

УДК 519.233.3

ВЫБОР СТАТИСТИЧЕСКОГО КРИТЕРИЯ ДЛЯ ПРОВЕРКИ ГИПОТЕЗ

© 2008 г. А. М. Гржибовский

Национальный институт общественного здоровья, г. Осло, Норвегия

В предыдущих выпусках рубрики «Практикум» были кратко рассмотрены способы описания количественных и качественных данных [13, 20], проверка нормальности распределения количественных данных [20], статистические критерии для проверки гипотез о различиях между двумя, тремя и более независимыми группами как для случаев нормального распределения количественных данных, так и для случаев, когда данные имеют распределение, отличающееся от нормального [12, 15]. Кроме того, были рассмотрены критерии для сравнения количественных признаков для двух, трех и более групп, которые не являются независимыми, например, в исследованиях типа «до — после», когда измерение признака проводится у одной и той же группы участников исследования в разные моменты времени [18]. Основные статистические критерии, применяемые для анализа номинальных [13] и порядковых [14] данных, также были представлены. Помимо вышеперечисленного было показано, каким образом можно оценивать силу и направление взаимосвязи между двумя количественными переменными [17], а также прогнозировать значение зависимой переменной по известным значениям переменной-предиктора [19]. В данной статье предпринимается попытка представить алгоритм выбора адекватного способа описания и анализа данных в зависимости от их типа и распределения, а также от количества сравниваемых групп.

Перед тем как решить вопрос о выборе того или иного статистического критерия, следует ответить на вопрос о типе собранных данных. В принципе ответ на этот вопрос должен быть дан еще на стадии планирования исследования. Номинальные данные следует представлять в виде долей или частот, а количественные данные — с помощью мер центральных тенденций и мер рассеяния, причем какие меры центральной тенденции и какие меры рассеяния лучше применять, зависит от распределения данных и целей исследования. Так, если исследование ставит целью оценить генеральное (популяционное) среднее значение количественного признака, как в большинстве описательных исследований, а распределение признака в популяции (генеральной совокупности) подчиняется закону нормального распределения, то целесообразно представлять результат в виде среднего арифметического (M) и его 95 % доверительного интервала (ДИ). Также можно использовать 90 или 99 % ДИ в зависимости от уровня доверительной вероятности, определенного исследователями на стадии планирования. Если же оценка генерального среднего не является основной целью, как в большинстве аналитических исследований, то целесообразнее представлять данные, имеющие нормальное распределение, с помощью среднего арифметического и среднеквадратического (стандартного) отклонения (SD). Например, если бы в неоднократно упоминавшемся

В статье рассматривается алгоритм выбора статистического критерия для проверки статистических гипотез в зависимости от типа имеющихся данных и их распределения. Изложенный материал дает общие рекомендации о выборе наиболее адекватного способа анализа данных для наиболее часто встречающихся ситуаций и призван вызвать интерес читателей журнала к прочтению специализированной литературы перед началом работы над будущими публикациями.

Ключевые слова: тип данных, статистические критерии, распределение.

в рубрике «Практикум» Северодвинском когортном исследовании [20, 50] основной целью являлась оценка среднего значения массы тела новорожденных детей, то результат следовало бы представить в виде: $M = 3\,388$ (95 % ДИ: 3359–3417) г. Если же целью исследования является проверка гипотезы о различиях в массе тела детей первобеременных и повторнобеременных первородящих женщин, то описательную статистику лучше представить в виде среднего арифметического и среднеквадратического отклонения (табл. 1).

Таблица 1
Средние значения массы тела новорожденных, рожденных в срок от первородящих женщин в г. Северодвинске в 1999 году

	Масса тела новорожденных, г M (SD)
Первобеременные	3361 (427)
Повторобеременные	3432 (448)

Представлять нормально распределенные данные в виде $M \pm SD$, равно как и в традиционном виде $M \pm m$, где m — стандартная ошибка среднего арифметического, также можно, но тогда следует четко указывать, что подразумевается под каждым условным обозначением. Часто, однако, использование этого метода представления данных маскирует разброс данных и не позволяет сразу заподозрить существенные отклонения от нормальности. Конечно, можно взять квадратный корень из объема выборки и умножить его на стандартную ошибку среднего арифметического и получить таким образом среднеквадратическое отклонение, после чего умножить его на 2, затем вычесть получившееся произведение из среднего арифметического, и если результат будет отрицательным числом (чего не может быть для многих биологических показателей), то, значит, распределение смещено вправо и его представление в виде $M \pm m$ не является корректным ввиду отклонения распределения от нормального. Личный опыт показывает, что, видя данные, представленные в виде $M \pm m$, одни исследователи понимают под m стандартную ошибку среднего, другие — среднеквадратическое отклонение, а третьи — полуширину доверительного интервала. Представление данных, как рекомендуется выше, помогает избежать подобной путаницы.

Если же известно, что распределение количественного признака в генеральной совокупности не подчиняется закону нормального распределения, то представление данных в виде среднего арифметического и стандартной ошибки среднего или среднеквадратического отклонения будет некорректным. Для признаков, распределение которых значительно отличается от нормального, в качестве меры центральной тенденции целесообразнее использовать медиану (Me), а в качестве мер рассеяния — нижний (Q_1) и верхний (Q_3) квартили (25 и 75 проценти-ли), а также размах вариации, где это может быть

необходимо. Для интервальной оценки медианы ДИ можно рассчитать, как описано в [35]. Можно также попытаться преобразовать имеющиеся данные таким образом, чтобы их распределение после трансформации приближалось к нормальному. В первом выпуске «Практикума» упоминалось несколько способов преобразования данных, а также рассматривался пример трансформации данных с помощью десятичного логарифма и использования средней геометрической как меры центральной тенденции [20].

