

## АНАЛИЗ НОМИНАЛЬНЫХ ДАННЫХ (НЕЗАВИСИМЫЕ НАБЛЮДЕНИЯ)

© 2008 г. А. М. Гржибовский

Национальный институт общественного здоровья, г. Осло, Норвегия

*В статье рассматривается сравнение номинальных данных с помощью критерия хи-квадрат Пирсона, критерия хи-квадрат с поправкой на правдоподобие, критерия хи-квадрат с поправкой Йейтса на непрерывность, а также точного критерия Фишера. Для оценки силы взаимосвязи между номинальными переменными рассматриваются критерий  $\phi$  и критерий  $V$  Крамера, коэффициент сопряженности Пирсона,  $\gamma$ -критерий Гудмана – Крaskела, коэффициент неопределенности Тейла, а также относительный риск и отношение шансов. Изложенный материал дает общие сведения о статистических критериях, применяемых для проверки гипотез о равенстве частот в сравниваемых группах, а также силе взаимосвязи между номинальными переменными и призван вызвать интерес читателей журнала к прочтению специализированной литературы перед началом работы над будущими публикациями.*

**Ключевые слова:** номинальные данные, анализ, сила взаимосвязи, SPSS.

В предыдущем выпуске журнала были рассмотрены способы расчета доверительных интервалов для частот и долей [4]. В данной статье будут представлены статистические критерии для проверки гипотез о значимости различий между частотами, а также способы оценки силы взаимосвязи между качественными переменными, измеряемыми на номинальной и порядковой шкалах.

Интересно отметить, что отечественные исследователи наиболее часто сравнивают частоты и доли с помощью критерия Стьюдента, в то время как за рубежом для таких сравнений чаще всего используется критерий хи-квадрат ( $\chi^2$ ) Пирсона. Причина, вероятно, кроется в простоте применения критерия Стьюдента, слабой информированности исследователей об ограничениях применения данного критерия, а также в том, что именно с него начинается описание сравнения частот и долей в некоторых, например [8], отечественных пособиях по биостатистике для исследователей-медиков. Следует отметить, что большинство авторов учебников и пособий указывают, что нормальная аппроксимация биномиального распределения актуальна лишь при наличии больших выборок и при частотах, близких к 0,5 [6],

однако исследователями это в большинстве случаев игнорируется. Игнорирование ограничений дает излишне приближенные результаты и может приводить к обнаружению различий там, где их нет, так как оценка ошибки частоты по формуле  $m = \sqrt{(pq/n)}$  дает слишком «оптимистичные» результаты для ситуаций, когда частота события меньше 0,25 или больше 0,75 [7, 10, 11]. Несмотря на то, что в некоторых пособиях сообщается, что свободным от подобного рода ограничений, а значит и более универсальным, является способ проверки равенства частот, основанный на угловом преобразовании Фишера [7], он используется отечественными исследователями редко. Кстати говоря, во многих зарубежных пособиях по биостатистике для медиков применение аппроксимации нормального распределения, а также критерия Стьюдента для анализа качественных признаков вообще не упоминается, вероятно, чтобы не запутывать начинающих исследователей [12, 13, 15].

Кроме того, сравнение частот с помощью критерия Стьюдента возможно только для четырехпольных таблиц, то есть только в ситуациях, когда для каждой из изучаемых переменных может быть только два возможных значения (например, есть фактор риска/нет фактора риска, выжил/умер). В биомедицинских исследованиях нередки ситуации, когда объемы выборок и/или частоты событий очень малы, а также когда качественные переменные могут принимать более двух значений. Более универсальными способами сравнения частот и долей являются способы, основанные на идее сравнения фактических частот, полученных в результате исследования, с ожидаемыми частотами. К таким способам анализа качественных переменных относится критерий согласия  $\chi^2$  Пирсона, который свободен от вышеперечисленных ограничений.

Критерий  $\chi^2$  для таблиц сопряженности был предложен Карлом Пирсоном (1857–1936) еще в 1900 году (цит. по [1]). С помощью данного критерия оценивается значимость различий между фактическим (выявленным в результате исследования) количеством исходов или качественных характеристик выборки, попадающих в каждую категорию, и теоретическим количеством, которое можно ожидать в изучаемых группах при справедливости нулевой гипотезы. Для применения критерия  $\chi^2$  Пирсона необходимо соблюдение следующих условий:

1. Номинальные или порядковые данные (возможно создание категорий из непрерывных данных).
2. Независимость наблюдений (отбор участников исследования из генеральной совокупности производится независимо друг от друга).

3. Независимость групп (метод нельзя применять для исследований типа «до – после»).
4. Ожидаемое (не фактическое) число наблюдений в любой из ячеек должно быть не менее 5 (или 10) для четырехпольных таблиц.
5. Доля ячеек с ожидаемым числом наблюдений менее 5 не должна превышать 20 % для многопольных таблиц.
6. Для расчета критерия  $\chi^2$  используются только абсолютные фактические и ожидаемые числа (проценты и доли для расчетов не используются).

Рассмотрим принцип метода с помощью простого примера. Предположим, что проводится проспективное наблюдение за участниками исследования (когорты из 48 человек), причем у 23 из них имеется изучаемый фактор риска, а у остальных 25 этого фактора риска нет. После определенного времени изучаемый исход наблюдали у 10 человек (43,5 %) из тех, у кого имелся фактор риска, и у 4 человек (16,0 %) из тех, у кого изучаемого фактора риска не было. Результаты исследования можно отобразить в виде четырехпольной таблицы (табл. 1).

Для ответа на вопрос о наличии статистической взаимосвязи между фактором риска и исходом с помощью критерия  $\chi^2$  следует сначала рассчитать ожидаемое количество наблюдений в каждой из ячеек при условии справедливости нулевой гипотезы об отсутствии взаимосвязи.

*Таблица 1*

**Пример таблицы сопряженности**

|                       | Исход есть (1) | Исхода нет (0) | Всего    |
|-----------------------|----------------|----------------|----------|
| Фактор риска есть (1) | 10 (A)         | 13 (B)         | 23 (A+B) |
| Фактора риска нет (0) | 4 (C)          | 21 (D)         | 25 (C+D) |
| Всего                 | 14             | 34             | 48       |

Ожидаемое количество наблюдений для каждой ячейки рассчитывается путем перемножения сумм рядов и столбцов (маргинальных итогов) с последующим делением полученного произведения на общее число наблюдений. Так, для ячейки А (находится в первом ряду и в первом столбце) ожидаемое число будет равно  $23 \cdot 14 / 48$  (итог ряда)  $\cdot 14$  (итог столбца)  $/ 48$  (объем выборки)  $= 23 \cdot 14 / 48 = 6,7$ . Для ячейки С ожидаемое число будет равно  $14 \cdot 25 / 48 = 7,3$ . Для оставшихся ячеек эти значения будут равны 16,3 (В) и 17,7 (D). Как видно из расчетов, ожидаемые значения необязательно целые числа.

