НАПРАВЛЕННАЯ СТАТИСТИЧЕСКАЯ ОЦЕНКА НАБЛЮДАЕМЫХ РАЗЛИЧИЙ В СРАВНИТЕЛЬНОМ ДВУХВЫБОРОЧНОМ ИССЛЕДОВАНИИ

Обеснюк В.Ф.

Южно-Уральский институт биофизики (г. Озерск)

Введение. Двухвыборочные результаты регистрации некоторых специфических событий обычно заносят в таблицы сопряженности, простейшая разновидность которых имеет размер 2×2, или в более сложном случае — $2\times(m+1)$, где m – натуральное, превышающее единицу. Это могут быть таблицы “случай – контроль”, различающиеся по действию небольшого набора причинных факторов, или же результаты сравнения двух факторных групп, члены которых могут иметь два типа эффектов или более. На практике подобные сопоставления приходится производить в самых разных областях науки и техники, например, при изучении технологического брака, в социологических и психологических исследованиях, генетике, при анализе когорт в эпидемиологии и т.п., то есть всюду, где наблюдения за состоянием единичного объекта можно описать категориальной переменной 0 или 1. В теории обработки данных подобного типа введен специальный термин “статистика объектов нечисловой природы” [1]. Целью описанного анализа обычно является сопоставление долей наступления изучаемых специфических событий p_1 и p_2 в двух однородных выборках путем оценки разности $p_2 - p_1$ или отношения p_2/p_1. Однако до определения этих показателей, в первую очередь, необходима оценка статистической значимости наблюдаемых отличий. В рамках частотного подхода для этого используется критерий “хи-квадрат” Пирсона, с помощью которого оценивается вероятность регистрации значительного отклонения фактических наблюдений от некоторых средних или “ожидаемых” величин. При этом приходится производить расчет вероятности, оперируя ее “ожидаемым” распределением, которое отражает некую процедуру усреднения по обеим выборкам.

Структура таблицы сопряженности

	Эффект		
	Да	Нет	Σ
Выборка 1	$a_{1,1}$	$a_{1,2}$	N_1
Выборка 2	$a_{2,1}$	$a_{2,2}$	N_2

Примечание:
$N_1 = a_{1,1} + a_{1,2}$; $N_2 = a_{2,1} + a_{2,2}$; $p_1 \approx a_{1,1}/N_1$; $p_2 \approx a_{2,1}/N_2$.

Это нетрудно показать. Действительно, пусть некоторые целочисленные результаты наблюдений представлены четырехпольной таблицей ($m = 1$). В соответствии с известной практикой применения метода Пирсона к четырехпольным таблицам [2], статистика хи-квадрат с одной степенью свободы рассчитывается как сумма четырех зависимых слагаемых

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(a_{i,j}-\tilde{a}_{i,j})^2}{\tilde{a}_{i,j}}, \qquad (1)$$

где величины $\tilde{a}_{i,j}$ — введенные Пирсоном нецелочисленные “ожидаемые” абсолютные частоты наблюдений. Они формально определяются из нулевой гипотезы H_0 о возможном отсутствии различий в выборках (H_0: $p_2 = p_1$), предполагающем совпадение относительных долей изучаемых событий как в каждой отдельной выборке, так и после их объединения:

$$\frac{\tilde{a}_{i,j}}{N_i} = \frac{a_{1,j}+a_{2,j}}{N_1+N_2}. \qquad (2)$$

Используя это определение, нетрудно показать, что сумма четырех слагаемых (1), в действительности, может быть тождественно сведена к единственному слагаемому

