значимости для небольших данных, и с ней – в виде значений вычисленных статистик и соответствующих им уровней р-вероятностей (P). В целом, как это видно из таблицы, можно утверждать, что в данном конкретном примере все три критерия показали согласованный результат.

В последней из создаваемых приложением выходных форм (табл. 3) представлены результаты статистического сравнения по совокупности

показателей тех же двух матриц данных, что и в таблице 2, посредством знаково-рангового критерия Уилкоксона и критерия знаков. Теперь содержащиеся в этих матрицах выборки рассматриваются как зависимые. Как известно, эти критерии позволяют обнаружить наличие статистически значимой тенденции изменения показателя в ту или иную сторону в случае, когда его относительное изменение незначительно. Более мощный критерий Уилкоксона рассматривается как ведущий, однако, так же, как и в предшествующей таблице, к нему добавлен менее мощный, но более устойчивый по отношению к совпадающим рангам критерий знаков.

Как видно из таблицы, она содержит информацию об общем числе пар наблюдений, числе пар с несовпадающими значениями (поскольку только они учитываются при расчете), а также количестве последовательностей совпадающих рангов (серий) и их длинах. В столбцах таблицы содержатся суммы положительных и отрицательных рангов и статистика критерия Уилкоксона T (т.е. меньшая из двух предыдущих сумм), затем критическое число для T-статистики [13]. Еще два столбца, относящиеся к критерию Уилкоксона, – это значение нормированной статистики Уилкоксона Z, близкой к стандартному нормальному распределению при числе вариантов в вариационном ряду не менее 15 и соответствующая ей вероятность. Наконец, в последних трех столбцах таблицы представлен критерий знаков. Здесь указано число положительных и отрицательных разностей при парном сравнении наблюдений и соответствующее биномиальному распределению р-значение.

Как видно из таблицы, несмотря на разную мощность, при выявлении статистических различий между показателями оба критерия также показывают одинаковый результат. Сравнивая между собой таблицы 2 и 3, можно отметить, что в последней добавилось статистически значимая тенденция к изменению еще в одном показателе – количестве тромбоцитов в крови.

**Заключение.** Существует много работ, нацеленных на объяснение того, как нужно работать с отдельными статистическими процедурами в ППСР, однако проблема объединения этих методов в единую согласованную группу до сих пор не решена окончательно. Как видно, статистические вычисления, которые являются неотъемлемой частью клинических исследований, при обработке материала, связанного с острыми отравлениями, характеризуются существенной трудоемкостью в силу большого числа клинико-лабораторных признаков и значительного количества повторяющихся операций, что является следствием необходимости статистического сравнения нескольких выборок, описываю-

Таблица 2  
**Сравнение выборочных распределений показателей анализа крови при отравлениях веществами прижигающего действия в 1-е и 2-3-и сутки после поступления в клинику по критерию Манна-Уитни и медианному критерию**

Показатели	п1, п2	Критерий Манна-Уитни						P (Стьюдент)	Медианный критерий								
		Серии: число	размер	Сумма рангов		U	Укр.		Z	P	Медиана	A B C D	ТФ	$\chi^2$	P	$\chi^2$ (Йейтс)	P
				Выб. I	Выб. II												
Гемоглобин	25, 21	12	2 2 2 2 2 2 3 2 3 2 3 3	686,5	394,5	<b>164</b>	173	-2,183	<b>0,0290</b>	127,5	17 8 6 15	<b>0,017</b>	7,097	<b>0,0077</b>	5,608	<b>0,0179</b>	
Гематокрит	25, 21	6	2 3 2 2 2 2	693,5	387,5	<b>157</b>	173	-2,338	<b>0,0194</b>	39,4	16 9 7 14	<b>0,075</b>	4,293	<b>0,0382</b>	3,154	0,0757	
Эритроциты	25, 21	4	2 2 2 2	697,5	383,5	<b>153</b>	173	-2,426	<b>0,0152</b>	4,27	16 9 7 14	<b>0,075</b>	4,293	<b>0,0382</b>	3,154	0,0757	
Лейкоциты	25, 21	6	2 3 2 2 2 2	715	366	<b>135</b>	173	-2,812	<b>0,0049</b>	9,1	16 9 7 14	<b>0,075</b>	4,293	<b>0,0382</b>	3,154	0,0757	
Тромбоциты	25, 21	7	2 3 2 2 2 2 2	627	454	223	173	-0,871	0,3837	176,5	13 12 10 11	1,000	0,088	0,7672	0	1	
MCV	25, 21	10	2 2 4 4 3 3 4 3 4 2	581,5	499,5	257	173	-0,132	0,8947	92,5	12 13 11 10	1,000	0,088	0,7672	0	1	
MCH	25, 21	6	2 4 2 2 2 2	571	510	246	173	-0,364	0,7159	30,35	12 13 11 10	1,000	0,088	0,7672	0	1	
Средняя кон- центрация	25, 21	10	3 3 2 2 2 3 4 2 3 2	558,5	522,5	234	173	-0,64	0,5224	328,5	11 14 12 9	0,017	0,789	0,3745	0,35	0,5538	

