УДК 658.562.012.7, 519.233.33, 519.6
DOI: 10.17586/0021-3454-2024-67-9-751-758

ОПТИМАЛЬНОЕ ОБЪЕДИНЕНИЕ ИНТЕРВАЛОВ ГРУППИРОВАННОЙ ВЫБОРКИ ДЛЯ ПРИМЕНЕНИЯ КРИТЕРИЯ ТИПА χ^2

П. М. Винник[1]*, Т. В. Винник[2], Е. А. Еськова[1]

[1] *Балтийский государственный технический университет „ВОЕНМЕХ“ им. Д. Ф. Устинова, Санкт-Петербург, Россия*

** vinnik_pm@voenmeh.ru*

[2] *Санкт-Петербургский государственный технологический институт (технический университет), Санкт-Петербург, Россия*

Аннотация. Обсуждается применение интервалов равной длины или интервалов равной вероятности для использования критерия типа χ^2. При этом интервалы равной вероятности предопределяются проверяемым законом распределения. При формировании исходной выборки по данным реального производства она часто сразу является группированной с заранее заданными и неизменяемыми на производстве границами группировки и может не удовлетворять рекомендациям по применению критериев типа χ^2. Предложен способ построения набора оптимальных интервалов группировки путем объединения некоторых из имеющихся в исходной выборке интервалов. Под оптимальным набором таких интервалов понимается набор интервалов, имеющий наименьшее квадратическое отклонение взвешенных частот попадания от дискретного равномерного распределения, что позволяет не изменять набор интервалов при смене подбираемого закона распределения и автоматически решить проблему выбора оптимального числа интервалов. Перечислены некоторые свойства таких наборов, рассмотрены примеры возникающих при их построении ситуаций, приведен пример формирования оптимального набора.

Ключевые слова: *закон распределения, эмпирические данные, критерий типа* χ^2*, интервалы группировки, группированные выборки, оптимальность группировки*

Ссылка для цитирования: *Винник П. М., Винник Т. В., Еськова Е. А.* Оптимальное объединение интервалов группированной выборки для применения критерия типа χ^2 // Изв. вузов. Приборостроение. 2024. Т. 67, № 9. С. 751–758. DOI: 10.17586/0021-3454-2024-67-9-751-758.

OPTIMAL AGGREGATION OF CLUSTERED SAMPLE INTERVALS FOR APPLYING THE χ^2 TEST

P. M. Vinnik[1]*, T. V. Vinnik[2], E. A. Eskova[1]

[1] *D. F. Ustinov Baltic State Technical University VOENMEH, St. Petersburg, Russia*

** vinnik_pm@voenmeh.ru*

[2] *St. Petersburg State Institute of Technology, St. Petersburg, Russia*

Absctract. The use of intervals of equal length or intervals of equal probability for using the χ2-type criterion is discussed. In this case, intervals of equal probability are predetermined by the distribution law being tested. When forming the initial sample based on real production data, it is often immediately grouped with predetermined and unchangeable grouping boundaries in production and may not satisfy the recommendations for applying χ2-type criteria. A method is proposed for constructing a set of optimal grouping intervals by combining some of the intervals available in the initial sample. An optimal set of such intervals is understood to be a set of intervals that has the least square deviation of weighted frequencies of hits from a discrete uniform distribution, which makes it possible not to change the set of intervals when changing the selected distribution law and to automatically solve the problem of choosing the optimal number of intervals. Some properties of such sets are listed, examples of situations arising during their construction are considered, and an example of forming such an optimal set is given.

Key words: distribution law, empirical data, χ^2 test, grouping intervals, grouped samples, grouping optimality

For citation: Vinnik P. M., Vinnik T. V., Eskova E. A. Optimal aggregation of clustered sample intervals for applying the χ^2 test. *Journal of Instrument Engineering*. 2024. Vol. 67, N 9. P. 751–758 (in Russian). DOI: 10.17586/0021-3454-2024-67-9-751-758.

Введение. Одним из возможных путей повышения качества и снижения себестоимости продукции машиностроительных производств является оптимизация методов выборочного контроля характеристик полуфабрикатов и готовых изделий на этапе их изготовления.

На предприятиях широко применяется контроль по количественному признаку[1]. Для проверки гипотезы о законе распределения контролируемого количественного признака допускается [1, с. 363] проводить проверку по критериям согласия. Для группированных экспериментальных данных наиболее широко применяемыми являются критерии типа χ^2 [2, 3].

