

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Кафедра «Теории управления»

Яковенко Максим Сергеевич

Магистерская диссертация

**Оптимизация объема выборки в маркетинговых
исследованиях**

Направление 01.04.02

Прикладная математика и информатика

Методы прикладной математики и информатики в задачах
управления

Научный руководитель,
доктор физ.-мат. наук,
профессор

Захаров Виктор Васильевич

Рецензент,

доктор экономических наук,
профессор НИУ Высшая
школа экономики СПб

Лукинский Владислав Валерьевич

Санкт-Петербург

2016

Содержание

| | |
|--|----|
| Введение | 3 |
| Цели и задачи исследования | 5 |
| Классы рассматриваемых задач | 5 |
| Обзор литературы | 6 |
| Глава 1 . Определение объёма выборки на основании гипотезы, что две выборки подчиняются одному закону распределения..... | 9 |
| 1.1. Постановка задачи..... | 9 |
| 1.2.Способы выбора элементов для удаления..... | 11 |
| 1.3.Дополнительные ограничения..... | 12 |
| Глава 2: Определение минимального объёма выборки..... | 14 |
| 2.1. Нахождение минимального количества элементов..... | 14 |
| 2.2.Алгоритм приближения по значениям в каждом из интервалов..... | 16 |
| 2.3. Насыщенность выборки..... | 18 |
| 2.4. Вычисления..... | 19 |
| 2.5. Оценка алгоритма | 24 |
| 2.6. Применение алгоритма для других классов задач..... | 26 |
| Выводы | 27 |
| Заключение | 28 |
| Список литературы | 29 |
| Приложения | 32 |

Введение

В качестве введения стоит дать несколько основных определений, которые будут использоваться в дальнейшем.

Маркетинг – вид человеческой деятельности, направленной на удовлетворение нужд и потребностей посредством обмена [1]

Маркетинговое исследование – это систематический поиск, сбор, анализ и представление данных и сведений, относящихся к конкретной рыночной ситуации, с которой пришлось столкнуться предприятию.[1]

Любое маркетинговое исследование преследует одну или несколько целей. В связи с этим можно подразделить их на несколько групп:

1. Поисковые цели — поиск информации, предназначенный для получения представления о структуре проблемы;
2. Описательные цели — описание интересующих событий и факторов, оказывающих на них влияние;
3. Каузальные цели — поиск ответа на вопрос: «есть ли некая причинно-следственная связь между событиями?»;
4. Тестовые цели — определение предположительных вариантов развития и оценка принятых решений;
5. Прогнозные цели — построение прогноза события.

По существу, цели исследования можно сформулировать с помощью пяти основных вопросов: «кто?», «что?», «когда?», «где?» и «как?» вместе с сопутствующим им вопросом «почему?», который связывает исследование с социологией и психологией.

Генеральная совокупность — вся изучаемая выборочным методом статистическая совокупность объектов и/или явлений обществ, жизни (единиц

отбора), имеющих общие качественные признаки или количественные переменные. [2]

Выборка – это некоторое количество элементов, которое используется в наблюдении/эксперименте.

В силу того, что маркетинговое исследование предоставляет ценную и важную информацию для принятия решений предприятия о выпуске новой продукции или изменении характеристик уже существующей, оно должно обладать несколькими обязательными свойствами:

1. Репрезентативность выборки–свойство выборки, заключающееся в ее способности адекватно представлять состояние дел в генеральной совокупности.[2]Пример: В качестве опроса были выбраны все жители Санкт-Петербурга. Так как в Санкт-Петербурге процент женщин приблизительно равен 55%, а мужчин 45% по последней переписи населения, то, соответственно, такая же пропорция должна сохраниться и в исследовании.

2. Определение погрешности полученной информации. В силу того, что опросить каждого объекта (человека/клиента/покупателя), как правило, не представляется возможным, то в исследовании участвует лишь их часть. В некоторых случаях необходимо также оценивать достоверность полученных данных.

В маркетинговом исследовании, как и в любом другом бизнесе, есть затраты. К ним относятся время, затраченное на получение одной единицы информации (опроса одного человека), и денежное вознаграждение. Соответственно, чем больше данных необходимо получить для удовлетворения целей исследования, тем больше ресурсов необходимо затратить. Таким образом, определение объема выборки является одной из приоритетных задач маркетинговых исследований.

Общую задачу определения объёма выборки можно сформулировать как определение минимального необходимого количества данных в конкретном исследовании.

Далее будет сформулирована математическая постановка задачи.

Цели и задачи исследования:

1. Рассмотреть классы возможных задач связанных с оптимизацией выборки;
2. Исследовать возможность уменьшения объема выборки маркетингового исследования;
3. Произвести вычисления на примере реального маркетингового исследования.

Классы рассматриваемых задач

1. Маркетинговые исследования;
2. Оценка брака на заводе;
3. Оценка качества лесных угодий/сельскохозяйственных посевов;
4. Определения минимального количества показаний (температура, влажность, скорость ветра) региона, достаточно для дальнейшего статистического исследования и удовлетворяющего необходимому наперед заданному уровню доверия.

Обзор литературы

Минимизация объёма выборки является одним из основных направлений в маркетинговых исследованиях, так как позволяет уменьшить

затраты на проведение опросов, сохраняя при этом необходимую точность. Данная тематика включает широкий спектр вопросов.

Одним из основных двигателей этого направления в настоящее время является Орлов Ю.Н.[3][4]. В его работах рассматриваются вопросы оценки и прогнозирования, связанные с нестационарными временными рядами. Приведены примеры из фондовой и валютной бирж. В том числе рассмотрен ряд задач, связанных с оцениванием уровня хаоса в выборке, что позволяет определить моменты, когда в выборку были внесены какие-либо изменения, и на их основании сделать соответствующие выводы.

Оцениванию выборки и ее оптимизации посвящена работа Чичагова В. В.[5], в которой он сравнивает методы оценки выборки, такие как несмещенная оценка и оценка максимального правдоподобия. Продолжая анализировать выборки, О.А. Бакаева [6] выводит в своих работах формулы минимального объема выборки для различных типов выборок. Необычный метод сравнения выборок встречается в работе Мотренко А. П.[7], в которой автор рассматривает уменьшение объема выборки на основании сравнения гистограмм двух выборок и использования расстояния Кульбака-Лейблера.

Тематика минимизации объема выборки также была поднята в работах преподавателей и студентов СПбГУ. Примером тому может служить работа Крамаренко Е. Ю.[8].

Фактически маркетинговое исследование является частью более общего вопроса, а именно «социального исследования» или «социального опроса». Если рассматривать данную тематику, то обязательно придется столкнуться с таким понятием, как корректировка выборочных данных. Этот вопрос подробно освещен в работах Орехова О.В. [9]

Применение:

В настоящее время задача оценивания объема выборки имеет несколько вариаций и применима в различных сферах жизни, таких как:

1.Маркетинговые исследования. Одним из примеров исследований в этой области является работа Ondřej Vilikus[10], где автор рассматривает, как одну из важнейших частей исследования, формирование шаблона для опроса респондентов, который включает в себя параметры сложности анкеты и сложности каждого из вопросов в анкете. На основании этих данных делается вывод о необходимом объеме выборки респондентов. Также в работе рассматривается случай, когда респондент не смог выбрать в каком-то из вопросов ни одного варианта ответа.

Некоторые из авторов заостряют своё внимание на стратифицированном(многослойном) исследовании. Так в работе [11] предполагается разбиение генеральной совокупности на некоторое количество страт (слоев) с коэффициентом значимости каждой из них и минимизации линейного критерия оценки стоимости. Делается вывод о том, что линейный критерий оценки стоимости является наиболее оптимальным для большинства стратифицированных исследований. Продолжая анализ стратифицированных исследований в работе [12], приводятся примеры двух методов для оценивания каждой из страт. При допущении фиксированной стоимости с заданным числом слоев и размеров выборок.

2.Изучение лесного покрова. Так Christophe Coronac коллегами рассматривают в работе [13] оптимизацию объема выборки применительно к изучению лесного покрова конкретной территории, а также вводят параметр шума-погрешности вычислений. Балансируя между значениями объема выборки и шума, они делают вывод о необходимом количестве деревьев в исследовании.

3. Сфера IT. Задача определения объема обучающей выборки машинного обучения рассмотрена в работе Richard H. Byrd [14]. Где также предложен метод увеличения выборки по отношению к методу Ньютона, основываясь на вычислении периодического градиента.

4. Страхование. Рассмотрено в работе [15], где авторы, используя Bootstrapping и линейный критерий оптимизации расходов компании, приходят к выводу, что цена ошибок оператора зачастую обходится компании слишком дорого, особенно с учетом параметра того, что само исследование стоит немалых денег. Кроме того, в работе даются оценочные значения для объема выборки.

5. Медицинские исследования. Хорошим примером в данном направлении является работа John M. Castelleo [16]. В данной задаче объем выборки невозможно рассматривать отдельно от корреляции влияния препарата на параметры пациента. Соответственно, чем выше влияние какого-либо препарата, тем меньшее количество респондентов необходимо для изучения. Приведены расчеты и формулы оценивания объема выборки.

6. Климатология. В работе [17] изложен вопрос оценки и построения прогноза в выборке одинаково распределенных случайных величин, где заданы лишь количества измерений и их критические значения. Вопрос может быть актуален для геологии, гидрологии. Также в статье приведены 2 метода для оценки оптимального объема выборки.

Также есть фундаментальные наработки. В частности в работе A. Andropov [18] излагает метод нахождения объема выборки для сложной функции, включающей параллельное и последовательное соединение, путем разделения сложной функции на более простые с помощью дерева решений и нахождения оптимальных значений на его ветвях.

Глава 1 . Определение объёма выборки на основании гипотезы, что две выборки подчиняются одному закону распределения

1.1. Постановка задачи:

Предположим, что после проведения маркетингового исследования, есть необходимость проведения повторного сбора информации. Можно ли использовать для такого исследования меньший объём данных?