Затем рассчитывается значение критерия  $\chi^2$  по формуле:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

где  $i$  – номер ряда (строки, от 1 до  $r$ ),  $j$  – номер столбца (от 1 до  $c$ ),  $O_{ij}$  – фактическое количество наблюдений в ячейке  $ij$ ,  $E_{ij}$  – ожидаемое число наблюдений в ячейке  $ij$ . После подставления чисел в формулу получим:

$$\chi^2 = \frac{(10-6,7)^2}{6,7} + \frac{(13-16,3)^2}{16,3} + \frac{(4-7,3)^2}{7,3} + \frac{(21-17,7)^2}{17,7} = 4,378$$

Затем значение критерия  $\chi^2$  сравнивается с критическими значениями для  $(r - 1) \cdot (c - 1)$  числа степеней свободы по таблицам, которые имеются в большинстве пособий по статистике. Для данного примера число степеней свободы равно  $(2 - 1) \cdot (2 - 1)$ , то есть 1. Для 1 степени свободы (а значит, для всех четырехпольных таблиц) критическое значение критерия равно 3,841 при уровне значимости 0,05. Наше значение (4,378) превышает критическое, значит, на основании применения критерия  $\chi^2$  Пирсона нулевая гипотеза об отсутствии статистической взаимосвязи между изучаемым фактором риска и исходом может быть отвергнута при критическом уровне значимости 5 %. В целом, чем больше различия между фактическими и ожидаемыми числами в каждой из ячеек таблицы, тем больше будет значение критерия и тем меньше будет значение достигнутого уровня значимости ( $p$ ). При равенстве ожидаемых и фактических чисел значение критерия будет равно 0, а  $p = 1$ .

Хотелось бы подчеркнуть, что речь идет только о статистической взаимосвязи, поэтому выводы о наличии либо причинно-следственных, либо столь любимых многими авторами «достоверных» связей только на основании статистически значимых результатов были бы некорректны.

Вычисленное значение критерия  $\chi^2$  изменяется скачкообразно, так как основывается на частотах, которые являются целыми числами. В то же время табличные значения для распределения  $\chi^2$  составлены для непрерывной шкалы, поэтому в 1934 году английский статистик Фрэнк Йейтс (Frank Yates, 1902–1994) предложил поправку на непрерывность [19], которая сейчас известна под названием поправки Йейтса (Yates's correction). Поправка заключается в вычитании 0,5 из абсолютного значения разности между фактическим и ожидаемым количеством наблюдений в каждой ячейке, что ведет к уменьшению величины критерия:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(|O_{ij} - E_{ij}| - 0,5)^2}{E_{ij}}$$

Практически во всех отечественных пособиях по биостатистике отмечается, что применение поправки Йейтса целесообразно. В других оговаривается, что ее применение необходимо при небольших объемах выборки и/или при количестве ожидаемых наблюдений в любой из ячеек  $< 10$  или  $< 5$ . В третьих считается, что поправку на непрерывность следует применять всегда. В нашем примере в двух из четырех ячеек ожидаемое число наблюдений  $< 10$ , значит, применение поправки согласно большинству учебников целесообразно. Однако не все статистики согласны с необходимостью применять эту поправку, так как было показано, что она может приводить к получению заниженных значений критерия, а значит, увеличивать вероятность ошибки второго типа, то есть вероятность не найти различия там, где они есть. Уже через несколько лет после опубликования Йейтсом работы о применении поправки на непрерывность целесообразность ее применения была оспорена другим известным английским статистиком Эгоном Пирсоном (Egon Pearson, 1895–1980, сын Карла Пирсона) и другими исследователями [11, 16, 17]. В 1970–1980 годах число критических публикаций, основанных на результатах компьютерных симуляций, достигло нескольких десятков [9, 14]. В некоторых зарубежных учебниках прямо сообщается, что применение поправки Йейтса нецелесообразно ввиду излишней ее консервативности [13], однако в большинстве других пособий она по-прежнему рекомендуется, несмотря на то, что авторы обсуждают ее консервативность. С одной стороны, такая консервативность хороша, так как уменьшается вероятность ошибки 1 типа, то есть обнаружения различий там, где их нет. Такая консервативность могла бы предотвратить применение более дорогого, но не более эффективного метода лечения. Ведь логично предположить, что если более дорогой метод оказался бы более эффективным, многие сочли бы внедрение этого метода в практику целесообразным. С другой стороны, консервативная оценка уменьшает статистическую мощность (чувствительность), а значит, вероятность обнаружить различия там, где они на самом деле есть, тоже уменьшается, что может остановить разработку потенциально более эффективного лечения, особенно на ранних этапах его разработки, если строго следовать дихотомическому подходу к принятию или отвержению нулевой гипотезы.

При наличии больших выборок различия в значениях критерия  $\chi^2$ , получаемых с использованием поправки Йейтса и без нее незначительны, однако при малых

выборках различия могут быть существенными. Следует помнить, что поправка Йейтса применяется только для четырехпольных таблиц, то есть при анализе двух дихотомических переменных. Для нашего примера значение критерия  $\chi^2$  с поправкой Йейтса составляет 3,149, что меньше критического значения (3,841), а значит, мы не можем отвергнуть нулевую гипотезу об отсутствии статистической связи между фактором риска и исходом.

Использование поправки привело к получению противоположного результата! Данная ситуация весьма показательна, так как имеет прямое отношение к гипертрофированной вере многих исследователей в величину  $p$ , а точнее, в дихотомичность ее значений (либо  $p > 0,05$ , либо  $p < 0,05$ ), что может привести на основании одних и тех же данных к противоположным выводам всего исследования в зависимости от учебника, который читали (или какому больше верят) исследователи. В такой ситуации хотелось бы порекомендовать исследователям не заикливаться на значении 0,05 и уж тем более не докладывать результаты применения только того критерия, который помогает «найти то, что очень хотелось». Далее в статье будут рассмотрены методы, позволяющие смотреть более широко на результаты сравнения качественных переменных. Но прежде рассмотрим альтернативы критерию  $\chi^2$ .

Одной из таких альтернатив является расчет отношения правдоподобия ( $\Lambda\chi^2$ ), или критерия  $\chi^2$  с поправкой на правдоподобие. Расчет основан на методе максимального правдоподобия, при котором оценка неизвестного параметра производится путем максимизации функции правдоподобия. Расчет  $\Lambda\chi^2$  производится по формуле:

$$\Lambda\chi^2 = 2 \cdot \sum_{i=1}^r \sum_{j=1}^c O_{ij} \cdot \ln\left(\frac{O_{ij}}{E_{ij}}\right)$$

после чего полученные значения критерия  $\chi^2$  сравниваются с табличными значениями, как описано выше. При больших выборках значения  $\Lambda\chi^2$  и  $\chi^2$  приблизительно равны. При малых выборках значение  $\Lambda\chi^2$  обычно несколько меньше, а потому считается некоторыми авторами предпочтительнее [13]. Для нашего примера

$$\Lambda\chi^2 = 2 \cdot \left[ 10 \cdot \ln\left(\frac{10}{6,7}\right) + 13 \cdot \ln\left(\frac{13}{16,3}\right) + 4 \cdot \ln\left(\frac{4}{7,3}\right) + 21 \cdot \ln\left(\frac{21}{17,7}\right) \right] = 4,473,$$

что также превышает критическое значение  $\chi^2$  для 1 степени свободы (3,841), а значит, позволяет отвергнуть нулевую гипотезу.

