**КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ**
**ИНСТИТУТ УПРАВЛЕНИЯ, ЭКОНОМИКИ И ФИНАНСОВ**
***Кафедра эеономико-математического моделирования***

**А.М. ШИХАЛЁВ**

# ПРИМЕНЕНИЕ ТАБЛИЦ ВЗАИМНОЙ СОПРЯЖННОСТИ

**Учебно-методическое пособие**

**Казань – 2015**

**УДК (075.8)311**
**ББК 60.6я73**

*Принято на заседании кафедры экономико-математического моделирования*
*Протокол № 1 от 18 сентября 2014 года*

**Рецензенты:**
кандидат экономических наук,
доцент кафедры экономико-математического моделирования КФУ **Е.Л. Фесина;**
кандидат экономических наук,
доцент кафедры экономико-математического моделирования КФУ **Е.И. Кадочникова**

**Шихалёв А.М.**
**Применение таблиц взаимной сопряженности** / А.М. Шихалёв. – Казань: Казан. ун-т, 2015. – 32 с.

Учебно-методическое пособие предназначено для подготовки и выполнения студентами самостоятельной работы 4 из раздела «Общая теория статистики».

Важное место в изучении взаимосвязи социально-экономических явлений и процессов занимает исследование особенностей распределения единиц совокупности по двум признакам. По характеру распределения можно судить, случайно оно или неслучайно, есть ли зависимость между исследуемыми признаками, положенными в основу группировки, или нет. Если же связь между явлениями существует, исследователю предоставляется возможность оценки тесноты этой связи. Как правило, подобным исследованиям предшествуют исследования на предмет законов распределения исследуемых случайных величин. Если же такие исследования не проводились либо закон распределения неизвестен, используются непараметрические статистики.

Применение таблиц взаимной сопряженности рассматривается на модельном примере. В заключение предлагаются варианты выполнения самостоятельной работы.

Ключевые слова: статистические показатели, взаимосвязь, случайное и неслучайное распределения, эмпирические и теоретические частоты, критерий Пирсона, нулевые гипотезы, коэффициенты тесноты связи: ассоциации, контингенции, Пирсона, Чупрова.

**СОДЕРЖАНИЕ**

# 1. ОЦЕНКА НАЛИЧИЯ СВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ

## 1.1. Анализ таблиц сопряженности. Постановка задачи

Как нам уже известно, тесноту взаимосвязей между двумя наблюдаемыми социально-экономическими явлениями и процессами наиболее оперативно и просто можно, например, оценить с помощью вычислений коэффициентов корреляции (парной ранговой, линейной и др.) и парной регрессии. Однако на практике осуществить такое непосредственно не всегда возможно, так как все методы в качестве исходных данных подразумевают наличие как минимум двух статистических совокупностей и, что весьма важно, равной мощности (содержащие равное число элементов) и для получения удовлетворительных качественных характеристик, желательно, значительной мощности.

Так, если $X = \{x_i\}$, $i = 1,n$; $Y = \{y_j\}$, $j = 1,m$, то предполагается, что $n = m$. Если же размерности исследуемых статистических совокупностей неодинаковы, то исследователю необходимо как-то выровнять их: так, если, например, $n < m$, то надо либо сократить мощность множества Y до n, либо нарастить мощность множества X до m каким-либо способом – формализованным (путем экстраполяции) или любым другим, менее формализованным, например, одним из экспертных методов.

Кроме того, нетрудно показать на отдельных методах вычисления искомого коэффициента корреляции как объективного показателя тесноты связей между статистическими совокупностями X и Y, в частности, на примере вычисления коэффициента линейной корреляции, - его величина будет тем достовернее, чем больше объем $n = m$ - пар элементов множеств X и Y. То есть с улучшением статистики величина случайной ошибки закономерно уменьшается.

Однако на практике встречаются постановки задач определения тесноты связи для случая двух статистических совокупностей предельно малой мощности, когда число их элементов не превышает двух ($n = m = 2$). Тогда в

качестве аппарата исследования наличия между ними связи с оценкой степени ее тесноты используют метод таблиц взаимной сопряженности (ТВС). Если исследуется взаимосвязь каких-либо двух признаков в альтернативной шкале («признак присутствует» - «признак не присутствует»), то применяются ТВС, которые называют «четырехпольными», что и рассмотрим на модельном примере.

### 1.2. Заполнение четырехпольных таблиц на модельном примере

Предположим, что перед нами как исследователями администрацией института, ведающей заочным обучением в институте, поставлена задача оценить тесноту связи (если она есть) между успеваемостью слушателей заочного отделения (пусть для примера их будет 600; обозначим как N=600) от профиля их повседневной деятельности (характера работы) по соответствующим учебным дисциплинам (так или иначе связанными с содержанием повседневной работы). Понятно, что успеваемость слушателей, скажем, по такой специальной учебной дисциплине, как «Бухгалтерский учет», ожидается выше среди остальных обучаемых, если по характеру работы они исполняют обязанности бухгалтера какого-либо учреждения или являются работниками экономического отдела. Их успеваемость по данной дисциплине просто обязана быть выше. Однако данная оценка, несмотря на очевидность ситуации, является субъективной по своей природе. Поэтому наше содержание предвидения такого рода нуждается в известной объективации, что и должно выразиться в наличии (или отсутствии) связи между характером работы и успеваемостью по соответствующей учебной дисциплине в форме тесноты связи. Если же такая связь установлена, то требуется выяснить степень ее тесноты.