Имея выборку X состоящую из N элементов $X = \{x_1 \dots x_N\}$, требуется определить минимальное количество элементов в выборке, при котором с наперед заданным уровнем значимости α сохраняется распределение начальной выборки. Для этого необходимо определить критерий сравнения.

Так как выборка задана качественными (номинальными) параметрами (хорошо, средне, плохо/наименования товаров и т.д.), то имеет смысл использовать критерий Пирсона [20].

Пусть

$X = \{x_1 \dots x_N\}$, - заданная начальная выборка

$X_L = \{x_1 \dots x_L\}$, - выборка, полученная после удаления из нее части элементов

$$X = \{x_1 \dots x_N\},$$

$$X_L = \{x_1 \dots x_L\},$$

$$L < N$$

Требуется проверить, подчиняются ли выборки одному закону распределения:

- Гипотеза H_0 : две выборки подчиняются одному и тому же закону распределения на уровне.
- Гипотеза H_1 : выборки подчиняются разным законам распределения

Критерий проверки гипотез :

$$\chi_{ch}^2 = N \sum \frac{(P_i^{theor} - P_i^{empir})^2}{P_i^{theor}} \quad (1)$$

Где

χ_{ch}^2 – значение критерия согласия, полученное по формуле (1)

$$P_i^{theor} = \frac{n_i}{N}$$

$$P_i^{empir} = \frac{n_i^l}{N^l}$$

Где

n_i и n_i^l являются частотами попадания в i -й интервал для начального и эмпирического распределений.

χ_{ch}^2 – значение критерия согласия, полученное по формуле (1)

Введенная ранее величина χ_{ch}^2 тоже будет случайной величиной в силу случайности выборки X и потому подчиняющаяся распределению χ^2 .

Суть критерия: сравнение χ^2 и χ_{ch}^2 , где χ^2 задана с $(m - 1)$ - степенями свободы и уровнем значимости α . Где m -количество интервалов в интервальном ряде.

Гипотеза H_0 принимается верной на уровне значимости α , если $\chi^2 > \chi_{ch}^2$.

Будем называть выборки **согласованными**, если выполняется условие:
 $\chi^2 > \chi_{ch}^2$.

Примечание:

К недостаткам критерия Пирсона относится потеря части первоначальной информации при малом количестве наблюдений в интервалах, что приводит к необходимости объединения интервалов с малым числом наблюдений или исключению таковых.

В связи с этим, как правило, требуют, чтобы в каждом из интервалов было не менее 5 элементов.

1.2. Один из способов определения количества элементов в выборке

Есть разные способы выбирать элементы из выборки. Воспользуемся для начала алгоритмом, подробно рассмотренным в работе Крамаренко[8]:

Имеем начальную выборку, состоящую из N элементов. Генерируем из нее N выборок из $N-1$ элементов путем последовательного удаления из начальной выборки по одному элементу. Выбираем случайную выборку из образованных таким способом и проверяем ее согласованность с начальной выборкой из N элементов.

Если они согласованы, тогда считаем порожденную матрицу «основной» и путем убираем из нее еще 1 элемент, получаем $N-1$ выборку из $N-2$ элементов. Вновь выбираем случайную из порожденных выборок и проверяем на согласованность с начальной (состоящей из N -элементов). Если выборки оказались не согласованными, то стоит проверить на согласованность другую из порожденных выборок на этом шаге. Продолжаем до тех пор, пока выборки не окажутся несогласованными. В таком случае, мы нашли искомое n -минимальный объем выборки. Так как этот процесс может привести к нескольким различным ответам (в силу случайности исключения элементов), то необходимо проделать все действия несколько раз и посчитать среднее значение между полученными объемами выборки.

1.3. Дополнительные ограничения:

Рассмотрим выражение

$$\chi_{ch}^2 = N \Sigma \frac{(P_i^{theor} - P_i^{empir})^2}{P_i^{empir}}$$

а именно составляющую, отвечающую за согласованность в l -м интервале

$$\frac{(P_i^{theor} - P_i^{empir})^2}{P_i^{empir}}$$

распишем более детально

$$\frac{\left(\frac{n_i^l}{N^l} - \frac{n_i}{N}\right)^2}{\frac{n_i^l}{N^l}}$$

Или иначе

$$0 \leq \frac{N^l}{n_i^l} \left(\left(\frac{n_i^l}{N^l} - \frac{n_i}{N} \right)^2 \right) \leq \left(\frac{1}{R_i} \right) \left(\left(R_i - \frac{n_i}{N} \right)^2 \right)$$

Где

$$R_i = \frac{n_i^l}{N^l}$$

Стоит отметить очевидное ограничение: $n_i^l > 0$, так как в противном случае деление на 0-а, значит в каждом из l -х интервалов должно оставаться минимум по одному значению.

Если рассматривать отдельный l -й промежуток, то ясно, что минимальное значение $=0$ при $\frac{n_i^l}{N^l} = \frac{n_i}{N}$, что возможно в 2х случаях:

1) мы не удаляли ни одного значения в данном интервале или же

2) при $\frac{n_i^l}{n_i} = \frac{N^l}{N}$. Т.е. при удалении элементов из других интервалов случилась ситуация, когда частота попадания в данный интервал стала равна вероятности попадания в данный интервал значений из начальной выборки.

И максимальное значение при $n_i^l = 1$, и $(N - N^l) \rightarrow \min$ (или иначе $R_i = \frac{n_i^l}{N^l} \rightarrow \min$). Такое возможно, в случае, когда мы удаляли элементы исключительно из данного интервала.

Тогда для всех интервалов выполнено

$$0 \leq \sum \frac{(P_i^{theor} - P_i^{empir})^2}{P_i^{empir}} \leq \sum \left(\frac{1}{R_i} \right) \left(\left(R_i - \frac{n_i}{N} \right)^2 \right)$$

Глава 2: Определение минимального объёма выборки.

2.1.Нахождение минимального количества элементов

В описанном выше алгоритме есть неявные ограничения: в каком случае и из какого столбца убирать элементы. Постараемся оптимизировать процесс нахождения минимального количества элементов выборки.

Для начала рассмотрим несколько случаев:

1-й случай. Будем называть его «Четкий лидер»
Рассмотрим случай со 100 элементами, 3-мя интервалами и 1-м четко выраженным лидером.

Допустим, что есть интервалы $i_1..i_3$

| | i_1 | i_2 | i_3 |
|--|-------|-------|-------|
| Количество попаданий в интервал(частота) | 3 | 95 | 2 |
| Вероятность | 0,03 | 0,95 | 0,02 |

В данном случае при уменьшении количества элементов, вероятности в интервалах будутувеличиваться и уменьшаться до тех пор, пока не достигнут некого равновесного случая, а именно:

| | i_1 | i_2 | i_3 |
|--|--------|--------|--------|
| Количество попаданий в интервал(частота) | 1 | 31 | 1 |
| Вероятность | 0,0303 | 0,9394 | 0,0303 |

Схема поиска «равновесного» решения:

Начать стоит с интервала, в который попало меньше всего значений. Попало 2 значения, то есть его можно уменьшить в 2 раза, чтобы количество попаданий равнялось 1.

Полученная таблица частот и вероятностей:

| | i_1 | i_2 | i_3 |
|--|--------|--------|--------|
| Количество попаданий в интервал(частота) | 1,5 | 47 | 1 |
| Вероятность | 0,0303 | 0,9495 | 0,0202 |

Так как 1,5 не целое, а нам нужны только целые значения, получаем:

| | i_1 | i_2 | i_3 |
|--|---------|---------|---------|
| Количество попаданий в интервал(частота) | 1 | 36 | 1 |
| Вероятность | 0,02632 | 0,94737 | 0,02632 |

2-й случай «2 чётких лидера»

| | i_1 | i_2 | i_3 |
|--|-------|-------|-------|
| Количество попаданий в интервал(частота) | 2 | 49 | 49 |
| Вероятность | 0,02 | 0,49 | 0,49 |

В данном случае при уменьшении количества элементов, вероятности в интервалах будут увеличиваться и уменьшаться до тех пор, пока не достигнут некоего равновесного случая, а именно:

| | i_1 | i_2 | i_3 |
|--|---------|---------|---------|
| Количество попаданий в интервал(частота) | 1 | 23 | 23 |
| Вероятность | 0,02128 | 0,48936 | 0,48936 |

2.2.Алгоритм приближения по значениям в каждом из интервалов:

Рассмотрим следующий алгоритм:

Подготовительный этап:

Дано: количество интервалов для рассмотрения m . Объём начальной выборки N и процент попадания в каждый из m интервалов P_i , уровень значимости β .

Необходимо:

1. Посчитать частоты и вероятности попадания значений в каждый из интервалов.

$$p_i = \frac{P_i}{100}$$

$$k_i = p_i N$$

$$i = 1..m$$

Где p_i – вероятность попадания значения в i –интервал, k_i –количество попавших значений в i - интервал(частота).

2. Найти значение χ^2 (квантиль) с уровнем значимости β и $(m - 1)$ степенями свободы

Убрать из рассмотрения m_0 -интервалов, в которых $k_i = 1$. Значения частот в этих интервалах не будут меняться в дальнейшем.

Алгоритм:

1. Рассмотреть, есть ли среди интервалов, интервалы, в которых частота $k_i = 1$. Посчитать количество таких интервалов и обозначить через m_0 . Убрать из дальнейшего рассмотрения эти m_0 интервалов, т.е. значения частот в этих интервалах не будут меняться в дальнейшем.

2. Определить минимальную частоту k_i^* :

$$k_i^* = \min(k_1 \dots k_i), k_i > 1, i = 1 \dots i - m_0,$$

Зафиксировать интервал/интервалы с минимальной частотой. Пусть количество таких интервалов m_1 , принять $a = 1$.

3. Сократить частоты в $(i - m_0)$ -интервалах в (k_i^*/a) раз. Так как частоты уменьшились, то уменьшилось и количество элементов в выборке и поменялись вероятности.

Посчитать частоты, вероятности и количество элементов в новой выборке.

Проверить новую и начальную выборки на соответствие критерию Пирсона $\chi^2 > \chi_{ch}^2$ (см. параграф 1.1 «постановка задачи»). Если критерий согласия не выполняется, то перейти к пункту 4, если критерий выполняется, перейти к пункту 5.