Все вышеперечисленные методы дают приблизительную (асимптотическую, asymptotic) оценку вероятности распределения чисел по ячейкам таблицы так, как было получено в результате исследования, если бы была верна нулевая гипотеза об отсутствии взаимосвязи между фактором риска и исходом. Точную (exact) вероятность для всевозможных четырехпольных таблиц с совпадающими маргинальными итогами можно рассчитать с помощью точного критерия Фишера (Fisher's exact test) по формуле

$$P = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{A!B!C!D!N!},$$

где ! – факториал, который равен произведению числа на последовательность чисел, каждое из которых меньше предыдущего на 1. Например, факториал числа 5 рассчитывается как  $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ . Для нашего примера

$$P = \frac{(10+13)!(4+21)!(10+4)!(13+21)!}{10!13!4!21!48!} = 0,057,$$

что на уровне доверительной вероятности 95 % (критический уровень значимости 5 %) не позволяет отвергнуть нулевую гипотезу об отсутствии статистической взаимосвязи между фактором риска и исходом.

Этот метод вызывает меньше споров, чем поправка Йейтса, хотя некоторыми исследователями также высказываются сомнения в целесообразности его применения для малых выборок ввиду его консервативности [16]. Большинство статистиков, однако, по-прежнему придерживается мнения, что точный критерий Фишера следует применять при количестве ожидаемых наблюдений  $<5$  (некоторые говорят о числе  $<10$ ) в любой из ячеек четырехпольной таблицы. Более того, некоторые исследователи рекомендуют применять этот критерий даже в ситуациях, когда объем выборки равен нескольким сотням [15]. Использование компьютеров позволяет рассчитывать точную вероятность и при больших выборках, однако по мере увеличения числа наблюдений значение  $p$ , полученное с помощью точного критерия Фишера, будет приближаться к таковому, полученному с помощью критерия  $\chi^2$ . Необходимые условия для применения точного критерия Фишера соответствуют условиям для применения критерия  $\chi^2$  за исключением пунктов 4 и 5, подразумевается также

гипергеометрическое распределение значения в левой верхней ячейке четырехпольной таблицы, чего мы проверить не можем.

Итак, два из четырех статистических критериев говорят о том, что нулевую гипотезу можно отвергнуть, а два других – наоборот. Читатели, вероятно, помнят, что значение уровня значимости ( $p$ ) во многом зависит от объема выборки. Даже сильную статистическую связь сложно выявить при малом числе наблюдений, в то время как при больших выборках даже слабая и клинически маловажная связь становится статистически значимой. Поэтому ошибочно было бы делать вывод о силе взаимосвязи между переменными только на основании достигнутого уровня значимости, а также сравнивать по значениям  $p$  силу взаимосвязи между признаками в совокупностях с разным числом наблюдений.

Большинство редакций зарубежных журналов рекомендует авторам не только представлять в научных работах достигнутые уровни значимости при проверке статистических гипотез, но и оценивать величину эффекта (effect size), то есть силу связи между признаками [18].

Критерии, оценивающие силу связи между номинальными переменными, могут принимать значения от 0 до 1. Они не могут иметь отрицательных значений, так как данные, измеряемые на номинальной шкале, не имеют порядкового отношения, что не позволяет изучать направление зависимости. Учитывая, что данная публикация ориентирована в основном на пользователей программного пакета SPSS, рассматриваться будут только те опции, которые предлагает этот пакет, что, разумеется, не означает, что других способов оценки величины эффекта не существует.

Критерий  $\phi$  (фи, phi) предназначен для взаимосвязи только для четырехпольных таблиц. Для многопольных таблиц целесообразнее применять критерий V Крамера (Cramer's V). Значения обоих критериев варьируют от 0 до 1 (за исключением критерия  $\phi$  для многопольных таблиц, поэтому для них его применение и не рекомендуется). Оба критерия основаны на критерии  $\chi^2$  и могут быть рассчитаны вручную по формулам:

$$\phi = \sqrt{\frac{\chi^2}{n}} \quad \text{и} \quad V = \sqrt{\frac{\chi^2}{n \cdot (r-1) \cdot (c-1)}}$$

Для четырехпольных таблиц значения обоих критериев будут совпадать (0,302 для нашего примера). Если интерпретировать полученные значения критериев  $\phi$  и V Крамера согласно рекомендациям Rea & Parker (табл. 2), то можно сказать, что имеется средней силы связь между фактором риска и исходом.



Таблица 2

**Интерпретация значений критериев  $\phi$  и V Крамера согласно рекомендациям Rea & Parker**

| Значение критериев<br>$\phi$ или V Крамера | Сила взаимосвязи     |
|--|----------------------|
| <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        |

Коэффициент сопряженности представляет собой меру оценки силы взаимосвязи, основанной на критерии  $\chi^2$ . Зарубежные исследователи чаще применяют коэффициент сопряженности Пирсона (С), в то время как в отечественных пособиях [5] сообщается, что для малых таблиц (не более 5 x 5) более точную оценку дает критерий Чупрова (К), который в зарубежной литературе фигурирует как Tshuprow's T. Расчет коэффициентов сопряженности может быть выполнен вручную по формулам

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad \text{и} \quad K = \sqrt{\frac{\chi^2}{n \sqrt{(r-1) \cdot (c-1)}}},$$

где n – объем выборки, r – количество рядов (строк), c – количество столбцов, а  $\chi^2$  – значение критерия хи-квадрат. Коэффициенты сопряженности принимают значения от 0 (нет взаимосвязи) до значений, приближающихся к 1, но не достигающих ее (сильная взаимосвязь). Максимально возможное значение С зависит от размера таблицы, поэтому для симметричных таблиц можно вручную рассчитать нормированное или скорректированное значение С' по формуле Sakoda:

$$C' = \frac{C}{\sqrt{\frac{r-1}{r}}},$$

где r – количество рядов (или столбцов, так как формула предназначена только для симметричных таблиц). В знаменателе рассчитывается максимально возможное значение С. Для таблицы 2 x 2, как в нашем примере, максимально возможное значение С равно 0,707, фактическое значение С = 0,289, а нормированное значение С' = 0,409 (или 0,41). Нормированное значение С и критерий Чупрова в SPSS не рассчитываются.

Следующие два критерия отнести к мерам силы взаимосвязи признаков можно лишь условно, но они рассматриваются здесь потому, что SPSS предоставляет возможность рассчитывать их при анализе номинальных переменных. Критерий  $\lambda$  Гудмана – Краскела основан на принципе относительного уменьшения ошибки при прогнозировании значений зависимой переменной с помощью независимой переменной. Критерий  $\lambda$  принимает значения от 0 до 1, где 0 означает, что наличие информации о независимой переменной никак не улучшает возможности прогнозирования значений зависимой переменной. Вручную  $\lambda$ -критерий Гудмана – Краскела рассчитывается по формуле

$$\lambda = \frac{\sum f_i - f_d}{n - f_d},$$

где  $f_i$  – наибольшие числа в ячейках в каждом из классов независимой переменной;  $f_d$  – наибольший из маргинальных итогов (сумм) зависимой переменной, а  $n$  – объем выборки. Для нашего примера зависимой переменной является исход, так как мы хотим прогнозировать его на основании наличия или отсутствия фактора риска. Значит,

$$\lambda = \frac{(13 + 21) - 34}{48 - 34} = 0,$$

это означает, что знания о наличии фактора риска не уменьшают ошибки предсказания исхода. В отличие от всех рассматриваемых выше,  $\lambda$ -критерий асимметричен, то есть его значение зависит от того, какая переменная является зависимой, а какая независимой. Если бы мы хотели прогнозировать наличие фактора риска, зная исход, то значение критерия было бы равно 0,261, то есть знание исхода снизило бы количество неверно предсказанных значений для фактора риска на 26,1 %. Данный критерий очень чувствителен к значениям маргинальных итогов для независимой переменной.