Для решения поставленной даже такой простой для понимания и представления условий задачи необходимо каким-то образом эти условия эксплицировать (формализовать).

Итак, по условию задачи нам необходимо прежде всего формализовать оба признака - «характер работы» и «успеваемость».

С «характером работы» - экспликация вполне проста и очевидна: первая позиция (i = 1) – это «работающие по профилю института»; вторая позиция (i = 2) – «работающие не профилю института» (см. табл. 1). Однако «успеваемостью» однозначно формализовать сложнее.

Можно за показатели успеваемости с одной стороны взять только отличные оценки (положительная успеваемость), а с другой стороны – только отрицательные оценки (отрицательная успеваемость) – уже два варианта формализации. При этом к положительной успеваемости можно отнести не только оцени «отлично», но и «хорошо» - еще вариант. А можно «успеваемость» эксплицировать и так: «сдавшие сессию без неудовлетворительных оценок» (j = 1) и «получившие неудовлетворительные оценки» (j = 2) – тоже вариант, на чем свой выбор вариантов на данном этапе формализации исходных данных можно завершить.

Далее необходимо обеспечить сформулированные нами градации «характера работы» и «успеваемости» соответствующими статистическими данными (статистическими показателями) и придать им форму расчетной таблицы следующего вида (табл. 1), которую как рабочую таблицу на первом этапе заполнения легко частично заполнить (последняя графа), воспользуясь данными, предоставленными учебной частью (отделом кадров) института, даже в начале семестра.

Для этого на первом этапе заполнения табл. 1 надо просто сосчитать число «работающих по профилю института» и работающих «не по профилю института». Это организовать несложно, поскольку для достижения такой частной цели даже списка слушателей специально составлять на данном этапе не требуется. Но такой список потребуется на следующем, послесессионном, периоде наблюдений (второй этап заполнения таблицы исходных данных).

Для лучшего понимания процесса заполнения таблиц взаимной сопряженности продемонстрируем этот процесс поэтапно.

Начало заполнения подобных таблиц (первый этап) следует начинать с отображения информационной обстановки, а именно: корректно внести признаки по их градации в соответствующие строки и столбцы таблицы, как это показано на примере табл. 1. Кроме признаков, уже в начале семестра можно заполнить числовой информацией последний столбец табл. 1.

*Таблица 1*

Характер работы и успеваемость слушателей заочного отделения (заготовка)

| № | | Успеваемость | | Всего |
|---|---|---|---|---|
| i | Характер работы | Сдавшие сессию без неудовлетворительных оценок j = 1 | Получившие неудовлетворит. оценки j = 2 = n | слушателей |
| i = 1 | Работающие по профилю института | | | 320 |
| i = 2 | Работающие не профилю института | | | 280 |
| Всего слушателей | | | | 600 |

Для заполнения третьей и четвертой графы (второй этап) понадобиться дождаться окончания сессии (см. табл. 2). В качестве статистической информации необходимо воспользоваться экзаменационными ведомостями по учебным группам (с обязательными пометками «работает по профилю» «работает не по профилю») и привести обобщенные результаты в табл. 2.

Несмотря на изменчивость информационной обстановки (кто-то из списков исследуемых групп выбыл, кто-то зачислен дополнительно), в итоге необходимо обеспечить баланс содержания рабочих клеток сформированной нами таблицы взаимной сопряженности вида табл. 1 по строкам и столбцам, так как последующие расчеты связаны с вычислением пропорций при определении значений «теоретических» частот по имеющимся значениям «эмпирических» частот, полученных в результате статистических наблюдений.

*Таблица 2*

Характер работы и успеваемость слушателей заочного отделения

| № i | Характер работы | Успеваемость: Сдавшие сессию без неудовлетворительных оценок, j = 1, X | Успеваемость: Получившие неудовлетворит. оценки, j = 2, Y | Всего слушателей |
|---|---|---|---|---|
| i = 1 | Работающие по профилю института | 270 | 50 | 320 |
| i = 2 | Работающие не профилю института | 150 | 130 | 280 |
| Всего слушателей | | 420 | 180 | 600 |

Поскольку появились заполненные третий и четвертый графы табл. 2, получим матрицу размером 2 х 2. При этом третью графу можно представить как статистическую совокупность X = {270; 150}, четвертую графу - как совокупность Y = {50; 180}. Для установления тесноты связи между профилем работы и успеваемостью по специальным предметам в рамках ранее изученных методов мы можем воспользоваться методиками вычисления с помощью коэффициента парной ранговой корреляции, линейной корреляции и знаковой корреляции.

Остальные цифровые показатели в строке и графе «всего слушателей» нам просто не понадобятся, но были полезными при заполнении табл. 1 и 2 для контроля корректности заполнения (суммы по строкам и столбцам таблицы должны совпадать, чтобы избежать механических ошибок и ошибок регистрации).

При установлении связи между двумя явлениями мы, вообще говоря, можем воспользоваться, например, аппаратом вычисления коэффициента парной ранговой корреляции. Не прибегая к подробным вычислениям, непосредственно из табл. 2 видно, что совокупность X в ранговой шкале можно записать как X = {1; 2}, а совокупность У как У = {2; 1}. Очевидно, что коэффициент парной ранговой корреляции $\rho_{xy}$ = - 1. Большую точность

решения при данном методе мы получить вряд ли сможем: есть только две пары данных, больше данных нет.