Следует помнить, что первостепенное значение имеет распределение признака в генеральной совокупности (популяции), а не в имеющейся выборке, поэтому при принятии решения о типе распределения исследуемого признака следует всегда обращаться к результатам предыдущих исследований и прочей доступной информации о распределении признака в популяции. Если же такой информации обнаружить не удастся, а выборка репрезентативна и достаточно велика, то судить о распределении признака в генеральной совокупности можно по распределению признака в имеющейся выборке [51]. Проверка нормальности распределения по выборочным данным с помощью описательной статистики, квантильных диаграмм и гистограмм, а также с помощью критериев Колмогорова — Смирнова и Шапиро — Уилка с использованием SPSS была рассмотрена в [20].

Представление номинальных данных также зависит от целей исследования. Если основной целью исследования является оценка распространенности заболевания в определенной генеральной совокупности (популяции) по данным выборочного исследования, то целесообразно представлять результаты в виде доли участников исследования с изучаемым заболеванием и 95 % ДИ для этой доли. Например, при исследовании распространенности бронхиальной астмы (БА) среди детей в возрасте 0–18 лет (объем выборки 1494 человека), проживающих в Василеостровском районе Санкт-Петербурга, была получена распространенность заболевания 7,4 (95 % ДИ: 6,2–8,8) %. Это свидетельствует о том, что при условии репрезентативности выборки можно с 95 % уверенностью говорить, что истинное значение распространенности БА среди детей, проживавших на исследуемой территории в исследуемый год, находилось в интервале от 6,2 до 8,8 % [49]. Если же основной целью исследования является лишь проверка статистических гипотез о различиях между долями или частотами, как во многих аналитических исследованиях, то, может быть, достаточно указать доли в процентах. Широко распространенная в отечественной биомедицинской литературе форма представления $M \pm m$ не является адекватной для небольших выборок, особенно для долей, значительно отличающихся от 50 %, поскольку в таких ситуациях ДИ не являются симметричными, а рассчитанные с помощью m (тоже не всегда понятно, как рассчитанной) границы доверительных интервалов могут превышать 1 (верхняя граница) или принимать отрицательные значения (нижняя граница), чего не

может быть по определению. Интересно, что во многих отечественных учебниках по статистике для медиков эта проблема освещается [32, 33, 46], но, к сожалению, пока не находит отражения во многих биомедицинских публикациях. Кроме того, часто возникает такая же путаница относительно того, что подразумевается под мерой рассеяния m , как в случае со средним арифметическим. Различные способы расчета ДИ для долей кратко представлены в [16] и детально в [48, 52].

Порядковые (ранговые) признаки можно представлять в виде долей, как номинальные признаки, например образование, а можно как количественные данные, которые не подчиняются закону нормального распределения, то есть с помощью медианы и квартилей или процентилей, например распределение оценки состояния новорожденных по шкале Апгар.

Кроме того, некоторые признаки, которые изначально являются количественными и непрерывными, можно группировать и рассматривать либо в виде порядковых, либо в виде номинальных признаков в зависимости от целей и задач исследования. Например, индекс массы тела можно использовать в исследовании как количественную переменную, а можно как порядковую (недостаточная масса тела, норма, избыточная масса тела, ожирение) и даже как дихотомическую (не превышает норму / превышает норму). Признаки же, которые изначально являются порядковыми или номинальными, в виде количественных представлять не следует. Так, представление порядковых переменных, например результаты оценки состояния новорожденных по шкале Апгар, в виде среднего арифметического и его стандартной ошибки будет некорректно.

После решения вопроса о том, как представить данные, следует приступить к выбору статистического критерия для проверки той или иной гипотезы. Для этого необходимо знать 1) данные какого типа будут использоваться для сравнений, 2) сколько групп планируется сравнивать и 3) являются ли сравниваемые группы независимыми (несвязанными). Поскольку наиболее часто в биомедицинских публикациях анализируются количественные признаки, с них и начнем. Для количественных признаков помимо того, что необходимо решить, сколько групп планируется сравнить и являются ли группы независимыми, необходимо знать распределение изучаемого признака в генеральной совокупности, а при отсутствии такой информации проверить распределение, как описано в [20]. В случае если распределение признака можно считать нормальным, а сравнить предполагается одну группу с каким-либо гипотетическим значением, то использовать следует t -критерий Стьюдента для одной выборки, который не рассматривался в предыдущих выпусках «Практикума». Для примера рассмотрим, как применить t -критерий Стьюдента для одной выборки с использованием SPSS для сравнения средней массы тела новорожденных г.

Северодвинска с гипотетическим значением. Файл Human_Ecology_1_2008.sav с оригинальными данными доступен на сайте журнала: http://www.nsmu.ru/nauka_sgmu/gio/eco_human/. Нулевой гипотезой будет являться отсутствие различий между средним значением массы тела новорожденных, полученным в результате исследования, и гипотетическим значением, скажем, 3 500 г. Альтернативной соответственно будет являться гипотеза об отличии среднего значения от 3 500 г. Для проверки этой гипотезы следует в выпадающем меню «Analyze» выбрать «Compare Means», после чего выбрать «One-Sample T-test». В появившемся диалоговом окне «One-Sample T-test» (рис. 1) нужно выделить левой кнопкой мыши в списке переменных в левой части окна интересующую нас переменную «ves» и путем нажатия левой кнопкой мыши на «>» переместить переменную в поле «Test Variable(s)». После этого в поле «Test Value» нужно ввести то гипотетическое значение, с которым предстоит произвести сравнение, в данном случае 3 500 (см. рис. 1). По умолчанию SPSS рассчитывает среднее арифметическое (Mean), среднеквадратическое отклонение (Std. Deviation) и стандартную ошибку среднего арифметического (Std. Error Mean) для анализируемой переменной (рис. 2).