Попробуем взглянуть на прогнозирование исхода с другой стороны. Наиболее вероятным будет предположение о том, что переменная «исход» будет равна «0» (нет исхода), так как она принимает значение «0» в 34 случаях из 48, или в 70,8 % случаев. Таким образом, если мы ничего не знаем о факторе риска, наиболее вероятным будет предположение об отсутствии исхода, однако в 14 случаях из 48 (29,2 %) это предположение будет неверным. Для повышения вероятности правильного прогноза попробуем учесть значения переменной «фактор риска». Из 23 человек, у кого имелся фактор риска, 13 имели значение исхода «0», а 10 – «1» (есть исход). Значит, можно

предположить, что значение переменной «исход» будет 0 и при этом ошибиться в 10 случаях. Для тех, у кого фактора риска не было, также вероятнее предположить, что исхода не будет (21 из 25), но при этом ошибиться в 4 случаях. Таким образом, для выборки в 48 человек получим  $10 + 4 = 14$  ошибочно классифицированных случаев, или 29,2 %. Несложно увидеть, что первоначальная доля неверно классифицированных исходов (29,2 %) не уменьшилась, так как доля неверно классифицированных исходов при наличии информации о факторе риска также составляет 29,2 %. Учитывая суть критерия, можно записать  $\lambda$  как

$$\lambda = \frac{\text{Ошибка1} - \text{Ошибка2}}{\text{Ошибка1}},$$

где ошибка 1 – доля неверно предсказанных значений зависимой переменной без учета значений независимой переменной; ошибка 2 – доля неверно предсказанных значений зависимой переменной с учетом значений независимой переменной. Таким образом, наглядно видно, что  $\lambda$  показывает процент снижения ошибок прогнозирования при наличии информации о независимой переменной.

Последним критерием, который предлагает SPSS для оценки силы взаимосвязи, является коэффициент неопределенности (uncertainty coefficient), в литературе он также встречается как коэффициент энтропии (entropy coefficient) или энтропийный коэффициент Тейла (Theils U-coefficient). Коэффициент неопределенности имеет несколько отличные от  $\lambda$ -критерия Гудмена – Краскела теоретические обоснования, но также относится к коэффициентам, показывающим значение относительного уменьшения ошибки. Обычно считают, что коэффициент неопределенности предпочтительнее  $\lambda$ -критерия Гудмена – Краскела. Значение коэффициента неопределенности может варьировать от 0 до 1 и интерпретируется так же, как и  $\lambda$ -критерий Гудмена – Краскела. Учитывая название коэффициента, говорят, что его значение отражает степень неточности прогноза. Значение 0 говорит о том, что зависимую переменную невозможно предсказать по значениям независимой переменной, а 1 – о том, что значения первой полностью предсказываются значениями второй. Расчет коэффициента Тейла более сложен, поэтому формула не приводится.

Итак, мы рассмотрели способы проверки гипотез о наличии статистической связи между номинальными переменными, а также способы оценки силы взаимосвязи между этими переменными. Тем не менее сообщение о том, что была обнаружена статистически значимая связь средней силы между фактором риска и исходом, для исследователей в области биомедицинских наук, заинтересованных в практическом

применении результатов исследования, недостаточно информативно. Гораздо продуктивнее было бы говорить о количественной оценке вероятности исхода, связанной с наличием фактора риска. Однако не все исследования позволяют говорить о риске и оценивать вероятность возникновения исхода в зависимости от наличия или отсутствия фактора риска. Подробнее об этом можно прочитать в литературе по эпидемиологии (например, [3]). Мы же остановимся только на некоторых расчетах, применимых к нашему примеру. Учитывая, что наше гипотетическое исследование было проспективным, мы можем рассчитать относительный риск (Relative Risk, RR). Поскольку в примере ничего не сообщается о времени наблюдения, но подразумевается, что оно было одинаковым для обеих групп (с наличием фактора риска и без него), относительный риск будет равен отношению рисков. Отношение рисков отражает, во сколько раз риск исхода при наличии фактора риска выше риска исхода при отсутствии фактора риска и рассчитывается применимо к табл. 1 следующим образом:

$$RR = \frac{\frac{A}{A+B}}{\frac{C}{C+D}} = \frac{A \cdot (C+D)}{C \cdot (A+B)} = \frac{10 \cdot (4+21)}{4 \cdot (10+13)} = 2,72,$$

что говорит о том, что фактор риска может увеличивать вероятность возникновения исхода в 2,7 раза или что риск исхода у тех, у кого есть фактор риска, в 2,7 раза выше, чем у тех, у кого фактора риска нет. Такой результат гораздо более информативен. Однако различия в 2,7 раза справедливы только для нашей выборочной совокупности. Даже если допустить, что наша выборка репрезентативна, систематические ошибки отсутствуют, а влияние вмешивающихся факторов (конфаундеров) минимально, относительный риск для генеральной совокупности может отличаться, поэтому всегда рекомендуется представлять интервальную оценку относительного риска с помощью 95 % доверительного интервала. Этот интервал представляет собой область, в которую попадает истинное значение доли в 95 % случаев. Другими словами, можно с 95 % надежностью сказать, что истинное значение частоты встречаемости признака в генеральной совокупности будет находиться в пределах 95 % доверительного интервала. Методы расчета доверительного интервала для частот и долей рассматривались в предыдущем номере журнала [4]. Для относительного риска 95 % доверительный интервал можно рассчитать по формуле:

$$x = \ln(RR) + 1,96 \cdot \sqrt{\frac{B}{A \cdot (A+B)} + \frac{D}{C \cdot (C+D)}}$$

Верхняя граница:  $e^x$ , где

$$\text{Нижняя граница: } e^x, \text{ где } x = \ln(RR) - 1,96 \cdot \sqrt{\frac{B}{A \cdot (A+B)} + \frac{D}{C \cdot (C+D)}}$$

а  $e$  – основание натурального логарифма (число Эйлера  $\sim 2,7$ ). Для данного примера можно с 95 % уверенностью сказать, что относительный риск будет находиться в промежутке от 1,0 до 7,5. Значительная ширина доверительного интервала вызвана малым объемом выборки. Хотелось бы предостеречь читателей от переоценки важности относительного риска. Например, относительный риск может быть равен 2,0 как в ситуации, когда абсолютные риски развития заболевания равны 1 на 1 000 000 и 2 на 1 000 000, а также 1 на 10 и 2 на 10. В первом случае абсолютная разница рисков будет не очень важна, так как составит 0,000 001. Во втором же разница рисков составит 0,1. Если взять обратные величины из полученных разностей рисков, то можно будет увидеть у скольких человек необходимо устранить фактор риск, чтобы предотвратить 1 исход. В первом случае надо устранить фактор риска у миллиона, а во втором – всего у 10 человек. Особенно актуальными такие расчеты становятся при оценке эффективности лечебного вмешательства. Рассчитанная величина будет называться числом пациентов, подвергаемых лечению, на один предотвращенный неблагоприятный исход (в англоязычной литературе NNT – Number Needed to Treat).