$$\chi^2 = \frac{(p_2-p_1)^2}{Var(\tilde{p}_2)+Var(\tilde{p}_1)}, \qquad (3)$$

где $p_2 = a_{2,1}/N_2$ – эмпирическая оценка доли специфических событий во второй выборке; $p_1 = a_{1,1}/N_1$ – в первой выборке; $Var(\tilde{p}_2)$, $Var(\tilde{p}_1)$ – дисперсии “ожидаемых” долей при условии статистической независимости выборок и нормальной аппроксимации биномиального распределения событий по клеткам таблицы $Var(\tilde{p}_i) = \tilde{p}_i \cdot (1-\tilde{p}_i)/N_i$. Исходя из формулы (3) можно говорить о единственной степени свободы для случайной величины χ^2. Для принятия решения о справедливости гипотезы H_0, в качестве меры традиционно принимается вероятность *p-value* событий с $\chi^2 > \chi^2_{кр}$, где $\chi^2_{кр}$ определяется по формуле, аналогичной (3), но с подстановкой конкретных данных, наблюдаемых в таблице сопряженности. При этом сама случайная величина χ^2 трактуется в качестве квадрата от центрированного, нормированного и нормально распределенного случайного отклонения. Таким образом, и множеству H_0, и дополняющему его множеству $\overline{H_0}$ формально ставится в соответствие одно и то же одномерное распределение величины χ^2, что позволяет утверждать об использовании постулата $p\text{-}value = 1 - P\left(data \middle| H_0\right) = P\left(data \middle| \overline{H_0}\right)$, принимаемого на веру.

Такова давняя и почти общепринятая практика [3]. Однако можно заметить, что поиск различий в выборках путем оценки их однородности по величине вероятности *p-value* в случае применения к таблицам сопряженности обладает рядом недостатков. Главные из них – ненаправленность тестируемой гипотезы, а также – оперирование ненаблюдаемой плотностью вероятности. Можно указать и на некоторое смещение оценок, вызванное отклонениями распределения долей p_1 и p_2 от обычно предполагаемого нормального распределения, а также — несоответствием числителя знаменателю в формуле (3).

На недостатки меры *p-value* указывают также и более простые, но в то же время более значимые доводы. Нетрудно видеть, что при обращении к метрологически более оправданной байесовой статистике, являющейся альтернативой по отношению к частотной (3), величина $P\left(data \middle| \overline{H_0}\right)$ по своему смыслу является условной. Более адекватной условной мерой отклонения была бы достаточно малая величина другой вероятности $P\left(\overline{H_0} \middle| data\right)$ при том, что $P\left(\overline{H_0} \middle| data\right)$ и $P\left(data \middle| \overline{H_0}\right)$ не совпадают по величине в общем случае. Между отмеченными показателями есть связь, определяемая известной теоремой Байеса:

$$P\left(\overline{H_0} \middle| data\right) \cdot P(data) = P\left(data \middle| \overline{H_0}\right) \cdot P\left(\overline{H_0}\right) . \tag{4}$$

В силу того, что множества H_0 и $\overline{H_0}$ не пересекаются и вместе образуют абсолютно достоверное событие, нетрудно точно вычислить априорные вероятности $P\left(\overline{H_0}\right)$=1 и $P(H_0)$ = 0, поскольку гипотеза (H_0: $p_2 = p_1$) описывается множеством меры ноль при любых непрерывных априорных распределениях величин p_1 и p_2 генеральной совокупности. Это хорошо видно в двумерном пространстве состояний, образующем квадрат со стороной 1. Гипотезе H_0 точно соответствует диагональ этого квадрата. По этой причине в согласии с формулой (4) и с учётом неравенства $P(data) < 1$ всегда имеем $P\left(\overline{H_0} \middle| data\right) > p\text{-}value$, то есть вероятность ошибки гипостазирования в действительности всегда превосходит оценку *p-value*. Более того, из второго (аналогичного) тождества Байеса в отношении гипотезы H_0

$$P\left(H_0 \middle| data\right) \cdot P(data) = P\left(data \middle| H_0\right) \cdot P(H_0) \tag{5}$$

следует, что вместе с $P(H_0)=0$ выполняется $P(H_0| \; data)=0$, а также $P(\overline{H_0}| \; data)=1$ независимо от достигнутого значения *p-value* ! Мы приходим к парадоксу, следствием которого является вывод о том, что *традиционно определяемая величина p-value в процессе оценки таблиц сопряженности никогда не является мерой истинной вероятности отклонения от нулевой гипотезы*, причем последняя всегда равна 100%.

