

УДК: 519.23-614.2

Мудров В.А.

АЛГОРИТМЫ СТАТИСТИЧЕСКОГО АНАЛИЗА КАЧЕСТВЕННЫХ ПРИЗНАКОВ В БИМЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ С ПОМОЩЬЮ ПАКЕТА ПРОГРАММ SPSS

*Федеральное государственное бюджетное образовательное учреждение
высшего образования «Читинская государственная медицинская академия»
Министерства здравоохранения Российской Федерации, 672000, г. Чита, ул. Горького, 39-а.*

Цель исследования. Предметом исследования явился статистический анализ качественных признаков. Темой исследования явился статистический анализ качественных признаков в биомедицинских исследованиях. Целью работы явилось описание подробных алгоритмов статистического анализа качественных признаков в биомедицинских исследованиях с помощью пакета программ SPSS.

Материалы и методы. Проведен научный обзор методов статистического анализа качественных признаков в биомедицинских исследованиях. Практические основы статистической обработки качественных переменных исследования рассмотрены на примере пакета программ «IBM SPSS Statistics Version 25.0» (International Business Machines Corporation, США).

Результаты. Определены оптимальные алгоритмы статистического анализа качественных признаков в биомедицинских исследованиях. Подробно описан принцип интеграции результатов исследования при описании статистических методов с помощью пакета программ SPSS.

Заключение. Использование описанных алгоритмов статистического анализа качественных признаков позволит повысить уровень представления результатов биомедицинских исследований.

Ключевые слова. Статистический анализ, алгоритм, качественные признаки, номинальные переменные, ранговые переменные, критерий хи-квадрат, критерий Пирсона, критерий Фишера, критерий Мак-Нимара, критерий Кокрена, относительный риск, отношение шансов.

Mudrov V.A.

STATISTICAL ANALYSIS ALGORITHMS OF QUALITATIVE FEATURES IN BIOMEDICAL RESEARCH USING THE SPSS SOFTWARE PACKAGE

Chita State Medical Academy, 39-a, Gorky street, Chita, Russia, 672000

The aim of the research. The subject of research was a statistical analysis of qualitative signs. The topic of the research was a statistical analysis of qualitative features in biomedical research. The aim of the study was to describe a detailed statistical analysis algorithms of qualitative features in biomedical research using the SPSS software package.

Materials and methods. A scientific review of a statistical analysis methods of qualitative features in biomedical research is carried out. The practical foundations of a statistical processing of qualitative research results are examined using the IBM SPSS Statistics (Version 25.0) software package (International Business Machines Corporation, USA) as an example.

Results. The optimal algorithms for statistical analysis of qualitative features in biomedical research is determined. The principle of research results integration in the description of statistical methods using the IBM SPSS Statistics (Version 25.0) software package is described in detail.

Conclusion. The use of the described algorithms for statistical analysis of qualitative features will increase a presentation's level of biomedical research results.

Keywords. Statistical analysis, algorithm, qualitative features, nominal variables, ranking variables, chi-square test, Pearson test, Fisher test, McNemar test, Cochran's test, relative risk, odds ratio.

В данной статье основное внимание уделено правилам выбора наиболее подходящего способа обработки качественных признаков в биомедицинских исследованиях, а также алгоритмам применения основных статистических процедур с помощью пакета программ SPSS. Создание подробных алгоритмов статистического анализа, представленных в доступном для читателя виде, является актуальным методом предупреждения научных заблуждений, что в перспективе позволит повысить уровень представления результатов биомедицинских исследований начинающими исследователями [1, 2, 3].

Следует указать, что качественные признаки, в отличие от количественных, не поддаются числовому измерению. В некоторых научных источниках их также называют категориальными. Качественные признаки делят на номинальные и порядковые (ранговые). К номинальным относят признаки, которые невозможно расположить в порядке убывания или возрастания, произведение арифметических действий между данными признаками не поддается логическому осмыслению (пол плода, семейное положение, исход родов через естественные родовые пути, факт развития конкретного осложнения процесса родов и т.д.). Бинарными (дихотомическими) называют переменные, которые могут принимать только одно из двух противоположных значений (наличие или отсутствие вредных привычек, пол плода, наличие или отсутствие анатомического сужения таза и т.д.). Признаки, которые можно ранжировать в логическом порядке, но которые также не поддаются арифметическим действиям, называют порядковыми (степень артериальной гипертензии, степень нарушения маточно-плодово-плацентарного кровотока, форма плацентарной недостаточности и т.д.). Следует отметить, что порядковые переменные могут быть сформированы на основании количественных данных: количественную переменную разделяют на ряд категорий (например: разделение возраста пациенток на возрастные группы). Данная технология позволяет провести некоторые виды статистического анализа, характерные исключительно для качественных переменных (логистическая регрессия, определение чувствительности и специфичности метода и т.д.). Основным преимуществом статистического анализа качественных признаков (в сравнении с анализом количественных признаков) является возможность оценки истинного клинически значимого исхода исследования [1, 2, 3].