Для нашего примера тоже можно рассчитать разность рисков:  $A(A+B)/C(C+D)$ , которая будет равна 0,275, или 27,5 %, а для того, чтобы предотвратить один исход, необходимо устранить фактор риска всего у 4 человек (NNT  $\sim 3,6$ ), что говорит о том, что потенциальный эффект от профилактических мер, направленных на устранение изучаемого фактора риска, очень велик (при условии, что распространенность фактора риска в генеральной совокупности такая же, как и в выборке).

Если бы наше исследование было типа «случай – контроль», было бы неверным рассчитывать относительный риск. В таких исследованиях в качестве меры эффекта выступает отношение шансов (Odds Ratio, OR). Представим на минуту, что наше исследование было исследованием типа «случай – контроль». Тогда

$$OR = \frac{A \cdot D}{B \cdot C} = \frac{10 \cdot 21}{13 \cdot 4} = 4,04,$$

то есть шансы на изучаемый исход были в 4 раза выше у тех участников исследования, у кого имелся фактор риска, чем у тех, у кого фактора риска не было. При проецировании результатов на генеральную совокупность также необходимо рассчитать 95 % доверительный интервал, в который попадут значения от

$$e^{\ln(OR)-1,96 \cdot \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}}} \quad \text{до} \quad e^{\ln(OR)+1,96 \cdot \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}}}$$

где  $e$  – основание натурального логарифма. Для нашего примера 95 % значений отношения шансов (ОШ) будут попадать в интервал от 1,0 до 15,6. Следует помнить, что вышеприведенные формулы для расчета доверительных интервалов предназначены только для независимых данных и неприменимы в исследованиях типа «до – после», а также в исследованиях типа «случай – контроль» по методу подобранных пар (Matched case-control study). Не стоит представлять в одном исследовании и относительный риск, и отношение шансов в одном и том же исследовании. Для исследований типа «случай – контроль» описанные выше расчеты относительного риска, разницы рисков и NNT провести невозможно.

Итак, рассчитанные значения критериев  $\phi$ ,  $V$  Крамера, коэффициентов сопряженности и особенно относительного риска позволяют заподозрить, что изучаемый фактор риска может оказать значительное влияние на вероятность возникновения исхода даже при значении пресловутого  $p > 0,05$ , который мы получим, если применим точный критерий Фишера согласно рекомендациям Н. Motulsky [15] или поправку Йейтса, как рекомендуется многими другими авторами. Если бы мы ограничились проверкой статистической гипотезы при критическом уровне значимости 5 % и дихотомическом подходе к трактовке результатов, то пришлось бы принять нулевую гипотезу об отсутствии связи между фактором риска и исходом и сделать вывод о безвредности изучаемого фактора. Еще интереснее, если бы исследование было повторено другими на выборке хотя бы в 70 человек. Тогда (при прочих равных условиях) по причине большей статистической мощности достигнутой уровень значимости был бы  $< 0,05$ , а значит, при аналогичном подходе к величине  $p$  вывод был бы противоположным, хоть величина эффекта была бы та же самая!

Рассуждения о том, что проверка статистических гипотез сообщает только часть информации, были опубликованы Пирсоном еще в 1901 году, продолжены Фишером и, наконец, нашли свое выражение в современных рекомендациях, сначала американской

психологической ассоциации [18], а затем и других, где четко говорится о том, что помимо результатов статистических тестов необходимо представлять меры силы взаимосвязи между изучаемыми факторами.

Еще более грамотной стратегией является принятие решения еще на этапе планирования исследования, то есть задолго до начала сбора данных, о том, какие значения относительного риска или отношения шансов будут считаться клинически важными, после чего рассчитывается необходимый объем выборки. Уже на этом этапе может стать очевидным, что некоторые исследования проводить нецелесообразно по причине невозможности набрать достаточное количество участников исследования (проб, лабораторных животных и т. д.) для того, чтобы выявить статистически значимые различия на желаемом уровне. Личный опыт показывает, что в большинстве случаев рассчитанное необходимое количество наблюдений значительно превышает то количество, которое изначально задумывалось исследователями или имеется в наличии.

Расчет всех вышеописанных критериев и коэффициентов с помощью пакета прикладных статистических программ SPSS довольно прост, и мы рассмотрим его с помощью данных вышеприведенного примера, сохраненных в файле Human\_Ecology\_2008\_6.sav, которые можно «скачать» с сайта журнала. Для анализа номинальных данных следует в меню «Analyze» выбрать «Descriptive Statistics», затем «Crosstabs». В открывшемся диалоговом окне «Crosstabs» независимую переменную (в данном случае фактор риска, FR, перенести в поле «Row» (ряд), а зависимую (исход, Ishod) в поле «Column» (столбец) как показано на рис. 1. Оставшаяся переменная id показывает идентификационный номер участника исследования. Затем, нажав на «Statistics», можно открыть окно «Crosstabs: Statistics», в левой половине которого поставить флажки напротив критерия  $\chi^2$  (Chi-square), коэффициента сопряженности (contingency coefficient),  $\phi$  и V Крамера (Phi and Cramer's V),  $\lambda$  Гудмана – Краскела (Lambda) и коэффициента неопределенности (Uncertainty coefficient). В правой нижней области окна «Crosstabs: Statistics» можно поставить флажок напротив оценки риска (Risk), что позволит рассчитать и относительный риск, и отношение шансов (рис. 2). Критерии, которые остались неотмеченными, будут рассмотрены в последующих выпусках журнала.

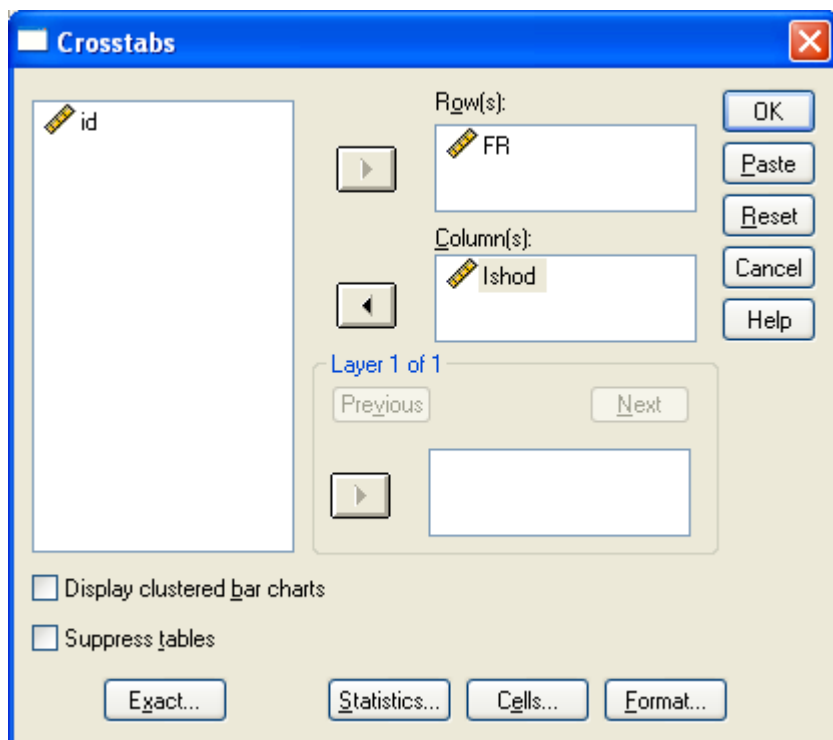


Рис. 1. Диалоговое окно «Crosstabs»