Для корректного применения критерия типа χ^2 исходные данные требуется сгруппировать по интервалам. В литературе предлагаются три основных варианта — интервалы равной длины [4, 5], интервалы равной вероятности [6, 7] и асимптотически оптимальные интервалы [8]. По выбору количества интервалов существуют различные рекомендации; см. ГОСТ Р 50.1-033-2001[2], а также [4, 5, 7, 9].

Интервалы равной длины, согласно [5], определяются выборкой и не зависят от принятия той или иной гипотезы H_0. Количество интервалов рекомендуется выбирать так, чтобы число элементов выборки во всех интервалах было взаимно сравнимо и составляло не менее 10 (иногда 5). Недостатки интервалов равной длины заключаются в неоднозначности выбора таких интервалов и том, что с точки зрения мощности критерия равные интервалы проигрывают асимптотически оптимальным [8].

Интервалы равной вероятности, согласно [6, 7], требуют до определения их количества и границ принятия полностью заданного распределения (со всеми параметрами) гипотезы H_0. Недостатки таких интервалов заключаются в зависимости от гипотезы H_0, наличии исходной негруппированной выборки [7] и в том, что они тоже проигрывают асимптотически оптимальным.

Асимптотические интервалы [8] обладают преимуществом по мощности критерия, но требуют фиксации класса распределения гипотезы H_0. Эта зависимость от гипотезы H_0 может считаться недостатком.

Отметим, что зависимость интервалов группировки от гипотезы представляется не вполне обоснованной и неудобной (при смене гипотезы необходимо строить новые интервалы).

Статистические данные, поступающие от предприятий промышленности и сельского хозяйства, во многих случаях предоставляются „как есть“. Иными словами, исследователь не имеет возможности изменить ни порядок извлечения образцов изделий, ни методику проведения их измерений, а также повлиять на точность измерения, тогда как погрешность любого измерительного инструмента предопределяет дискретность измерения любых физически непрерывных величин (например, геометрических и механических характеристик).

Поэтому указанные выборки часто следует полагать группированными с заданными границами интервалов группировки или в лучшем случае частично-группированными по терминологии [10]. Тем не менее возможно дробное перегруппирование исходных интервалов этих выборок (см. [11, ст. „Группировка“] и [12]), но в настоящей статье такие перегруппировки не рассматриваются.

Для группированных выборок с интервалами, границы которых исследователь получает „как есть“, существующие методы подготовки выборки для применения критерия типа χ^2 не являются полностью удовлетворительными.

Цель настоящей статьи — разработка такого способа перегруппировки исходной группированной выборки, который не зависит от выбора гипотезы H_0 и при котором интервалы сформированной группированной выборки наиболее близки к равновероятным интервалам.

Для достижения цели требуется решить две задачи:

[1] ГОСТ Р ИСО 3534-2-2019. Статистические методы. Словарь и условные обозначения. Ч. 2. Прикладная статистика. М.: Стандартинформ, 2019.

[2] ГОСТ Р 50.1-033-2001. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Критерии типа χ^2. М.: Стандартинформ, 2006.

1) определить, какую группированную выборку, построенную из исходной путем неоднократного объединения двух соседних интервалов в один, следует считать наиболее близкой к равновероятной;

2) разработать способ построения такой группированной выборки из исходной.

Выборки, формируемые из исходной группированной, наиболее близкие к равновероятным. Запишем исходную группированную выборку как $[k_1, \ldots, k_{j-1}, k_j, k_{j+1}, \ldots, k_M]$, где индексы обозначают номер интервала, а числа $k_1, \ldots, k_{j-1}, k_j, k_{j+1}, \ldots, k_M$ — количество элементов выборки в соответствующем интервале.

Объединяя два неких соседних (например, $(j-1)$-й и j-й) интервала в один, из исходной выборки получаем выборку $[k_1, \ldots, (k_{j-1}, k_j), k_{j+1}, \ldots, k_M]$, состоящую из $M-1$ интервалов. Проведя несколько раз такое объединение соседних интервалов в один, получим новую группированную выборку $[(k_1, \ldots, k_{i_1}), (k_{i_1+1}, \ldots, k_{i_2}), \ldots (k_{i_{L-1}+1}, \ldots, k_M)]$, назовем ее разбиением исходной выборки, а число L — количество интервалов сформированной выборки — рангом разбиения. Заметим, что разбиение, имеющее ранг L, однозначно определяется расстановкой $(L-1)$ „стенок", отделяющих части друг от друга.