Обозначения: Выб. I – первая выборка; Выб. II – вторая выборка; U – статистика Манна-Уитни; Uкр – критическое значение для статистики Манна-Уитни для выборки объемом не более 30; Z – нормированная и центрированная статистика Манна-Уитни; A / B / C / D – таблица сопряженности 2x2; ТФ – точный тест Фишера; MCV – средний объем эритроцита; MCH – среднее содержание гемоглобина.

Таблица 3

**Сравнение выборочных распределений показателей анализа крови при отравлениях веществами прижигающего действия в 1-е и 2-3-и сутки после поступления в клинику по критерию Уилкоксона и критерию знаков**

Показатели	n	n <sub>0</sub>	Критерий Уилкоксона										Критерий знаков		
			Серии:		Сумма рангов		T	T <sub>кр.</sub>	Z	P	n <sub>0</sub> (+)	n <sub>0</sub> (-)	P		
			число	размер	< 0	> 0									
Гемоглобин	21	20	6	3 2 2 2 2 2	-179	31	52	-2,764	<b>0,0057</b>	16	4	<b>0,012</b>			
Гематокрит	21	21	0		-204	27	58	-3,076	<b>0,0021</b>	17	4	<b>0,007</b>			
Эритроциты	21	21	1	2	-208,5	22,5	58	-3,233	<b>0,0012</b>	17	4	<b>0,007</b>			
Лейкоциты	21	20	2	3 4	-196	14	52	-3,401	<b>0,0007</b>	17	3	<b>0,003</b>			
Тромбоциты	21	20	3	2 2 2	-161,5	48,5	52	-2,110	<b>0,0349</b>	16	4	<b>0,012</b>			
MCV	21	15	2	11 4	-36	84	25	-1,431	0,1524	6	9	0,607			
MCH	21	21	5	3 3 3 2 3	-75,5	155,5	58	-1,392	0,1639	9	12	0,664			
Средняя концентрация	21	20	5	2 3 2 2 2	-82,5	127,5	52	-0,840	0,4006	10	10	1,176			

Обозначения: n – число пар наблюдений; n<sub>0</sub> – число пар наблюдений с не совпадающими; T – статистика Уилкоксона; T<sub>кр.</sub> – критическое значение для статистики Уилкоксона для числа пар не более 50; Z – нормированная и центрированная статистика Уилкоксона; n<sub>0</sub>(+) / n<sub>0</sub>(-) – число пар, в которых значение показателя, соответственно, увеличилось или уменьшилось; MCV – средний объем эритроцита; MCH – среднее содержание гемоглобина.



щих состояние, как правило, нескольких групп исследуемых больных в динамике. В то же время в распространенных прикладных пакетах статистических программ необходимость многократного повторного применения одного и того же метода при статистическом обеспечении клинических исследований не учитывается или учитывается недостаточно. Поэтому проблема применения статистических методов в клинических исследованиях связана не только с использованием широко распространенных статистических пакетов, но и с необходимостью повышения производительности труда при статистической обработке клинических исследований. Используя ППСП для целей многомерного статистического анализа [16, 17], мы обратили внимание на то, что хотя популярные статистические пакеты предлагают удобное для реализации большое количество статистических методов, из них интенсивно используется лишь небольшая часть, которая может быть достаточно легко реализована с помощью известных сред разработки программных кодов, одной из наиболее доступных и распространенных среди которых является среда программирования Visual Basic for Applications фирмы Microsoft.

В результате обобщения длительного опыта обработки статистических данных в клинической токсикологии нами на базе этой среды разработано специализированное программное приложение, реализующее распространенный сценарий обработки клиничко-лабораторных данных в клинических исследованиях, позволившее при этом существенно ускорить указанный процесс. Этому в значительной мере способствовало отсутствие необходимости перемещения данных из MS Excel в выбранный для проведения расчетов ППСП, применение специальной цветовой разметки для определения параметров решаемой задачи, объединение используемых методов в один неизменяемый комплекс, реализация статистического сравнения всех возможных сочетаний матриц данных по совокупности показателей как единой операции, не требующей вмешательства пользователя и, наконец, автома-

тическое формирование таблиц в формате MS Word. Таким образом, нами с помощью средств объектно-ориентированного программирования VBA Excel реализована жесткая фиксированная схема вычислений, позволяющая обрабатывать предоставленные данные по известному принципу «запустил и забыл», в известной мере отвечающему понятию искусственного интеллекта.