Каким же образом возможно устранение парадокса? Очевидно, он порождается исключительно условием $P(H_0)=0$. Априорная вероятность исследуемой гипотезы может не быть равной нулю только в двух случаях: 1) при смене самой гипотезы; 2) при отказе от непрерывности априорных распределений, то есть при отказе от понятия генеральной совокупности. С практической (метрологической) точки зрения представляется наиболее разумным изменить гипотезу. А именно, чаще всего истинной целью статистического исследования является обоснование значимости наблюдаемых различий между выборками, а не отсутствия таковых. Поэтому в процессе традиционной оценки стремятся обнаружить именно малость *p-value*. Так не лучше ли оперировать, например, гипотезами о наличии тренда $H_1 : p_2 > p_1$ и/или противоположной ей $\overline{H_1}$, если эмпирические данные указывают на аналогичное соотношение долей? Тогда для любых неинформативных априорных распределений $P(H_1)=P(\overline{H_1})=0{,}5$. Это видно на квадрате пространства состояний. Парадокс исчезает.

Предлагаемая методика исследования гипотезы H₁. Отмеченные выше недостатки традиционного подхода к оценке событий могут быть исправлены одновременно в рамках подхода Байеса, если рассматривать выборочные распределения долей p_1 и p_2, которые задаются на отрезке непрерывными бета-функциями [4], отражающими учет байесовой априорики. Подход менее удобен для вычислений, чем использование нормального распределения, но позволяет избегать неоправданных предположений о возможности наблюдения событий $p_i < 0$ и $p_i > 1$, свойственных традиционному. Тогда все целые числа эмпирической таблицы наблюдений становятся фиксированными параметрами распределений, а переменные p_1 и p_2 — либо переменными состояния в фазовом пространстве случайных реализаций, либо оценками последних. Например, если принять $\eta = p_2 - p_1$, то плотность вероятности по η будет определяться интегралами

$$\varphi(\eta)=\begin{cases}\int\limits_0^{1-|\eta|} f_1(\tau+|\eta|)\cdot f_2(\tau)\,d\tau, & -1\le\eta\le 0;\\ \int\limits_0^{1-|\eta|} f_1(\tau)\cdot f_2(\tau+|\eta|)\,d\tau, & 0\le\eta\le 1,\end{cases} \tag{6}$$

где $f_1(\cdot)$, $f_2(\cdot)$ – плотности распределения долей специфических событий в выборках. В этом случае областью естественной случайной изменчивости разности долей можно считать околомедианную зону с долей вероятности, например 90%, а края распределения вблизи $\eta\approx\pm 1$ будут соответствовать вероятностям по 5%. Выход за указанные квантили можно условно считать неслучайными отклонениями. Заметим, что только одна из двух неслучайных областей окажется соответствующей по знаку гипотезе H_1 и знаку эмпирически наблюдаемой разности. Например, при исследовании гипотезы H_1 о превышении $p_2>p_1$ по аналогии с мерой *p-value* можно ввести байесову вероятность ε для события $p_2<p_1$. Тогда при уровне значимости (уровне принятия решения) $\varepsilon<5\%$ условно допускается считать событие $p_2>p_1$ неслучайным. Нетрудно показать, что на практике при оценке случайности повышения риска справедлива численная формула:

$$\varepsilon=\int\limits_0^1 f_1(\tau)\cdot F_2(\tau)\,d\tau\,, \tag{7}$$

где $F_2(\tau)$ – кумулятивная функция бета-распределения специфической доли для второй выборки. Заметим, что выражение (7) не только вытекает непосредственно из соотношений (6), но и соответствует общим рекомендациям применения статистик интегрального типа при обработке “объектов нечисловой природы” [1]. По смыслу предыдущих обозначений $\varepsilon=P(\overline{H_1}|\,data)$. Если формула (7) дает оценку $\varepsilon>0{,}5$, то эмпирические данные не согласуются с предположением $p_2>p_1$ и следует воспользоваться дополняющей вероятностью $1-\varepsilon$, что всегда при данном определении ограничивает вероятность ошибки величиной 50%, в отличие от традиционной меры *p-value*, способной приближаться к парадоксальной отметке 100%. Широкое распространение средств вычислительной техники и соответствующего программного обеспечения делает проблему вычисления интеграла (7) разрешимой практически всегда. При необходимости интеграл может быть точно сведен к конечной сумме, если использовать тождество Клоппера–Пирсона, связывающее формулы биномиального и бета распределений [5]. Например, предполагая справедливым априорно равномерное распределение оценок долей до проведения измерений, получим апостериорную оценку

$$\varepsilon = \frac{1}{C_{N_1+N_2+2}^{N_1+1}} \sum_{i=a_{2,1}+1}^{N_2+1} C_{i+a_{1,1}}^{a_{1,1}} \cdot C_{N_1+N_2+1-a_{1,1}-i}^{N_1-a_{1,1}} . \qquad (8)$$