Определение 1. Пусть $[(k_1, \ldots, k_{i_1}), (k_{i_1+1}, \ldots, k_{i_2}), \ldots (k_{i_{L-1}+1}, \ldots, k_M)]$ — некоторое разбиение выборки, имеющее ранг L. Вычислим суммы элементов частей разбиения: $\Sigma_1 = k_1 + \ldots + k_{i_1}, \ldots, \Sigma_L = k_{i_{L-1}+1} + \ldots + k_M$. Величину N/L, где N — объем исходной группированной выборки, назовем оптимальным содержанием интервала (если N/L — натуральное число, а выборка получилась равновероятной, то в каждом из ее L интервалов будет по N/L элементов). Квадратичным отклонением разбиения назовем величину

$$\Delta L = (\Sigma_1 - N/L)^2 + \ldots + (\Sigma_L - N/L)^2.$$

Определение 2. Назовем разбиение, имеющее ранг L, наиболее близким к равновероятному, если для него величина ΔL минимальна среди всех разбиений исходной выборки, имеющих ранг L.

Так как для построения разбиения, имеющего ранг L, из исходной выборки необходимо установить $(L-1)$ „стенку" на неких из $(M-1)$ местах, то количество таких разбиений конечно и равно C_{M-1}^{L-1}. Поэтому минимум ΔL обязательно достигается.

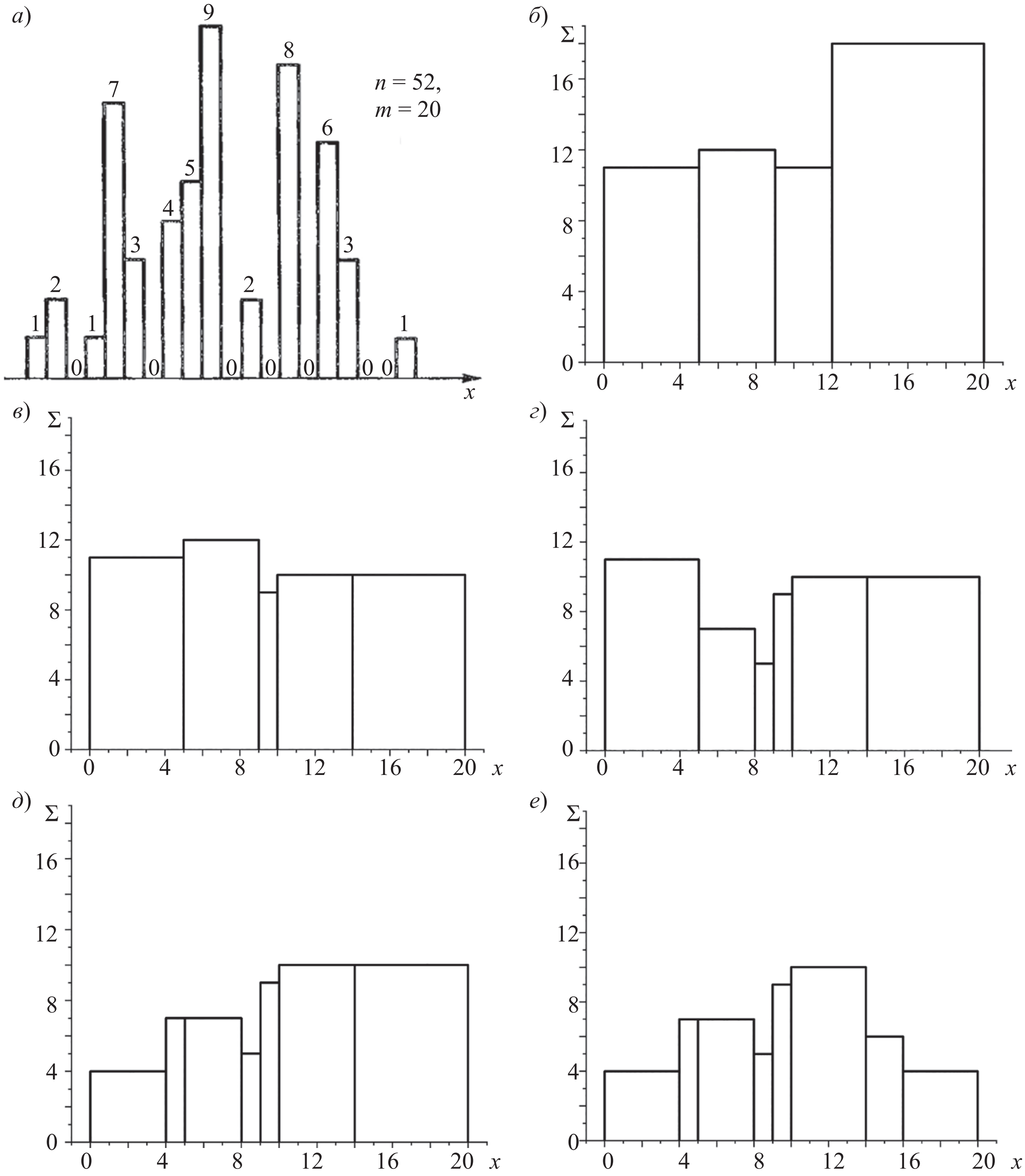
Пусть $M_{\min}$ — число интервалов разбиения, принятое в качестве минимально допустимого. В [4] $M_{\min} = 6$, в [9] $M_{\min} = 4$.