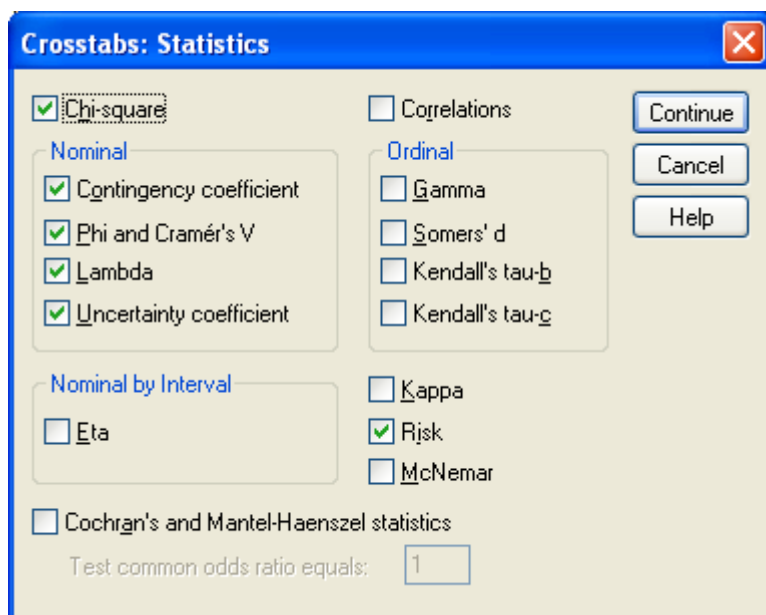


Рис. 2. Диалоговое окно «Crosstabs: Statistics»

Для того чтобы компьютер рассчитал относительный риск и отношение шансов правильно, необходимо удостовериться, что наличие признака (фактора риска, исхода) кодируется меньшим числом, чем отсутствие признака (фактора риска, исхода). Именно поэтому в файле Human\_Ecology\_2008\_6.sav наличие признака кодируется как 1, а отсутствие как 2, хотя во всех остальных случаях рекомендуется отсутствие признака кодировать как 0. Вернуться к диалоговому окну «Crosstabs» можно, нажав на кнопку «Continue». Далее следует открыть окно «Crosstabs: Cell Display» (рис. 3) путем



нажатия на кнопку «Cells», после чего можно поставить флажки напротив слов «Observed» и «Expected». Первое позволит увидеть фактическое количество наблюдений в каждой из ячеек четырехпольной таблицы, а второе – ожидаемое количество. Также можно отметить флажками «Row», «Column» и «Total» для получения долей, после чего следует вернуться к окну «Crosstabs» путем нажатия кнопки «Continue». Запуск расчетов осуществляется кнопкой «ОК».

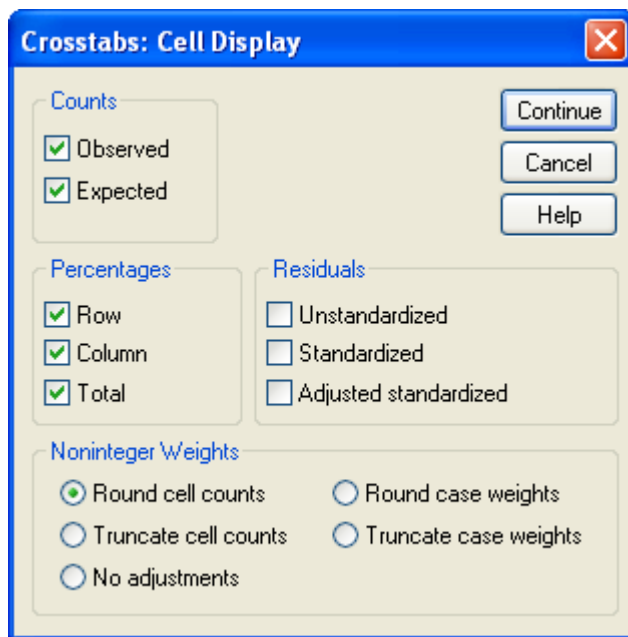


Рис. 3. Диалоговое окно «Crosstabs: Cell Display»

Результаты запрашиваемых расчетов SPSS выдаст в виде шести таблиц. Первая (рис. 4) представляет собой описание количества наблюдений с имеющимися (столбец Valid) и пропущенными (столбец Missing) значениями (последние в анализ не включаются).

**Case Processing Summary**

|            | Cases |         |         |         |       |         |
|------------|-------|---------|---------|---------|-------|---------|
|            | Valid |         | Missing |         | Total |         |
|            | N     | Percent | N       | Percent | N     | Percent |
| FR * Ishod | 48    | 100,0%  | 0       | ,0%     | 48    | 100,0%  |

Рис. 4. Данные об общем количестве наблюдений и количестве пропущенных значений

Вторая – четырехпольная таблица сопряженности с маргинальными итогами (Total). Она содержит всю информацию (рис. 5), которая использовалась для расчета всех критериев, рассмотренных в данной статье. Count обозначает фактическое количество наблюдений в каждой из ячеек, а Expected Count – ожидаемое, причем результаты совпадают с расчетами вручную, которые представлены в начале статьи.

Например, из 23 человек с имеющимся фактором риска у 10 (43,5 %) наблюдался изучаемый исход, а у 13 (56,5 %) – нет. Ожидаемые же значения в ячейках А и В составили 6,7 и 16,3 (наименование ячеек как в табл. 1).

**FR \* Ishod Crosstabulation**

|       |   |                | Ishod  |        | Total  |
|-------|---|----------------|--------|--------|--------|
|       |   |                | 1      | 2      |        |
| FR    | 1 | Count          | 10     | 13     | 23     |
|       |   | Expected Count | 6,7    | 16,3   | 23,0   |
|       |   | % within FR    | 43,5%  | 56,5%  | 100,0% |
|       |   | % within Ishod | 71,4%  | 38,2%  | 47,9%  |
|       |   | % of Total     | 20,8%  | 27,1%  | 47,9%  |
|       | 2 | Count          | 4      | 21     | 25     |
|       |   | Expected Count | 7,3    | 17,7   | 25,0   |
|       |   | % within FR    | 16,0%  | 84,0%  | 100,0% |
|       |   | % within Ishod | 28,6%  | 61,8%  | 52,1%  |
|       |   | % of Total     | 8,3%   | 43,8%  | 52,1%  |
| Total |   | Count          | 14     | 34     | 48     |
|       |   | Expected Count | 14,0   | 34,0   | 48,0   |
|       |   | % within FR    | 29,2%  | 70,8%  | 100,0% |
|       |   | % within Ishod | 100,0% | 100,0% | 100,0% |
|       |   | % of Total     | 29,2%  | 70,8%  | 100,0% |

Рис. 5. Таблица сопряженности

Результаты проверки статистических гипотез о наличии взаимосвязи между переменными представлены в третьей таблице (рис. 6). В первой строке можно увидеть значение (Value) критерия  $\chi^2$  (Pearson Chi-Square) с указанием числа степеней свободы (df, degrees of freedom) и уровнем значимости  $p$  для двустороннего теста (Asymp. Sig. (2-sided)), который равен 0,036. При представлении результатов применения критерия  $\chi^2$  рекомендуется указывать значение критерия, количество степеней свободы и достигнутый уровень значимости. Для данного примера результаты могут выглядеть так:  $\chi^2_{(1)} = 4,38$ ,  $p = 0,036$ . Результаты односторонних тестов использовать не рекомендуется за исключением редких случаев, когда именно односторонний тест был запланирован еще до начала сбора данных и зафиксирован в протоколе исследования.

