
СТАТИСТИКА И ЭКОНОМИЧЕСКОЕ ИЗМЕРЕНИЕ

УДК 311.12

СТАТИСТИЧЕСКАЯ ГРУППИРОВКА И РАСПОЗНАВАНИЕ НЕКОТОРЫХ ВИДОВ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

В.Г. Ионин

Новосибирский государственный университет
экономики и управления «НИНХ»
E-mail: stat@nsuem.ru

Задача статистической группировки результатов первичного наблюдения, рассматриваемая в данной статье, предполагает, что исследователь имеет некоторое представление о количестве типологических или структурных позиций, слоев. Заключение о количестве типов, образов или структурных слоев предлагается получать на основании того, в какой мере общая вариация уровней признака объясняется так называемой вариацией межгрупповой (межклассовой). Совокупность структурных позиций, обеспечивающих максимальную межгрупповую дисперсию, является как бы «структурным образом» изучаемого множества наблюдений. Степень точности такого структурного отображения задается количеством структурных позиций. Понятие структурного образа изучаемой совокупности наблюдений в данной статье рассматривается также применительно к совокупностям, подчиняющимся известным строгим законам распределения вероятностей.

В статье рассматривается оптимальное разделение статистической совокупности наблюдений на заданное число классов, групп в зависимости от вида функций вероятностей распределения. Определяются устойчивые структуры оптимальной группировки для равномерного, нормального, показательного и логарифмически нормального законов распределения. Показана возможность применения полученных стабильных структур (структурных образов) в качестве потенциальных для определения вида функции распределения вероятностей для статистических совокупностей с неизвестным законом распределения.

Ключевые слова: оптимальная группировка, межгрупповая дисперсия, потенциальная структура, структурный образ, функции распределения вероятностей.

THE STATISTICAL GROUPING AND DETECTION OF SOME TYPES OF PROBABILITY DISTRIBUTION

V.G. Ionin

Novosibirsk State University of Economics and Management
E-mail: stat@nsuem.ru

The statistical grouping the results of primary observations, discussed in this article, suggests that the researcher has some idea about the number of typological or structural positions, layers. Conclusion on the number of types of images or structural layers to be

based on the extent to which the total variance explained by the so-called tag levels variation of the Intergroup (interclass). A structural position, providing maximum cross-group variance is a «structural manner» of many observations. The degree of accuracy of the structural display is set to the number of structural positions. The concept of structural image of the investigated range of observations of this article is also applicable to pools, under the authority of the known strict laws of probability distributions.

Describes the optimal division of statistical set of observations by the given number of classes, groups, depending on the type of probability distribution functions. Defines the sustainable optimal grouping for uniform, normal, exponential and log-normal distribution laws. The possibility of application of stable structures (structural images) as a potential function to determine the type of probability distributions to statistical aggregates with unknown law of distribution.

Key words: optimal grouping, cross-group dispersion, potential structure, structural images, function of the probability distribution.

Статистическая группировка наблюдений, выполняемая как начальный этап статистического исследования, предполагает распознавание реально существующих типов (если группировка *типологическая*), или структурных слоев (если выполняется группировка *структурная*), а также оценку степени проявления каждой структурной или типологической позиции. И в том, и в другом случае исследователь должен иметь некоторое сложившееся представление о количестве типологических или структурных составляющих. Очевидно, что однозначного ответа на вопрос о количестве типов или групп, в принципе, не существует. Ответ определяется положением исследователя относительно изучаемой совокупности явлений. На первом, низшем, уровне исследования каждый отдельный объект есть явление самостоятельное, представляющее как бы самостоятельный тип или самостоятельную структурную позицию. И лишь по мере «удаления» от отдельных объектов наблюдения, по мере повышения статуса исследователя начинают проявляться отдельные сгущения точек-объектов, обладающих общими или подобными характеристиками.