Определение 3. Разбиение исходной выборки, имеющей ранг $M_{\min} \le L < M$, назовем наиболее близким к равновероятному, если для него усредненный ранг Δ_L/L разбиения минимален. Разбиение, наиболее близкое к равновероятному, его ранг и его части, коротко будем называть оптимальными.

Приведем пример построения оптимального разбиения. Рассмотрим исходную выборку, изображенную на рисунке, *а*, — [1, 2, 0, 1, 7, 3, 0, 4, 5, 9, 0, 2, 0, 8, 0, 6, 3, 0, 0, 1] (это выборка из работы [4, рис. 5.2, *а*)]). Всего в выборке $n = 52$ наблюдения, распределяющихся на $m = 20$ интервалов. Построим ее оптимальные разбиения, имеющие ранги 4–8. Так, при 52 наблюдениях оптимальным содержанием интервала для ранга 4 (см. рисунок, *б*) является число 13. Непосредственным перебором находим оптимальное разбиение: [(1, 2, 0, 1, 7), (3, 0, 4, 5), (9, 0, 2), (0, 8, 0, 6, 3, 0, 0, 1)]. При этом объединенные интервалы содержат [11, 12, 11, 18] наблюдений. Оценим качество разбиения, вычислив квадратическое отклонение от оптимального содержания интервала:

$$\Delta_4 = (11-13)^2 + (12-13)^2 + (11-13)^2 + (18-13)^2 = 4 + 1 + 4 + 25 = 35.$$

Разбивая исходную выборку на 4, 5, 6, 7, 8 оптимальных интервалов, получаем результаты, приведенные в таблице и на рисунке, *б*–*е*. Таким образом, оптимальным является разбиение, имеющее ранг 5.



Число интервалов, L	Оптимальное разбиение	Оптимальное содержание интервала	Δ_L	Δ_L/L
4	[(1, 2, 0, 1, 7), (3, 0, 4, 5), (9, 0, 2), (0, 8, 0, 6, 3, 0, 0, 1)]	13	35	8,75
5	[(1, 2, 0, 1, 7), (3, 0, 4, 5), (9), (0, 2, 0, 8), (0, 6, 3, 0, 0, 1)]	10,4	5,2	1,04
6	[(1, 2, 0, 1, 7), (3, 0, 4), (5), (9), (0, 2, 0, 8), (0, 6, 3, 0, 0, 1)]	26/3	25,33	4,22
7	[(1, 2, 0, 1), (7), (3, 0, 4), (5), (9), (0, 2, 0, 8), (0, 6, 3, 0, 0, 1)]	52/7	33,71	4,82
8	[(1, 2, 0, 1), (7), (3, 0, 4), (5), (9), (0, 2, 0, 8), (0, 6), (3, 0, 0, 1)]	6,5	34	4,25

Способ формирования оптимальной выборки. Простейший (но вычислительно затратный) способ построения оптимальной выборки — перебор всех вариантов расстановки „стенок“ для разбиения, имеющего фиксированный ранг; затем выбор оптимального распределения из всех рангов. С ростом числа ненулевых интервалов исходной выборки перебор быстро становится неприемлемым.

Для разработки конструктивного способа построения оптимальной выборки необходимо определить максимально возможное количество свойств оптимальных выборок. Два таких свойства доказываются в утверждениях 1 и 2.

Утверждение 1. Пусть среди чисел $k_1, \ldots, k_M$ нет равных нулю. Пусть требуется разделить выборку $k_1, \ldots, k_M$ на две части в соотношении $p : q$, где $p > 0$, $q > 0$ и $p + q = 1$. Обозначим через j номер интервала выборки, такой что $k_1 + \ldots + k_j < pN$, а $k_1 + \ldots + k_{j+1} \geq pN$.

Тогда наименьшее отклонение разбиения от соотношения $p : q$ будет иметь место либо при разбиении $(k_1, \ldots, k_j)$ и $(k_{j+1}, \ldots, k_M)$, либо при разбиении $(k_1, \ldots, k_{j+1})$ и $(k_{j+2}, \ldots, k_M)$.