В отечественной практике статистического анализа биомедицинских данных наиболее часто с целью сравнения качественных признаков до недавнего времени использовался критерий Стьюдента. Вероятно, данный факт обусловлен низким уровнем информированности о существующих ограничениях применения данного критерия. Игнорирование ограничений метода приводит к обнаружению «статистически значимых» различий там, где их на самом деле нет [4]. В современных рекомендациях по проведению статистического анализа критерий Стьюдента не рекомендуется для сравнения качественных признаков в принципе, что опосредовано увеличением частоты «ложных» умозаключений, в частности наблюдаемых в работах начинающих исследователей. Сравнение частот с помощью критерия Стьюдента возможно только в ситуациях, когда для каждой из изучаемых переменных может быть только два возможных значения, при этом объем выборки должен быть значительным. Для большинства биомедицинских исследований характерна низкая частота событий и объем выборки, а также значительное число вариантов, которые могут принимать качественные переменные, что исключает возможность применения критерия Стьюдента. Более объективными возможностями сравнения частот и долей обладают способы, основанные на сравнении фактических частот с ожидаемыми [3, 4, 5].

Целью работы явилось описание подробных алгоритмов статистического анализа качественных признаков в биомедицинских исследованиях с помощью пакета программ SPSS.

Материалы и методы. Научный обзор методов статистического анализа качественных признаков в биомедицинских исследованиях проведен на основании анализа современных зарубежных и отечественных источников литературы. В качестве основополагающих рассмотрены общеизвестные источники, имеющие наибольший авторитет и широкое распространение [1, 2, 3, 6]. Практические основы статистической обработки качественных результатов исследования рассмотрены на примере пакета программ «IBM SPSS Statistics Version 25.0» (International Business Machines Corporation, США).

Результаты и их обсуждение. При проведении исследования следует руководствоваться принципами Международного комитета редакторов медицинских журналов (ICMJE):

1. Используемые статистические методы следует описывать максимально подробно: при описании результатов статистического анализа не следует ограничиваться только уровнем значимости P , так как он не дает информации о размере наблюдаемого эффекта.

2. Результаты исследования должны быть представлены достаточно подробно, что позволит включить их во вторичный анализ: требуется приведение описательной статистики, на основании которой рассчитываются другие статистические показатели – числители и знаменатели в процентных величинах, рисках, шансах и отношениях угроз. Уровня значимости P недостаточно для выполнения вторичного анализа – необходимо указать объемы исследуемых групп, связанную с уровнем значимости величину эффекта и оценку его точности [7].

Перед проведением статистического анализа следует указать любые формальные процедуры, использованные для преобразования исходных данных (преобразование непрерывных данных в категориальные, объединение категорий и т.д.) [6].

Согласно рекомендациям САМПЛ при проведении первичного анализа следует указать:

1. Цель проведения анализа;
2. Описательную статистику для каждого анализируемого признака;
3. Наименьшие различия, являющиеся клинически значимыми;
4. Основные статистические методы, используемые для достижения основных целей исследования;
5. Статистические методы, используемые для описания каждого конкретного анализа;
6. Условия применимости использованного метода (теста);
7. Поправки при проведении множественных сравнений (при наличии);
8. Метод обработки выбросов (при наличии);
9. Обоснование использования одностороннего или двустороннего теста;
10. Пороговую величину α -ошибки (уровня значимости P), определяющую статистическую значимость;
11. Полное наименование использованной статистической программы;
12. Статистические методы, используемые для дополнительного анализа, включая анализ чувствительности, заполнение пропущенных значений или проверку условий применимости [6].

Представление описательной статистики качественного признака проводится в соответствии со следующими правилами [6]:

1. В первую очередь следует указать наименование показателя, который Вы приводите: показатель темпа событий (заболеваемость, перинатальная смертность и т.д.), показатель отношения (отношение шансов, отношение угроз), показатель риска (абсолютный риск, разность относительных рисков и т.д.);
2. При описании качественного признака необходимо указать абсолютную величину для каждого его значения, а также процентную долю в структуре всей совокупности;
3. При описании относительных величин следует указать значения числителя и знаменателя (например: отношение числа женщин с раком молочной железы к числу женщин, у которых можно было предположить наличие рака);
4. При описании показателей, относящихся к популяционным, следует указать соответствующие множители (например: $\times 100\ 000$ – при расчете показателя материнской смертности);
5. При описании величины темпа (скорости наступления) событий следует указать соответствующий временной отрезок;
6. Также следует привести меру точности для оценок рисков, темпов и отношений (доверительный интервал).

Алгоритм описания анализа сопряженности качественных признаков в биомедицинских исследованиях [6]:

1. Необходимо описать изучаемую сопряженность (ассоциацию).
2. Необходимо представить описательную статистику для всех признаков, участвующих в анализе.
3. Следует указать название теста сопряженности.

4. Следует указать, был ли тест двусторонним или односторонним. При использовании одностороннего теста необходимо обосновать его применение: результаты односторонних тестов в биомедицинских исследованиях использовать не рекомендуется за исключением редких случаев, когда именно односторонний тест был запланирован еще до начала сбора данных и зафиксирован в протоколе исследования.
5. Для тестов сопряженности (типичным примером служит хи-квадрат) необходимо привести вычисленное значение уровня значимости P , так как сопряженность выявляется в результате проверки статистической гипотезы.
6. Для мер сопряженности (типичным примером служит ϕ -критерий Крамера) необходимо привести вычисленное значение этой меры и доверительный интервал.
7. Не следует описывать сопряженность как слабую, среднюю или сильную до момента определения диапазонов значений этих категорий сопряженности.
8. При первичном анализе данных следует включить полную таблицу сопряженности в отчет (табл. 1).