Следует сказать, что иногда позиционирование исследователя относительно изучаемой совокупности как бы предусматривается статистической методикой. Можно сослаться, например, на работу [1], где рассматриваются проблемы типологии, или работу [2], где рассматривается конкретное применение аппарата статистического анализа канонических корреляций для построения многомерной статистической оценки инвестиционной привлекательности фирм. В последнем случае методика многомерного канонического корреляционного анализа предполагает ограниченное количество уровней (порядков) таких канонических корреляций. Аналогичным образом формируются уровни ограниченного числа главных компонент и факторов в компонентном и факторном анализе. Но в общем случае при анализе первичного статистического материала исследователь должен обосновать свое суждение о количестве реально существующих типов или структурных позиций, групп.

Заключение о количестве типов, образов или структурных слоев, реально образующих изучаемое множество наблюдений, следует получать на основании того, в какой мере общая вариация уровней признака объясняется (определяется) так называемой вариацией межгрупповой (межклассовой).

Межгрупповая вариация, как известно, образуется за счет различия среднего уровня признака в группах (типах, классах, структурных слоях) и общего среднего значения для совокупности наблюдений в целом. Существенным является то, чтобы выделяемые в процессе анализа слои или группы были однородны по уровню наблюдаемого показателя. Таким образом, для каждого варианта распределения совокупности на « k » однородных групп ($k = 2, 3, \dots$) должно обеспечиваться максимальное значение межгрупповой дисперсии признака. Задача максимизации межгрупповой дисперсии имеет строгое математическое решение как одна из задач динамического программирования. В частности, именно так выполнялось исследование по распределению населения области на группы по величине денежного дохода (см., напр., [3–6]). Программная реализация алгоритма максимизации межгрупповой дисперсии рассматривается в [7].

Применение точного алгоритма для оптимизации состава структурных слоев, групп имеет, конечно, свои преимущества. Но, учитывая статистический характер первичного числового материала, достаточно, по-видимому, использовать и менее строгие подходы. Во всяком случае, можно предположить, что для наиболее значимых видов вероятностных распределений образуются свои собственные достаточно устойчивые соотношения по количественному наполнению выделяющихся типов, групп, слоев, если процедура группировки базируется на оптимизации межгрупповой дисперсии. Эти соотношения мы определим для выборок, полученных методом статистического моделирования соответствующих вероятностных законов.

Рассмотрим результаты оптимальной группировки случайных статистических совокупностей на k групп ($k = 2, 3, 5$ и 7) при различных функциях распределения вероятностей. Моделировались статистические совокупности, распределенные по равномерному, нормальному, показательному законам и логарифмически нормальному распределению вероятностей.

В процессе оптимального деления на группы равномерно распределенной совокупности получены результаты, представленные в табл. 1.

Таким образом, если массив первичных статистических данных подчиняется закону равномерного распределения вероятностей, то практически оптимальное распределение на заданное число классов, слоев, групп наблюдается, когда статистическая совокупность, упорядоченная по возрастанию признака, разделяется на равно наполненные по количеству наблюдений группы. Некоторое предпочтение по количеству объектов отмечается для групп, находящихся в средней части совокупности. Отметим также, что оптимальное распределение обеспечивает достаточно весомый вклад межгрупповой дисперсии в общую вариацию признака. Так, при делении на 2 группы доля межгрупповой дисперсии в общей вариации формируется на уровне $(70 \div 75)\%$. Для $k = 3$ эта доля составляет $(80 \div 90)\%$, для $k=5 - (95 \div 96)\%$. При распределении на $k = 7$ групп эта доля варьирует около 98% . В качестве иллюстрации: массив 1000 случайных чисел, распределенных по равномерному закону, при делении на 7 равно наполненных групп (по 143 наблюдения) обеспечил долю межгрупповой дисперсии на уровне $97,7\%$. Использование процедуры оптимального распределения на 7 групп того же массива чисел обеспечило удельный вес межгрупповой дисперсии на уровне $98,0\%$.