Доказательство. Заметим, что для первого разбиения выполняются неравенства $k_1 + \ldots + k_j < pN$ и $k_{j+1} + \ldots + k_M > qN$, поэтому отклонение этого разбиения от соотношения $p : q$ будет следующим:

$$\Delta = (pN - (k_1 + \ldots + k_j))^2 + (k_{j+1} + \ldots + k_M - qN)^2, \tag{1}$$

где содержимое выражений в обеих скобках положительно.

Рассмотрим произвольное разбиение $(k_1, \ldots, k_m)$ и $(k_{m+1}, \ldots, k_M)$, где $m < j$. Для него сумма $(k_1 + \ldots + k_j)$ в левых скобках выражения (1) заменится на $(k_1 + \ldots + k_m)$, т. е. будет состоять из меньшего числа слагаемых, и, в силу их положительности, уменьшится, следовательно, само выражение в левых скобках (1) увеличится. При этом в правых скобках выражения (1) сумма $(k_{j+1} + \ldots + k_M)$ заменится на $(k_{m+1} + \ldots + k_j + \ldots + k_M)$, т. е. будет состоять из большего числа слагаемых, а поэтому увеличится. Следовательно, само выражение в правых скобках (1) увеличится.

Таким образом, отклонение, аналогичное формуле (1), для выборок такого типа будет заведомо больше, чем для выборки с $m = j$.

Заметим, что для второго разбиения будут выполняться неравенства $k_1 + \ldots + k_{j+1} \geq pN$ и $k_{j+2} + \ldots + k_M \leq qN$, а поэтому отклонение этого разбиения от соотношения $p : q$ будет следующим:

$$\Delta = ((k_1 + \ldots + k_{j+1}) - pN)^2 + (qN - (k_{j+2} + \ldots + k_M))^2, \tag{2}$$

где содержимое выражений в обеих скобках неотрицательно.

Рассмотрим теперь произвольное разбиение $(k_1, \ldots, k_M)$ и $(k_{m+1}, \ldots, k_M)$, где $m > j + 1$. Для него сумма $(k_1 + \ldots + k_{j+1})$ в левых скобках (2) заменится на $(k_1 + \ldots + k_{j+1} + \ldots + k_m)$, т. е. будет состоять из большего числа слагаемых, и, в силу их положительности, увеличится, следовательно, само выражение в левых скобках (2) увеличится. При этом в правых скобках (2) сумма $(k_{j+2} + \ldots + k_M)$ заменится на $(k_{m+1} + \ldots + k_M)$, т. е. будет состоять из меньшего числа слагаемых, а поэтому уменьшится. Следовательно, само выражение в правых скобках (2) увеличится.

Таким образом, отклонение, аналогичное формуле (2), для выборок такого типа будет заведомо больше, чем для выборки с $m = j + 1$. ■

Замечание 1. На примере выборок [1,3,2] и [2,3,1], которые оптимально разбиваются на две части как [(1, 3), (2)] и [(2), (3, 1)], видно, что оба варианта разбиения, указанные в утверждении 1, могут обеспечивать оптимальное разбиение на две части.

Замечание 2. На примере выборки [1, 1, 1, 1, 8] видно, что второй из вариантов разбиения, указанных в утверждении 1, является вырожденным [(1, 1, 1, 1, 8)()]. Для выборки [8, 1, 1, 1, 1] вырожденным будет первый вариант.

Утверждение 2. Пусть для выборки $k_1, \ldots, k_M$ разбиение ее на L наиболее равновероятных частей $[(k_1, \ldots, k_{m_1}), (k_{m_1+1}, \ldots, k_{m_2}), \ldots, (k_{m_{L-1}+1}, \ldots, k_{m_L})]$ является оптимальным.

Тогда для выборки $[k_{m_i+1}, \ldots, k_{m_{i+1}}, \ldots, k_{m_{i+1}+1}, \ldots, k_{m_{i+2}}]$ образованной двумя соседними частями, ее разбиение $[(k_{m_i+1}, \ldots, k_{m_{i+1}}), (k_{m_{i+1}+1}, \ldots, k_{m_{i+2}})]$ на две равновероятные части является оптимальным.