4. Увеличить a , $a = a + 1$. Вернуться в начало пункта 3 (с теми же частотами, что были до последнего сокращения).

Примечание: продолжать увеличивать a до тех пор, пока условие критерия не выполнится.

5. Убрать из дальнейшего рассмотрения m_1 интервалов, определенных в пункте 2.

Обновить $m_0 = m_0 + m_1$.

6. Продолжать пункты 2-5 до тех пор, пока $m_0 < m$. То есть пока из рассмотрения не будут убраны все интервалы.

Итогом работы алгоритма является новая выборка с количеством элементов $N^l < N$.

2.3.Насыщенность выборки

Исследование насыщенности выборки – это определение такого количества экспериментов, после которого последующие эксперименты не приносят новой информации. Часто подобные исследования проводятся в социологии или психологии, используя открытые вопросы.

Пример: необходимо узнать, какие проблемы вызывают больше всего беспокойств у людей в возрасте от 16 до 24. После проведения ряда исследований(предположим 30), удалось установить, что наиболее ценными и уникальными в своём роде было всего 15 ответов. Провели еще 30 опросов, выяснилось, что 2й опрос дал еще несколько уникальных ответов, допустим 2. Проводя еще ряд опросов, обнаружилось, что в них вообще не оказалось уникальных ответов, либо вероятность таковых слишком мала и, в целях исследования, их можно не принимать во внимание. Таким образом, на некоем шаге было определено, что после n -экспериментов новые опросы не дают статистически важной информации. Тогда n – точка насыщения.

Полагаем, что в силу обстоятельств мы знаем, какие должны быть параметры у распределения(пример: аналогичное маркетинговое исследование, проводившееся в прошлый период). И задачей является проверить, появились ли значимые изменения по данному вопросу.

В данной работе предлагается начать проверку(повторное исследование) с рассмотрения малого количества анкет. Затем проверить статистические параметры: распределение, среднее и дисперсию. Если есть значимые отличия от известных параметров, то добавить новые элементы в выборку.

Начать предполагается с размера выборки, определенного с помощью алгоритма в пункте 2.2.

2.4. Вычисления.

Проводится маркетинговое исследование с целью узнать, какими критериями руководствуются жители Москвы в возрасте от 15 до 60 при выборе сока, а также установить популярность марок сока.

В ходе 300 опросов[21] были получены следующие ответы на вопросы:

Результаты ответа на вопрос: «Какой самый важный критерий при выборе сока?»

| Качественный параметр | Вероятность, % | Вероятность | Частота |
|--|----------------|-------------|---------|
| Наличие скидок, акций | 1 | 0,01 | 3 |
| Отсутствие сахара | 3 | 0,03 | 9 |
| Производитель | 5 | 0,05 | 15 |
| 100% сок | 18 | 0,18 | 54 |
| Из каких фруктов или ягод сделан сок | 24 | 0,24 | 72 |
| Марка сока | 19 | 0,19 | 57 |
| Отсутствие красителей, ароматизаторов, | 15 | 0,15 | 45 |
| Рекомендовано к употреблению детям | 7 | 0,07 | 21 |
| Цена | 6 | 0,06 | 18 |
| Страна и место производства | 2 | 0,02 | 6 |

Результаты по первому ответу на вопрос: «Какие марки сока вы знаете?»

| | Вероятность, % | Вероятность | Частота |
|---------------|----------------|-------------|---------|
| Добрый | 23 | 0,23 | 69 |
| J-7 | 18 | 0,18 | 54 |
| Я | 9 | 0,09 | 27 |
| Моя семья | 6 | 0,06 | 18 |
| Фруктовый сад | 7 | 0,07 | 21 |
| Рич | 5 | 0,05 | 15 |
| Тонус | 5 | 0,05 | 15 |
| Сады придонья | 7 | 0,07 | 21 |
| Любимый | 3 | 0,03 | 9 |
| 100% Gold | 2 | 0,02 | 6 |
| Santal | 5 | 0,05 | 15 |
| Фруто-няня | 6 | 0,06 | 18 |
| Тропикана | 1 | 0,01 | 3 |
| Агуша | 3 | 0,03 | 9 |

Говорить о функции распределения, средней или о среднеквадратическом отклонении качественных параметров, можно лишь «условно» в силу номинальности (не являются числовыми) самих параметров. Это означает, что для оценки параметров выборки необходимо ввести индексацию.

Допустим, что в данном случае индексация введена так, чтобы данные принадлежали нормальному закону распределения.

Примечание 1: стоит отметить, что если используется индексация, то для сравнения данных в других аналогичных исследованиях, индексация должна остаться той же или быть пересмотрена и, соответственно, пересчитаны статистические показатели.

Примечание 2: предполагается, что индексация в большинстве случаев будет подбираться так, чтобы выборка соответствовала нормальному закону распределения.

Результаты проверки на нормальность при использовании критерия Пирсона на уровне значимости 0,05 приведены в таблице 1.(См приложение «Таблица 1»).

Итог: принимая во внимание индексацию с уровнем значимости в 5%, можно «условно» утверждать, что выборка подчиняется нормальному закону распределения.

Далее будут приведены расчёты для определения минимального объёма выборки с помощью алгоритма из пункта 2.2 при уровне значимости $\alpha=0,05$ и значением степеней свободы равного $(r - 1)$ (где r – количество интервалов). Вычисления разделены на 2 типа:

1. Предполагается, что будет использован механизм, применяющийся для определения насыщенности выборки. То есть, когда на первом шаге берется малая исследуемая группа, далее к ее рассмотрению добавляется еще одна группа такого же количества и так далее. При добавлении каждой новой группы происходит проверка на согласованность с генеральной выборкой. В данной работе будем называть такой подход «не менее 1 элемента в интервале».
2. В силу того, что критерий Пирсона может давать не очень точные результаты при частоте в интервале равной 5, предполагается, что в каждом из интервалов будет не менее 5 элементов. Интервалы, которые изначально содержат менее 5 элементов, отбрасываются. В данной работе будем называть такой подход «не менее 5 элементов в интервале».

Расчёты «не менее 1 элемента в интервале»:

- Выкладки для первого вопроса представлены в таблице 2. (См. приложение «Таблица 2»).

Итог: минимально возможное количество элементов выборки, при котором может сохраняться первоначальное распределение =28.

- Выкладки для второго вопроса, представлены в таблице 3. (См. приложение «Таблица 3»).

Итог: минимально возможное количество элементов выборки, при котором может сохраняться первоначальное распределение =21.

Итог 1: В силу того, что оба вопроса относятся к одному исследованию, то нужно выбрать максимальное значение из полученных объёмов. А именно: 28.

Итог 2: Количество анкет, от которого стоит «отталкиваться» при проведении повторного опроса 28. То есть в первую исследуемую группу стоит включить 28 человек. Затем создать еще одну группу из 28 человек из оставшихся 272 и проанализировать изменение вероятностей при добавлении 2й группы к исследованию. Нужно добавлять новые группы до тех пор, пока не выполнится условие критерия согласия Пирсона на заданном уровне значимости.

Расчёты «не менее 5 элементов в интервале»:

- Выкладки для первого вопроса представлены в таблице 4. (См. приложение «Таблица 4»).

Итог: минимальный объём выборки = 130.

- Выкладки для второго вопроса, представлены в таблице 5. (См. приложение «Таблица 5»).

Итог: минимальный объём выборки = 115.

Итог: в силу того, что оба вопроса относятся к одному исследованию, нужно выбрать максимальное значение из полученных объёмов и этого объёма будет достаточно для проведения исследования. А именно: 130.

2.5. Оценка алгоритма:

Так как для рассмотрения выбраны качественные данные и исследование проводится повторно, то для оценки минимального объёма используется «Бесповторный выбор»[19]. При бесповторном отборе элемент, попавший в выборочную совокупность, в генеральную совокупность не возвращается. То есть объём генеральной выборки в процессе исследования сокращается.

Основная формула для определения объёма новой выборки:

$$n = \frac{1}{\frac{\Delta^2}{t^2 y} + 1/N}$$

где n – объём новой выборки;

N – объём генеральной совокупности (в приведенном примере равен 300);

t – коэффициент соответствия доверительной вероятности;

Δ – допустимая ошибка;

y – доля исследуемого признака в генеральной совокупности

| | | | | |
|------------------------------|------|------|------|-------|
| t | 1 | 1,96 | 2 | 3 |
| Уровень значимости, α | 0,68 | 0,95 | 0,99 | 0,997 |

Используя «Бесповторный выбор», рассмотрим различные значения для уровня значимости α , допустимой ошибки при условии, что доля исследуемого признака равной 0,24 (вероятность самого популярного ответа). Получаем таблицы:

| | | | | | | | | |
|----------|--------|--------|---------|--------|--------|--------|-------|--------|
| Δ | 0,17 | 0,1 | 0,063 | 0,05 | 0,17 | 0,1 | 0,063 | 0,05 |
| t | 1,96 | 1,96 | 1,96 | 1,96 | 2,00 | 2,00 | 2,00 | 2,00 |
| y | 0,24 | 0,24 | 0,24 | 0,24 | 0,24 | 0,24 | 0,24 | 0,24 |
| N | 300 | 300 | 300 | 300 | 300 | 300 | 300 | 300 |
| n | 28,836 | 70,524 | 130,921 | 165,43 | 29,907 | 72,727 | 161,4 | 168,42 |

| | | | | | | | | |
|----------|-------|--------|--------|---------|--------|--------|--------|--------|
| Δ | 0,17 | 0,1 | 0,063 | 0,05 | 0,17 | 0,1 | 0,063 | 0,05 |
| t | 1,96 | 1,96 | 1,96 | 1,96 | 2,00 | 2,00 | 2,00 | 2,00 |
| y | 0,23 | 0,23 | 0,23 | 0,23 | 0,23 | 0,23 | 0,23 | 0,23 |
| N | 297 | 297 | 297 | 297 | 297 | 297 | 297 | 297 |
| n | 27,72 | 68,098 | 127,24 | 161,383 | 28,752 | 161,38 | 130,19 | 164,35 |

Итог 1: для расчётов «не менее 1 элемента в интервале» при учете доли признака в генеральной совокупности = 0,24 (24%), уровне значимости 5%, с доверительными границами $\pm 17\%$ достаточный объем = 29 анкетам.