Таблица 1

Структура равномерно распределенной совокупности при различных вариантах оптимального деления на группы (на 10 опытах), %

Номер группы в порядке возрастания признака	Номер опыта										Средняя структура
	1	2	3	4	5	6	7	8	9	10	
<i>Оптимальное деление на 2 группы</i>											
Группа 1	48,7	49,7	45,4	50,7	52,2	49,3	47,6	47,2	47,1	49,7	48,8
Группа 2	51,3	50,3	54,6	49,3	47,8	50,7	52,4	52,8	52,9	50,3	51,2
<i>Оптимальное деление на 3 группы</i>											
Группа 1	33,9	33,9	29,5	34,4	29,9	34,9	32,3	32,3	29,5	34,8	32,5
Группа 2	34,0	32,2	36,7	31,8	35,6	33,3	34,6	34,6	45,4	32,6	35,1
Группа 3	32,1	33,9	33,8	33,8	34,5	31,8	33,1	33,1	25,1	32,6	32,4
<i>Оптимальное деление на 5 групп</i>											
Группа 1	20,5	22,2	20,8	20	21	20,6	21	8,9	19,4	19,5	19,4
Группа 2	17,6	22,9	18,8	18,4	21,3	20,3	18,7	24	17,7	19,7	19,9
Группа 3	18,8	20,3	22,6	21	18,8	20,4	20,3	27,9	19,7	19,5	20,9
Группа 4	19,5	18,6	19,1	18,4	18,4	22	21,2	25,3	20,9	21,2	20,5
Группа 5	23,6	16	18,7	22,2	20,5	16,7	18,8	13,9	22,3	20,1	19,3
<i>Оптимальное деление на 7 групп</i>											
Группа 1	14,2	15,2	13,5	18,5	13,9	15,6	14,6	6,5	14,2	13	13,9
Группа 2	13,6	14,6	14,7	15,3	12,6	13,7	15,2	16,3	15,7	14,5	14,6
Группа 3	12,3	15,6	15,3	14,3	13,6	17,6	11,5	18,3	15,3	13,4	14,7
Группа 4	14,4	13,1	17,8	13,9	13,1	14	15,5	19,7	14,4	14,8	15,1
Группа 5	15,4	12,3	13,8	11,3	16,9	13,5	14,3	19,5	15,7	15,2	14,8
Группа 6	15,8	15,8	12,6	12,9	15,9	13,2	14,6	12,8	12,4	15,7	14,2
Группа 7	14,3	13,4	12,3	13,8	14	12,4	14,3	6,9	12,3	13,4	12,7

Для закона нормального распределения вероятностей складываются соотношения, показанные в табл. 2.

Для определения тенденции в оптимальных структурных уровнях совокупности, подчиненной закону нормального распределения, была использована методика прогнозирования структуры, также изложенная в работе [7]. Применительно к рассматриваемой задаче были определены средние уровни оптимальных структур для 10 (как и выше) опытов, затем для 20 и, наконец, для 30 опытов. На основе динамики полученных трех структур выполнялся поиск стабильной структуры, ожидаемой в перспективе при увеличении количества опытов. Итоги такого исследования приведены в табл. 3.

При общей симметрии процентных соотношений наполненности групп отмечается некоторое смещение в пользу групп, старших по уровню признака. Для нормально распределенных совокупностей указанные выше соотношения по численности наблюдений в группах обеспечивают удельный вес межгрупповой дисперсии в общей вариации на уровнях (табл. 4).

Например, оптимальная процедура группировки совокупности 1000 нормально распределенных чисел – $N(0; 1)$ – на 7 классов обеспечила долю межгрупповой дисперсии на уровне 96 %. Если определить наполненность

Таблица 2

Структура нормально распределенной совокупности при различных вариантах оптимального деления на группы (на 10 опытах), %