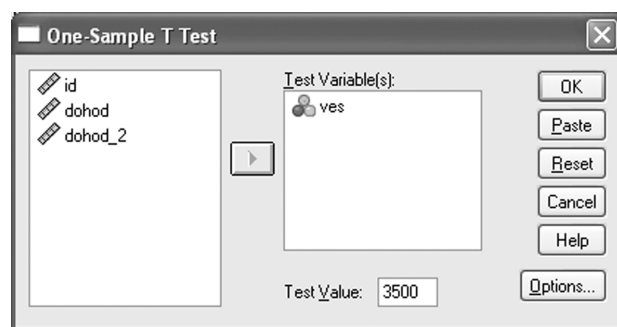


Рис. 1. Диалоговое окно «One-Sample T-test»

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
ves	869	3388,20	435,806	14,784

Рис. 2. Описательная статистика для переменной «ves»

One-Sample Test						
	Test Value = 3500					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
ves	-7.563	868	.000	-111.804	-140.82	-82.79

Рис. 3. Результат применения одностороннего t -критерия Стьюдента

На рис. 3 представлен результат применения одностороннего t -критерия Стьюдента с указанием гипотетического значения, с которым проводилось сравнение (Test Value = 3500), значение критерия (столбец t), число степеней свободы (столбец df), достигнутый уровень значимости для двустороннего

теста (столбец Sig 2-Tailed), а также среднее значение разности между выборочным средним арифметическим и гипотетическим значением, с которым производится сравнение (Mean Difference), а также нижняя и верхняя границы 95 % ДИ для полученной разности. На основании результатов можно отвергнуть нулевую гипотезу о равенстве между средним значением массы тела новорожденных в г. Северодвинске и гипотетическим значением 3 500 г. Северодвинские новорожденные в среднем на 112 г легче, причем можно сказать с 95 % уверенностью, что средняя разность находится в интервале от 83 до 141 г. При представлении результатов одностороннего *t*-критерия рекомендуется представлять значение критерия, число степеней свободы и значение достигнутого уровня значимости (в данном случае $t_{(868)} = -7,56, p < 0,001$).

Если необходимо сравнить две независимые выборки (например, средние значения изучаемого признака для мужчин и женщин), данные в которых подчиняются закону нормального распределения, то необходимо применить *t*-критерий Стьюдента для непарных (независимых, несвязанных) выборок, который обозначен в SPSS как Independent Samples T-test и рассматривался в [13]. Следует помнить, что *t*-критерий не предназначен для сравнения трех и более групп между собой. При проведении множественных попарных сравнений с помощью *t*-критерия необходимо применять поправку Бонферрони, меняя тем самым критический уровень значимости. Например, если имеются 3 группы и проводятся все возможные сравнения (1–2, 2–3, 1–3), то статистически значимыми можно будет считать только те различия, для которых достигнутый уровень значимости (*p*) будет менее 0,017, а не 0,05. Подробнее о проблеме множественных сравнений можно прочитать в [15]. Там же можно кратко ознакомиться с однофакторным дисперсионным анализом (One-Way ANOVA) и его применением с помощью SPSS. Именно дисперсионный анализ, а не критерий Стьюдента целесообразно применять для сравнения средних при наличии трех или более независимых групп. При применении критерия Стьюдента и дисперсионного анализа для независимых групп следует помнить, что помимо нормальности распределения необходимо еще соблюдение условия равенства дисперсий, которое можно проверить в SPSS с помощью критерия Левена (Levene test).

Если же сравниваемые группы не являются независимыми, как в исследованиях типа «до — после» или «случай — контроль» по методу подобранных пар, но распределение признака подчиняется закону нормального распределения, то следует использовать либо *t*-критерий Стьюдента для парных выборок (Paired *t*-test), либо однофакторный дисперсионный анализ повторных наблюдений (Repeated Measures ANOVA). Парный критерий Стьюдента следует применять, если сравниваются только две группы, в

то время как дисперсионный анализ повторных наблюдений можно использовать при наличии трех и более групп. Из дополнительных условий, которые должны выполняться, следует упомянуть условие нормального распределения разности между значениями признака для парного критерия Стьюдента и условие сферичности для дисперсионного анализа повторных наблюдений. Краткое описание применения обоих методов с использованием SPSS дано в [18].

Если известно, что распределение изучаемого количественного признака в генеральной совокупности отличается от нормального, или в отсутствие такой информации данные выборки четко на это указывают, то тогда необходимо применять непараметрические критерии, которые не предполагают подчинения данных какому-либо закону распределения, в том числе и нормальному. SPSS не дает возможности прямо, то есть по методу «выбери и кликни», сравнить медиану исследуемой выборки с гипотетическим значением, как в случае с критерием Стьюдента для одной выборки. Однако такое сравнение в SPSS все-таки возможно, только для этого придется создать новую переменную, значения которой для всех наблюдений будут равны гипотетическому числу, с которым производится сравнение. После этого можно проводить сравнение изучаемой переменной и вновь созданной переменной с помощью критерия Вилкоксона для парных выборок (Wilcoxon signed rank test), как представлено в [18].

Если необходимо сравнить две независимые совокупности, данные в которых не подчиняются закону нормального распределения, можно воспользоваться критерием Манна — Уитни, который иногда называется критерием Вилкоксона для непарных выборок (Wilcoxon rank sum test) [12]. Если же сравниваемых групп больше, то можно, конечно, проводить попарные сравнения с помощью критерия Манна — Уитни, предварительно заменив традиционный критический уровень значимости на новый, который во столько раз меньше традиционного, сколько производится сравнений, хотя более целесообразно воспользоваться критерием Краскела — Уоллиса [15].