Поэтому расчеты коэффициентов корреляции и другими методами не смогут привести к надежным результатам вследствие малого числа исходных статистических данных. Следовательно, для работы с небольшим числом исходных данных необходимо применение какого-либо принципиально иного метода. Таковым в изложенной постановке и является метод таблиц взаимной сопряженности (здесь – четырехпольных).

### 1.3. Анализ четырехпольных таблиц на модельном примере

Для применения данного метода необходимо несколько по-иному оценить имеющиеся исходные данные, приведенные в табл. 2, в плане поставленной проблемы. Для этого будем рассматривать цифровые показатели (табл. 2) как некоторые эмпирические частоты $m_{ij}$ (полученные нами путем организации наблюдений) появления соответствующего события, что и покажем в табл. 3.

*Таблица 3*

Распределение эмпирических частот

| № i | Характер работы | Успеваемость | | Всего слушателей |
|---|---|---|---|---|
| | | Сдавшие сессию без неудовлетворительных оценок, j = 1 | Получившие неудовлетворит. оценки, j = 2 = n | |
| i = 1 | Работающие по профилю института | $m_{11}$ = 270 | $m_{12}$ = 50 | 320 |
| i = 2 = m | Работающие не профилю института | $m_{21}$ = 150 | $m_{22}$ =130 | 280 |
| Всего слушателей | | 420 | 180 | 600 |

То есть вместо двух коротких статистических совокупностей X и У ту же информацию рассмотрим иначе – как матрицу распределения различных

состояний объекта исследования размером 2 х 2. Так, число «сдавших сессию без неудовлетворительных оценок» и при этом «работающие по профилю института» равно 270 слушателям; число «получивших не3довлетворительные оценки» и «работающих по профилю института» равно 50 слушателям; число «сдавших сессию без неудовлетворительных оценок» и «работающих не по профилю института» равно 150 слушателям; число «получивших неудовлетворительные оценки» и «работающих не по профилю института» - 130 слушателям.

Тогда на данном этапе исследования задача сводится принципиально к двум моментам:

1. Определить, случайно или неслучайно распределение основной информации в табл. 3 и сделать вывод о наличии или отсутствии зависимости успеваемости слушателей заочного отделения института от соответствия характеру работы.

2. Измерить тесноту данной зависимости, если она есть.

Ответ на первый вопрос «да»-«нет» носит принципиальный характер, на второй вопрос – частично очевиден по результатам ответа на вопрос первый: если зависимость есть, то и ее теснота должна быть заметно выражена; если зависимости нет, то и степень тесноты следует ожидать незначительной. Поэтому ответ на второй вопрос по отношению к ответу на первый и главный вопрос носит скорее иллюстративный, пояснительный характер. Причем оценку тесноты связи по формулам, приведенным далее, можно проводить вне зависимости, существует ли связь между исследуемыми признаками или не существует: формулы для оценки тесноты связи такого ограничения не содержат.

Поскольку анализ таблиц взаимной сопряженности (на примере четырехпольной таблицы) мы рассматриваем на так называемом «модельном примере», то ответ на первый вопрос в общем-то просматривается на умозрительном уровне по содержанию табл. 2 и 3. Пожалуй, связь между характером работы и успеваемостью, видимо, судя по нашей исходной

информации, все же существует: работающие по профилю имеют 270 положительных оценок и только 50 отрицательных (успешность обучения выражена как 270 к 50-ти), тогда как у работающих не по профилю успешность обучения выражена менее заметно: 150 против 130-ти.

Для *объективированного анализа* (не зависящего от наших личных впечатлений) необходимо и достаточно воспользоваться критерием Пирсона $\chi^2$ («хи-квадрат»), для чего необходимо вычислить $\chi^2_р$ (расчетный), а затем сравнить его значение с $\chi^2_т$ (табличными; их три для разных уровней значимости – для ошибок в 10%, 5% и 1%) по известному правилу:

$$\chi^2_р = \sum_{i,j=1}^{m,n} \frac{(m_{ij} - m_{ij}^т)^2}{m_{ij}^т}, \qquad (1)$$

где $m_{ij}$ - эмпирические частоты табл. 3 и, альтернативные им, так называемые теоретические частоты $m_{ij}^т$, рассчитанные по формуле (2) для первого квадрата и приведенные в скобках в новой рабочей табл. 4.

*Таблица 4*

Характер работы и успеваемость слушателей заочного отделения в виде отдельного значения теоретической частоты

| № i | Характер работы | Успеваемость: Сдавшие сессию без неудовлетворительных оценок, j = 1 | Успеваемость: Получившие неудовлетворительные оценки, j = 2 = n | Всего слушателей |
|---|---|---|---|---|
| i =1 | Работающие по профилю института | $m_{11}$ = 270 ($m_{11}^т$ = 224) | 50 | 320 |
| i = 2 = m | Работающие не профилю института | 150 | 130 | 280 |
| Всего слушателей | | 420 | 180 | 600 |

Правило (2) показано на примере вычисления первой «теоретической» частоты. Значение $m_{11}^т$ = 224 получено из следующих соображений. Это – так

называемое «теоретическое значение частоты успешных сдач работающих по профилю» как справедливое среднестатистическое, и вот почему. Всего работающих по профилю = 320; положительные оценки получили как среди работающих по профилю, так и неработающих по профилю = 420; всего слушателей = 600 (как раз эмпирическое значение $m_{11}$=270 находится на пересечении суммарных значений строк и столбцов)