Итог 2: для расчётов «не менее 5 элементов в интервале» при учете доли признака в генеральной совокупности = 0,24 (24%), уровне значимости 5%, с доверительными границами $\pm 6,3\%$ достаточный объем = 127 анкет.

2.6. Применение алгоритма для других классов задач:

Как отмечалось ранее, есть множество различных классов задач, где возможно использование алгоритма. В качестве еще одного примера применения рассмотрим задачу определения необходимого количества измерений в задаче мониторинга температур океанического течения.

Задача: имеются выборка из температур океанического течения. Необходимо определить объём выборки температур для повторного исследования на уровне значимости 5%.

Используя алгоритм из пункта 2.2. и расчёты типа «не менее 1 элемента в интервале» получаем, что минимальное количество измерений, при котором может выполняться критерий согласия Пирсона на уровне значимости 5% равно 178. (См. «Приложение 3»).

При расчётах «не менее 5 элементов в интервале» получаем объём выборки равный 788 элементам. (См. «Приложение 3»).

Выводы о проделанной работе:

1. Задача уменьшения объема выборки при повторном исследовании свойств генеральной совокупности является актуальной, так как позволяет сократить затраты на маркетинговое исследование.
2. Разработанный алгоритм приближения по значениям в каждом из интервалов позволяет сокращать объем выборки при повторном исследовании
3. Сравнение данного алгоритма с методом «бесповторный выбор» показывает его эффективность.
4. Возможности использования алгоритма приближения по значениям в каждом из интервалов не ограничиваются только маркетинговыми исследованиями. Область его применения может быть расширена.

Заключение

В данной работе рассмотрены классы возможных задач, связанных с оптимизацией выборки. Также был разработан алгоритм для определения минимального объёма выборки при повторном исследовании. Алгоритм может быть применён как к численным, так и к качественным параметрам (хорошо/средне/плохо), что продемонстрировано на примерах из маркетингового исследования.

Список литературы

1. Котлер Ф., Армстронг Г., Сондерс Д., Вонг В. Основы маркетинга: Пер. с англ. 2е Европ. Изд. М.; СПб.; К.: Издат. Дом "Вильямс", 2005: 502–512.
2. Сост. А.А. Грицанов, В.Л. Абушенко, Г.М. Евелькин, Г.Н. Соколова, О.В. Терещенко. Социология: Энциклопедия//Мн.: Книжный Дом, 2003.— 1312 с
3. Орлов Ю.Н., Шагов Д.О. Индикативные статистики для нестационарных временных рядов // Препринты ИПМим. М.В.Келдыша. 2011. № 53. 20 с
4. Орлов Ю.Н., Осминин К.П. Методы статистического анализа литературных текстов. // М.: Editorial URSS, Книжный дом «ЛИБРОКОМ», 2012.– 312 с.
5. Чичагов В. В. «Определение объема выборки, необходимого для построения асимптотических интервалов с использованием несмещенных оценок» // «Вестник» Пермского университета «Математика. Механика Информатика.» 2008 Вып. 4(20)
6. Бакаева О.А. Определение минимального объема выборки // «Вестник» Мордовского университета 2010 №4 Серия «физико-математические науки»
7. Мотренко А.П. «Оценка объема выборки в задачах прогнозирования» // ВКР Московский физико-технический институт. Москва 2014г.
8. Крамаренко Е.Ю. «Минимизация объема выборки в маркетинговых исследованиях» // СПбГУ 2014.
9. Орехов А.В. Рандомизированный алгоритм корректировки выборочных данных // ВЕСТНИК САНКТ-ПЕТЕРБУРГСКОГО УНИВЕРСИТЕТА. СЕРИЯ 10: ПРИКЛАДНАЯ МАТЕМАТИКА.

ИНФОРМАТИКА. ПРОЦЕССЫ УПРАВЛЕНИЯ. Издательство: Санкт-Петербургский государственный университет (Санкт-Петербург) ISSN: 1811-9905, 2015. 3. 96-104.

10. Ondřej Vilík «Optimization of sample size and number of tasks per respondent in conjoint studies using simulated datasets»

11. Saman Khowaja, Shazia Ghufra & M. J. Ahsan «Multi-objective optimization for optimum allocation in multivariate stratified sampling with quadratic cost» // Journal of Statistical Computation and Simulation, 82:12, 1789-1798 14 Jul 2011

12. Timur Keskin Turk, Sebnem Er «A genetic algorithm approach to determine stratum boundaries and sample sizes of each stratum in stratified sampling» // Computational Statistics & Data Analysis 52 (2007) 53 – 67. Available online 18 April 2007

13. Christophe Corona, Jérôme Lopez Saez, Markus Stoffel «Defining optimal sample size, sampling design and thresholds for dendrogeomorphic landslide reconstructions» // journal Quaternary Geochronology homepage: // www.elsevier.com/locate/quageo

14. Richard H. Byrd, Gillian M. Chiny, Jorge Nocedal, Yuchen Wux «Sample Size Selection in Optimization Methods for Machine Learning» // January 16, 2012

15. Nicholas Facchiano, Ashley Kingman, Amanda Olore, David Zuniga «Sampling Strategies for Error Rate Estimation and Quality Control»

16. John M. Castelloe «Sample Size Computations and Power Analysis with the SAS System» SAS Institute Inc., Cary, NC

17. Barranco-Chamorro I., J.L. Moreno-Rebollo, M.D. Jiménez-Gamero, M.V. Alba-Fernández «Estimation of the sample size n based on record values» // Mathematics and Computers in Simulation 118 (2015) 58–72. Available online 6 January 2015

18. Andronov A, Merkuryev Y. «Optimization of statistical sample sizes in simulation»// Journal of Statistical Planning and Inference 85 (2000) 93–102
19. Кендалл М., Стьюарт А. «Статистические выводы и связи» М.: Наука, 1973. С 595-604.
20. Кобзарь А.И. «Прикладная математическая статистика для инженеров и научных работников» Издание второе, исправленное. Москва «Физматлит» 2012г
21. Данные по маркетинговым исследованиям были взяты с сайта компании FDFGroup: <http://www.fdfgroup.ru/?id=261>

Приложение 1: таблица проверки гипотезы о нормальном распределении выборки.

| Качественный параметр | Индекс(x_i) | Процент($p_i\%$) | Границы интервалов x_i | | Интегральная функция $F(x)$ распределения на границах интервалов | | Вероятность попадания в интервал, p_i | $n \cdot p_i$ | $p_i\% - n \cdot p_i$ | $(p_i\% - n \cdot p_i)^2 / n \cdot p_i$ |
|---|-----------------|--------------------|--------------------------|--------------|--|-----------------|---|---------------|-----------------------|---|
| | | | $x_{i \min}$ | $x_{i \max}$ | $F(x_{i \min})$ | $F(x_{i \max})$ | | | | |
| Наличие скидок, акций | 1 | 1 | 0,5 | 1,5 | 0,00235 | 0,0114269 | 0,009077037 | 0,913174 | 0,086826 | 0,008255 |
| Отсутствие сахара | 2 | 3 | 1,5 | 2,5 | 0,011427 | 0,0422806 | 0,030853765 | 3,103972 | -0,10397 | 0,003483 |
| Производитель | 3 | 5 | 2,5 | 3,5 | 0,042281 | 0,1202468 | 0,077966116 | 7,843602 | -2,8436 | 1,030913 |
| 100% сок | 4 | 18 | 3,5 | 4,5 | 0,120247 | 0,2667422 | 0,146495444 | 14,73784 | 3,262162 | 0,722067 |
| Из каких фруктов или ягод сделан сок | 5 | 24 | 4,5 | 5,5 | 0,266742 | 0,4714452 | 0,204702968 | 20,59367 | 3,406327 | 0,563429 |
| Марка сока | 6 | 19 | 5,5 | 6,5 | 0,471445 | 0,6841803 | 0,212735152 | 21,40173 | -2,40173 | 0,269526 |
| Отсутствие красителей, ароматизаторов, консервантов | 7 | 15 | 6,5 | 7,5 | 0,68418 | 0,8486074 | 0,164427079 | 16,54181 | -1,54181 | 0,143707 |
| Рекомендовано к употреблению детям | 8 | 7 | 7,5 | 8,5 | 0,848607 | 0,9431224 | 0,094514982 | 9,508463 | -2,50846 | 0,661767 |
| Цена | 9 | 6 | 8,5 | 9,5 | 0,943122 | 0,983521 | 0,040398661 | 4,064215 | 1,935785 | 0,922014 |
| Страна и место производства | 10 | 2 | 9,5 | 10,5 | 0,983521 | 0,9963589 | 0,012837835 | 1,291521 | 0,708479 | 0,388645 |
| Дополнительные параметры расчета | $\sum p_i$ | 100 | среднее | 5,63 | | $\sum p_i =$ | 0,99400904 | | $\chi^2_{теор} =$ | 4,713806 |
| | $n \cdot$ | 100,6 | ср кв откл | 1,815 | | | | | $\chi^2_{крит} =$ | 14,06714 |

Приложение 2: Таблицы результатов вычислительных экспериментов для задачи определения необходимого количества анкет при повторном маркетинговом исследовании.

Таблица 2: применение алгоритма из пункта 2.2 к результатам исследования по вопросу: «Какой самый важный критерий при выборе сока?» («не менее 1 элемента в интервале»).