Если же необходимо сравнить количественные признаки, не подчиняющиеся закону нормального распределения, в связанных группах, то при наличии двух групп следует воспользоваться критерием Вилкоксона (Wilcoxon signed rank test), а если групп больше, то критерием Фридмана (Friedman test) [18].

Общий алгоритм выбора статистического критерия для анализа количественных данных представлен на рис. 4.

Если задачей исследователя является сравнение номинальных переменных, прежде чем выбрать статистический критерий, также необходимо определиться с количеством сравниваемых групп. Если имеется всего одна выборка, размер которой составляет как минимум несколько десятков наблюдений, и планируется сравнить частоту встречаемости признака



Рис. 4. Алгоритм выбора статистического критерия для анализа количественных данных

или заболевания с заданным значением (например, оценить, является ли распространенность астмы у детей Василеостровского района Санкт-Петербурга выше или ниже, например, 5 %), то можно применить критерий Z, о котором подробно написано в большинстве пособий по статистике, например в [32]. Расчет критерия несложен и может быть произведен «вручную» с помощью формулы:

$$Z = \frac{p - \pi}{\sqrt{\pi \cdot (1 - \pi)}} \cdot \sqrt{n},$$

где Z — абсолютное значение критерия, p — выборочное значение доли, полученное в результате исследования, π — заданная доля, с которой планируется провести сравнение, а n — объем выборки, на основании которой было получено значение p. После подставления всех значений в формулу необходимо сравнить Z с критическим значением, которое для традиционного уровня значимости 0,05 составляет 1,96, если $p > \pi$, и -1,96, если $p < \pi$ (читатель уже, вероятно, догадался, что используется аппроксимация нормального распределения). Если $Z > 1,96$ или $Z < -1,96$, то нулевую гипотезу об отсутствии различий между полученной в результате исследования частотой и заданным значением можно отвергнуть. Применим критерий Z для вышеописанного примера с распро-

страненностью астмы у детей Василеостровского района Санкт-Петербурга, для которого нулевой будет гипотеза об отсутствии различий между выборочным значением распространенности астмы и заданным значением 5 %. Альтернативной соответственно будет гипотеза о наличии таких различий. Подставив фактические данные в формулу, получим:

$$Z = \frac{0,074 - 0,05}{\sqrt{0,05 \cdot (1 - 0,05)}} \cdot \sqrt{1464} = 13,3.$$

Поскольку 13,3 значительно превышает критическое значение (1,96), можно отвергнуть нулевую гипотезу о том, что распространенность бронхиальной астмы среди детей Василеостровского района Санкт-Петербурга равна 5 %. Более точное значение уровня значимости для полученных значений Z можно найти в таблицах в любом учебнике по статистике. Расчет данного критерия вручную представлен для того, чтобы читатель мог проверять гипотезы с помощью критерия Z по материалам публикаций на основании представленных значений p и π .

Критерий Z можно также применять с помощью SPSS. Для примера рассмотрим гипотетическое исследование, проведенное на выборке объемом 32 человека и посвященное оценке частоты встречаемости фактора риска в исследуемой группе до

вмешательства (переменная T1), сразу после вмешательства (переменная T2) и через 6 месяцев после вмешательства (переменная T3). Данные сохранены в файле «Human_Ecology_11_2008.sav» на веб-сайте журнала http://www.nsmu.ru/nauka_sgmu/rto/eco_human/. Допустим, что нам необходимо сравнить распространенность фактора риска в исследуемой группе до вмешательства с гипотетическим значением 50 %. Для этого в выпадающем меню «Analyze» необходимо выбрать «Nonparametric Tests», а затем «Binomial». В открывшемся диалоговом окне «Binomial Test» (рис. 5) нужно выделить левой кнопкой мыши в списке переменных в левой части окна интересующую нас переменную «T1» и путем нажатия левой кнопкой мыши на «>» переместить переменную в поле «Test Variable List». После этого в поле «Test Proportion» нужно ввести то гипотетическое значение, с которым предстоит произвести сравнение, в данном случае 0,5 (см. рис. 5).

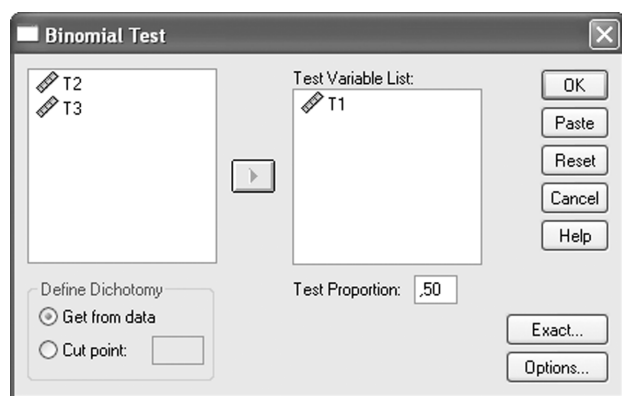


Рис. 5. Диалоговое окно «Binomial Test»

Запуск анализа осуществляется нажатием на кнопку «OK» (см. рис. 5) левой кнопкой мыши. Результаты применения критерия Z представлены на рис. 6. В первом столбце обозначены как Group 1 и Group 2 имеющиеся категории, которые в столбце «Category» представлены как «FR-» и «FR+». Первая обозначает отсутствие изучаемого фактора риска, а вторая — его наличие. В следующих столбцах (N и Observed Prop.) представлено абсолютное количество участников исследования без и с наличием фактора риска и их доли соответственно. Далее указана доля, с которой будет производиться сравнение, в данном случае 0,5. В последнем столбце представлен достигнутый уровень значимости, который для данного примера равен 0,377, что говорит о том, что нулевую гипотезу об отсутствии различий между выборочным и заданным значениями отвергнуть нельзя.