Таблица 1

Пример таблицы сопряженности

	Исход есть (1)	Исхода нет (0)	Всего
Фактор риска есть (1)	A	B	A + B
Фактора риска нет (0)	C	D	C + D
Всего	A + C	B + D	A + B + C + D

9. Необходимо указать название статистического программного продукта (желательно также указать номер лицензии соответствующего программного продукта), использованного для проведения анализа.

Алгоритм выбора статистического критерия для анализа качественных данных [3, 8]:

1. При сравнении частоты встречаемости качественного признака в одной выборке с заданным значением используется критерий Z .
2. При сравнении двух независимых групп номинальных данных используется критерий хи-квадрат Пирсона.
 - 2.1. Если количество ожидаемых наблюдений хотя бы в одной из ячеек четырехпольной таблицы составляет менее 10, для сравнения двух независимых групп номинальных данных используется критерий хи-квадрат с поправкой Йейтса на непрерывность.
 - 2.2. Если количество ожидаемых наблюдений хотя бы в одной из ячеек четырехпольной таблицы составляет менее 5, для сравнения двух независимых групп номинальных данных используется точный критерий Фишера.
3. При сравнении двух зависимых групп номинальных данных используется критерий Мак-Нимара (в частности при сравнении относительных показателей в динамике при двухэтапном измерении).
4. При сравнении номинальных признаков трех и более независимых групп следует использовать критерий хи-квадрат Пирсона.
 - 4.1. Если в многопольной таблице доля ячеек с ожидаемым числом наблюдений меньше 5 превысит 20% от общего числа или хотя бы одно ожидаемое значение будет меньше единицы, то для расчета критерия хи-квадрат следует объединить группы, чтобы увеличить количество наблюдений в клетках таблицы.
 - 4.2. При малых выборках предпочтительным является определение критерия хи-квадрат Пирсона с поправкой на правдоподобие.
 - 4.3. Для определения силы взаимосвязи целесообразным является определение критерия V Крамера.
5. При сравнении номинальных признаков трех и более зависимых групп следует использовать Q -критерий Кокрена. При обнаружении статистически значимых различий с помощью критерия Кокрена следует провести попарное сравнение признаков с помощью критерия Мак-Нимара с поправкой Бонферрони.
6. Сравнение порядковых признаков:

6.1. В случае наличия большого числа возможных категорий сравнение порядковых признаков следует проводить в соответствии с правилами анализа количественных переменных, не подчиняющихся закону нормального распределения.

6.2. В случае наличия малого числа возможных категорий для сравнения порядковых признаков можно использовать критерии, предназначенные для сравнения номинальных признаков.

6.2.1. Критерий хи-квадрат для линейного тренда следует использовать для проверки гипотезы о наличии статистической взаимосвязи между переменными.

6.2.2. Критерий γ Гудмана-Краскела, критерии тау-b и тау-c Кендалла, d-критерий Сомера следует применять для оценки величины эффекта.

Алгоритм сравнения частоты встречаемости номинального признака в одной выборке с заданным значением с помощью критерия Z в программе IBM SPSS Statistics Version 25.0 [3]:

1. После внесения номинальных признаков исследуемой выборки в исходную форму исследователю необходимо в меню «Анализ» выбрать раздел «Непараметрические критерии», в подразделе «Устаревшие диалоговые окна» выбрать «Биномиальный критерий». В открывшемся диалоговом окне из списка переменных, расположенных слева, необходимо выбрать исследуемую переменную и перенести ее в поле «Список проверяемых переменных». После этого в поле «Проверяемая доля» необходимо ввести гипотетическое значение, с которым предстоит произвести сравнение (программой автоматически задается значение равное 0,5). Запуск статистического анализа осуществляется нажатием кнопки «ОК».

2. В табличном заключении программы SPSS в первом столбце обозначены имеющиеся категории (Группа 1 и Группа 2), которые в столбце «Категории» представлены как «0» и «1» («0» обозначает отсутствие изучаемого фактора риска, «1» – его наличие). В столбце «N» представлена структура абсолютного числа исследуемых: исследуемые с фактором риска и без, а также общее количество участников исследования. В столбце «Наблюдаемая доля» представлены соответствующие абсолютным числам доли. В последнем столбце представлен точный уровень двухсторонней значимости. Если достигнутый уровень значимости меньше критического значения (0,05), можно отвергнуть нулевую гипотезу об отсутствии различий между выборочным и заданным значениями. Т.е. можно сделать вывод о наличии статистически значимых различий между выборочным и заданным значениями в исследуемой группе.

3. Сравнительную статистику при описании номинального признака необходимо представить в виде абсолютной величины для каждого его значения, а также процентной доли в структуре всей совокупности. Следует использовать формат: утвердительное выражение с указанием уровня значимости ($p = \dots$).

4. Для наглядного представления в данной ситуации следует использовать круговую диаграмму.