Номер группы в порядке возрастания признака	Номер опыта										Средняя структура
	1	2	3	4	5	6	7	8	9	10	
<i>Оптимальное деление на 2 группы</i>											
Группа 1	51,9	51,8	50,6	51,3	48,6	48,6	51,1	47,2	52,8	43,2	49,7
Группа 2	48,1	48,2	49,4	48,7	51,4	51,4	48,9	52,8	47,2	56,8	50,3
<i>Оптимальное деление на 3 группы</i>											
Группа 1	32,1	30,8	27,8	24,7	29	26,1	28,3	29,5	27,7	27,2	28,3
Группа 2	40,1	43,6	47,2	48,2	48,7	49,5	45,3	45,4	44,8	49	46,2
Группа 3	27,8	25,6	25	27,1	22,3	24,4	26,4	25,1	27,5	23,8	25,5
<i>Оптимальное деление на 5 групп</i>											
Группа 1	9,1	8,7	13,8	13,1	7,7	11,1	9,7	8,9	13,1	9,5	10,5
Группа 2	29,9	25,4	27,4	21,7	21,3	21,9	22,3	24	26,6	23,4	24,4
Группа 3	29,9	27,9	27,9	27	28,5	28,5	29,3	27,9	26,3	27,4	28,1
Группа 4	22,6	25,9	22,7	26	28,6	26,2	25,4	25,3	24,9	26,8	25,4
Группа 5	8,5	12,1	8,2	12,2	13,9	12,3	13,3	13,9	9,1	12,9	11,6

Таблица 3

Структура нормально распределенной совокупности при оптимальном делении на 7 групп, %

Номер группы в порядке возрастания признака	Количество опытов			
	10	20	30	Стабильная структура (прогноз)
Группа 1	5,4	5,2	5,1	5,0
Группа 2	13,6	13,3	13,1	12,8
Группа 3	19,4	19,5	19,6	19,6
Группа 4	21,0	21,3	21,4	21,6
Группа 5	20,6	20,3	20,4	20,4
Группа 6	14,3	14,4	14,4	14,6
Группа 7	5,7	6,0	6,0	6,0
Итого	100	100	100	100

Таблица 4

Удельный вес межгрупповой дисперсии в общей вариации при оптимальном делении на группы нормально распределенной совокупности, %

Показатель	Число групп			
	2	3	5	7
Доля межгрупповой дисперсии в общей вариации	≈ 63	≈ 81	≈ 92	≈ 96

Таблица 5

Структура показательно распределенной совокупности при различных вариантах оптимального деления на группы (на 10 опытах), %

Номер группы в порядке возрастания признака	Номер опыта										Средняя структура
	1	2	3	4	5	6	7	8	9	10	
<i>Оптимальное деление на 2 группы</i>											
Группа 1	79,3	80,2	77,6	81,7	80,5	85	78,5	86,4	78,4	80,4	80,8
Группа 2	20,7	19,8	22,4	18,3	19,5	15	21,7	13,6	21,6	19,6	19,2
<i>Оптимальное деление на 3 группы</i>											
Группа 1	58,7	62,8	56,9	65,4	56,2	66,4	63	62,4	63,6	66,2	62,2
Группа 2	30,4	30,6	34,2	27,3	34,7	28,3	30,3	29,7	30,1	28,9	30,4
Группа 3	10,9	6,6	8,9	7,3	9,1	5,3	6,7	7,9	6,3	4,9	7,4
<i>Оптимальное деление на 5 групп</i>											
Группа 1	43,4	46,5	45,7	51,6	45,3	54,4	47,1	48,3	45,2	44,2	47,1
Группа 2	25,7	30,9	29,1	27,6	31,5	29	29,7	32	30,2	30,9	29,7
Группа 3	18,8	15,3	15,0	13,5	15,2	11,3	16,1	13,5	17,6	17,4	15,4
Группа 4	9,2	6,1	8,2	5,8	6,7	4,1	6,2	5,7	5,9	6,1	6,4
Группа 5	2,9	1,2	2,0	1,5	1,3	1,2	0,9	0,5	1,1	1,4	1,4
<i>Оптимальное деление на 7 групп</i>											
Группа 1	38,3	31,3	32,5	37,8	36,0	38,1	36,6	39,1	35,8	33,5	35,9
Группа 2	21,7	26,1	24,1	27,3	26,3	28,3	25,9	24,9	26,7	26,1	25,7
Группа 3	17,1	20,8	20,4	16,6	18,7	18,9	18,6	17,2	17,6	18,0	18,4
Группа 4	10,8	13,1	12,1	9,4	10,9	9,1	11,6	10,0	12,8	12,4	11,2
Группа 5	6,8	5,4	8,0	5,3	5,4	3,7	5,5	5,7	4,5	6,3	5,7
Группа 6	4,3	2,6	2,4	2,8	2,2	1,5	1,5	2,7	2,2	2,5	2,5
Группа 7	1,0	0,7	0,5	0,8	0,5	0,4	0,3	0,4	0,4	1,2	0,6