Метод подсчета вероятности ошибки суммированием гипергеометрических вероятностей напоминает “точный” односторонний тест Фишера для таблиц сопряженности [6], но принципиально отличается от него тем, что нами проверяется иная гипотеза, причем в рамках подхода Байеса. В качестве примера на рисунке 1 приведены расчетные кривые двух бета-распределений оцениваемых долей и распределения их разности (см. преобразование (6)) для следующей таблицы данных: пять случаев из 12 против пяти случаев из 38. Очевидно, ε-ошибка здесь довольно мала. Это ясно по величине площади под кривой “в”, которая соответствует отрицательной абсциссе. Подсчет интеграла (7) приводит к оценке $\varepsilon = 1{,}89\,\% < 5\,\%$, что указывает на неслучайность наблюдаемых различий.

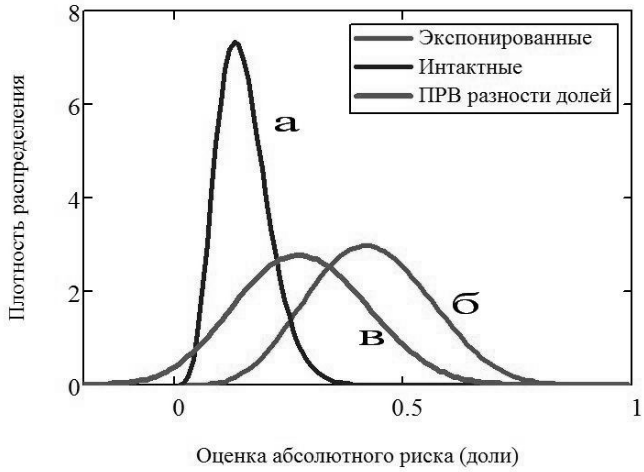

Рис. 1. Распределения оценок долей в выборках (кривые “а” и “б”). Плотность распределения вероятности (ПРВ) наблюдения разности долей в выборках (кривая “в”; композиция).

Заметим, что процедура оценки по формуле (7) формально выполнима даже в тех случаях, когда количество наблюдаемых событий меньше 5, включая нуль, либо приближается к объему выборки. Обычно тест-конкурент “хи-квадрат” в таких ситуациях выполнять вообще не рекомендуется ([6], со ссылкой на работу Кокрэна), так как дискретное биномиальное распределение наблюдаемых событий при использовании частотного подхода уже не может быть хорошо приближено непрерывным нормальным распределением. Наш подход свободен от этого недостатка. Незначительные проблемы с численной оценкой величины $\varepsilon = P\left(\overline{H_1}\middle|\, data\right)$ могут возникнуть только при работе с очень большими выборками, когда численные алгоритмы подсчета бета-распределения дают погрешность из-за переполнения разрядной сетки при вычислении факториальных величин. Эта трудность преодолевается благодаря тому, что для достаточно объемной статистики наблюдений центральная часть всякого бета-распределения стремится к нормальному. Аналогичная тенденция наблюдается и в отношении разности случайных бета-величин. В частности, это хорошо видно на рис. 1 (кривая "в") даже для редких событий. Учитывая независимость долей p_1 и p_2 для двух выборок, математическое ожидание и дисперсия разности могут быть вычислены точно даже для бета-величин и их композиции:

$$\begin{aligned}
&E\{p_2 - p_1\} = \overline{p_2 - p_1} = E\{p_2\} - E\{p_1\};\\
&Var\{p_2 - p_1\} = Var\{p_2\} + Var\{p_1\};\\
&E\{p_i\} = \frac{a_i + 1}{N_i + 2}; \quad Var\{p_i\} = \frac{a_i + 1}{N_i + 2} \cdot \left(\frac{a_i + 2}{N_i + 3} - \frac{a_i + 1}{N_i + 2}\right).
\end{aligned} \tag{9}$$

С этой точки зрения в случае большого объема собранной статистики вполне работоспособна аппроксимация

$$\varepsilon \approx Norm\left(0, \overline{p_2 - p_1}, Var\{p_2 - p_1\}\right), \tag{10}$$

где $Norm\left(x, \bar{x}, Var\{x\}\right) -$ кумулятивное нецентральное нормальное распределение.