В результате опыта (наблюдений) нами было зафиксировано 270 слушателей, сдавших сессию успешно из числа работающих по профилю института. А вот если бы все оказалось строго *пропорционально* согласно полученных сумм по столбцам и строкам, равных 420 и 320 – соответственно, то число сдавших успешно из числа работающих по профилю было бы результатом решения простой пропорции (из 5 - го класса школы):

$$m_{ij}^{\text{т}} \ — \ 320 \qquad (2)$$
$$420 \ — \ 600,$$

откуда следует, что неизвестное теоретическое (средне статистически ожидаемое $m_{11}^{\text{т}}$) значение сдавших успешно из числа работающих по профилю института из правила решения пропорций (2) будет следующим:

$$m_{11}^{\text{т}} = \frac{320 \cdot 420}{600} = 224 \text{ (чел.)}. \qquad (3)$$

Аналогично составленным пропорциям вида (2) и их решению вида (3) находятся и остальные теоретические частоты событий, заявленных в исходной таблице: $m_{12}^{\text{т}}$ = 96; $m_{21}^{\text{т}}$ = 196; $m_{22}^{\text{т}}$ = 84 (чел.) – см. табл. 4.

Нетрудно заметить, что суммы по строкам и столбцам в табл. 4 для теоретических значений частот строго равны соответствующим итоговым показателям, вычисленным на основании эмпирических, то есть наблюдаемых нами событий (частот).

Тогда табл. 4 примет окончательный вид табл. 5, наконец-то пригодной для организации дальнейших расчетов по формуле (1).

*Таблица 5*

Характер работы и успеваемость слушателей заочного отделения в виде окончательно заполненной четырехпольной таблицы сопряженности

| № i | Характер работы | Успеваемость: Сдавшие сессию без неудовлетворительных оценок, j = 1 | Успеваемость: Получившие неудовлетворительные оценки, j = 2 = n | Всего слушателей |
|---|---|---|---|---|
| i = 1 | Работающие по профилю института | $(m_{11}^{т} = 224)$ $m_{11} = 270$ a | $(m_{12}^{т} = 96)$ $m_{12} = 50$ b | 320 |
| i = 2 = m | Работающие не профилю института | $(m_{21}^{т} = 196)$ $m_{21} = 150$ c | $(m_{22}^{т} = 84)$ $m_{22} = 130$ d | 280 |
| Всего слушателей | | 420 | 180 | 600 |

Разумеется, табл. 5 заполнена на основе табл. 1 (обозначим расчетные поля в табл. 5 дополнительными символами a,b,c,d - принятыми обозначениями тех же эмпирических частот для полей четырехпольных таблиц $m_{ij}$).

При решении практических задач на основе табл. 1 можно, пусть и неодномоментно, получить табл. 5. Поэтому табл. 2 – 4 приведены здесь затем, чтобы продемонстрировать именно *поэтапность* наполнения наших исходных представлений дополнительными знаниями, полезными в дальнейшем для ответов на два, ранее заданных, вопроса:

1. Определить, случайно или неслучайно распределение основной информации в табл. 5 и сделать вывод о наличии или отсутствии зависимости успеваемости слушателей заочного отделения института от соответствия характера работы соответствующим учебным дисциплинам.

2. Измерить тесноту данной зависимости, если она есть. Если даже ответ на предыдущий вопрос будет отрицательным, вычисление тесноты связи не утрачивает смысла, что важно при проведении мониторинга объекта исследования.

## 2. ЗАДАЧА О НАЛИЧИИ СВЯЗИ И СТЕПЕНИ ЕЕ ТЕСНОТЫ

### 2.1. Оценка случайности или неслучайности распределения информации

Как уже отмечалось ранее, прежде всего необходимо дать ответ на вопрос – случайно или неслучайно распределение исходной информации в виде эмпирических частот табл. 5.

Нашу готовность ответить на этот вопрос продемонстрируем умением применять формулу Пирсона (1), расписав ее на четыре слагаемых (по числу информационных клеток табл. 5):

$$\chi^2_p = \sum_{i,j=1}^{m,n} \frac{(m_{ij} - m_{ij}^т)^2}{m_{ij}^т} = \frac{(270 - 224)^2}{224} + \frac{(150 - 196)^2}{196} + \frac{(50 - 96)^2}{96} + \frac{(130 - 84)^2}{84} = 67{,}48.$$

Фактически мы получили *меру рассогласования* между эмпирическими и теоретическими значениями частот случайной величины успешности / неуспешности успеваемости некоего абстрактного слушателя в зависимости от его профессиональной принадлежности. Таких слушателей зафиксировано ровно 600, что можно записать как N = 600.

Для того, чтобы сделать вывод о случайности или неслучайности распределения, полученного нами в результате поэтапного заполнения, начиная от табл. 1 и до самой табл. 5, необходимо полученное нами расчетное значение «хи-квадрат» $\chi^2_p$ , равного ≈ 67,5 сравнить с его же табличным значением $\chi^2_т$ , рассчитанного автором метода (Пирсоном) для заданных уровней значимости (α) в соответствии со степенями свободы (df), чтобы оценить, *допустимо ли* случайное расхождение между эмпирическими ($m_{ij}$) и теоретическими (случайными) частотами ($m_{ij}^т$).