Доказательство. Пусть $W = N/L$ — „среднее содержание“ каждой из L частей исходной выборки. Наличие в оптимальном разбиении соседних частей $(k_{m_i+1}, \ldots, k_{m_{i+1}})$, $(k_{m_{i+1}+1}, \ldots, k_{m_{i+2}})$ означает, что сумма квадратов

$$(k_{m_{i+1}+1}, \ldots, k_{m_{i+1}} - W)^2 + (k_{m_{i+1}+1}, \ldots, k_{m_{i+2}} - W)^2 \tag{3}$$

является минимальной из возможных. В другом случае можно было бы, не изменяя остальные части разбиения исходной выборки, „подвинуть стенку“ между данными частями влево или вправо и уменьшить вклад этих двух частей в квадратическое отклонение, а следовательно, и общее квадратическое отклонение разбиения исходной выборки, что противоречило бы его оптимальности.

Временно введем обозначения:

$$x = k_{m_i+1} + \ldots + k_{m_{i+1}},\ Q = k_{m_i+1} + \ldots + k_{m_{i+1}} + k_{m_{i+1}+1} + \ldots + k_{m_{i+2}}.$$

Тогда $k_{m_{i+1}+1} + \ldots + k_{m_{i+2}} = Q - x$. С учетом введенных обозначений сумму (3) можно переписать в виде $(x - W)^2 + (Q - x - W)^2 = F$. Для поиска минимального значения F на отрезке $x \in [0; Q]$ найдем производную и приравняем ее к нулю: $2(x - W) + 2(Q - x - W)(-1) = 0$ откуда точка минимума $x = Q/2$. Заметим, что точка минимума не зависит от „среднего содержания“ W каждой части. Так как значения переменной x изменяются дискретно, то искомое наименьшее значение будет достигаться при ближайшем к $x = Q/2$ значении переменной x из ряда 0, k_{m_i+1}, $k_{m_i+1} + k_{m_i+2}$, … $k_{m_i+1} + \ldots + k_{m_{i+1}} + k_{m_{i+1}+1} + \ldots k_{m_{i+2}} = Q$.

При переходе от исходной выборки к выборке $[k_{m_i+1}, \ldots, k_{m_{i+1}}, k_{m_{i+1}+1}, \ldots, k_{m_{i+2}}]$ „среднее содержание“ части будет равным $W_0 = Q/2$. Однако, как установлено выше, точка минимума не зависит от „среднего содержания“ части, поэтому минимальное значение будет достигнуто в той же точке, т. е. разбиение $[(k_{m_i+1}, \ldots, k_{m_{i+1}}), (k_{m_{i+1}+1}, \ldots, k_{m_{i+2}})]$ останется оптимальным. ■

Следует отметить, что при переходе от оптимального разбиения к разбиению, имеющему больший ранг, может происходить дробление сразу нескольких интервалов. Например, при переходе от разбиения выборки [2, 0, 1, 2, 4, 1, 2, 5, 6, 2, 3, 4, 2, 1, 8, 0, 2, 1, 1, 1, 0, 2] на 5 оптимальных интервалов [(2, 0, 1, 2, 4, 1, 2), (5, 6), (2, 3, 4), (2, 1, 8), (0, 2, 1, 1, 1, 0, 2)] к разбиению на 6 оптимальных интервалов [(2, 0, 1, 2, 4), (1, 2, 5), (6, 2), (3, 4, 2), (1, 8), (0, 2, 1, 1, 1, 0, 2)] дробятся все объединенные интервалы, кроме последнего.

Для разбиения исходной выборки на части, вероятности которых наиболее приближены к числам $p_1, p_2, \ldots p_L$, где $p_1 + p_2 + \ldots p_L = 1$, можно привести пример, что при переходе к подвыборке „стенка“ может быть смещена. Действительно, деля выборку [27, 6, 30, 30, 147] на три части в соотношении $p_1 = 1/6$, $p_2 = 1/3$, $p_3 = 1/2$, находим оптимальное разбиение [(27), (6, 30, 30), (147)], а при переходе к подвыборке [27, 6, 30, 30] при ее делении на две части, вероятности $\tilde{p}_1 = 1/3$, $\tilde{p}_2 = 2/3$ которых пропорциональны исходным вероятностям $p_1 = 1/6$, $p_2 = 1/3$, получаем оптимальное разбиение [(27, 6), (30, 30)].

На основании утверждения 1 представляется очевидным следующее решение задачи: накопление суммы элементов группированной выборки; для любого $j = 1, \ldots, L - 1$ установление двух предполагаемых „стенок“ в точках $i(j)$ и $i(j) + 1$, таких что сумма $k_1 + \ldots + k_{i(j)} < jN/L$, а $k_1 + \ldots + k_{i(j)+1} < jN/L$; затем проверка всех вариантов расстановки „стенок“ с выбором по одной „стенке“ из каждой пары. Однако на примере выборки [1, 1, 1, 1, 8] видно, что оптимальное разбиение ее на три части будет таким: [(1, 1), (1, 1), (8)], тогда как предполагаемые „стенки“ [1, 1, 1, 1, ||8||] будут расположены по две с обеих сторон от 8. Перебор вариантов расстановки таких „стенок“ не позволит получить оптимальное разбиение.