Во второй строке можно увидеть значение критерия  $\chi^2$  с поправкой Йейтса на непрерывность (Continuity Correction) с указанием числа степеней свободы и уровнем значимости (0,076). В третьей представлены результаты применения критерия  $\chi^2$  с поправкой на правдоподобие (Likelihood Ratio). В четвертой видим результаты для точного критерия Фишера (Fisher's Exact Test). Для него нет значения критерия и числа степеней свободы, так как суть точного критерия Фишера заключается в прямом

расчете вероятности распределения признаков по ячейкам таблицы так, как было получено в результате исследования, если бы была верна нулевая гипотеза об отсутствии взаимосвязи между фактором риска и исходом. Результаты, представленные в пятой строке, будут рассматриваться в следующем выпуске журнала. Для того чтобы помочь исследователю ориентироваться в результатах и выбрать подходящий критерий для проверки гипотез, в самой нижней строке под таблицей говорится, сколько ячеек (cells, и их доля в скобках) имеет ожидаемые значения (expected count) <5. Кроме того, сообщается значение минимального ожидаемого числа наблюдений (minimum expected count). Напомню, что если в четырехпольной таблице хоть в одной ячейке ожидаемое значение будет < 5, то рекомендуется применять точный критерий Фишера. Если для многопольной таблицы доля ячеек с ожидаемыми значениями <5 будет превышать 20 %, то критерий  $\chi^2$  применять не рекомендуется.

**Chi-Square Tests**

|                                    | Value              | df | Asymp. Sig.<br>(2-sided) | Exact Sig.<br>(2-sided) | Exact Sig.<br>(1-sided) |
|------------------------------------|--------------------|----|--------------------------|-------------------------|-------------------------|
| Pearson Chi-Square                 | 4,378 <sup>a</sup> | 1  | ,036                     |                         |                         |
| Continuity Correction <sup>b</sup> | 3,149              | 1  | ,076                     |                         |                         |
| Likelihood Ratio                   | 4,473              | 1  | ,034                     |                         |                         |
| Fisher's Exact Test                |                    |    |                          | ,057                    | ,037                    |
| Linear-by-Linear<br>Association    | 4,287              | 1  | ,038                     |                         |                         |
| N of Valid Cases                   | 48                 |    |                          |                         |                         |

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6,71.

Рис. 6. Таблица результатов применения критерия  $\chi^2$  Пирсона в различных модификациях и точного критерия Фишера

В четвертой таблице результатов представлены значения  $\lambda$ -критерия Гудмена – Краскела, а также не упоминавшийся ранее критерий  $\tau$  (tau) Гудмена – Краскела и коэффициент неопределенности (рис. 7). Напомню, что эти критерии асимметричны (имеют направленный характер), то есть их значения будут отличаться в зависимости от того, какая из переменных является зависимой. SPSS не знает, какая из переменных зависимая, и рассчитывает значения критериев для обоих случаев. В нашем примере зависимой переменной является исход, поэтому смотрим результаты в строках «Ishod Dependent». Обращать внимания на ряд «Symmetric», а также на ряд «FR Dependent», в котором за зависимую принимается наша независимая переменная, не стоит.

Результаты говорят о том, что наличие информации о факторе риска по данным коэффициента  $\lambda$  (Lambda) нисколько не улучшает прогнозирование исхода ( $\lambda = 0$ ). Чуть большее значение принимает коэффициент неопределенности (Uncertainty Coefficient,  $U = 0,077$ ), согласно которому ошибка прогнозирования исхода при наличии данных о факторе риска может сократиться на 7,7 % ( $U = 0,077$ ,  $p = 0,034$ ).

**Directional Measures**

|                         |        |                 | Value | Asymp. Std. Error <sup>a</sup> | Approx. T <sup>b</sup> | Approx. Sig.      |
|-------------------------|--------|-----------------|-------|--------------------------------|------------------------|-------------------|
| Nominal by Nominal      | Lambda | Symmetric       | ,162  | ,087                           | 1,648                  | ,099              |
|                         |        | FR Dependent    | ,261  | ,140                           | 1,648                  | ,099              |
|                         |        | Ishod Dependent | ,000  | ,000                           | . <sup>c</sup>         | . <sup>c</sup>    |
| Goodman and Kruskal tau | Lambda | FR Dependent    | ,091  | ,081                           |                        | ,038 <sup>d</sup> |
|                         |        | Ishod Dependent | ,091  | ,082                           |                        | ,038 <sup>d</sup> |
| Uncertainty Coefficient | Lambda | Symmetric       | ,072  | ,066                           | 1,091                  | ,034 <sup>e</sup> |
|                         |        | FR Dependent    | ,067  | ,062                           | 1,091                  | ,034 <sup>e</sup> |
|                         |        | Ishod Dependent | ,077  | ,070                           | 1,091                  | ,034 <sup>e</sup> |

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Cannot be computed because the asymptotic standard error equals zero.

d. Based on chi-square approximation

e. Likelihood ratio chi-square probability.

**Рис. 7. Результаты оценки силы взаимосвязи (асимметричные критерии)**

В пятой таблице приведены симметричные критерии, то есть критерии, показывающие силу взаимосвязи между переменными независимо от того, какая из них является зависимой (рис. 8). Значения критериев (Value) соответствуют взаимосвязи средней силы (см. табл. 2). Кроме того, представлены приблизительные уровни значимости ( $p$ ) для всех критериев (Approx. Sig). Результаты можно представить как  $\phi = 0,30$ ,  $p = 0,036$ , хотя уровень значимости здесь не так важен. Для многопольных таблиц использование критерия V Крамера предпочтительнее. Также следует помнить, что приведенный в таблице коэффициент сопряженности Пирсона (Contingency Coefficient) не достигает максимума, поэтому лучше производить коррекцию этого коэффициента, как было рассмотрено выше.

### Symmetric Measures

|                  |                         | Value | Approx. Sig. |
|------------------|-------------------------|-------|--------------|
| Nominal by       | Phi                     | ,302  | ,036         |
| Nominal          | Cramer's V              | ,302  | ,036         |
|                  | Contingency Coefficient | ,289  | ,036         |
| N of Valid Cases |                         | 48    |              |

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.

Рис. 8. Результаты оценки силы взаимосвязи (симметричные критерии).