Первый шаг:

| | Индекс | Начальные вероятности | Начальные частоты | Новые частоты | Новые вероятности | Разница вероятностей | |
|-------------------------|--------|-----------------------|-------------------|---------------|-------------------|----------------------|-----------------|
| Наличие скидок, акций | 1 | 0,01 | 3 | 1 | 0,01 | 0 | |
| Отсутствие сахара | 2 | 0,03 | 9 | 3 | 0,03 | 0 | |
| Производитель | 3 | 0,05 | 15 | 5 | 0,05 | 0 | |
| 100% сок | 4 | 0,18 | 54 | 18 | 0,18 | 0 | |
| Из каких фруктов.. | 5 | 0,24 | 72 | 24 | 0,24 | 0 | |
| Марка сока | 6 | 0,19 | 57 | 19 | 0,19 | 0 | |
| Отсутствие красителей.. | 7 | 0,15 | 45 | 15 | 0,15 | 0 | |
| Рекомендовано..детям | 8 | 0,07 | 21 | 7 | 0,07 | 0 | |
| Цена | 9 | 0,06 | 18 | 6 | 0,06 | 0 | Хи крит-Хи теор |
| Страна..производства | 10 | 0,02 | 6 | 2 | 0,02 | 0 | 3,325112864 |
| | СУММА | 1 | 300 | 100 | 1 | 0 | 0 |
| | | | | Хи крит | 3,325112864 | Хи теор | 0 |

Завершающий шаг метода:

| | Индекс | Начальные вероятности | Начальные частоты | Новые частоты | Новые вероятности | Разница вероятностей | |
|-------------------------|--------|-----------------------|-------------------|---------------|-------------------|----------------------|-----------------|
| Наличие скидок, акций | 1 | 0,01 | 3 | 1 | 0,035714286 | 0,066122449 | |
| Отсутствие сахара | 2 | 0,03 | 9 | 1 | 0,035714286 | 0,001088435 | |
| Производитель | 3 | 0,05 | 15 | 2 | 0,071428571 | 0,009183673 | |
| 100% сок | 4 | 0,18 | 54 | 5 | 0,178571429 | 1,13379E-05 | |
| Из каких фруктов.. | 5 | 0,24 | 72 | 5 | 0,178571429 | 0,015722789 | |
| Марка сока | 6 | 0,19 | 57 | 5 | 0,178571429 | 0,000687433 | |
| Отсутствие красителей.. | 7 | 0,15 | 45 | 4 | 0,142857143 | 0,000340136 | |
| Рекомендовано..детям | 8 | 0,07 | 21 | 2 | 0,071428571 | 2,91545E-05 | |
| Цена | 9 | 0,06 | 18 | 2 | 0,071428571 | 0,002176871 | Хи крит-Хи теор |
| Страна..производства | 10 | 0,02 | 6 | 1 | 0,035714286 | 0,012346939 | теор |
| | СУММА | 1 | 300 | 28 | 1 | 0,107709218 | 0,309254766 |
| | | | | Хи крит | 3,325112864 | Хи теор | 3,015858098 |

Примечание: здесь и далее под разницей вероятностей понимается соотношение типа

$$\frac{(\text{Начальные вероятности} - \text{Новые вероятности})^2}{\text{Начальные вероятности}}$$

Таблица 3: применение алгоритма из пункта 2.2 к результатам исследования по вопросу: «Какие марки сока вы знаете?» («не менее 1 элемента в интервале»).

Первый шаг:

| | Индекс | Вероятности | Начальные вероятности | Начальные частоты | Новые частоты | Новые вероятности | Разница вероятностей | |
|---------------|--------|-------------|-----------------------|-------------------|---------------|-------------------|----------------------|----------|
| Добрый | 1й | 23 | 0,23 | 69 | 23 | 0,23 | 0 | |
| J-7 | 2й | 18 | 0,18 | 54 | 18 | 0,18 | 0 | |
| Я | 3й | 9 | 0,09 | 27 | 9 | 0,09 | 0 | |
| Моя семья | 4й | 6 | 0,06 | 18 | 6 | 0,06 | 0 | |
| Фруктовый сад | 5й | 7 | 0,07 | 21 | 7 | 0,07 | 0 | |
| Рич | 6й | 5 | 0,05 | 15 | 5 | 0,05 | 0 | |
| Тонус | 7й | 5 | 0,05 | 15 | 5 | 0,05 | 0 | |
| Сады придонья | 8й | 7 | 0,07 | 21 | 7 | 0,07 | 0 | |
| Любимый | 9й | 3 | 0,03 | 9 | 3 | 0,03 | 0 | |
| 100% Gold | 10й | 2 | 0,02 | 6 | 2 | 0,02 | 0 | |
| Santal | 11й | 5 | 0,05 | 15 | 5 | 0,05 | 0 | |
| Фруто-няня | 12й | 6 | 0,06 | 18 | 6 | 0,06 | 0 | |
| Тропикана | 13й | 1 | 0,01 | 3 | 1 | 0,01 | 0 | Хи крит- |
| Агуша | 14й | 3 | 0,03 | 9 | 3 | 0,03 | 0 | Хи теор |
| | СУММА | 100 | 1 | 300 | 100 | 0,44 | 0 | 5,891864 |
| | | | | | ХИ крит | 5,89186437 | Хи теор | 0 |

Завершающий шаг метода:

| | Индекс | Вероятности | Начальные вероятности | Начальные частоты | Новые частоты | Новые вероятности | Разница вероятностей | |
|---------------|--------|-------------|-----------------------|-------------------|---------------|-------------------|----------------------|----------|
| Добрый | 1й | 23 | 0,23 | 69 | 3 | 0,14285714 | 0,033016859 | |
| J-7 | 2й | 18 | 0,18 | 54 | 3 | 0,14285714 | 0,007664399 | |
| Я | 3й | 9 | 0,09 | 27 | 2 | 0,0952381 | 0,000304863 | |
| Моя семья | 4й | 6 | 0,06 | 18 | 1 | 0,04761905 | 0,0025548 | |
| Фруктовый сад | 5й | 7 | 0,07 | 21 | 2 | 0,0952381 | 0,009099449 | |
| Рич | 6й | 5 | 0,05 | 15 | 1 | 0,04761905 | 0,000113379 | |
| Тонус | 7й | 5 | 0,05 | 15 | 1 | 0,04761905 | 0,000113379 | |
| Сады придонья | 8й | 7 | 0,07 | 21 | 2 | 0,0952381 | 0,009099449 | |
| Любимый | 9й | 3 | 0,03 | 9 | 1 | 0,04761905 | 0,010347695 | |
| 100% Gold | 10й | 2 | 0,02 | 6 | 1 | 0,04761905 | 0,03814059 | |
| Santal | 11й | 5 | 0,05 | 15 | 1 | 0,04761905 | 0,000113379 | |
| Фруто-няня | 12й | 6 | 0,06 | 18 | 1 | 0,04761905 | 0,0025548 | |
| Тропикана | 13й | 1 | 0,01 | 3 | 1 | 0,04761905 | 0,141519274 | Хи крит- |
| Агуша | 14й | 3 | 0,03 | 9 | 1 | 0,04761905 | 0,010347695 | Хи теор |
| | СУММА | 100 | 1 | 300 | 21 | 1 | 0,264990008 | 0,327074 |
| | | | | | ХИ крит | 5,89186437 | Хи теор | 5,56479 |

Таблица 4: применение алгоритма из пункта 2.2 к результатам исследования по вопросу: «Какой самый важный критерий при выборе»

сока?» («не менее 5 элементов в интервале»)

Завершающий шаг:

| | Индекс | Вероятно | Частота | Новые частоты | Вероятно | Разница вероятностей | |
|--------------------|--------|----------|---------|---------------|----------|----------------------|----------|
| Отсутствие сахара | 2 | 0,030303 | 9 | 5 | 0,038462 | 0,002386588 | |
| Производитель | 3 | 0,050505 | 15 | 7 | 0,053846 | 0,000295858 | |
| 100% сок | 4 | 0,181818 | 54 | 23 | 0,176923 | 5,2597E-05 | |
| ягод сделан сок | 5 | 0,242424 | 72 | 30 | 0,230769 | 0,00035503 | |
| Марка сока | 6 | 0,191919 | 57 | 24 | 0,184615 | 0,0001526 | |
| красителей, | 7 | 0,151515 | 45 | 19 | 0,146154 | 9,86193E-05 | |
| употреблению детям | 8 | 0,070707 | 21 | 9 | 0,069231 | 8,45309E-06 | |
| Цена | 9 | 0,060606 | 18 | 8 | 0,061538 | 3,94477E-05 | Хи крит- |
| производства | 10 | 0,020202 | 6 | 5 | 0,038462 | 0,01704142 | Хи теор |
| | СУММА | 1 | 297 | 130 | 1 | 0,020430613 | 0,076657 |
| | | | | ХИ крит | 2,732637 | | 2,65598 |

Таблица 5: применение алгоритма из пункта 2.2 к результатам исследования по вопросу: «Какие марки сока вы знаете?» («не менее 5 элементов в интервале»)

Завершающий шаг

| | Индекс | Вероятно | Частота | Новые частоты | Вероятно | Разница вероятностей | |
|-------------|--------|----------|---------|---------------|----------|----------------------|----------|
| Добрый | 1й | 0,232323 | 69 | 25 | 0,217391 | 0,000691214 | |
| J-7 | 2й | 0,181818 | 54 | 18 | 0,156522 | 0,003062382 | |
| Я | 3й | 0,090909 | 27 | 10 | 0,086957 | 0,00010292 | |
| Моя семья | 4й | 0,060606 | 18 | 6 | 0,052174 | 0,001020794 | |
| Фруктовый с | 5й | 0,070707 | 21 | 8 | 0,069565 | 2,70051E-06 | |
| Рич | 6й | 0,050505 | 15 | 6 | 0,052174 | 9,4518E-05 | |
| Тонус | 7й | 0,050505 | 15 | 6 | 0,052174 | 9,4518E-05 | |
| Сады придон | 8й | 0,070707 | 21 | 8 | 0,069565 | 2,70051E-06 | |
| Любимый | 9й | 0,030303 | 9 | 5 | 0,043478 | 0,006055451 | |
| 100% Gold | 10й | 0,020202 | 6 | 5 | 0,043478 | 0,027561437 | |
| Santal | 11й | 0,050505 | 15 | 6 | 0,052174 | 9,4518E-05 | |
| Фруто-няня | 12й | 0,060606 | 18 | 7 | 0,06087 | 1,26024E-05 | Хи крит- |
| Агуша | 14й | 0,030303 | 9 | 5 | 0,043478 | 0,006055451 | Хи теор |
| | СУММА | 1 | 297 | 115 | 1 | 0,044851204 | 0,068141 |
| | | | | ХИ крит | 5,226029 | Хи теор | 5,157888 |

Приложение 3: Таблицы результатов вычислительных экспериментов для других рассмотренных задач. Пример реализации алгоритма в Excel.