классов на уровнях, как указано в табл. 3 для стабильной структуры (5 %; 12,8; 19,6; 21,6; 20,4; 14,6 и 6 %), то удельный вес межгрупповой дисперсии в общей вариации будет равным 95,4 %.

Результаты моделирования оптимального распределения на группы совокупностей чисел, следующих показательному распределению, например, распределению времени ожидания с $\lambda = 1$, представлены в табл. 5.

Как видно, оптимальное размещение по группам показательно распределенной совокупности выражается в большей наполненности групп с меньшими значениями изучаемого показателя и относительно небольшими долями в общем количестве наблюдений групп со старшими значениями признака. Для показательно распределенных совокупностей указанные выше соотношения по численности наблюдений в группах обеспечивают удельный вес межгрупповой дисперсии в общей вариации на уровнях (табл. 6).

Как и ранее для иллюстрации покажем результаты разделения на группы совокупности из 1000 случайных чисел, подчиняющихся закону показательного распределения вероятностей. Оптимальная процедура группировки этой совокупности на 7 классов обеспечила долю межгрупповой дисперсии на уровне 96,6 %. Если определить наполненность классов на

Таблица 6

Удельный вес межгрупповой дисперсии в общей вариации при оптимальном делении на группы показательно распределенной совокупности, %

Показатель	Число групп			
	2	3	5	7
Доля межгрупповой дисперсии в общей вариации	62–67	80–83	92–93	96–97

уровнях, как указано в табл. 5 (35,9 %; 25,7; 18,4; 11,2; 5,7; 2,5 и 0,6 %), то удельный вес межгрупповой дисперсии в общей вариации окажется равным 96,1 %.

Приведем результаты исследования еще одного важного вида вероятностного распределения – логарифмически нормального. Именно такая форма распределения свойственна, например, показателю денежных доходов населения. В процессе анализа моделировались совокупности случайных чисел, подчиняющихся логнормальному закону с параметрами $\mu = 9,56699$ и «сигма» = 0,840975, что соответствует массиву населения региона со средним уровнем денежного дохода порядка 20 тыс. руб. на 1 чел. в месяц.

Результаты моделирования оптимального распределения на группы совокупностей чисел, следующих логарифмически нормальному распределению, представлены в табл. 7.

Таблица 7

Структура логарифмически нормально распределенной совокупности при различных вариантах оптимального деления на группы (на 10 опытах), %

Номер группы в порядке возрастания признака	Номер опыта										Средняя структура
	1	2	3	4	5	6	7	8	9	10	
<i>Оптимальное деление на 2 группы</i>											
Группа 1	89,5	85,4	88,6	89,2	86,8	86,0	91,1	83,7	95,6	88,4	88,4
Группа 2	10,5	14,6	11,4	10,8	13,2	14,0	8,9	16,3	4,4	11,6	11,6
Межгрупповая дисперсия	57,2	64,6	61,9	60,0	59,3	62,0	53,9	60,0	46,5	58,2	
<i>Оптимальное деление на 3 группы</i>											
Группа 1	76,0	70,1	70,7	69,5	68,8	71,3	71,2	73,4	88,4	78,1	73,7
Группа 2	21,5	24,8	24,4	26,8	26,3	25,0	24,9	23,3	11,3	20,4	22,9
Группа 3	2,5	5,1	4,9	3,7	4,9	3,7	3,9	3,3	0,3	1,5	3,4
Межгрупповая дисперсия	80,4	82,1	81,2	82,2	79,5	81,1	74,5	80,3	72,8	80,1	
<i>Оптимальное деление на 5 групп</i>											
Группа 1	59,6	42,5	51,8	43,7	49,7	42,6	62,8	49,5	69,6	57,8	53,0
Группа 2	28,1	31,0	30,4	30,7	31,7	29,7	28,3	30,3	23,6	28,6	29,2
Группа 3	9,8	16,3	12,5	16,5	13,4	17,7	7,2	15,3	6,0	10,7	12,5
Группа 4	1,9	7,4	4,0	6,7	4,2	7,7	1,6	4,1	0,7	2,3	4,1
Группа 5	0,6	2,8	1,3	2,4	1,0	2,3	0,1	0,8	0,1	0,6	1,2
Межгрупповая дисперсия	92,7	94,0	92,1	93,2	93,0	92,3	90,7	93,1	91,2	92,7	