Алгоритм сравнения двух независимых групп номинальных данных с помощью критерия хи-квадрат Пирсона в программе IBM SPSS Statistics Version 25.0 [3]:

1. После внесения номинальных признаков в исходную форму исследователю необходимо в меню «Анализ» выбрать раздел «Описательные статистики», подраздел «Таблицы сопряженности». В открывшемся диалоговом окне «Таблицы сопряженности» из списка переменных, расположенных слева, необходимо независимую переменную перенести в поле «Строки», а зависимую – в поле «Столбцы». При нажатии на кнопку «Статистики» открывается диалоговое окно «Таблицы сопряженности: Статистики», в левой половине которого напротив «Хи-квадрат» необходимо поставить флажок. В диалоговом окне «Таблицы сопряженности» следует указать необходимость вывода кластеризованных столбчатых диаграмм. Далее при нажатии на кнопку «Ячейки» открывается диалоговое окно «Таблицы сопряженности: Вывод в ячейках», в левой половине которого в блоке «Количества» можно поставить флажки напротив слов «Наблюдаемые» (позволит оценить фактическое количество наблюдений в каждой из ячеек четырехпольной таблицы) и

«Ожидаемые» (позволит оценить ожидаемое количество наблюдений в каждой из ячеек четырехпольной таблицы). Для определения соответствующих долей в блоке «Проценты» необходимо отметить флажками «По строке», «По столбцу» и «Всего». После этого необходимо вернуться в диалоговое окно «Таблицы сопряженности». Запуск статистического анализа осуществляется нажатием кнопки «ОК».

2. В табличном заключении программы SPSS в первой таблице представлено описание количества наблюдений с валидными (имеющимися) и пропущенными значениями. Во второй таблице содержится информация, которая использовалась для расчета необходимых критериев. В третьей таблице представлены результаты проверки статистических гипотез о наличии взаимосвязи между переменными.
 - 2.1. В первой строке третьей таблицы отражено значение хи-квадрата Пирсона, числа степеней свободы и уровня асимптотической двусторонней значимости.
 - 2.2. В строке под третьей таблицей указывается число ячеек и их доля, которые имеют ожидаемые значения менее 5, а также значение минимального ожидаемого числа наблюдений.
 - 2.3. Если количество ожидаемых наблюдений хотя бы в одной из ячеек четырехпольной таблицы составляет менее 10, для сравнения двух независимых групп номинальных данных используется критерий хи-квадрат с поправкой Йейтса на непрерывность. Значение данного критерия, числа степеней свободы и уровня значимости представлены во второй строке третьей таблицы.
 - 2.4. Если количество ожидаемых наблюдений хотя бы в одной из ячеек четырехпольной таблицы составляет менее 5, для сравнения двух независимых групп номинальных данных используется точный критерий Фишера. Точная значимость данного критерия представлена в четвертой строке третьей таблицы.
3. При представлении результатов применения критерия хи-квадрат необходимо указать значение критерия, число степеней свободы и достигнутый уровень значимости. Следует использовать формат: утвердительное выражение ($\chi^2 = \dots$, $df = \dots$, $p = \dots$).
4. Для наглядного представления в данной ситуации следует использовать столбиковую диаграмму (в заключении анализа в программе SPSS располагается под третьей таблицей).

Алгоритм сравнения двух зависимых групп номинальных данных с помощью критерия Мак-Нимара в программе IBM SPSS Statistics Version 25.0 [3]:

1. После внесения номинальных признаков в исходную форму исследователю необходимо в меню «Анализ» выбрать раздел «Непараметрические критерии», в подразделе «Устаревшие диалоговые окна» выбрать «Для двух связанных выборок». В открывшемся диалоговом окне «Критерии для двух связанных выборок» необходимо поставить флажок напротив критерия Мак-Нимара. Исследуемые признаки из списка переменных, расположенных слева, необходимо перенести в окно «Тестовые пары». Запуск статистического анализа осуществляется нажатием кнопки «ОК».
2. В табличном заключении программы SPSS в первой таблице представлено абсолютное число наблюдений выборочной совокупности с наличием и отсутствием изучаемого признака в изучаемых группах (например, наличие болевого синдрома в группах «до» и «после» оперативного вмешательства). Во второй таблице представлено число наблюдений с валидными данными по изучаемым переменным и достигнутый уровень точной двухсторонней значимости. Если достигнутый уровень значимости меньше критического значения (0,05), можно отвергнуть нулевую гипотезу об отсутствии различий между изучаемыми группами. Т.е. можно сделать вывод о наличии статистически значимых различий между изучаемыми группами.
3. При представлении результатов применения критерия Мак-Нимара следует использовать формат: утвердительное выражение ($p = \dots$).
4. Для наглядного представления в данной ситуации следует использовать столбиковую диаграмму.

Алгоритм сравнения номинальных признаков трех и более независимых групп с помощью критерия хи-квадрат Пирсона в программе IBM SPSS Statistics Version 25.0 [3]:

1. Для упрощения способа ввода номинальных признаков изучаемых групп исследователю следует ввести данные в исходную форму программы SPSS в виде многопольной таблицы (табл. 2).