Таблица 6: начальные данные по температурам океанического течения.

| Номер интервала | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | |
|-------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Интервал, °С | 16-17 | 17-18 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 | 24-25 | 25-26 | 26-27 | 27-28 | 28-29 | 29-30 | 30-31 | 31-32 | СУММА |
| Начальные вероят | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 |
| Начальные частоты | 0 | 0 | 0,01 | 0,03 | 0,03 | 0,07 | 0,08 | 0,09 | 0,1 | 0,15 | 0,17 | 0,12 | 0,08 | 0,05 | 0 | 1 |

Таблица 7: расчеты по алгоритму из пункта 2.2 типа «не менее 1 элемента в интервале»

| ЭКСПЕРИМЕНТ 1 удаление из 1го интервала | | | | | | | | | | | | | | | | СУММА | | |
|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота нс | 1 | 4 | 12,66667 | 33 | 43 | 85 | 107 | 114,3333 | 129 | 190 | 220 | 158,3333 | 101,6667 | 65 | 3,666667 | 1267,667 | 6,570631 | |
| Вероятнос | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | СУММА*N | |
| Разница в | 0 | 0 | 3,01E-34 | 4,62E-34 | 0 | 2,87E-33 | 2,28E-33 | 0 | 1,89E-33 | 0 | 4,44E-33 | 0 | 0 | 0 | 0 | 1,22E-32 | 1,55E-29 | |
| | | | | | | | | | | | | | | | | | ХИ крит-ХИ теор | 6,570631 |

| ЭКСПЕРИМЕНТ 2 удаление из 15го интервала | | | | | | | | | | | | | | | | СУММА | | |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота нс | 1 | 1,090909 | 3,454545 | 9 | 11,72727 | 23,18182 | 29,18182 | 31,18182 | 35,18182 | 51,81818 | 60 | 43,18182 | 27,72727 | 17,72727 | 1 | 346,4545 | 6,570631 | |
| Вер-ти нов | 0,002886 | 0,003149 | 0,009971 | 0,025977 | 0,033849 | 0,066912 | 0,08423 | 0,090003 | 0,101548 | 0,149567 | 0,173183 | 0,124639 | 0,080031 | 0,051168 | 0,002886 | 1 | СУММА*N | |
| Разн вер-т | 0,005577 | 1,39E-08 | 4,4E-08 | 1,15E-07 | 1,49E-07 | 2,95E-07 | 3,72E-07 | 3,97E-07 | 4,48E-07 | 6,6E-07 | 7,65E-07 | 5,5E-07 | 3,53E-07 | 2,26E-07 | 1,27E-08 | 0,005582 | 1,933796 | |
| | | | | | | | | | | | | | | | | | ХИ крит-ХИ теор | 4,636835 |

| ЭКСПЕРИМЕНТ 3 удаление из 2го интервала | | | | | | | | | | | | | | | | СУММА | | |
|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота нс | 1 | 1 | 3,166667 | 8,25 | 10,75 | 21,25 | 26,75 | 28,58333 | 32,25 | 47,5 | 55 | 39,58333 | 25,41667 | 16,25 | 1 | 317,75 | 6,570631 | |
| Вер-ти нов | 0,003147 | 0,003147 | 0,009966 | 0,025964 | 0,033832 | 0,066876 | 0,084186 | 0,089955 | 0,101495 | 0,149489 | 0,173092 | 0,124574 | 0,07999 | 0,051141 | 0,003147 | 1 | СУММА*N | |
| Разн вер-т | 0,00705 | 2,17E-08 | 6,87E-08 | 1,79E-07 | 2,33E-07 | 4,61E-07 | 5,81E-07 | 6,2E-07 | 7E-07 | 1,03E-06 | 1,19E-06 | 8,59E-07 | 5,52E-07 | 3,53E-07 | 2,24E-05 | 0,007079 | 2,24947 | |
| | | | | | | | | | | | | | | | | | ХИ крит-ХИ теор | 4,321162 |

| ЭКСПЕРИМЕНТ 4 удаление из 3го интервала | | | | | | | | | | | | | | | | СУММА | | |
|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота нс | 1 | 1 | 2,605263 | 3,394737 | 6,710526 | 8,447368 | 9,026316 | 10,18421 | 15 | 17,36842 | 12,5 | 8,026316 | 5,131579 | 1 | 102,3947 | 6,570631 | | |
| Вер-ти нов | 0,009766 | 0,009766 | 0,009766 | 0,025443 | 0,033153 | 0,065536 | 0,082498 | 0,088152 | 0,09946 | 0,146492 | 0,169622 | 0,122077 | 0,078386 | 0,050116 | 0,009766 | 1 | СУММА*N | |
| Разн вер-т | 0,102163 | 0,01385 | 5,11E-06 | 1,33E-05 | 1,74E-05 | 3,43E-05 | 4,32E-05 | 4,61E-05 | 5,21E-05 | 7,67E-05 | 8,88E-05 | 6,39E-05 | 4,1E-05 | 2,62E-05 | 0,016335 | 0,132856 | 13,60372 | |
| | | | | | | | | | | | | | | | | | ХИ крит-ХИ теор | -7,03309 |

| ЭКСПЕРИМЕНТ 4.1 сокращение 3го интервала до 2х | | | | | | | | | | | | | | | | СУММА | | |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота нс | 1 | 1 | 2 | 5,210526 | 6,789474 | 13,42105 | 16,89474 | 18,05263 | 20,36842 | 30 | 34,73684 | 25 | 16,05263 | 10,26316 | 1 | 201,7895 | 6,570631 | |
| Вер-ти нов | 0,004956 | 0,004956 | 0,009911 | 0,025822 | 0,033646 | 0,06651 | 0,083725 | 0,089463 | 0,100939 | 0,14867 | 0,172144 | 0,123891 | 0,079551 | 0,050861 | 0,004956 | 1 | СУММА*N | |
| Разн вер-т | 0,02201 | 0,001027 | 6,53E-07 | 1,7E-06 | 2,22E-06 | 4,38E-06 | 5,52E-06 | 5,9E-06 | 6,65E-06 | 9,8E-06 | 1,13E-05 | 8,17E-06 | 5,24E-06 | 3,35E-06 | 0,001472 | 0,024573 | 4,958641 | |
| | | | | | | | | | | | | | | | | | ХИ крит-ХИ теор | 1,611991 |

| ЭКСПЕРИМЕНТ 4.2 сокращение 4го интервала до 5х | | | | | | | | | | | | | | | | СУММА | | |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота нс | 1 | 1 | 2 | 5 | 6,515152 | 12,87879 | 16,21212 | 17,32323 | 19,54545 | 28,78788 | 33,33333 | 23,9899 | 15,40404 | 9,848485 | 1 | 193,8384 | 6,570631 | |
| Вер-ти нов | 0,005159 | 0,005159 | 0,010318 | 0,025795 | 0,033611 | 0,066441 | 0,083637 | 0,089369 | 0,100834 | 0,148515 | 0,171965 | 0,123762 | 0,079468 | 0,050808 | 0,005159 | 1 | СУММА*N | |
| Разн вер-т | 0,024209 | 0,001272 | 1,06E-05 | 2,16E-06 | 2,82E-06 | 5,58E-06 | 7,02E-06 | 7,5E-06 | 8,46E-06 | 1,25E-05 | 1,44E-05 | 1,04E-05 | 6,67E-06 | 4,26E-06 | 0,001776 | 0,02735 | 5,301475 | |
| | | | | | | | | | | | | | | | | | ХИ крит-ХИ теор | 1,269157 |

| ЭКСПЕРИМЕНТ 4.3 сокращение 5го интервала до 6 | | | | | | | | | | | | | | | | СУММА | | |
|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота нс | 1 | 1 | 2 | 5 | 6 | 11,86047 | 14,93023 | 15,95349 | 18 | 26,51163 | 30,69767 | 22,09302 | 14,18605 | 9,069767 | 1 | 179,3023 | 6,570631 | |
| Вер-ти нов | 0,005577 | 0,005577 | 0,011154 | 0,027886 | 0,033463 | 0,066148 | 0,083268 | 0,088975 | 0,100389 | 0,14786 | 0,171206 | 0,123217 | 0,079118 | 0,050584 | 0,005577 | 1 | СУММА*N | |
| Разн вер-т | 0,029065 | 0,001859 | 0,000135 | 0,000132 | 6,17E-06 | 1,22E-05 | 1,54E-05 | 1,64E-05 | 1,85E-05 | 2,73E-05 | 3,16E-05 | 2,27E-05 | 1,46E-05 | 9,33E-06 | 0,002492 | 0,033857 | 6,070648 | |
| | | | | | | | | | | | | | | | | | ХИ крит-ХИ теор | 0,499984 |

ЭКСПЕРИМЕНТ 4.4 сокращение 14го интервала до 9

| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит |
| частота но | 1 | 1 | 2 | 5 | 6 | 11,76923 | 14,81538 | 15,83077 | 17,86154 | 26,30769 | 30,46154 | 21,92308 | 14,07692 | 9 | 1 | 178,0462 | 6,570631 |
| Вер-ти нов | 0,005617 | 0,005617 | 0,011233 | 0,028083 | 0,033699 | 0,066102 | 0,083211 | 0,088914 | 0,10032 | 0,147758 | 0,171088 | 0,123131 | 0,079063 | 0,050549 | 0,005617 | 1 | СУММА*N |
| Разн вер-т | 0,029545 | 0,00192 | 0,000154 | 0,000162 | 1,45E-06 | 1,35E-05 | 1,7E-05 | 1,81E-05 | 2,04E-05 | 3,01E-05 | 3,49E-05 | 2,51E-05 | 1,61E-05 | 1,03E-05 | 0,002565 | 0,034532 | 6,148343 |
| | | | | | | | | | | | | | | | | | ХИ крит-Хи теор |
| | | | | | | | | | | | | | | | | | 0,422288 |