#### Выводы:

1. Разработанное нами приложение вычисляет необходимые в каждом исследовании описательные выборочные характеристики и выполняет обязательную автоматическую оценку отклонений выборочных распределений от нормального закона по критериям асимметрии, эксцесса, Шапиро–Уилка и Эпса–Палли, размещая результаты в предложенном нами, удобном для визуального анализа виде. При необходимости результаты расчетов легко верифицировать путем сравнения их с выходными формами пакета Statistica.

2. Для статистического сравнения матриц данных по совокупности показателей (далее – сравнение), с использованием наиболее распространенных статистических критериев, программно реализован автоматический цикл таких сравнений, позволяющий рассматривать все возможные сравнения матриц данных в текущей задаче как единую операцию, не требующую вмешательства пользователя.

3. В предлагаемом приложении последовательно реализуется принцип автоматической генерации специально спроектированных оригинальных таблиц в формате Word, дающих пользователю возможность легко сопоставить результаты статистического сравнения по различным критериям.

4. Разработанное нами приложение позволяет значительно (в 5–10 раз) увеличить производительность труда при статистической обработке результатов клинических исследований за счет повышения уровня интеграции необходимых исследователю статистических алгоритмов и результатов расчетов, а также автоматической генерации таблиц в формате MS Word.

## СПИСОК ЛИТЕРАТУРЫ:

1. Макарова Н.В. Статистический анализ медико-биологических данных с использованием пакетов статистических программ Statistica, SPSS, NCSS, SYSTAT. Санкт-Петербург: Политехника-сервис; 2012.
2. Боровиков В.П. Популярное введение в современный анализ данных в системе STATISTICA. Москва: Телеком; 2013.
3. Халафян, А.А. Современные статистические методы медицинских исследований. 3-е изд. Москва: Ленанд; 2014.
4. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистические методы в медико-биологических исследованиях с использованием Excel. Киев: МОРИОН; 2001.
5. Глушаков А.И. Применение статистических программ и методов в медико-биологических научных исследованиях. Казанский медицинский журнал. 2009; 90(4): 550–555.
6. Просветов Г.И. Анализ данных с помощью Excel: задачи и решения. Москва: Альфа-Пресс; 2018.
7. Унгурияну Т.Н., Гржибовский А.М. Программное обеспечение для статистической обработки данных Stata: введение. Экология человека. 2014; (1):60–63.
8. Михеев Р.Н. VBA и программирование в MS OFFICE для пользователей. Санкт-Петербург: БХВ-Петербург; 2006.
9. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения: Постановление Госстандарта России от 22 января 2002 г. № 25-ст. Москва: Госстандарт России; 2002.
10. Лемешко Б.Ю., Лемешко С.Б. Сравнительный анализ критериев проверки отклонения распределения от нормального закона. Метрология. 2005; (2):3–23.
11. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. 2-е изд., испр. Москва: Физматлит; 2012.
12. Гайдышев И. Анализ и обработка данных. Специальный справочник. Санкт-Петербург: Питер; 2001.
13. Рунион Р. Справочник по непараметрической статистике: современный подход. Москва: Финансы и статистика; 1982.
14. Брандт З. Анализ данных. Статистические и вычислительные методы для научных работников и инженеров: пер. с англ. Москва: Мир; АСТ; 2003.
15. Mann H.B., Whitney D.R. On a test of whether one of two random variables is stochastically larger than the other. Ann.



Math. Statist. 1947; 18(1): 50-60.

16. Бадалян А.В., Гольдфарб Ю.С., Ельков А.Н., Биткова Е.Е., Боровкова Н.В., Клычникова Е.В. Использование

факторного анализа для оценки эффективности лечения острых отравлений на этапе реабилитации. Токсикологический вестник. 2017; (6): 17-30.

17. Бадалян А.В., Гольдфарб Ю.С., Поткхверия М.М., Годков М.А., Ельков А.Н., Биткова Е.Е. и др. Использование кластерного анализа для оценки эффек-

тивности реабилитационного лечения острых отравлений. Токсикологический вестник. 2018; (3): 2-17.