Таблица 2

Пример многопольной таблицы сопряженности номинальных признаков

№ строки	Признак 1	Признак 2	Число пациентов
1	A	A	N ₁
2	A	B	N ₂
3	A	C	N ₃
4	B	A	N ₄
5	B	B	N ₅
6	B	C	N ₆
7	C	A	N ₇
8	C	B	N ₈
9	C	C	N ₉

2. После внесения данных многопольной таблицы в исходную форму исследователю необходимо в меню «Данные» выбрать раздел «Взвесить наблюдения». В открывшемся диалоговом окне «Взвесить наблюдения» необходимо указать о необходимости взвесить наблюдения, перенести переменную «Число пациентов» в строку «Взвешивающая переменная».
3. Далее в меню «Анализ» необходимо выбрать раздел «Описательные статистики», подраздел «Таблицы сопряженности». В открывшемся диалоговом окне «Таблицы сопряженности» из списка переменных, расположенных слева, необходимо независимую переменную перенести в поле «Строки», а зависимую – в поле «Столбцы». При нажатии на кнопку «Статистики» открывается диалоговое окно «Таблицы сопряженности: Статистики», в левой половине которого напротив «Хи-квадрат» необходимо поставить флажок, в блоке «Номинальная» напротив V-критерия Крамера также следует поставить флажок. В диалоговом окне «Таблицы сопряженности» следует указать необходимость вывода кластеризованных столбчатых диаграмм. Далее при нажатии на кнопку «Ячейки» открывается диалоговое окно «Таблицы сопряженности: Вывод в ячейках», в левой половине которого в блоке «Количества» можно поставить флажки напротив слов «Наблюдаемые» (позволит оценить фактическое количество наблюдений в каждой из ячеек многопольной таблицы) и «Ожидаемые» (позволит оценить ожидаемое количество наблюдений в каждой из ячеек многопольной таблицы). Для определения соответствующих долей в блоке «Проценты» необходимо отметить флажками «По строке», «По столбцу» и «Всего». После этого необходимо вернуться в диалоговое окно «Таблицы сопряженности». Запуск статистического анализа осуществляется нажатием кнопки «ОК».
4. В табличном заключении программы SPSS в первой таблице представлено описание количества наблюдений с валидными (имеющимися) и пропущенными значениями. Во второй таблице содержится информация, которая использовалась для расчета необходимых критериев. В третьей таблице представлены результаты проверки статистических гипотез о наличии взаимосвязи между переменными.
 - 4.1. В первой строке третьей таблицы отражено значение хи-квадрата Пирсона, числа степеней свободы и уровня асимптотической двусторонней значимости.
 - 4.2. В строке под третьей таблицей указывается число ячеек и их доля, которые имеют ожидаемые значения менее 5, а также значение минимального ожидаемого числа наблюдений.
 - 4.3. Если в многопольной таблице доля ячеек с ожидаемым числом наблюдений меньше 5 превысит 20% от общего числа или хотя бы одно ожидаемое значение будет меньше

единицы, то для расчета критерия хи-квадрат следует объединить группы, чтобы увеличить количество наблюдений в клетках таблицы.

- 4.4. При малых выборках предпочтительным является определение критерия хи-квадрат Пирсона с поправкой на правдоподобие. В третьей строке третьей таблицы отражено значение хи-квадрата Пирсона с поправкой на правдоподобие, числа степеней свободы и уровня асимптоматической двусторонней значимости.
- 4.5. Для определения силы взаимосвязи целесообразным является определение критерия V Крамера, значение которого и его приблизительная значимость указываются во второй строке четвертой таблицы (табл. 3).

Таблица 3

Интерпретация значений V-критерия Крамера

Значение V-критерия Крамера	Сила взаимосвязи
менее 0,1	Несущественная
[0,1 – 0,2)	Слабая
[0,2 – 0,4)	Средняя
[0,4 – 0,6)	Относительно сильная
[0,6 – 0,8)	Сильная
[0,8 – 1,0]	Очень сильная

5. При представлении результатов применения описанных критериев необходимо указать значение критерия, число степеней свободы (для хи-квадрата), достигнутый уровень значимости и силу взаимосвязи (для V-критерия Крамера).
6. Для наглядного представления в данной ситуации следует использовать столбиковую диаграмму (в заключении анализа в программе SPSS располагается под четвертой таблицей).

Алгоритм сравнения номинальных признаков трех и более зависимых групп с помощью Q-критерия Кокрена в программе IBM SPSS Statistics Version 25.0 [3]:

1. После внесения номинальных признаков в исходную форму исследователю необходимо в меню «Анализ» выбрать раздел «Непараметрические критерии», в подразделе «Устаревшие диалоговые окна» выбрать «Для K связанных выборок». В открывшемся диалоговом окне «Критерии для нескольких связанных выборок» необходимо поставить флажок напротив критерия Q Кокрена. Исследуемые признаки из списка переменных, расположенных слева, необходимо перенести в окно «Проверяемые переменные». Запуск статистического анализа осуществляется нажатием кнопки «ОК».
2. В табличном заключении программы SPSS в первой таблице представлено абсолютное число наблюдений выборочной совокупности с наличием и отсутствием изучаемого признака для каждой группы, входящей в исследование. Во второй таблице представлено число наблюдений, значение Q-критерия Кокрена, число степеней свободы и асимптотическое значение достигнутого уровня значимости. Если достигнутый уровень значимости меньше критического значения (0,05), можно отвергнуть нулевую гипотезу об отсутствии различий между изучаемыми группами. Т.е. можно сделать вывод о наличии статистически значимых различий между изучаемыми группами.
3. При представлении результатов применения Q-критерия Кокрена следует использовать формат: утвердительное выражение ($Q = \dots, df = \dots, p = \dots$).
4. При обнаружении статистически значимых различий с помощью критерия Кокрена следует провести попарное сравнение признаков с помощью критерия Мак-Нимара с поправкой Бонферрони (алгоритм сравнения описан ранее).
5. Для наглядного представления следует использовать столбиковую диаграмму.