Binomial Test					
	Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
T1	Group 1	19	,59	,50	,377 ^a
	Group 2	13	,41		
	Total	32	1,00		

a. Based on Z Approximation.

Рис. 6. Результаты применения критерия Z

Если предполагается сравнение двух независимых групп номинальных данных, то можно воспользоваться критерием хи-квадрат Пирсона. Кроме того, этот критерий можно применять и при сравнении большего количества групп, однако следует помнить, что в ситуациях, когда сравниваемых групп две и они малы, лучше пользоваться критерием хи-квадрат с поправкой Йейтса на непрерывность (Yates' continuity correction), если количество ожидаемых наблюдений хотя бы в одной из ячеек менее 10, или точным критерием Фишера (Fisher's exact test), если количество ожидаемых наблюдений хотя бы в одной из ячеек менее 5. Каким образом определять количество ожидаемых наблюдений, а также как рассчитывать вышеупомянутые критерии с помощью SPSS, было кратко представлено в [13].

Если же группы не являются независимыми, то при наличии двух групп можно воспользоваться критерием Мак-Нимара (MacNemar test), а если групп больше, то Q-критерием Кокрена (Cochran's Q-test). Вернемся к файлу «Human_Ecology_11_2008.sav» и предположим, что нам необходимо сравнить доли участников исследования с наличием фактора риска до вмешательства и сразу после него. Поскольку это одна и та же выборка, в которой измерения производятся в различные моменты времени, необходимо производить сравнения с помощью критерия Мак-Нимара, применение критерия хи-квадрат было бы в данной ситуации некорректным. Для применения критерия Мак-Нимара следует в выпадающем меню «Analyze» выбрать «Descriptive Statistics», затем «Crosstabs», после чего в открывшемся диалоговом окне «Crosstabs» нужно выбрать сравниваемые переменные (в данном случае T1 и T2) и поместить их в соответствующие поля, как показано на рис. 7. Читатель может заметить, что это то же самое диалоговое окно, что и для расчета критерия хи-квадрат, как описывалось в [13].

Отличие заключается лишь в том, что в меню «Statistics», открываемом путем нажатия на одноименную кнопку в нижней части окна «Crosstabs», нужно отметить только критерий Мак-Нимара, как показано на рис. 8.

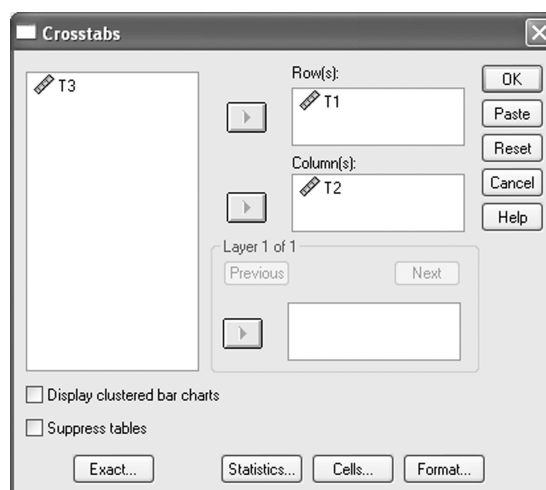


Рис. 7. Диалоговое окно «Crosstabs»

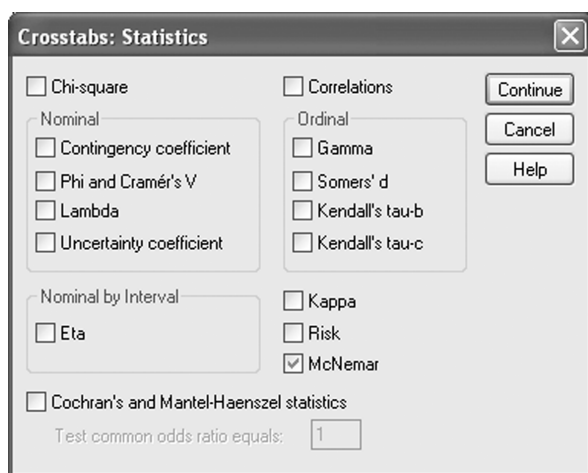


Рис. 8. Окно «Crosstabs: Statistics» с отмеченным критерием Мак-Нимара

Результаты применения критерия Мак-Нимара показаны на рис. 9. Несмотря на то, что над таблицей, которую выдает SPSS, написано Chi-Square Test, результат в таблице рассчитан с помощью критерия Мак-Нимара. В столбце «Exact Sig (2-sided)» представлен точный уровень значимости с учетом биномиального распределения признака. Напомним, что для расчета критерия Z использовалась нормальная аппроксимация. Результат показывает, что нулевую гипотезу о различиях между частотой встречаемости фактора риска в исследуемой группе до и сразу после вмешательства отвергнуть нельзя, так как уровень значимости ($p = 0,227$) выше критического ($0,05$).

Chi-Square Tests		
	Value	Exact Sig. (2-sided)
McNemar Test		,227 ^a
N of Valid Cases	32	

a. Binomial distribution used.

Рис. 9. Результат применения критерия Мак-Нимара

Если необходимо сравнить частоты в трех связанных группах (и более), то лучше применять Q-критерий Кокрена. Для этого в выпадающем меню «Analyze» нужно выбрать «Nonparametric Tests», затем «K-Related Samples», после чего в открывшемся диалоговом окне «Tests for Several Related Samples» следует выбрать из списка переменных в левой области сравниваемые переменные и перенести их в правую область, также нужно убрать отмеченный по умолчанию критерий Фридмана и отметить «Cochran's Q».

Результаты применения критерия Кокрена представлены на рис. 11. В таблице показано количество человек в выборке, которые сравниваются между собой (N), абсолютное значение критерия Q, количество степеней свободы (df) и асимптотическое (приблизительное) значение уровня значимости (Asymp. Sig.).