Для варианта разделения на 7 групп логарифмически нормально распределенной совокупности была применена процедура, аналогичная использованной для нормального закона. Результаты приведены в табл. 8.

Таблица 8

Структура логарифмически нормально распределенной совокупности при оптимальном делении на 7 групп, %

Номер группы в порядке возрастания признака	Количество опытов			Стабильная структура (прогноз)
	10	20	30	
Группа 1	38,1	37,2	37,0	37,2
Группа 2	29,6	29,6	29,3	28,9
Группа 3	16,9	17,2	17,4	17,1
Группа 4	8,9	9,0	9,1	9,0
Группа 5	4,1	4,4	4,6	5,2
Группа 6	1,9	2,0	2,0	2,0
Группа 7	0,5	0,6	0,6	0,6
Итого	100,0	100,0	100,0	100,0

Оптимальное размещение по группам логнормально распределенной совокупности весьма похоже на результаты, полученные для показательного распределения, но при заметно большем наполнении групп с небольшими уровнями признака, и относительно небольшими долями в общем количестве наблюдений групп с наибольшими значениями признака. Отметим также, что оптимальные группировки логарифмически нормально распределенных совокупностей обеспечивают удельный вес межгрупповой дисперсии в общей вариации на уровнях (табл. 9).

Таблица 9

Удельный вес межгрупповой дисперсии в общей вариации при оптимальном делении на группы логарифмически нормально распределенной совокупности, %

Показатель	Число групп			
	2	3	5	7
Доля межгрупповой дисперсии в общей вариации	58–59	80–81	92–93	≈ 96

Как и ранее покажем для иллюстрации результаты расслоения на группы выборки из 1000 случайных чисел, подчиняющихся закону логарифмически нормального распределения вероятностей. Оптимальная процедура группировки этой совокупности на 7 классов обеспечила долю межгрупповой дисперсии в 96,7 %. Если же определить наполненность классов на стабильных уровнях, как указано в табл. 8 для логарифмически нормального распределения (38,1 %; 29,6; 16,9; 8,9; 4,1; 1,9 и 0,5 %), то удельный вес межгрупповой дисперсии в общей вариации оказывается равным 95,8 %, что вполне сопоставимо с результатами строго оптимального распределения.

Решая задачу определения количества типов или слоев, на которые распадается исследуемая совокупность наблюдений, полезно проследить за изменением величины межгрупповой дисперсии, образующейся при последовательном увеличении числа групп. Так, например, последний 10-й вариант моделирования случайных чисел, подчиняющихся логнормальному распределению, при размещении совокупности последовательно на 2, 3, ..., 15 групп, представлен на рисунке.