Таким образом, разработка способа построения оптимальной выборки без перебора всех вариантов далека от завершения.

Математические аналоги задачи построения оптимальной выборки. В теории чисел существует направление задач, связанных с разбиением натурального числа в сумму натуральных чисел (см., например, [13]). Задача разбиения выборки на оптимальные интервалы дает некоторое разбиение натурального числа M, равного объему исходной выборки, в сумму натуральных чисел $\Sigma_1, \Sigma_2, \ldots, \Sigma_L$ — объемов объединенных интервалов:

$$\Sigma_1 + \Sigma_2 + \ldots + \Sigma_L = M.$$

В дискретной математике известна задача разбиения набора натуральных чисел (в том числе повторяющихся) на несколько подмножеств, имеющих как можно более близкие суммы [14]. В таких задачах качество разбиения определяется максимумом разности сумм подмножеств. Отличие задачи оптимального разбиения выборки от такой задачи заключается в том, что при построении оптимального разбиения нельзя переставлять части (интервалы).

В [15] доказывается предпочтительность метода равных частот относительно метода равных интервалов при использовании функции максимального правдоподобия, что тесно связано с рассматриваемой в настоящей статье задачей.

Заключение. Определено, какую группированную выборку, построенную из исходной группированной выборки путем неоднократного объединения двух соседних интервалов в один, следует считать наиболее близкой к равновероятной. Показано, что наиболее близкую к равновероятной выборку можно построить посредством перебора конечного числа вариантов, однако способ построения такой выборки без полного перебора не найден. Установлен ряд свойств выборок, что должно способствовать разработке такого способа в дальнейшем.

Использование полученных результатов позволяет из исходной группированной выборки, не удовлетворяющей рекомендациям по применению критерия типа χ^2, сформировать удовлетворяющую рекомендациям группированную выборку, границы интервалов которой не зависят от принятия той или иной гипотезы H_0.

СПИСОК ЛИТЕРАТУРЫ

1. *Миттаг Х.-Й., Ринне Х.* Статистические методы обеспечения качества. М.: Машиностроение, 1995. 616 с.
2. *Кобзарь А. И.* Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.
3. *Лемешко Б. Ю., Постовалов С. Н.* Статистический анализ одномерных наблюдений по частично группированным данным // Изв. вузов. Физика. 1995. № 9. С. 39–45.
4. *Новицкий П. В., Зограф И. А.* Оценка погрешностей результатов измерений Л.: Энергоатомиздат, 1991. 304 с.
5. *Хальд А.* Математическая статистика с техническими приложениями. М.: Изд-во иностр. лит., 1956. 664 с.
6. *Mann H. B., Wald A.* On the choice of the number of class intervals in the application of the chi square test // Ann. Math. Stat. 1942. Vol. 13. P. 306–317.
7. *Williams C. A. Jr.* On the Choice of the Number and Width of Classes for the Chi-Square Test of Goodness of Fit // Journal of the American Statistical Association. 1950. Vol. 45, N 249. P. 77–86.
8. *Лемешко Б. Ю., Постовалов С. Н.* О зависимости предельных распределений статистик хи-квадрат Пирсона и отношения правдоподобия от способа группирования данных // Заводская лаборатория. 1998. Т. 64, № 5. С. 56–63.
9. *Лемешко Б. Ю., Чимитова Е. В.* О выборе числа интервалов в критериях согласия типа χ^2 // Заводская лаборатория. Диагностика материалов. 2003. Т. 69. С. 61–67.
10. *Куллдорф Г.* Введение в теорию оценивания по группированным и частично группированным выборкам. М.: Наука, 1966. 176 с.
11. Энциклопедия статистических терминов. Т. 1. Методологические основы статистики. М.: Федеральная служба государственной статистики, 2011 [Электронный ресурс]: https://03.rosstat.gov.ru/storage/mediabank/05_tom1(1).pdf, 18.03.2024.
12. *Винник П. М., Винник Т. В., Еськова Е. А.* Математические задачи, возникающие при статистическом контроле технологических процессов машиностроительных производств // Вестник образования и развития науки Российской академии естественных наук. 2022. № 4. С. 24–30. DOI: 10.26163/RAEN.2022.32.23.003.