Алгоритм сравнения порядковых признаков в случае наличия малого числа возможных категорий в программе IBM SPSS Statistics Version 25.0 [3]:

1. Для упрощения способа ввода порядковых признаков изучаемых групп исследователю следует ввести данные в исходную форму программы SPSS в виде многопольной таблицы (табл. 4).

Таблица 4

Пример многопольной таблицы сопряженности порядковых признаков

№ строки	Признак 1	Признак 2	Число пациентов
1	A	D	N ₁
2	A	E	N ₂
3	A	F	N ₃
4	B	D	N ₄
5	B	E	N ₅
6	B	F	N ₆
7	C	D	N ₇
8	C	E	N ₈
9	C	F	N ₉

2. После внесения данных многопольной таблицы в исходную форму исследователю необходимо в меню «Данные» выбрать раздел «Взвесить наблюдения». В открывшемся диалоговом окне «Взвесить наблюдения» необходимо указать о необходимости взвесить наблюдения, перенести переменную «Число пациентов» в строку «Взвешивающая переменная».
3. В меню «Анализ» следует выбрать раздел «Описательные статистики», подраздел «Таблицы сопряженности». В открывшемся диалоговом окне «Таблицы сопряженности» из списка переменных, расположенных слева, необходимо независимую переменную перенести в поле «Строки», а зависимую – в поле «Столбцы». При нажатии на кнопку «Статистики» открывается диалоговое окно «Таблицы сопряженности: Статистики», в левой половине которого напротив «Хи-квадрат» необходимо поставить флажок, в блоке «Порядковая» справа необходимо поставить флажки напротив γ -критерия Гудмана-Краскела, критериев τ -b и τ -c Кендалла, d-критерия Сомера. В диалоговом окне «Таблицы сопряженности» следует указать необходимость вывода кластеризованных столбчатых диаграмм. Далее при нажатии на кнопку «Ячейки» открывается диалоговое окно «Таблицы сопряженности: Вывод в ячейках», в левой половине которого в блоке «Количества» можно поставить флажки напротив слов «Наблюдаемые» (позволит оценить фактическое количество наблюдений в каждой из ячеек многопольной таблицы) и «Ожидаемые» (позволит оценить ожидаемое количество наблюдений в каждой из ячеек многопольной таблицы). Для определения соответствующих долей в блоке «Проценты» необходимо отметить флажками «По строке», «По столбцу» и «Всего». После этого необходимо вернуться в диалоговое окно «Таблицы сопряженности». Запуск статистического анализа осуществляется нажатием кнопки «ОК».
4. В табличном заключении программы SPSS в первой таблице представлено описание количества наблюдений с валидными (имеющимися) и пропущенными значениями. Во второй таблице содержится информация, которая использовалась для расчета необходимых критериев. В третьей таблице представлены результаты проверки статистических гипотез о наличии взаимосвязи между переменными.
 - 4.1. В первой строке третьей таблицы отражено значение хи-квадрата Пирсона, числа степеней свободы и уровня асимптотической двусторонней значимости.
 - 4.2. В четвертой строке третьей таблицы представлены результаты применения критерия хи-квадрат для линейного тренда, указана степень свободы и асимптотическая двухсторонняя значимость для данного критерия.

- 4.3. Значение критерия d Сомера представлено в четвертой таблице. Если достигнутый уровень значимости критерия меньше критического значения (0,05), гипотезу о равенстве критериев нулю можно отвергнуть.
- 4.4. Значения γ -критерия Гудмана-Краскела, критериев тау-b и тау-c Кендалла представлены в пятой таблице. Помимо абсолютных значений рассчитанных критериев SPSS рассчитывает стандартную ошибку и асимптотический уровень значимости, допуская нормальное распределение значений данных критериев. Если достигнутый уровень значимости критериев меньше критического значения (0,05), гипотезу о равенстве критериев нулю можно отвергнуть.
5. При представлении результатов применения представленных критериев необходимо указать значение конкретного критерия и достигнутый уровень значимости.
6. Для наглядного представления в данной ситуации следует использовать столбиковую диаграмму (в заключении анализа в программе SPSS располагается под пятой таблицей).

Для биомедицинских исследований наряду с проверкой гипотезы о наличии статистической связи между исследуемыми качественными признаками определен интерес составляет количественная оценка вероятности исхода, связанная с наличием фактора риска. Поэтому целесообразно рассмотреть алгоритм расчета отношения шансов и относительного риска [1, 2, 3, 4].

Отношение шансов – статистический показатель, позволяющий оценить связь между определенным исходом и фактором риска. Кроме того, результатом расчета отношения шансов является не только определение статистической значимости связи между фактором и исходом, но и ее количественная оценка.

Условия для применения расчета отношения шансов:

1. Результативные и факторные показатели должны быть измерены в номинальной шкале.
2. Отношение шансов позволяет проводить анализ только четырехпольных таблиц (фактор и исход обязательно являются бинарными переменными).
3. Сопоставляемые группы должны быть независимыми.
4. Показатель отношения шансов используется в ретроспективных исследованиях (исследования вида: «случай – контроль») [1, 2, 3, 4].