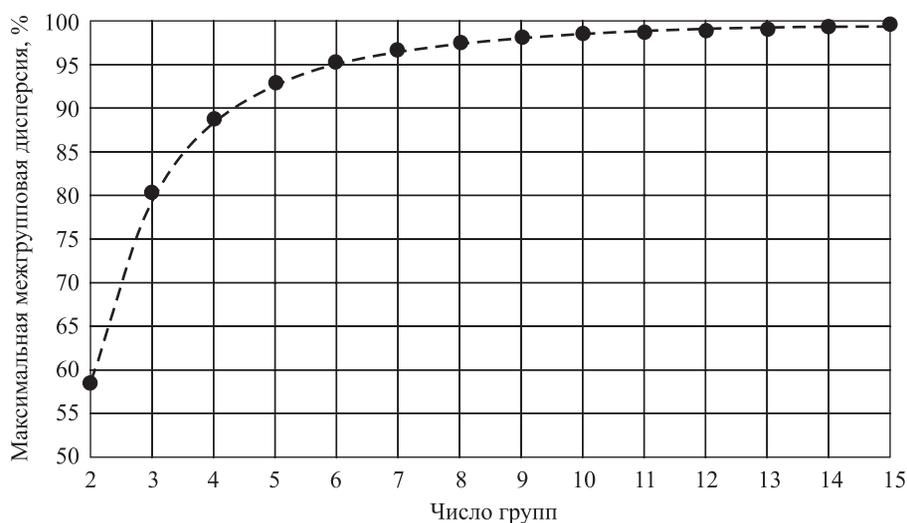


График максимальных значений межгрупповой дисперсии при распределении совокупности на заданное количество групп (классов)

Из графика видно, что максимальная доля межгрупповой дисперсии, которой можно достичь при разделении совокупности на 2 класса, это ориентировочно 57 %. При распределении совокупности на 3 класса максимальная межгрупповая дисперсия выходит на уровень 80 %.

Дальнейшее увеличение числа групп до 5 повышает дисперсию до (92÷93)%, что может считаться достаточно приемлемым уровнем. Увеличение числа групп до 7 влечет возрастание межгрупповой дисперсии до уровня (96÷97)%. Последующее дробление изучаемой совокупности едва ли целесообразно; возрастание межгрупповой дисперсии становится незначимым.

Полученные выше расчетные структуры, приближающие межгрупповую дисперсию к максимальному значению, могут иметь не только практическую направленность как средство достаточно быстрого разделения совокупности наблюдений на заданное количество слоев, классов, типов. В качестве гипотезы предположим, что каждая из полученных выше структур является структурой потенциальной, способной распознать тип вероятностного распределения неизвестной статистической совокупности. Потенциальная структура это квантильная характеристика формы распределения. Так, в частности, для $k = 7$ образы потенциальных структур представлены в табл. 10.

Таблица 10

**Потенциальная структура оптимального деления совокупности на 7 групп
для распознавания типа вероятностного распределения, %**

Номер группы в порядке возрастания изучаемого показателя	Тип вероятностного распределения			
	равномерное	нормальное	показательное	логарифмически нормальное
1	13,9	5,0	35,9	37,2
2	14,6	12,8	25,7	28,9
3	14,7	19,6	18,4	17,1
4	15,1	21,6	11,2	9,0
5	14,8	20,4	5,7	5,2
6	14,2	14,6	2,5	2,0
7	12,7	6,0	0,6	0,6
Итого	100,0	100,0	100,0	100,0

Примечание. Для равномерного вероятностного распределения уровни потенциальной структуры были также рассчитаны и по методике прогнозирования по аналогии с нормальным и логарифмически нормальным законами. Предельная структура оптимального распределения совокупности на 7 классов определилась как: 14,3 %; 14,2; 14,7; 15,0; 14,4; 13,8 и 13,6 %, что в целом совпадает со структурой, которая приведена выше (13,9 %; 14,6; 14,7; 15,1; 14,8; 14,2 и 12,7 %).

Предложим для испытания 4 набора случайных чисел, взятых как выборки из генеральных совокупностей, подчиненных, соответственно, распределению вероятностей: равномерного, нормального, показательного и логарифмически нормального законов. Для каждой выборки оценим качество каждой из указанных выше потенциальных структур. Показателем такого качества установим удельный вес соответствующей межгрупповой дисперсии в общей вариации (табл. 11).