13. *Эндрюс Г.* Теория разбиений. М.: Наука, 1982. 256 с.
14. *Korf R*. Multi-way number partitioning // IJCAI'09: Proc. of the 21st Intern. Joint Conf. on Artificial Intelligence, Pasadena, California, USA, July 11–17, 2009. P. 538–543.
15. *Бардасов С. А*. Предпочтительность метода равных частот относительно метода равных интервалов при построении вариационных рядов // Вестн. Тюменск. гос. ун-та. 2003. № 5. С. 217–219.

СВЕДЕНИЯ ОБ АВТОРАХ

Петр Михайлович Винник — д-р техн. наук, доцент; Балтийский государственный технический университет „ВОЕНМЕХ“ им. Д. Ф. Устинова, кафедра высшей математики; заведующий кафедрой; E-mail: vinnik_pm@voenmeh.ru

Татьяна Викторовна Винник — канд. техн. наук; Санкт-Петербургский государственный технологический институт (технический университет), кафедра математики; доцент; E-mail: vinnik.tv92@gmail.com

Екатерина Александровна Еськова — Балтийский государственный технический университет „ВОЕНМЕХ“ им. Д. Ф. Устинова, кафедра высшей математики; ассистент; E-mail: eskova_ea@voenmeh.ru

Поступила в редакцию 29.03.2024; одобрена после рецензирования 12.05.2024; принята к публикации 23.07.2024.

REFERENCES

1. Mittag H.-J., Rinne H. *Statistische Methoden der Qualitatssicherung*, Munchen, Wien, 1993.
2. Kobzar' A.I. *Prikladnaya matematicheskaya statistika. Dlya inzhenerov i nauchnykh rabotnikov* (Applied Mathematical Statistics. For Engineers and Scientists), Moscow, 2006, 816 p. (in Russ.)
3. Lemeshko B.Yu., Postovalov S.N. *Izvestiya vuzov*. Fizika, 1995, no. 9, pp. 39–45. (in Russ.)
4. Novitskiy P.V., Zograf I.A. *Otsenka pogreshnostey rezul'tatov izmereniy* (Evaluation of Errors in Measurement Results), Leningrad, 1991, 304 p. (in Russ.)
5. Hald A. *Statistical Theory with Engineering Applications*, NY, Wiley, 1952, 783 p.
6. Mann H.B., Wald A. *Ann. Math. Stat.*, 1942, vol. 13, pp. 306–317.
7. Williams C.A., jr., *Journal of the American Statistical Association*, 1950, no. 249(45), pp. 77–86.
8. Lemeshko B.Yu., Postovalov S.N. *Industrial Laboratory. Diagnostics of Materials*, 1998, no. 5(64), pp. 56–63. (in Russ.)
9. Lemeshko B.Yu., Chimitova E.V. *Industrial Laboratory. Diagnostics of Materials*, 2003, vol. 69, pp. 61–67. (in Russ.)
10. Kulldorff G. *Contributions to the theory of estimation from grouped and partially grouped samples*, NY, John Wiley, 1963, 144 p.
11. https://03.rosstat.gov.ru/storage/mediabank/05_tom1(1).pdf.
12. Vinnik P.M., Vinnik T.V., Eskova E.A. *Bulletin of Education and Development of Science of the Russian Academy of Natural Sciences*, 2022, no. 4, pp. 24–30, DOI: 10.26163/RAEN.2022.32.23.003. (in Russ.)
13. Andrews G.E. *The theory of partitions*, Addison-Wesley Pub. Co., 1976, 255 p.
14. Korf R. IJCAI'09, *Proceedings of the 21st International Joint Conference on Artificial Intelligence Pasadena*, California, USA, July 11-17, 2009, pp. 538–543.
15. Bardasov S.A. *Bulletin of Tyumen State University*, 2003, no. 5, pp. 217–219. (in Russ.)

DATA ON AUTHORS

Petr M. Vinnik — Dr. Sci., Associate Professor; D. F. Ustinov Baltic State Technical University VOENMEH, Department of Higher Mathematics; Head of the Department; E-mail: vinnik_pm@voenmeh.ru

Tatyana V. Vinnik — PhD; St. Petersburg State Institute of Technology, Department of Mathematics; Associate Professor; E-mail: vinnik.tv92@gmail.com

Ekaterina A. Eskova — D. F. Ustinov Baltic State Technical University VOENMEH, Department of Higher Mathematics; Assistant; E-mail: eskova_ea@voenmeh.ru

Received 29.03.2024; approved after reviewing 12.05.2024; accepted for publication 23.07.2024.