## REFERENCES:

1. Makarova N.V. Statistical analysis of biomedical data using statistical software packages Statistica, SPSS, NCSS, SYSTAT. Saint-Petersburg: Politekhniko-servis; 2012 (in Russian).
2. Borovikov V.P. A popular introduction to modern data analysis in the system STATISTICA. Moscow: Telekom; 2013 (in Russian).
3. Khalafyan A.A. Modern statistical methods of medical research. 3rd ed.. Moscow: LENAND; 2014 (in Russian).
4. Lapach S.N., Chubenko A.V., Babich P.N. Statistical methods in biomedical research using Excel. Kiev: MORION; 2001 (in Russian).
5. Glushakov A.I. The use of statistical programs and methods in biomedical research. Kazan Medical Journal. 2009; 90(4): 550-555 (in Russian).
6. Prosvetov G.I. Data Analysis with Excel: Tasks and Solutions. Moscow: Alfa-Press; 2018 (in Russian).
7. Unguryanu T.N., Grzhibovskiy A.M. Stata Statistical Processing Software: Introduction. Human Ecology. 2014; (1):60-63 (in Russian).
8. Mikheev R.N. VBA and programming in MS OFFICE for users. Saint-Petersburg: BHV-Petersburg; 2006 (in Russian).
9. Statistical methods. Checking the deviation of the probability distribution from the normal distribution: Resolution of the State Standard of the Russian Federation of January 22, 2002 No. 25-art. Moscow: Gosstandart of the Russian Federation; 2002 (in Russian).
10. Lemesenko B.Yu., Lemesenko S.B. Comparative analysis of the criteria for checking the deviation of the distribution from the normal law. Metrology. 2005; (2):3-23. (in Russian)
11. Kobzar' A.I. Applied Mathematical Statistics. For engineers and scientists. 2nd ed., rev. Moscow: Fizmatlit; 2012 (in Russian).
12. Gaydyshev I. Analysis and data processing. Special reference. Saint-Petersburg: Piter; 2001 (in Russian).
13. Runion R. Handbook of Nonparametric Statistics: Modern approach. Moscow: Finance and Statistics; 1982 (in Russian).
14. Brandt Z. Statistical and computational methods for scientists and engineers: transl. from english. Moscow: Mir; AST; 2003 (in Russian).
15. Mann H.B., Whitney D.R. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 1947; 18(1): 50-60.
16. Badalyan A.V., Gol'dfarb Yu.S., El'kov A.N., Bitkova E.E., Borovkova N.V., Klychnikova E.V. Using factor analysis to assess the effectiveness of the treatment of acute poisoning at the rehabilitation stage. Toxicological Review. 2017; (6): 17-30 (in Russian).
17. Badalyan A.V., Gol'dfarb Yu.S., Potkshveriya M.M., Godkov M.A., El'kov A.N., Bitkova E.E., et al. The use of cluster analysis to assess the effectiveness of the rehabilitation treatment of acute poisoning. Toxicological Review. 2018; (3): 2-17 (in Russian).

A.N. El'kov<sup>1,2</sup>, Yu.S. Gol'dfarb<sup>1,2</sup>, M.M. Potkshveriya<sup>1,2</sup>, A.V. Badalyan<sup>1,2</sup>

## OPTIMIZATION OF STATISTICAL PROCESSING OF CLINICAL AND LABORATORY DATA IN ACUTE POISONING

<sup>1</sup>Sklifosovsky Research Institute for Emergency Medicine, Public Healthcare Institution of Moscow Healthcare Department, 129090, Moscow, Russian Federation

<sup>2</sup>Russian Medical Academy of Continuous Professional Education, RF Ministry of Health, 125993, Moscow, Russian Federation

The article describes a specialized software application developed by the authors that summarizes the long-term experience of analyzing clinical and laboratory data in clinical toxicology using Microsoft's Visual Basic for Applications object-oriented programming tools in Excel. The main purpose of the application is a significant (5-10 times) acceleration of the implementation of the complex of the most commonly used statistical algorithms. The application calculates descriptive sample characteristics, identifies deviations of sample distributions from the normal law by several generally accepted criteria, performs statistical comparison of data matrices by several criteria, and automatically generates a number of tables ready for placement in scientific texts for various purposes in Microsoft Word format. The presentation of the output data makes it easy to verify the results of calculations by comparing them with the output forms of the Statistica application software package. The acceleration of obtaining the final result is achieved: 1 – due to the absence of the need to move data from Excel to the package of applied statistical programs; 2 – as a result of automating the determination of task parameters by means of special color markup; 3 – by combining a limited number of the most commonly used statistical methods into one immutable complex; 4 – as a result of the software implementation of statistical comparison of all possible combinations of data matrices involved in the study for a set of indicators as a single operation that does not require user intervention; 5 – due to the application's use of syntactic transformations of output data and automatic table formation by filling in ready-made templates with calculation results.

**Keywords:** acute poisoning, computational statistics, packages of applied statistical programs, Microsoft Excel, Visual Basic for Applications.

Quote: A.N. El'kov, Yu.S. Gol'dfarb, M.M. Potkshveriya, A.V. Badalyan. Optimization of statistical processing of clinical and laboratory data in acute poisoning. Toxicological Review. 2021; 2:14-22.

Материал поступил в редакцию 05.07.2020 г.