Аналог и прототип рассмотренного тестирования. У данного способа оценки гипотезы о наблюдаемых различиях существует прототип [7], известный как статистический критерий $\varphi *$, построенный на принципе нормализующего углового преобразования Фишера, а также аналог [8] — метод Уэлча сравнения двух средних (формула 10). Однако по сравнению с прототипом изложенный метод точнее работает вплоть до нулевых и предельных чисел "случаев" в одной из выборок. Это обстоятельство может представлять особый интерес для практикующих статистиков. При $a_{i,1} \to N_i/2 >> 1$ оба метода сопоставимы по качеству. Метод Уэлча позволяет воспользоваться таблицами нормального распределения или распределения Стьюдента. При этом так же, как и в методе Уэлча, нет необходимости контролировать совпадение дисперсией, в отличие от подхода Стьюдента. Есть и различие — в предлагаемом методе центральная оценка разности долей стремится к медиане, а не среднему арифметическому, поэтому соотношение (10) является только аппроксимацией, но не отражает существо подхода. Интересно также отметить, что в предельном случае при анализе четырехпольных таблиц при условии плохой достоверности различий и достаточно объемной статистике наблюдений величина ε примерно равна половине *p-value* или немного превосходит её, что соответствует так называемой “one-tailed *p-value*”. Необходимость отличать эту величину от обычного “double-tailed *p-value*” подчеркивалась рядом авторов программного обеспечения [9]. При достоверном различии выборок оценки *p-value* и ε уже заметно не совпадают, имея расхождение до нескольких порядков величины. Существенные различия между *p-value* и ε наблюдаются также для тех таблиц, где анализируется сравнение выборок более, чем по двум факторам или эффектам.

Работоспособность ε-оценок при множественном сравнении. Рассмотрим преимущество корректного определения гипотезы и метода ее проверки на конкретном примере. Пусть в результате некоторого ограниченного исследования “случай-контроль” в двух разных выборках, отличающихся по некоторому эффекту, выполнялся поиск связи с пятью независимыми факторами (таблица 1). Результаты занесены в таблицу так, что для каждого фактора указано количество случаев его наблюдения в соответствующей выборке. Требуется установить, достоверно ли отличие двух выборок по пяти факторам одновременно. Это обстоятельство устанавливается по малой величине групповой вероятности ошибки в отношении изучаемой гипотезы.

По правилам применения теста “хи-квадрат” в силу независимости факторов суммарная величина “хи-квадрат” равна арифметической сумме аналогичных величин $\chi_i^2(1)$ с одной степенью свободы для каждого фактора. Общая практика работы с таблицами [2, 3, 6] диктует, что если распределение $\chi_\Sigma^2(m)$ известно, можно найти $p\text{-}value = 1 - Chisq\left(\chi_\Sigma^2(m), m\right)$, где $Chisq(\cdot\,, m) -$ кумулятивная функция распределения “хи-квадрат” с m степенями свободы.

Таблица 1. Результаты множественных сравнений

	Численность	Независимые факторы				
	N_i	№1	№2	№3	№4	№5
Выборка 1	100	50	33	97	64	62
Выборка 2	100	70	53	86	81	44
“Ожидаемые”	100	60	43	91,5	72,5	53
p-value(1)	—	0,39%	0,43%	0,53%	0,71%	1,1%
$\varepsilon(1)$	—	0,20%	0,22%	0,26%	0,36%	0,55%

Несколько иным является алгоритм оценки ε-ошибки. Отнесение всех наблюдений по всем факторам к выделенной группе будет достоверно только в том случае, если оно достоверно по каждому в отдельности. Тогда общая достоверность оценивается произведением всех $1-\varepsilon_i$, где $i=\overline{1,5}$. Следовательно, $\varepsilon_\Sigma = 1-\prod_{i=1}^{5}(1-\varepsilon_i)$.