ЭКСПЕРИМЕНТ 4.5 сокращение бго интервала до 11

| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит |
| частота но | 1 | 1 | 2 | 5 | 6 | 11 | 13,84706 | 14,79608 | 16,69412 | 24,58824 | 28,47059 | 20,4902 | 13,15686 | 9 | 1 | 168,0431 | 6,570631 |
| Вер-ти нов | 0,005951 | 0,005951 | 0,011902 | 0,029754 | 0,035705 | 0,065459 | 0,082402 | 0,088049 | 0,099344 | 0,146321 | 0,169424 | 0,121934 | 0,078295 | 0,053558 | 0,005951 | 1 | СУММА*N |
| Разн вер-т | 0,033779 | 0,002477 | 0,000365 | 0,000532 | 9,39E-05 | 3,78E-05 | 4,76E-05 | 5,09E-05 | 5,74E-05 | 8,46E-05 | 9,79E-05 | 7,05E-05 | 4,53E-05 | 0,000102 | 0,003234 | 0,041074 | 6,902162 |
| | | | | | | | | | | | | | | | | | ХИ крит-Хи теор |
| | | | | | | | | | | | | | | | | | -0,33153 |

ЭКСПЕРИМЕНТ 4.6 сокращение бго интервала до 12

| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит |
| частота но | 1 | 1 | 2 | 5 | 6 | 12 | 15,10588 | 16,14118 | 18,21176 | 26,82353 | 31,05882 | 22,35294 | 14,35294 | 9 | 1 | 181,0471 | 6,570631 |
| Вер-ти нов | 0,005523 | 0,005523 | 0,011047 | 0,027617 | 0,033141 | 0,066281 | 0,083436 | 0,089155 | 0,100591 | 0,148158 | 0,171551 | 0,123465 | 0,079277 | 0,049711 | 0,005523 | 1 | СУММА*N |
| Разн вер-т | 0,028416 | 0,001777 | 0,000111 | 9,65E-05 | 1,79E-05 | 8,87E-06 | 1,12E-05 | 1,19E-05 | 1,35E-05 | 1,98E-05 | 2,3E-05 | 1,65E-05 | 1,06E-05 | 4,77E-05 | 0,002393 | 0,032975 | 5,970097 |
| | | | | | | | | | | | | | | | | | ХИ крит-Хи теор |
| | | | | | | | | | | | | | | | | | 0,600535 |

ЭКСПЕРИМЕНТ 4.7 сокращение 13го интервала до 14

| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит |
| частота но | 1 | 1 | 2 | 5 | 6 | 12 | 14,73443 | 15,74426 | 17,76393 | 26,16393 | 30,29508 | 21,80328 | 14 | 9 | 1 | 177,5049 | 6,570631 |
| Вер-ти нов | 0,005634 | 0,005634 | 0,011267 | 0,028168 | 0,033802 | 0,067604 | 0,083009 | 0,088698 | 0,100076 | 0,147398 | 0,170672 | 0,122832 | 0,078871 | 0,050703 | 0,005634 | 1 | СУММА*N |
| Разн вер-т | 0,029755 | 0,001946 | 0,000163 | 0,000175 | 4,15E-07 | 4,53E-06 | 2,32E-05 | 2,48E-05 | 2,79E-05 | 4,11E-05 | 4,76E-05 | 3,43E-05 | 2,2E-05 | 6,39E-06 | 0,002598 | 0,034869 | 6,189472 |
| | | | | | | | | | | | | | | | | | ХИ крит-Хи теор |
| | | | | | | | | | | | | | | | | | 0,381159 |

ЭКСПЕРИМЕНТ 4.8 сокращение 7го интервала до 15

| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит |
| частота но | 1 | 1 | 2 | 5 | 6 | 12 | 15 | 16,02804 | 18,08411 | 26,63551 | 30,84112 | 22,19626 | 14 | 9 | 1 | 179,785 | 6,570631 |
| Вер-ти нов | 0,005562 | 0,005562 | 0,011124 | 0,027811 | 0,033373 | 0,066746 | 0,083433 | 0,089151 | 0,100587 | 0,148152 | 0,171544 | 0,12346 | 0,077871 | 0,05006 | 0,005562 | 1 | СУММА*N |
| Разн вер-т | 0,028884 | 0,001836 | 0,000128 | 0,000122 | 8,83E-06 | 1,4E-06 | 1,12E-05 | 1,2E-05 | 1,36E-05 | 2E-05 | 2,31E-05 | 1,66E-05 | 6,76E-05 | 2,88E-05 | 0,002464 | 0,033637 | 6,047362 |
| | | | | | | | | | | | | | | | | | ХИ крит-Хи теор |
| | | | | | | | | | | | | | | | | | 0,52327 |

ЭКСПЕРИМЕНТ 4.9 сокращение 8го интервала до 16

| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит |
| частота но | 1 | 1 | 2 | 5 | 6 | 12 | 15 | 16 | 18,05248 | 26,58892 | 30,78717 | 22,15743 | 14 | 9 | 1 | 179,586 | 6,570631 |
| Вер-ти нов | 0,005568 | 0,005568 | 0,011137 | 0,027842 | 0,03341 | 0,06682 | 0,083525 | 0,089094 | 0,100523 | 0,148057 | 0,171434 | 0,123381 | 0,077957 | 0,050115 | 0,005568 | 1 | СУММА*N |
| Разн вер-т | 0,028958 | 0,001845 | 0,000131 | 0,000126 | 7,68E-06 | 8,03E-07 | 9,21E-06 | 1,34E-05 | 1,51E-05 | 2,22E-05 | 2,57E-05 | 1,85E-05 | 6,27E-05 | 2,62E-05 | 0,002476 | 0,033738 | 6,058787 |
| | | | | | | | | | | | | | | | | | ХИ крит-Хи теор |
| | | | | | | | | | | | | | | | | | 0,511844 |

| ЭКСПЕРИМЕНТ 4.10 сокращение 9го интервала до 18 | | | | | | | | | | | | | | | | | | |
|---|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|---------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота но | 1 | 1 | 2 | 5 | 6 | 12 | 15 | 16 | 18 | 26,51163 | 30,69767 | 22,09302 | 14 | 9 | 1 | 179,3023 | 6,570631 | |
| Вер-ти но | 0,005577 | 0,005577 | 0,011154 | 0,027886 | 0,033463 | 0,066926 | 0,083658 | 0,089235 | 0,100389 | 0,14786 | 0,171206 | 0,123217 | 0,07808 | 0,050195 | 0,005577 | 1 | СУММА*N | |
| Разн вер-т | 0,029065 | 0,001859 | 0,000135 | 0,000132 | 6,17E-06 | 2,38E-07 | 6,65E-06 | 1,02E-05 | 1,85E-05 | 2,73E-05 | 3,16E-05 | 2,27E-05 | 5,6E-05 | 2,28E-05 | 0,002492 | 0,033885 | 6,075659 | |
| | | | | | | | | | | | | | | | | | Хи крит-Хи теор | 0,494972 |

| ЭКСПЕРИМЕНТ 4.11 сокращение 12го интервала до 22 | | | | | | | | | | | | | | | | | | |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота но | 1 | 1 | 2 | 5 | 6 | 12 | 15 | 16 | 18 | 26,4 | 30,56842 | 22 | 14 | 9 | 1 | 178,9684 | 6,570631 | |
| Вер-ти но | 0,005588 | 0,005588 | 0,011175 | 0,027938 | 0,033525 | 0,067051 | 0,083814 | 0,089401 | 0,100576 | 0,147512 | 0,170803 | 0,122927 | 0,078226 | 0,050288 | 0,005588 | 1 | СУММА*N | |
| Разн вер-т | 0,029192 | 0,001875 | 0,00014 | 0,00014 | 4,6E-06 | 2,89E-11 | 4,17E-06 | 6,93E-06 | 1,38E-05 | 3,75E-05 | 4,34E-05 | 3,12E-05 | 4,86E-05 | 1,9E-05 | 0,002511 | 0,034066 | 6,096786 | |
| | | | | | | | | | | | | | | | | | Хи крит-Хи теор | 0,473845 |

| ЭКСПЕРИМЕНТ 4.12 сокращение 10го интервала до 26 | | | | | | | | | | | | | | | | | | |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|----------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота но | 1 | 1 | 2 | 5 | 6 | 12 | 15 | 16 | 18 | 26 | 30,10526 | 22 | 14 | 9 | 1 | 178,1053 | 6,570631 | |
| Вер-ти но | 0,005615 | 0,005615 | 0,011229 | 0,028073 | 0,033688 | 0,067376 | 0,08422 | 0,089835 | 0,101064 | 0,145981 | 0,169031 | 0,123522 | 0,078605 | 0,050532 | 0,005615 | 1 | СУММА*N | |
| Разн вер-т | 0,029522 | 0,001917 | 0,000153 | 0,00016 | 1,6E-06 | 1,56E-06 | 4,15E-07 | 1,42E-06 | 4,79E-06 | 0,000102 | 0,000118 | 1,52E-05 | 3,17E-05 | 1,08E-05 | 0,002562 | 0,0346 | 6,162509 | |
| | | | | | | | | | | | | | | | | | Хи крит-Хи теор | 0,408123 |

| ЭКСПЕРИМЕНТ 4.13 сокращение 11го интервала до 30 | | | | | | | | | | | | | | | | | | |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------------|--------|
| | 1й | 2й | 3й | 4й | 5й | 6й | 7й | 8й | 9й | 10й | 11й | 12й | 13й | 14й | 15й | СУММА | | |
| | 0,000789 | 0,003155 | 0,009992 | 0,026032 | 0,033921 | 0,067052 | 0,084407 | 0,090192 | 0,101762 | 0,149882 | 0,173547 | 0,124901 | 0,0802 | 0,051275 | 0,002892 | 1 | | |
| | 3 | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3803 | ХИ крит | |
| частота но | 1 | 1 | 2 | 5 | 6 | 12 | 15 | 16 | 18 | 26 | 30 | 22 | 14 | 9 | 1 | 178 | 6,570631 | |
| Вер-ти но | 0,005618 | 0,005618 | 0,011236 | 0,02809 | 0,033708 | 0,067416 | 0,08427 | 0,089888 | 0,101124 | 0,146067 | 0,168539 | 0,123596 | 0,078652 | 0,050562 | 0,005618 | 1 | СУММА*N | |
| Разн вер-т | 0,029563 | 0,001922 | 0,000155 | 0,000163 | 1,33E-06 | 1,97E-06 | 2,24E-07 | 1,03E-06 | 4E-06 | 9,71E-05 | 0,000145 | 1,37E-05 | 2,99E-05 | 9,93E-06 | 0,002568 | 0,034674 | 6,171932 | |
| | | | | | | | | | | | | | | | | | Хи крит-Хи теор | 0,3987 |