Обычно в статистике оперируют уровнями надежности в 90, 95 и 99% (уровень *значимости* α, ошибки, соответственно – 10%, 5% и 1%, что в относительных единицах (поделить на сто) – 0,10; 0,05 и 0,01, которые как

правило используется в стандартных таблицах, в частности, в таблице для $\chi^2_т$ - «табличных» значений.

Что касается числа степеней свободы df, то оно оценивается для таблиц взаимной сопряженности следующим образом:

$$df = (m - 1)\cdot(n - 1), \quad (4)$$

где m – число исходных строк рабочей таблицы (табл. 1 – 5); n – число столбцов тех же таблиц. В данном случае число строк с исходной информацей = 2, число столбцов с исходной информацией = 2 (если переставим число столбцов и таблиц в формуле (4), то ничего не изменится: как известно, при перестановке множимого и множителя произведение не изменяется. Тогда по формуле (4) получим:

$df = (2 - 1)\cdot(2 - 1) = 1$ (число степеней свободы).

Так, для различных *уровней значимости* α и *числу степеней свободы* df заранее рассчитана таблица Пирсона, фрагмент которой приводится в табл. 6 (нам необходимо знать число степеней свободы, чтобы войти в эту таблицу и получить *три* табличных значения $\chi^2_т$ , чтобы сравнить их с результатами вычислений по формуле (1) – расчетным значением $\chi^2_р$ .

Итак, обращаясь к содержанию табл. 6, по параметру db = 1 мы всегда будем располагать тремя табличными значениями для уровней ошибки в 10%, 5% и 1% (выделено шрифтом) и равны соответственно 2,71; 3,84 и 6,63. Запишем в удобном виде еще раз:

$$\alpha = 0{,}10 \quad 0{,}05 \quad 0{,}01$$

$$\chi^2_т = 2{,}71 \quad 3{,}84 \quad 6{,}63. \quad (5)$$

Таким образом, «хи-квадрат» расчетное ($\chi^2_р$) мы определяем сами, тогда как «хи-квадрат» табличное ($\chi^2_т$) находим из табл. 6. Получается, чтобы получить $\chi^2_р$, мы оперируем только лишь эмпирическими и «теоретическими» значениями исходной таблицы - $m_{ij}$ и $m_{11}{}^т$ , тогда как для получения $\chi^2_т$ мы ориентируемся на имеющиеся в нашем распоряжении «степени свободы» db и уровнем значимости α в10%, 5% и 1%.

*Таблица 6*

Значения $\chi^2_т$ – критерия Пирсона при уровне значимости α = 0,10; 0,05 и 0,01 (фрагмент статистической таблицы)

| Степени свободы, | Уровни значимости, α | | |
|---|---|---|---|
| df | 0,10 | 0,05 | 0,01 |
| **1** | **2,71** | **3,84** | **6,63** |
| 2 | 4,61 | 5,99 | 9,21 |
| 3 | 6,25 | 7,81 | 11,34 |
| 4 | 7,78 | 9,49 | 13,28 |
| 5 | 9,24 | 11,07 | 15,09 |
| 6 | 10,64 | 12,59 | 16,81 |
| 7 | 12,02 | 14,07 | 18,84 |
| 8 | 13,36 | 15,51 | 20,09 |
| 9 | 14,68 | 16,92 | 21,67 |
| 10 | 15,99 | 18,31 | 23,21 |

Следовательно, в качестве формализации можно записать, что эти разные функции Пирсона «хи-квадрат» - расчетная и табличная, - могут быть *в общем виде* записаны так:

$\chi^2_р$, = f ($m_{ij}$ , $m_{11}{}^{т}$) – функция эмпирических и теоретических частот (см. табл. 5); (6)

$\chi^2_т$ = f (db, α ) – функция степеней свободы и уровня значимости (см. табл. 6). (7)

Две сравниваемые между собой функции различаются по факторам, действующих на них (см. правые части выражений 6 и 7). Однако совместить их можно на содержательном уровне, поскольку расчет табличных значений осуществляется с использованием независимым от результатов частных экспериментов специальных разделов теории вероятностей и математической статистики. Данное совмещение выражений (6) и (7) и осуществляется в процессе *верификации* проведенных расчетов.

## 2.2. Создание и верификация «нулевой гипотезы»

Для реализации процедуры верификации - совмещения выражений (6) и (7), в статистике разработана специальная *процедура верификации*, которую проводят в рамках так называемой «нулевой гипотезы» $H_0$ – по сути своей - содержательное совмещение выражений (7) и (8)

Гипотеза как стандартный примем – это предположение, нулевой ее назвали, видимо, потому, что она изначально констатирует отсутствие предполагаемой (искомой нами) связи:

$H_0$. : «Профильность работы и успеваемость *не связаны*». Если $\chi^2_р, < \chi^2_т$, то нулевая гипотеза *нами принимается*.

$H_0$ *отвергается*, если $\chi^2_р, \geq \chi^2_т$, То есть профильность работы и успеваемость *связаны*.

С позиции теории логического познания триада «понятия – суждения – умозаключение» в процессе принятия или непринятия «нулевой гипотезы» проявляется следующим образом. На основе понятий «хи-квадрат» расчетного и «хи-квадрат» табличных (в обычных статистических таблицах их три) строятся суждения по правилам (6) и (7), в результате чего «нулевая гипотеза» принимается (один вариант умозаключения) или опровергается (другой вариант умозаключения). Для выражения ситуации (6) расчетное значение «хи-квадрат» лежит левее выписки из табл. 6, приведенной в записи (8), тогда как для выражения одной из ситуаций (7), как и показано в выписке (8), расчетное значение «хи-квадрат» (равное полученному ранее 67,48) лежит много правее.