Последняя таблица наиболее информативна (рис. 9). Поскольку компьютер не знает, какого типа было наше исследование, он рассчитывает и отношение шансов (строка «Odds Ratio»), и относительный риск (строки «For cohort»), а также нижнюю (Lower) и верхнюю (Upper) границы доверительного интервала (Confidence Interval). Если бы наше исследование было типа «случай – контроль», мы не могли бы использовать относительный риск и должны были бы ограничиться отношением шансов. Но поскольку наш пример представляет собой проспективное (когортное) исследование, следует использовать относительный риск. Поскольку нас интересуют различия в исходах (вспомним, что наличие исхода закодировано в виде «1») относительно фактора риска, а не наоборот, результат следует искать в строке «For cohort Ishod=1». При описании результатов необходимо указывать не только относительный риск, но и доверительный интервал: RR = 2,72, 95 % CI: 0,99–7,48 или по-русски: ОР = 2,72, 95 % ДИ: 0,99–7,48. Видно, что интервал включает в себя 1, а значит, результат не является статистически значимым (на уровне доверительной вероятности 95 %). Несмотря на это, ширина интервала дает четкое представление о том, какие значения RR (ОР) может принимать в 95 % случаев, и это не позволяет отнести результаты к маловажным, что наглядно демонстрирует бóльшую информативность интервальной оценки силы взаимосвязи по сравнению с проверкой статистических гипотез.

| Risk Estimate             |       |                         |        |
|---------------------------|-------|-------------------------|--------|
|                           | Value | 95% Confidence Interval |        |
|                           |       | Lower                   | Upper  |
| Odds Ratio for FR (1 / 2) | 4,038 | 1,047                   | 15,581 |
| For cohort Ishod = 1      | 2,717 | ,988                    | 7,475  |
| For cohort Ishod = 2      | ,673  | ,452                    | 1,001  |
| N of Valid Cases          | 48    |                         |        |

Рис. 9. Отношение шансов и относительный риск с 95 % доверительными интервалами

Можно ли использовать отношение шансов для когортных исследований? В принципе можно, так как отношение шансов будет достаточно точно отражать относительный риск при редких исходах (скажем, до 10 %), однако если исход не является редким, значение отношения шансов будет существенно превышать значение относительного риска, что может привести к неверному толкованию проблемы. Например, для обеих ситуаций, представленных в табл. 3, относительный риск составляет 2,0, а отношение шансов – 2,1 в первом случае и 11,0 во втором (проверьте, используя вышеприведенные формулы). Поэтому если для второй ситуации в когортном исследовании рассчитать только отношение шансов, можно сделать неверный вывод о чрезвычайной вредности изучаемого фактора, если в качестве зависимой переменной используется заболевание или летальный исход (более чем пятикратное завышение!). Поэтому для когортных исследований рекомендуется представлять только относительный риск, а для исследований типа «случай-контроль» - только отношение шансов, причем не следует интерпретировать последнее с точки зрения рисков. Например, при классическом исследовании «случай – контроль» с соотношением количества случаев к количеству контролей 1:1 отношение шансов 4,0 не означает, что изучаемый фактор риска увеличивает вероятность исхода именно в 4 раза, но говорит о наличии сильной взаимосвязи.

Таблица 3

**Различия между относительным риском и отношением шансов (объяснения в тексте)**

| Ситуация | Вероятность исхода для тех, у кого нет фактора риска | Вероятность исхода для тех, у кого есть фактор риска | Относительный риск | Отношение шансов |
|----------|--|--|--------------------|------------------|
| 1        | 0,05, или 5 %  | 0,1, или 10 %  | 2,0                | 2,1              |
| 2        | 0,45, или 45 %                                       | 0,90, или 90 %                                       | 2,0                | 11,0             |

Несмотря на то, что мы рассмотрели пример только для дихотомических переменных и четырехпольных таблиц, читатели могут применять все описанное выше и для ситуаций, когда номинальные переменные принимают более чем два значения (для многопольных таблиц), за исключением точного критерия Фишера, поправки Йейтса и критерия  $\phi$ . Более подробно о методах сравнения номинальных данных можно прочитать в специализированной литературе (например, [2, 6]).

### Список литературы

1. Банержи А. Медицинская статистика понятным языком: вводный курс / А. Банержи. – М. : Практическая медицина, 2007. – 287 с.
2. Браунли К. А. Статистическая теория и методология в науке и технике / А. К. Браунли. – М. : Наука, 1980. – С. 376–417.
3. Власов В. В. Эпидемиология : учебное пособие для вузов / В. В. Власов. – М. : ГЭОТАР-МЕД, 2004. – 464 с.
4. Гржибовский А. М. Доверительные интервалы для частот и долей / А. М. Гржибовский // Экология человека. – 2008. – № 5. – С. 57–60.
5. Зайцев В. М. Прикладная медицинская статистика / В. М. Зайцев, В. Г. Лифляндский, В. И. Маринкин. – СПб. : Фолиант, 2003. – 428 с.
6. Медик В. А. Математическая статистика в медицине / В. А. Медик, М. С. Токмачев. – М. : Финансы и статистика, 2007. – 798 с.
7. Сергиенко В. И. Математическая статистика в клинических исследованиях / В. И. Сергиенко, И. Б. Бондарева. – М. : ГЭОТАР-МЕД, 2001. – 256 с.
8. Юнкеров В. И. Медико-статистическая обработка данных медицинских исследований / В. И. Юнкеров, С. Г. Григорьев. – СПб. : ВмедА, 2002. – 266 с.
9. Adler F. Yates's correction and the statisticians / F. Adler // Journal of the American Statistical Association. – 1951. – Vol. 12. – P. 490–501.
10. Brown L. D. Interval estimation for a binomial proportion / L. D. Brown, T. T. Cai, A. Dasgupta // Statistical Science. – 2001. – N 2. – P. 101–133.
11. Conover W. J. Some reasons for not using the Yates continuity correction on  $2 \times 2$  contingency tables / W. J. Conover // Journal of the American Statistical Association. – 1974. – Vol. 69. – P. 374–376.

12. Dawson B. Basic and clinical biostatistics / B. Dawson, R. G. Trapp. – Lange Medical Books : McGraw-Hill, 2001. – 399 p.
13. *Field A.* Discovering statistics using SPSS / A. Field. – SAGE Publications, 2005. – 779 p.
14. *Haviland M. G.* Yates's correction for continuity and the analysis of 2x2 contingency tables / M. G. Haviland // *Statistics in Medicine*. – 1990. – Vol. 9. – P. 363–367.
15. *Motulsky H.* Intuitive biostatistics // H. Motulsky. – Oxford : Oxford University Press, 1995. – 386 p.
16. *Overall J. E.* Power of chi-square tests for 2x2 contingency tables with small expected frequencies / J. E. Overall // *Psychological Bulletin*. – 1980. – Vol. 87. – P. 132–135.
17. *Pearson E. S.* The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table. / E. S. Pearson // *Biometrika*. – 1947. – Vol. 34. – P. 139–167.
18. *Wilkinson L.* Statistical methods in psychology journals: guidelines and explanations // L. Wilkinson // *American Psychologist*. – 1999. – Vol. 54. – P. 594–604.
19. *Yates F.* Contingency tables involving small numbers and the chi-square test / F. Yates // *Supplement to the Journal of the Royal Statistical Society*. – 1934. – Vol. 1. – P. 222.

## **ANALYSIS OF NOMINAL DATA (INDEPENDENT OBSERVATIONS)**

**A. M. Grjibovski**

*National Institute of Public Health, Oslo, Norway*

The article describes analysis of nominal data by Pearson's chi-squared test, likelihood ratio test, chi-squared test with Yates's continuity correction and Fisher's exact test. Methods for estimation of the effect size or strength of the association, such as  $\phi$ , Kramer's  $V$ , Pearson's contingency coefficient, Goodman-Kruskall  $\lambda$ , uncertainty coefficient, relative risk and odds ratio are presented. The paper provides only general introduction to the methods of analysis of nominal data. The readers are encouraged to consult statistical literature prior to analysing own data and preparing manuscripts.

**Key words:** nominal data, ordinal data, analysis, SPSS.

**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](mailto:angr@fhi.no)

Статья поступила 16.05.2008 г.