Оба подхода, очевидно, должны давать некоторые расхождения в оценках вероятности хотя бы потому, что соответствуют разным множествам в пространстве состояний. Если в случае m = 1 такая разница невелика, то с ростом размерности таблиц она не может оставаться незамеченной. Для приведенной выше таблицы результаты анализа составили: $\chi^2_\Sigma \approx 38{,}02$; $p\text{-}value = 3{,}7\cdot 10^{-5}$ %; $\varepsilon_\Sigma = 1{,}6$ %. Достоверность различий на уровне значимости 5% здесь наблюдается по обоим тестам, однако пять порядков разницы заставляют задуматься о том, какой же из тестов неадекватен? Очевидно, отличие вероятностей отражает здесь не только разницу гипотез, если исходно ставилась одна и та же цель оценки достоверных различий.

Заметим, что анализ значений *p-value* по каждому фактору в отдельности, то есть для одной степени свободы, указывает на близкую к пограничной достоверность теста для рассмотренного примера с учетом так называемого правила Бонферрони [10]. В качестве яркого примера уместно указать и на имеющийся опыт практических изысканий комбинаций генетических маркеров, предопределяющих выработку патогенного белка LDL-C [11]. В них установлено, что фактическая относительная частота 3,1…6,5 % для лиц с высоким уровнем белка LDL-C в популяции почему-то сильно превосходит математические оценки $p\text{-}value \sim 10^{-5} \div 10^{-7}$ (феномен утерянной наследуемости), то есть практически такие же, как и в нашем примере. Все это указывает на реалистичность ε-оценок и излишний оптимизм тестирования по показателю *p-value* для многопольных таблиц сопряженности признаков. Однако сама величина “хи-квадрат” по определению работы [12] все же сохраняет смысл меры близости двух точек в фазовом пространстве большой размерности.

К истории вопроса о применении теста “хи-квадрат” к оценке кросс-табулированных данных. Уместно отметить, что в самой первой работе [12], посвященной статистике “хи-квадрат”, нет ни одного упоминания о возможности ее применения к оценке таблиц сопряженности. И это понятно, поскольку автором теста была проанализирована устойчивость показателя, как некого среднего по множеству наблюдений координатных пар $\{x_i, y_i\}$, где одна из координат имеет строго нормальное распределение. Тестировалась при этом гипотеза о наличии линейного тренда, но вовсе не нулевая гипотеза, практика оценки которой в дальнейшем парадоксальным образом сложилась при исследовании таблиц сопряженности. Пирсон не предлагал применять свой критерий для одной такой пары, так как сумма нормированных квадратов отклонений данных от наилучшей линейной модели была бы в таком случае равна нулю, бессмысленно приводя к якобы безошибочному принятию трендовой гипотезы. Практика применения теста "хи-квадрат" для нулевой гипотезы к кросстабулированным данным связана, по-видимому, с работами У. Юла (см. обзор [6], глава 33) именно по этой причине. Позднее она была поддержана также авторитетом Р. Фишера [13]. В дальнейшем мера "хи-квадрат" использовалась С. Кульбаком [2]. В частности, в его работах и в публикациях Пирсона встречается выражение (1). Оценка Кульбака определена как разность экстремумов функционалов максимального правдоподобия, то есть при вычислении статистики "хи-квадрат" Кульбак фактически опирался не только на гипотезу H_0, но и на некую трендовую гипотезу, из-за чего "хи-квадрат" уже не являлся исключительно мерой нулевой гипотезы. К сожалению, чтобы сохранить преемственность с публикациями [6, 12, 13], он был вынужден использовать понятие "ожидаемых" распределений. Использование "ожидаемых" абсолютных частот $\tilde{a}_{i,j}$ приводит к тому, что оценки $a_{i,j} - \tilde{a}_{i,j}$ в формуле (1) уже не являются центрированными случайными наблюдениями, то есть статистика χ^2 в приложении к таблицам сопряженности не имеет распределения хи-квадрат, несмотря на совпадение названий. Роль нецентральности возрастает с увеличением размера таблиц, поэтому смещение оценки *p-value* по сравнению с "истинной" вероятностью может очень большим для таблиц $2\times(m+1)$ уже при $m > 3$. Это нетрудно заметить, например, в эпидемиологических исследованиях выборки, разбитой на m независимых страт, отличающихся уровнем воздействия одного и того же фактора.