Вообще говоря, если расчетное значение «хи-квадрат» равно или превышает значение минимальной табличной величины (для нашего числа степеней свободы), равное 2,71, то это иллюстрируется ситуацией (7) – отвержением «нулевой гипотезы» о предполагаемой ранее несвязанности исследуемых признаков.

Если же расчетное значение «хи-квадрат» меньше минимального значения табличной величины (для нашего числа степеней свободы), равное 2,71, то это иллюстрируется ситуацией (6) – принятием «нулевой гипотезы» о предполагаемой ранее несвязанности исследуемых признаков.

Если же расчетное значение «хи-квадрат» занимает при данном числе степеней свободы промежуточное положение между минимальным табличным значением (здесь 2,71 для $\alpha = 0{,}10$) и максимальным табличным значением (здесь 6,63 для $\alpha = 0{,}01$), то «нулевая гипотеза» опровергается с разной степенью надежности.

Таким образом, для решаемого модельного примера, располагая результатами сравнения рабочего значения параметра Пирсона «хи-квадрат» и трех табличных значений того же параметра, мы вынуждены *опровергнуть* $H_0$ : профильность профессии и успеваемость *связаны*, так как условие (7) выполняется ( $\chi^2_р, \geq \chi^2_т$). Формируя такое умозаключение, мы рискуем совершить ошибку не более, чем в одном случае из ста.

Для наглядности приведенных суждений в процессе верификации (поверки) «нулевой гипотезы» запишем в виде (8):

| Уровень ошибки α: | 0,10 (10%) | 0,05 (5%) | 0,01 (1%) | | |
|---|---|---|---|---|---|
| Из табл. 6 $\chi^2_т$: | 2,71 | 3,84 | 6,63 | | (8) |
| Расчетное $\chi^2_р$: | | | ….. | 67,48 | |

Следовательно, мы можем прийти к искомому умозаключению, например, следующего вида: «Профильность работы и успеваемость по соответствующим учебным дисциплинам связаны – с риском ошибиться не более, чем в одном случае из ста». Или: «Профильность работы и успеваемость связаны с надежностью не менее 99%». Верны оба утверждения. Если же будем утверждать обратное, что «Профильность работы и успеваемость не связаны», то «будем рисковать ошибиться более, чем в 99-ти случаях из ста».

Для большей ясности при выработке умозаключений приведем также наглядный пример (9) для некоторого промежуточного значения «хи-квадрат» расчетного, равного, например, 3,90:

| | | | | |
|---|---|---|---|---|
| Уровень ошибки α: | 0,10 (10%) | 0,05 (5%) | 0,01 (1%) | |
| Из табл. 6 $\chi^2_т$: | 2,71 | 3,84 | 6,63 | (9) |
| Расчетное $\chi^2_р$: | | 3,90 | | |

Умозаключение будет следующим: «нулевая гипотеза» о несвязанности профильности работы и успеваемости нами *опровергается* (то есть они связаны) с риском ошибиться не более, чем в 5-ти случаях из ста. Или таким: «нулевая гипотеза» о несвязанности профильности работы опровергается нами с достоверностью не менее 95% (точнее – от 95% до 99%).

Таким образом, эмпирическое распределение частот в рабочей таблице (табл. 5) *неслучайно* и скорее связано с зависимостью между признаками, положенными в основу группировки. То есть можно говорить о зависимости между характером работы слушателей и результатами сдачи ими экзаменов по специальным предметам. Именно таким является ответ на первый, главный вопрос: есть ли связь между признаками, положенными в основу рабочей табл. 5. Далее необходимо установить степень тесноты установленной связи.

Отметим также, что для четырехпольных ТВС вычисляется четыре коэффициента связи исследуемых признаков. В отличие от модельного примера с четырехпольной ТВС, для многопольных ТВС вычисляется только два параметра связи, что следует из характера исходной информации.

### 2.3. Оценка тесноты связи между признаками

Для измерения тесноты связи в общем случае (когда полей рабочей таблицы более четырех) используется коэффициент взаимной сопряженности Пирсона С и коэффициент взаимной сопряженности Чупрова $К_ч$.

Коэффициент взаимной сопряженности Пирсона

$$C = \left(\frac{\chi^2_p}{\chi^2_p + N}\right)^{1\backslash 2}, \text{ или } C = \left(\frac{\varphi^2}{\varphi^2 + 1}\right)^{1/2}, \text{ где параметр } \varphi^2 = \frac{\chi^2_p}{N}. \quad (10)$$

Коэффициент взаимной сопряженности Чупрова:

$$К_ч = \left(\frac{\varphi^2}{(db)^{1\backslash 2}}\right)^{1/2}. \quad (11)$$

Тогда на модельных данных табл. 5 согласно (10) и (11) получим:

$$C = \left(\frac{67{,}48}{67{,}48 + 600}\right)^{1/2} = 0{,}318. \quad \varphi^2 = \frac{67{,}48}{600} = 0{,}1125; \quad C = \left(\frac{0{,}1125}{1{,}1125}\right)^{1/2} = 0{,}318.$$

$$К_ч = \left(\frac{0{,}1125}{(1)^{1\backslash 2}}\right)^{1/2} = 0{,}335.$$

Теснота связи измерена. Однако для четырехпольных таблиц разработаны такие параметры связи, как коэффициент ассоциации $k_{acc}$ и коэффициент контингенции $k_{континг}$.