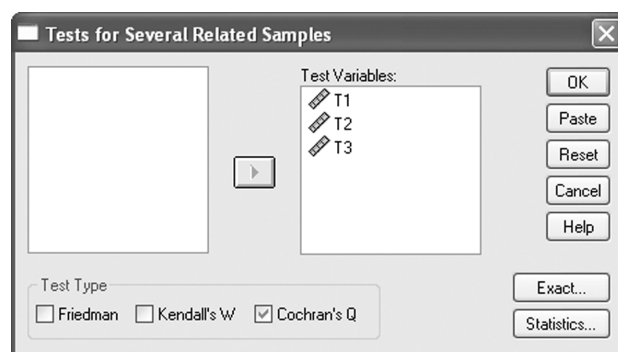


Рис. 10. Диалоговое окно «Tests for Several Related Samples»

Test Statistics	
N	32
Cochran's Q	2,375 ^a
df	2
Asymp. Sig.	,305

a. 0 is treated as a success.

Рис. 11. Результаты применения критерия Кокрена

Критерий Кокрена, как и критерии Краскела — Уоллиса, Фридмана, хи-квадрат, и т. д., является глобальным критерием, то есть он проверяет, есть ли различия между тремя или более группами, но не сообщает, где эти различия. Поэтому при обнаружении статистически значимых различий с помощью критерия Кокрена можно провести попарные сравнения с помощью критерия Мак-Нимара с поправкой Бонферрони. В данном случае нельзя отвергнуть нулевую гипотезу об отсутствии различий между группами ($p = 0,305$), поэтому дальнейший анализ можно не проводить.

Алгоритм выбора адекватного статистического критерия для номинальных данных представлен на рис. 12.

Что касается порядковых (ранговых) признаков, то их можно анализировать как количественные переменные, которые не подчиняются закону нормального распределения, особенно если имеется много возможных категорий, как, например, при анализе оценки состояния здоровья новорожденных по шкале Апгар. Если рангов мало (как, например, в переменной образование), можно использовать критерии, предназначенные для сравнения номинальных переменных. И те и другие примеры встречаются в литературе.

Представлением алгоритмов для выбора статистических критериев мы заканчиваем знакомить читателей журнала с наиболее простыми способами обработки данных, которые могут применяться в биомедицинских исследованиях с использованием пакета статистических программ SPSS. Следует помнить, что представленные алгоритмы и рассматриваемые в «Практикуме» критерии [12–20] являются далеко не единственными критериями для проверки статистических гипотез как для количественных, так и для качественных данных. Также хотелось бы отметить,