Сомнения в универсальности меры *p-value* в оценке табличных данных уже неоднократно были опубликованы в ряде недавних работ [14–17], но пока что в разделах "дебаты" и "точка зрения" [16, 17]. Недостатки предлагалось исправить путем применения подхода Байеса, который в данном случае, по мнению авторов публикаций, должен практически сводиться к ужесточению уровня принятия решений с 5 % до 0,1 % или сильнее. Из нашей публикации следует, что дело здесь не в уровне значимости. Сама нулевая гипотеза противоречит подходу Байеса, благодаря чему не следует принимать на веру сами расчеты вероятности, а также интервальные оценки, предложенные Пирсоном для таблиц сопряженности.

Выводы. Таким образом, в работе показано, что при исследовании кросс-табулированных категориальных наблюдений тест “хи-квадрат” не следует без ограничений использовать для доказательства статистической значимости больших отклонений от нулевой гипотезы, так как принятие последней приводит к внутренним противоречиям при попытке учета априорных вероятностных свойств объекта измерения. Чтобы сохранить смысл и внутреннюю непротиворечивость статистического теста, основную гипотезу следует адекватно изменить. При исследовании таблиц размера $2\times(m+1)$, где $m\geq 1$ – количество независимых факторов или эффектов, в качестве иной гипотезы практически всегда может быть использована гипотеза о значимости наблюдаемого соотношения долей в выборках. В частности, она вполне может быть оценена в рамках подхода Байеса. Для многопольных таблиц сопряженности возможны и другие гипотезы, что приведет к смене критерия. Особенно это касается таблиц размера $n\times(m+1)$, когда выполняется сравнение n выборок с m эффектами или факторами, например, когда эффекты являются зависимыми или конкурирующими. Предлагаемая методика может быть использована в различных областях науки и техники, опирающихся на статистические методы обработки категоризированных данных: в социологии, биологии и медицине, а также при исследовании причин технологического брака любых сложных производств.

ЛИТЕРАТУРА

1. **Орлов А.И.** Статистика объектов нечисловой природы (Обзор) / Журнал «Заводская лаборатория», 1990, Т.56, No.3, С.76–83.
2. **Кульбак С.** Теория информации и статистика. – М: Наука, 1967. – 408 с.
3. **Upton G.J.G.** The Analysis of Cross-tabulated Data. – NY: Wiley, 1978. – 160 p.
4. **Hey J.D.** Data in Doubt: An Introduction to Bayesian Statistical Inference for Economists. – Martin Robertson, 1983. – 320 p.
5. **Большев Л.Н., Смирнов Н.В.** Таблицы математической статистики. – М.: Наука, 1983. – 416 с.

6. **Кендалл М., Стюарт А.** Статистические выводы и связи, т.2. – М. Наука, 1973. – 900 с.
7. **Сидоренко Е.В.** Методы математической обработки в психологии. – СПб: ООО “Речь”, 2007. – 350 с.
8. **Welch B.L.** The Generalization of Student problem when several different population variances are involved / Biometrika, 1947, vol.34, No.1/2, P.28–35.
9. **Abramson J.H.** WINPEPI updated: computer programs for epidemiologists, and their teaching potential / Epidemiologic Perspectives & Innovations 2011, 8, P.1.
10. **Shaffer J.P.** Multiple hypothesis testing / Annual Review of Psych., 1995, vol.46, P.561–584.
11. **Sanna S. et al.** Fine mapping of five loci associated with low density lipoprotein cholesterol detects variants that double the explained heritability / PLOS Genetics, 2011, vol.7, No.7, P.1–10.
12. **Pearson K.** On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling / Philosophical Magazine Series, 1900, **5**, No.50 (302). – P.157–175.
13. **Fisher R.A.** On the interpretation of χ^2 from contingency tables, and the calculation of P / J. Roy. Statist. Soc., 1922, **85**, P.87 – 94.
14. **Goodman S.** A dirty dozen: Twelve P-value misconceptions / Seminars in hematology, 2008, **45**. – P.135–140.
15. **Hubbard R., Lindsay R.M.** Why P-values are not a useful measure of evidence in statistical significance test / Theory Psychology, 2008, Vol.18(1). – P.69–88.
16. **Sterne J.A.C., Smith J.D.** Sifting the evidence – what’s wrong with significance test? / British Medical Journal, 2001, 322. – P.226–231.
17. **Moran J.L., Solomon P.J.** A farewell to P-values? / Critical care and resuscitation, 2004, **6**. – P.130–137.