$$k_{acc} = \frac{ad - bc}{ad + bc} = \frac{270 \cdot 30 - 50 \cdot 150}{270 \cdot 130 + 50 \cdot 150} = 0{,}648. \quad (12)$$

$$k_{континг} = \frac{ad - bc}{[(a + b)(c + d)(a + c)(b + d)]^{1/2}} = 0{,}335. \quad (13)$$

Все рассчитанные показатели по (10), (11), (12) и (13), кроме коэффициента ассоциации, свидетельствует о том, что зависимость между

характером работы у слушателей и результатами сдачи экзаменов по специальным предметам примерно где-то в районе чуть «ниже средней».

Понятно, что для многопольных таблиц взаимной сопряженности, таблиц любых размеров, коэффициенты связи – ассоциации и контингенции (12) и (13), - не имеют смысла; рассчитываются только коэффициенты взаимной сопряженности Пирсона и Чупрова.

## 3. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ 4

Каждый студент получает вариант многопольной таблицы взаимной сопряженности (9 полей) в виде табл. 7, содержащей необходимые эмпирические результаты наблюдений $m_{ij}$. Необходимо:

1. Найти значения теоретических частот $m_{ij}^{т}$ , используя построенные пропорции вида (2), которые решить по правилу (3) и провести дополнение содержания табл. 3 до вида рабочей таблицы (табл. 5).

2. Рассчитать по формуле (1) значение критерия Пирсона $\chi^2_р$.

3. Определить число степеней свободы db по формуле (4) и войти в таблицу Пирсона (табл. 6), выписать значения $\chi^2_т$ для уровней значимости (уровней ошибки) 0,10; 0,05 и 0,01 (в относительных единицах) так, как это показано на фрагменте (8).

4. Поместить полученное расчетное значение $\chi^2_р$ сообразно его величине по отношению к табличным значениям с соответствующими уровнями значимости так, как это показано на выписках (8) и (9).

5. «Нулевая гипотеза» уже сформирована (она одна и та же): связи между изучаемыми признаками нет. Студентам по созданному им фрагменту вида (8) или (9) предстоит либо принять ее, либо опровергнуть.

Выполнение пп. 1 – 5 завершается реализация первой части поставленной задачи и свидетельствует о ее решении по установлению наличия связи между изучаемыми признаками или ее отсутствии.

6. Вторая часть задачи – получение значений тесноты связи между исследуемыми признаками по формулам Пирсона (10) и Чупрова (11) с попуьной оценкой тесноты связи в лингвистической шкале (сильная, средняя, ниже средней, слабая).

Для справки: до значений рассчитанных коэффициентов 0,2 – 0,3 тесноту связи относят к «ниже средней»; при значениях, приблизительно равных 0,5 – к «средней».

## 4. СОДЕРЖАНИЕ ВАРИАНТОВ САМОСТОЯТЕЛЬНОЙ РАБОТЫ 4

Студентам на базе полученных исходных данных предлагается участие в решении проблемы: связана ли урожайность (ее градации – высокая, средняя, низкая) со степенью полива (степени градации данного статистического показателя – обильный, средний, низкий). Для каждого варианта приведено *распределение участков* по степени полива (причина) и степени урожайности (следствие).

Понятно, что наличие (или отсутствие) связи производится при так называемых «прочих равных условиях» (модельный вариант). В действительности подобная задача, конечно же, должна решаться в несколько этапов, последовательно учитывающих такие факторы, как качество грунта посевных площадок, степень из освещенности, геолого-минералогические особенности мест расположения опытных участков, география (они необязательно должны находиться в непосредственной близости друг от друга) и ряд других.

И урожайность, измеряемая обычно ц/га (или т/га; кг/кв. метр; кг/на «сотку» и пр.), здесь измеряется в лингвистической шкале: «высокая», «средняя», «низкая» (основание деления – уровень урожайности). А сама конкретная урожайность известна лишь постановщикам задачи и специалистам, проводящим конкретные статистические наблюдения. То же относится и к поливу: не в тоннах на гектар, литрах на «сотку» и пр., а также в

лингвистической шкале: «обильный», «средний», низкий». Все эти наблюдения и сведены в отдельные варианты (см. ниже).

В данном случае на примерах распределения участков по степени полива и урожайности предлагается решить модельную (упрощенную по своим условиям) задачу на предмет:

1) есть ли *объективная связь* (по критерию «хи-квадрат» Пирсона) между поливом и урожайностью на опытных участках;

2) оценить тесноту связи (в независимости от ее наличия или отсутствия) по коэффициентам Пирсона и Чупрова с комментариями в предложенной лингвистической шкале.

Вариант № 1

Необходимо оценить, существует ли зависимость между двумя признаками – поливом и урожайностью на опытных участках. Распределение участков приведены в таблице 7.

*Таблица 7*

Распределение опытных участков по степени полива и урожайности

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 40 | 10 | 5 |
| Средний | 20 | 7 | 3 |
| Слабый | 0 | 5 | 10 |

1. Определить, случайно ли данное распределение, т.е. есть ли зависимость между степенью полива и уровнем урожайности.