Интерпретация значения отношения шансов:

1. При значении отношения шансов более 1, фактор имеет прямую связь с вероятностью наступления исхода.
2. При значении отношения шансов менее 1, фактор имеет обратную связь с вероятностью наступления исхода.
3. При значении отношения шансов, равном 1, фактор не оказывает никакого влияния на вероятность исхода.
4. Если 95% доверительный интервал отношения шансов не включает 1, выявленная связь между фактором и исходом статистически значима при уровне значимости $p < 0,05$.
5. Если 95% доверительный интервал отношения шансов включает 1, выявленная связь между фактором и исходом статистически незначима при уровне значимости $p > 0,05$.
6. Величина 95% доверительного интервала обратно пропорциональна уровню значимости связи фактора и исхода [1, 2, 3, 4].

Алгоритм расчета отношения шансов в программе IBM SPSS Statistics Version 25.0 [3]:

1. После внесения бинарных признаков в исходную форму исследователю необходимо в меню «Анализ» выбрать раздел «Описательные статистики», подраздел «Таблицы сопряженности». В открывшемся диалоговом окне «Таблицы сопряженности» из списка переменных, расположенных слева, необходимо независимую переменную перенести в поле «Строки», а зависимую – в поле «Столбцы». При нажатии на кнопку «Статистики» открывается диалоговое окно «Таблицы сопряженности: Статистики», в правой половине которого напротив «Риск» необходимо поставить флажок. Далее при нажатии на кнопку «Ячейки» открывается диалоговое окно «Таблицы сопряженности: Вывод в ячейках», в левой половине которого в блоке «Количества» можно поставить флажки напротив слов

«Наблюдаемые» (позволит оценить фактическое количество наблюдений в каждой из ячеек четырехпольной таблицы) и «Ожидаемые» (позволит оценить ожидаемое количество наблюдений в каждой из ячеек четырехпольной таблицы). Для определения соответствующих долей в блоке «Проценты» необходимо отметить флажками «По строке», «По столбцу» и «Всего». После этого необходимо вернуться в диалоговое окно «Таблицы сопряженности». Запуск статистического анализа осуществляется нажатием кнопки «ОК».

2. В табличном заключении программы SPSS в первой таблице представлено описание количества наблюдений с валидными (имеющимися) и пропущенными значениями. Во второй таблице содержится информация, которая использовалась для расчета необходимых критериев. В третьей таблице представлены результаты оценки риска. В первой строке третьей таблицы отражено значение отношения шансов, а также 95% доверительный интервал.
3. При представлении результатов расчета отношения шансов необходимо указать значение отношения шансов, а также 95% доверительный интервал. Следует использовать формат: утвердительное выражение – $OR = \dots (95\% CI_{OR_{lower}} - OR_{upper}), p < /> 0,05$.

Относительный риск – это статистический показатель, предназначенный для сравнения вероятности исхода в зависимости от наличия фактора риска. Кроме того, результатом расчета относительного риска является не только определение статистической значимости связи между вероятностью исхода и определенным фактором риска, но и ее количественная оценка [1, 2, 3, 4].

Условия для применения расчета относительного риска:

1. Показатели фактора и исхода должны быть измерены в номинальной шкале.
2. Расчет относительного риска позволяет проводить анализ только четырехпольных таблиц (фактор и исход обязательно являются бинарными переменными).
3. Относительный риск используется в проспективных исследованиях [1, 2, 3, 4].

Интерпретация значения относительного риска:

1. При значении относительного риска более 1, исследуемый фактор повышает частоту исходов (прямая связь).
2. При значении относительного риска менее 1, исследуемый фактор снижает частоту исходов (обратная связь).
3. При значении относительного риска, равном 1, исследуемый фактор не влияет на вероятность исхода.
4. Если 95% доверительный интервал относительного риска не включает 1, выявленная связь между фактором и исходом статистически значима при уровне значимости $p < 0,05$.
5. Если 95% доверительный интервал относительного риска включает 1, выявленная связь между фактором и исходом статистически незначима при уровне значимости $p > 0,05$ [1, 2, 3, 4].

Алгоритм расчета относительного риска в программе IBM SPSS Statistics Version 25.0 [3]:

1. После внесения бинарных признаков в исходную форму исследователю необходимо в меню «Анализ» выбрать раздел «Описательные статистики», подраздел «Таблицы сопряженности». В открывшемся диалоговом окне «Таблицы сопряженности» из списка переменных, расположенных слева, необходимо независимую переменную перенести в поле «Строки», а зависимую – в поле «Столбцы». При нажатии на кнопку «Статистики» открывается диалоговое окно «Таблицы сопряженности: Статистики», в правой половине которого напротив «Риск» необходимо поставить флажок. Далее при нажатии на кнопку «Ячейки» открывается диалоговое окно «Таблицы сопряженности: Вывод в ячейках», в левой половине которого в блоке «Количества» можно поставить флажки напротив слов «Наблюдаемые» (позволит оценить фактическое количество наблюдений в каждой из ячеек четырехпольной таблицы) и «Ожидаемые» (позволит оценить ожидаемое количество наблюдений в каждой из ячеек четырехпольной таблицы). Для определения соответствующих долей в блоке «Проценты» необходимо отметить флажками «По строке», «По столбцу» и «Всего». После этого необходимо вернуться в диалоговое окно «Таблицы сопряженности». Запуск статистического анализа осуществляется нажатием кнопки «ОК».