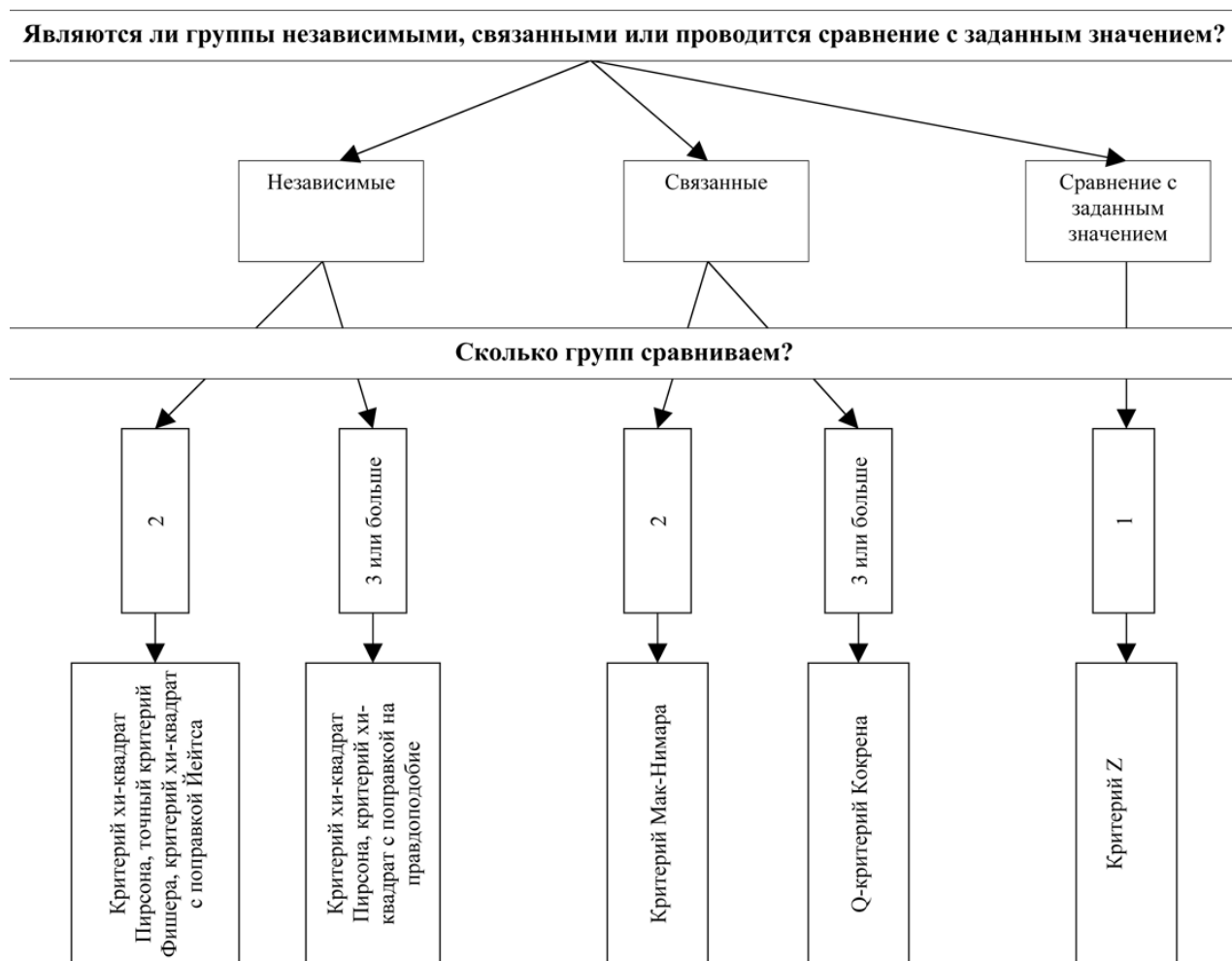


Рис. 12. Алгоритм выбора статистического критерия для анализа номинальных данных

что информации, представленной в «Практикуме», достаточно для более или менее обдуманного «нажимания на кнопки», но недостаточно для профессионального статистического анализа, поэтому читателям настоятельно рекомендуется ознакомиться с литературой по статистике. В настоящее время существует большое количество литературы по биостатистике для начинающих исследователей [3, 5, 11, 24, 25, 36], однако большинство этих книг являются переводными и представляют материал в покоем на рубрику «Практикум» виде, то есть максимально упрощенно. В отечественных источниках материал зачастую представлен более подробно, с теоретическим обоснованием и с большей математической составляющей, что может отпугнуть читателей без математической подготовки [10, 23, 31, 32, 37, 40, 45]. Из англоязычных источников можно порекомендовать [46, 51], которые написаны простым языком, но вместе с тем предоставляют несколько больше информации, чем литература по биостатистике для начинающих. Многие книги описывают те или иные критерии с привязкой к определенному программному обеспечению. Для пользователей программы STATISTICA можно рекомендовать [37], MS Excel – [10], SPSS – [5, 34, 41, 47]. Встречаются

и не очень удачные примеры [24]. Для детального изучения статистики применимо к биомедицинским исследованиям лучше всего ознакомиться с статистической литературой, опубликованной еще в советское время [1, 2, 4, 6, 9, 22, 26–31, 33, 38, 39, 42, 44]. Кроме того, очень полезными как для критического осмысления научной литературы, так и для подготовки собственных исследований являются книги по доказательной медицине и эпидемиологии [7, 8, 21, 43].

Заключительный выпуск «Практикума» будет посвящен анализу наиболее типичных ошибок применения и представления результатов статистического анализа в публикациях.

Список литературы

1. Айвазян С. А. Прикладная статистика : Основы моделирования и первичная обработка данных : справочное издание / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. — М. : Финансы и статистика, 1983. — 471 с.
2. Аптон Г. Анализ таблиц сопряженности / Г. Аптон. — М. : Мир, 1982. — 126 с.
3. Банержи А. Медицинская статистика понятным языком: вводный курс / А. Банержи. — М. : Практическая медицина, 2007. — 287 с.