2. Измерить тесноту зависимости между степенью полива и уровнем урожайности.

*Примечание: далее по вариантам приводятся только таблицы распределения опытных участков в соответствии со степенью полива и их урожайностью. Постановка задачи изложена на примере варианта 1.*

Вариант № 2

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 38 | 12 | 7 |
| Средний | 18 | 6 | 2 |
| Слабый | 0 | 4 | 9 |

Вариант № 3

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 36 | 14 | 9 |
| Средний | 16 | 5 | 2 |

| | | | |
|---|---|---|---|
| Слабый | 0 | 3 | 8 |

Вариант № 4

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 34 | 12 | 6 |
| Средний | 14 | 4 | 2 |
| Слабый | 0 | 2 | 6 |

Вариант № 5

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 32 | 10 | 8 |
| Средний | 12 | 3 | 2 |
| Слабый | 0 | 1 | 5 |

Вариант № 6

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 32 | 12 | 7 |
| Средний | 12 | 4 | 3 |
| Слабый | 1 | 2 | 6 |

Вариант № 7

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 32 | 14 | 8 |
| Средний | 12 | 5 | 3 |
| Слабый | 1 | 3 | 6 |

Вариант № 8

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 32 | 14 | 9 |
| Средний | 14 | 6 | 3 |
| Слабый | 1 | 4 | 7 |

Вариант № 9

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 32 | 15 | 10 |
| Средний | 15 | 7 | 4 |
| Слабый | 1 | 5 | 8 |

Вариант № 10

| Полив | Урожайность | | |
|---|---|---|---|
| | Выс | Сре | Низ |

| | окая | дняя | кая |
|---|---|---|---|
| Обильный | 33 | 16 | 10 |
| Средний | 16 | 8 | 5 |
| Слабый | 1 | 6 | 9 |

Вариант № 11

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 30 | 9 | 11 |
| Средний | 16 | 9 | 6 |
| Слабый | 2 | 7 | 10 |

Вариант № 12

| Полив | Урожайность | | |
|---|---|---|---|
| | Выс | Сре | Низ |

| | окая | дняя | кая |
|---|---|---|---|
| Обильный | 30 | 18 | 12 |
| Средний | 15 | 10 | 7 |
| Слабый | 3 | 8 | 11 |

Вариант № 13

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 29 | 19 | 13 |
| Средний | 14 | 12 | 8 |
| Слабый | 3 | 9 | 12 |

Вариант № 14

| Полив | Урожайность | | |
|---|---|---|---|
| | Выс | Сре | Низ |

| | | | |
|---|---|---|---|
| | окая | дняя | кая |
| Обильный | 28 | 20 | 14 |
| Средний | 13 | 12 | 9 |
| Слабый | 4 | 10 | 14 |

Вариант № 15

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 27 | 21 | 15 |
| Средний | 12 | 11 | 10 |
| Слабый | 5 | 11 | 15 |

Вариант № 16

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 27 | 21 | 15 |

| | | | |
|---|---|---|---|
| Средний | 11 | 12 | 11 |
| Слабый | 5 | 12 | 16 |

Вариант № 17

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 27 | 21 | 16 |
| Средний | 10 | 13 | 12 |
| Слабый | 6 | 13 | 17 |

Вариант № 18

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 27 | 22 | 18 |

| | | | |
|---|---|---|---|
| Средний | 11 | 14 | 13 |
| Слабый | 6 | 14 | 18 |

Вариант № 19

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 26 | 23 | 18 |
| Средний | 12 | 15 | 14 |
| Слабый | 5 | 15 | 19 |

Вариант № 20

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 26 | 24 | 19 |

| | | | |
|---|---|---|---|
| Средний | 14 | 16 | 15 |
| Слабый | 5 | 16 | 20 |

Вариант № 21

| Полив | Урожайность | | |
|---|---|---|---|
| | Высокая | Средняя | Низкая |
| Обильный | 26 | 25 | 20 |
| Средний | 15 | 17 | 16 |
| Слабый | 5 | 17 | 15 |

## ЛИТЕРАТУРА

1. Громыко Г.Л. Общая теория статистики: Практикум. М.: ИНФРА-М, 1999. 139 с. (Высшее образование).

2. Ниворожкина Л.И., Морозова З.А. Математическая статистика с элементами теории вероятностей в задачах и решениях: Учебное пособие. Москва: ИКЦ «МарТ»; Ростов-н/Д: Издательский центр «МарТ», 2005. 608 с. (Серия «Учебный курс»).

3. Рудакова Р.П., Букин Л.Л., Гаврилов В.И. Статистика. 2-е изд. СПб: Питер, 2007. 288 с.: ил. (Серия «Учебное пособие»).

*Учебное издание*

**Шихалёв** Анатолий Михайлович

ТПРИМЕНЕНИЕ ТАБЛИЦ ВЗАИМНОЙ СОПРЯЖЕННОСТИ

Дизайн обложки
***М.А. Ахметов***

Подписано в печать 15.06.2015.
Бумага офсетная. Печать цифровая.
Формат 60x84 1/16. Гарнитура «Times New Roman». Усл. печ. л. .
Тираж экз. Заказ

Отпечатано с готового оригинал-макета
в типографии Издательства Казанского университета

420008, г. Казань, ул. Профессора Нужина, 1/37
тел. (843) 233-73-59, 233-73-28