2. В табличном заключении программы SPSS в первой таблице представлено описание количества наблюдений с валидными (имеющимися) и пропущенными значениями. Во второй таблице содержится информация, которая использовалась для расчета необходимых критериев. В третьей таблице представлены результаты оценки риска. Во второй строке третьей таблицы отражено значение относительного риска, а также 95% доверительный интервал.
3. При представлении результатов расчета относительного риска необходимо указать значение относительного риска, а также 95% доверительный интервал. Следует использовать формат: утвердительное выражение – $RR = \dots (95\% CIRR_{lower} - RR_{upper})$, $p < /> 0,05$.

Для количественного представления степени влияния фактора на исход используются и другие расчетные показатели, однако, данная работа включала описание алгоритмов расчета основных статистических показателей, что в перспективе может стать для читателя «первой ступенью» к самостоятельному изучению биомедицинской статистики [8].

Заключение. Использование подробных алгоритмов статистического анализа качественных признаков в перспективе позволит повысить уровень представления результатов биомедицинских исследований. Статистическая обработка качественных результатов исследования с помощью пакета программ «IBM SPSS Statistics Version 25.0» доступна для широкого круга исследователей, отличается своей простотой и наглядностью [9].

Список литературы:

1. Гржибовский А.М., Иванов С.В., Горбатова М.А. Анализ номинальных и ранговых переменных данных с использованием программного обеспечения STATISTICA и SPSS. Наука и здравоохранение. 2016. 6. 5-39.
2. Кочетов А.Г., Лянг О.В., Масенко В.П., Жиров И.В., Наконечников С.Н., Терещенко С.Н. Методы статистической обработки медицинских данных: Методические рекомендации для ординаторов и аспирантов медицинских учебных заведений, научных работников. М. Российский кардиологический научно-производственный комплекс. 2012.
3. Гржибовский А.М., Унгурияну Т.Н. Анализ биомедицинских данных с использованием пакета статистических программ SPSS: учебное пособие. Архангельск. Издательство Северного государственного медицинского университета. 2017.
4. Левин И.А., Манухин И.Б., Пономарева Ю.Н., Шуметов В.Г. Методология и практика анализа данных в медицине: монография. М. Тель-Авив. АПЛИТ. 2010.
5. Peacock J.L., Peacock P.J. Oxford Handbook of Medical Statistics. OxfordUniversityPress. 2011.
6. Lang T.A., Altman D.G. Basic statistical reporting for articles published in Biomedical Journals: The “Statistical Analyses and Methods in the Published Literature” or the SAMPL Guidelines. International Journal of Nursing Studies. 2015. 52 (1). 5-9. doi:10.1016/j.ijnurstu.2014.09.006.
7. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication. 2011. Available from: <http://www.icmje.org>.
8. Унгурияну Т.Н., Гржибовский А.М. Краткие рекомендации по описанию, статистическому анализу и представлению данных в научных публикациях. Экология человека. 2011. 5. 55-60.
9. Наследов А.Д. SPSS 19. Профессиональный статистический анализ данных. Санкт-Петербург. Питер. 2011.

References:

1. Grzhibovsky A.M., Ivanov S.V., Gorbatoва M.A. Analysis of nominal and rank variable data using STATISTICA and SPSS software. Nauka i zdravookhranenie. 2016. 6. 5-39.in Russian.
2. Kochetov A.G., Lyang O.V., Masenko V.P., Zhiron I.V., Nakonechnikov S.N., Tereshchenko S.N. Methods of statistical processing of medical data: Guidelines for residents and graduate students of medical schools, scientists. Moscow. Russian Cardiology Research and Production Complex. 2012.in Russian.

3. Grzhibovsky A.M., Unguryanu T.N. Analysis of biomedical data using the statistical software package SPSS: a training manual. Arkhangelsk. Publishing House of the Northern State Medical University. 2017. in Russian.
4. Levin I.A., Manukhin I.B., Ponomareva Yu.N., Shumetov V.G. Methodology and practice of data analysis in medicine: monograph. Moscow. Tel-Aviv. APLIT. 2010. in Russian.
5. Peacock J.L., Peacock P.J. Oxford Handbook of Medical Statistics. Oxford University Press. 2011.
6. Lang T.A., Altman D.G. Basic statistical reporting for articles published in Biomedical Journals: The “Statistical Analyses and Methods in the Published Literature” or the SAMPL Guidelines. International Journal of Nursing Studies. 2015. 52 (1). 5-9. doi:10.1016/j.ijnurstu.2014.09.006.
7. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication. 2011. Available from: <http://www.icmje.org>.
8. Unguryanu T.N., Grzhibovsky A.M. Brief recommendations on the description, statistical analysis and presentation of data in scientific publications. Ekologiya cheloveka. 2011. 5. 55-60. in Russian.
9. Nasledov A.D. SPSS 19. Professional statistical data analysis. Sankt-Peterburg. Piter. 2011. in Russian.