Таблица 8: расчеты по алгоритму из пункта 2.2 типа «не менее 5 элементов в интервале»

Первый шаг:

| Интервал, °С | 17-18 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 | 24-25 | 25-26 | 26-27 | 27-28 | 28-29 | 29-30 | 30-31 | 31-32 | СУММА | |
|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------------|
| Начальные вероят | 0 | 0,01 | 0,03 | 0,03 | 0,07 | 0,08 | 0,09 | 0,1 | 0,15 | 0,17 | 0,13 | 0,08 | 0,05 | 0 | 1 | |
| Начальные частоты | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3800 | ХИ крит |
| Новые частоты | 5,45 | 17,3 | 45 | 58,6 | 116 | 146 | 156 | 176 | 259 | 300 | 216 | 139 | 88,6 | 5 | 1727 | 5,8919 |
| Новые вероятности | 0 | 0,01 | 0,03 | 0,03 | 0,07 | 0,08 | 0,09 | 0,1 | 0,15 | 0,17 | 0,13 | 0,08 | 0,05 | 0 | 1 | хи теор |
| Разница вероятност | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | | | | | | | | | | Хи крит-Хи т |
| | | | | | | | | | | | | | | | | 5,8919 |

Формулы, используемые для вычисления первого шага в Excel. Наиболее интересны первые два столбца и последние два, так как столбцы между ними образуются с помощью копирования строк.

(Первые 3 столбца первого шага)

| AE132 | | f_x | 19-20 |
|-------|-----------------------|---------------------|---------------------|
| | AC | AD | AE |
| 32 | Интервал, °С | 17-18 | 19-20 |
| 33 | Начальные вероятности | =AD34/\$AR\$34 | =AE34/\$AR\$34 |
| 34 | Начальные частоты | 12 | 38 |
| 35 | Новые частоты | =5*AD34/\$AQ\$34 | =5*AE34/\$AQ\$34 |
| 36 | Новые вероятности | =AD35/\$AR\$35 | =AE35/\$AR\$35 |
| 37 | Разница вероятностей | =(AD33-AD36)^2/AD33 | =(AE33-AE36)^2/AE33 |

(Последние 3 столбца первого шага)

| | AQ | AR | AS |
|----|---------------------|------------------|----------------------------------|
| 32 | 31-32 | СУММА | |
| 33 | =AQ34/\$AR\$34 | =СУММ(AD33:AQ33) | |
| 34 | 11 | =СУММ(AD34:AQ34) | ХИ крит |
| 35 | =5*AQ34/\$AQ\$34 | =СУММ(AD35:AQ35) | =ХИ2ОБР(0,95;13) |
| 36 | =AQ35/\$AR\$35 | =СУММ(AD36:AQ36) | хи теор |
| 37 | =(AQ33-AQ36)^2/AQ33 | =СУММ(AD37:AQ37) | =СУММ(AD37:AQ37)*СУММ(AD35:AQ35) |
| 38 | Хи крит-Хи теор | | =AS35-AS37 |

Второй шаг

| Интервал, °С | 17-18 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 | 24-25 | 25-26 | 26-27 | 27-28 | 28-29 | 29-30 | 30-31 | 31-32 | СУММА | |
|----------------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------------------------|
| Начальные вероят | 0,0032 | 0,01 | 0,03 | 0,03 | 0,07 | 0,08 | 0,09 | 0,1 | 0,15 | 0,17 | 0,13 | 0,08 | 0,05 | 0,003 | 1 | |
| Начальные частоты | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3800 | ХИ крит |
| Новые частоты | 5 | 15,83 | 41,3 | 53,8 | 106 | 134 | 143 | 161 | 238 | 275 | 198 | 127 | 81,3 | 5 | 1584 | 5,89186 |
| Новые вероятности | 0,0032 | 0,01 | 0,03 | 0,03 | 0,07 | 0,08 | 0,09 | 0,1 | 0,15 | 0,17 | 0,12 | 0,08 | 0,05 | 0,003 | 1 | Хи теор |
| Разница вероятностей | 2E-10 | 7E-10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2E-05 | 2E-05 | 0,03776 |
| | | | | | | | | | | | | | | | | Хи крит-Хи теó 5,85411 |

(Первые 3 столбца второго шага. Формулы в Excel.)

| AF122 | | fx | | =((AF118-AF121)^2)/AF118 | |
|-------|-----------------------|---------------------|--|--------------------------|--|
| | AC | AD | | AE | |
| 42 | Начальные вероятности | =AD43/\$AR\$34 | | =AE43/\$AR\$34 | |
| 43 | Начальные частоты | 12 | | 38 | |
| 44 | Новые частоты | =5* AD35/\$AD\$35 | | =5* AE35/\$AD\$35 | |
| 45 | Новые вероятности | =AD44/\$AR\$44 | | =AE44/\$AR\$44 | |
| 46 | Разница вероятностей | =(AD42-AD45)^2/AD42 | | =(AE42-AE45)^2/AE42 | |

(Последние 3 столба первого шага. Формулы в Excel. Цветом отмечен столбец, который был зафиксирован после прохождения первого шага)

| AU53 | | fx | | |
|------|---------------------|-----------------|---------------------------------|--|
| | AQ | AR | AS | |
| 41 | 31-32 | СУММА | | |
| 42 | =AQ43/\$AR\$34 | =СУММ(AD42:AQ42 | | |
| 43 | 11 | =СУММ(AD43:AQ43 | ХИ крит | |
| 44 | 5 | =СУММ(AD44:AQ44 | =ХИ2ОБР(0,95;13) | |
| 45 | =AQ44/\$AR\$44 | =СУММ(AD45:AQ45 | Хи теор | |
| 46 | =(AQ42-AQ45)^2/AQ42 | =СУММ(AD46:AQ46 | =СУММ(AD46:AQ46)*СУММ(AD44:AQ44 | |
| 47 | Хи крит-Хи теор | | =AS44-AS46 | |

Завершающий шаг:

| Интервал, °C | 17-18 | 19-20 | 20-21 | 21-22 | 22-23 | 23-24 | 24-25 | 25-26 | 26-27 | 27-28 | 28-29 | 29-30 | 30-31 | 31-32 | СУММА |
|-----------------------|---------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|-------------------------|
| Начальные вероятности | 0,00316 | 0,01 | 0,026 | 0,034 | 0,067 | 0,084 | 0,09 | 0,102 | 0,15 | 0,174 | 0,125 | 0,08 | 0,051 | 0,002895 | 1 |
| Начальные частоты | 12 | 38 | 99 | 129 | 255 | 321 | 343 | 387 | 570 | 660 | 475 | 305 | 195 | 11 | 3800 Хи крит |
| Новые частоты | 5 | 8 | 21 | 27 | 53 | 66 | 71 | 80 | 117 | 134 | 98 | 63 | 40 | 5 | 788 5,8919 |
| Новые вероятности | 0,00635 | 0,0102 | 0,027 | 0,034 | 0,067 | 0,084 | 0,09 | 0,102 | 0,148 | 0,17 | 0,124 | 0,08 | 0,051 | 0,006345 | 1 Хи теор |
| Разница вероятностей | 0,00322 | 2E-06 | 1E-05 | 3E-06 | 4E-07 | 6E-06 | 3E-07 | 1E-06 | 2E-05 | 8E-05 | 3E-06 | 1E-06 | 6E-06 | 0,004113 | 0,007 5,877194 |
| | | | | | | | | | | | | | | | Хи крит-Хи теор 0,01467 |

(Первые 3 столбца завершающего шага. Формулы в в Excel.)

| AF122 | | fx | | =((AF118-AF121)^2)/AF118 | |
|-------|-----------------------|--------------------------|--|--------------------------|--|
| | AC | AD | | AE | |
| 132 | Интервал, °C | 17-18 | | 19-20 | |
| 133 | Начальные вероятности | =AD134/\$AR\$34 | | =AE134/\$AR\$34 | |
| 134 | Начальные частоты | 12 | | 38 | |
| 135 | Новые частоты | 5 | | 8 | |
| 136 | Новые вероятности | =AD135/\$AR\$135 | | =AE135/\$AR\$135 | |
| 137 | Разница вероятностей | =((AD133-AD136)^2)/AD133 | | =((AE133-AE136)^2)/AE133 | |

(Последние 3 столба завершающего шага. Формулы в в Excel.)

| | AQ | AR | AS |
|-----|--------------------------|--------------------|--------------------|
| 132 | 31-32 | СУММА | |
| 133 | =AQ134/\$AR\$34 | =СУММ(AD133:AQ133) | |
| 134 | 11 | =СУММ(AD134:AQ134) | Хи крит |
| 135 | 5 | =СУММ(AD135:AQ135) | =ХИ2ОБР(0,95;13) |
| 136 | =AQ135/\$AR\$135 | =СУММ(AD136:AQ136) | Хи теор |
| 137 | =((AQ133-AQ136)^2)/AQ133 | =СУММ(AD137:AQ137) | =СУММ(AD137:AQ137) |
| 138 | Хи крит-Хи теор | | =AS135-AS137 |