4. Браунли К. А. Статистическая теория и методология в науке и технике / А. К. Браунли. — М. : Наука, 1980. — С. 376–417.
5. Бююль А. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей / А. Бююль П. Цёфель. — СПб. : Диасофт, 2002. — 608 с.
6. Владимирский Б. М. Математические методы в биологии / Б. М. Владимирский. — Ростов, 1983. — 304 с.
7. Власов В. В. Введение в доказательную медицину / В. В. Власов. — М. : Медиа Сфера, 2001. — 392 с.
8. Власов В. В. Эпидемиология : учебное пособие для вузов / В. В. Власов. — М. : ГЭОТАР-МЕД, 2004. — 464 с.
9. Гаек Я. Теория ранговых критериев / Я. Гаек, З. Шидак. — М. : Наука, 1971. — 376 с.
10. Герасимов А. Н. Медицинская статистика / А. Н. Герасимов. — М. : Медицинское информационное агентство, 2007. — 475 с.
11. Гланц С. Медико-биологическая статистика / С. Гланц. — М. : Практика, 1998. — 459 с.
12. Гржибовский А. М. Анализ количественных данных для двух независимых групп / А. М. Гржибовский // Экология человека. — 2008. — № 2. — С. 54–61.
13. Гржибовский А. М. Анализ номинальных данных (независимые наблюдения) / А. М. Гржибовский // Экология человека. — 2008. — № 6. — С. 58–68.
14. Гржибовский А. М. Анализ порядковых данных / А. М. Гржибовский // Экология человека. — 2008. — № 8. — С. 56–62.
15. Гржибовский А. М. Анализ трех и более групп количественных данных / А. М. Гржибовский // Экология человека. — 2008. — № 3. — С. 50–58.
16. Гржибовский А. М. Доверительные интервалы для частот и долей / А. М. Гржибовский // Экология человека. — 2008. — № 5. — С. 57–60.
17. Гржибовский А. М. Корреляционный анализ / А. М. Гржибовский // Экология человека. — 2008. — № 9. — С. 50–60.
18. Гржибовский А. М. Одномерный анализ повторных измерений / А. М. Гржибовский // Экология человека. — 2008. — № 4. — С. 51–60.
19. Гржибовский А. М. Однофакторный линейный регрессионный анализ / А. М. Гржибовский // Экология человека. — 2008. — № 10. — С. 55–64.
20. Гржибовский А. М. Типы данных, проверка распределения и описательная статистика / А. М. Гржибовский // Экология человека. — 2008. — № 1. — С. 52–58.
21. Гринхальх Т. Основы доказательной медицины / Т. Гринхальх. — М. : ГЭОТАР-МЕД, 2004. — 240 с.
22. Гублер Е. В. Применение непараметрических критериев статистики в медико-биологических исследованиях / Е. В. Гублер, А. А. Генкин. — Л. : Медицина, 1973. — 141 с.
23. Елисеева И. И. Общая теория статистики / И. И. Елисеева, М. М. Юзбашев. — М. : Финансы и статистика, 1998. — 368 с.
24. Жижин К. С. Медицинская статистика : учебное пособие / К. С. Жижин. — Ростов н/Д : Феникс, 2007. — 160 с.
25. Зайцев В. М. Прикладная медицинская статистика / В. М. Зайцев, В. Г. Лифляндский, В. И. Маринкин. — СПб. : Фолиант, 2003. — 428 с.
26. Кант В. И. Математические методы и моделирование в здравоохранении / В. И. Кант. — М. : Медицина, 1987. — 224 с.
27. Кендалл М. Многомерный статистический анализ и временные ряды / М. Кендалл, А. Стьюарт. — М. : Наука, 1976. — 736 с.
28. Кендалл М. Статистические выводы и связи / М. : Кендалл, А. Стьюарт. — М. : Наука, 1973. — 899 с.
29. Кендалл М. Теория распределений / М. Кендалл, А. Стьюарт. — М. : Наука, 1966. — 588 с.
30. Кендэл М. Ранговые корреляции / М. Кендэл. — М. : Статистика, 1975. — 214 с.
31. Лакин Г. Ф. Биометрия / Г. Ф. Лакин. — М. : Высшая школа, 1990. — 350 с.
32. Медик В. А. Математическая статистика в медицине / В. А. Медик, М. С. Токмачев. — М. : Финансы и статистика, 2007. — 798 с.
33. Миркин Б. Г. Анализ качественных признаков / Б. Г. Миркин. — М. : Статистика, 1976. — 168 с.
34. Наследов А. Д. SPSS: Компьютерный анализ данных в психологии и социальных науках / А. Д. Наследов. — СПб. : Питер, 2007 — 416 с.
35. Орлов А. И. Непараметрическое точечное и интервальное оценивание характеристик распределения / А. И. Орлов // Заводская лаборатория. — 2004. — Т. 70, № 5. — С. 65–70.
36. Петри А. Наглядная статистика в медицине / А. Петри, К. Сэбин. — М. : ГЭОТАР-МЕД, 2003. — 144 с.
37. Реброва О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA / О. Ю. Реброва. — М. : Медиа Сфера, 2002. — 312 с.
38. Рокицкий П. Ф. Биологическая статистика / П. Ф. Рокицкий. — Минск : Вышэйшая школа, 1973. — 320 с.
39. Сепетлиев Д. Статистические методы в научных медицинских исследованиях / Д. Сепетлиев. — М. : Медицина, 1968. — 419 с.
40. Сергиенко В. И. Математическая статистика в клинических исследованиях / В. И. Сергиенко, И. Б. Бондарева. — М. : ГЭОТАР-МЕД, 2001. — 256 с.
41. Таганов Д. SPSS: Статистический анализ в маркетинговых исследованиях / Д. Таганов. — СПб. : Питер, 2005. — 192 с.
42. Урбах В. Ю. Биометрические методы. Статистическая обработка опытных данных в биологии, сельском хозяйстве и медицине / В. Ю. Урбах. — М. : Наука, 1964. — 416 с.
43. Флетчер Р. Клиническая эпидемиология: Основы доказательной медицины / Р. Флетчер, С. Флетчер, Э. Вагнер. — М. : Медиа Сфера, 1998. — 345 с.
44. Холлендер М. Непараметрические методы статистики / М. Холлендер, Д. Вулф. — М. : Финансы и статистика, 1983. — 518 с.
45. Юнкеров В. И. Медико-статистическая обработка данных медицинских исследований / В. И. Юнкеров, С. Г. Григорьев. — СПб. : ВмедА, 2002. — 266 с.
46. Dawson B. Basic and clinical biostatistics / B. Dawson, R. G. Trapp. — Lange Medical Books : McGraw-Hill, 2001. — 399 p.
47. Field A. Discovering statistics using SPSS / A. Field. — SAGE Publications, 2005. — 779 p.
48. Garcia-Perez M. A. On the confidence interval for the binomial parameter / M. A. Garcia-Perez // Quality and quantity. — 2005. — N 39. — P. 467–481.
49. Glushkova A. V. Prevalence and correlates of asthma among children in central St. Petersburg, Russia: a cross-sectional study / A. V. Glushkova, A. M. Grjibovski // Croatian Medical Journal. — 2008, in press.

50. *Grjibovski A. M.* Social variations in fetal growth in Northwest Russia: an analysis of medical records / A. M. Grjibovski, L. O. Bygren, B. Svartbo, P. Magnus // *Annals of Epidemiology*. — 2003. — N 9. — P. 599–605.

51. *Motulsky H.* Intuitive biostatistics // H. Motulsky. — Oxford : Oxford University Press, 1995. — 386 p.

52. *Newcombe R. G.* Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods / R. G. Newcombe // *Statistics in Medicine*. — 1998. — N 17. — P. 857–872.

CHOOSING A STATISTICAL TEST FOR HYPOTHESIS TESTING

A. M. Grjibovski

National Institute of Public Health, Oslo, Norway

The article presents an algorithm for choosing a statistical test for hypothesis testing depending on type and distribution

of the data. The article provides only general recommendations on how to choose an adequate method of data analysis for the most common situations. The readers are encouraged to consult statistical literature prior to analysing own data and preparing manuscripts.

Key words: type of data, statistical tests, distribution.

Контактная информация:

Гржибовский Андрей Мечиславович — старший советник Национального института общественного здоровья, г. Осло, Норвегия

Адрес: Nasjonalt folkehelseinstitutt, Pb 4404 Nydalen, 0403 Oslo, Norway

Тел.: +47 21076392, +47 45268913; e-mail: angr@fhi.no

Статья поступила 20.10.2008 г.