Таблица 11

Оценка качества потенциальной структуры вероятностных распределений

Тип потенциальной структуры	Удельный вес межгрупповой дисперсии ($k = 7$) в общей вариации для различных вероятностных законов, %			
	равномерный	нормальный	показательный	логарифмически нормальный
Для равномерного распределения	97,9	93,8	87,6	80,3
Для нормального распределения	96,9	95,4	91,4	87,8
Для показательного распределения	93,0	87,3	96,2	96,0
Для логарифмически нормального распределения	93,0	86,4	96,2	96,4

Как видим, каждый вариант потенциальной структуры достаточно уверенно распознает свой собственный тип вероятностного распределения. Поясним это обстоятельство на примере. Предположим, что для распозна-

вания закона распределения вероятностей предлагается некоторый массив случайных чисел с неизвестной формой распределения. Применим к этому числовому массиву процедуру группировки – 13,9 %; 14,6; 14,7; 15,1; 14,8; 14,2 и 12,7 % – для распознавания равномерного распределения. В результате удельный вес межгрупповой дисперсии определился на уровне 94,2 %.

При использовании структуры – 5,0 %; 12,8; 19,6; 21,6; 20,4; 14,6 и 6,0 % – для распознавания закона нормального распределения удельный вес межгрупповой дисперсии определился на уровне 95,7 %.

При использовании структуры – 35,9 %; 25,7; 18,4; 11,2; 5,7; 2,5 и 0,6 % – для распознавания закона показательного распределения удельный вес межгрупповой дисперсии определился на уровне 87,1 %.

При использовании структуры – 37,2 %; 28,9; 17,1; 9,0; 5,2; 2,0 и 0,6 % – для распознавания логарифмически нормального распределения удельный вес межгрупповой дисперсии определился на уровне 86,2 %.

Закключаем, что исследуемый числовой массив является выборкой из нормально распределенной генеральной совокупности (наибольший уровень межгрупповой дисперсии – 95,7 %). И это действительно так, поскольку для исследования предлагался именно массив нормально распределенных случайных чисел, полученных в системе STATISTICA.

Результаты распознавания вида распределения были получены выше для варианта распределения совокупности на 7 типов. По-видимому, меньшее количество типов или структурных слоев применять едва ли целесообразно. При уменьшении количества типов сглаживаются особенности параметров законов распределения вероятностей. Так, можно ограничиться построением оптимальных структур для пяти типов (табл. 12).

Таблица 12

Потенциальная структура оптимального деления совокупности на 5 групп для распознавания типа вероятностного распределения, %

Номер группы в порядке возрастания изучаемого показателя	Типа вероятностного распределения			
	равномерное	нормальное	показательное	логарифмически нормальное
1	19,4	10,5	47,1	53,08
2	19,9	24,4	29,7	29,2
3	20,9	28,1	15,4	12,5
4	20,5	25,4	6,4	4,1
5	19,3	11,6	1,4	1,2
Итого	100,0	100,0	100,0	100,0

Результаты распознавания приведены в табл. 13.

Как видно, распознавательные возможности потенциальных структур понизились; логарифмически нормальное распределение вполне может быть опознано как обычное «нормальное» распределение, хотя, конечно, следует принять во внимание и то, что данные потенциальные структуры

Таблица 13

Оценка качества потенциальной структуры вероятностных распределений

Тип потенциальной структуры	Удельный вес межгрупповой дисперсии ($k = 5$) в общей вариации для различных вероятностных законов, %			
	равномерный	нормальный	показательный	логарифмически нормальный
Для равномерного распределения	97,0	89,4	81,2	71,1
Для нормального распределения	94,0	91,9	85,9	80,9
Для показательного распределения	87,1	80,5	92,9	93,2
Для логарифмически нормального распределения	83,5	77,4	92,7	92,5

были получены лишь в первом приближении как результат простого осреднения структур по 10 опытам. В заключение отметим, что для распознавания других функций распределения вероятностей целесообразно создать библиотеку потенциальных структур для нескольких вариантов количества типов или структурных слоев и на основе прогнозирования ожидаемой оптимальной структуры.

Литература

1. Глинский В.В. Статистические методы поддержки управленческих решений. Новосибирск: НГУЭУ, 2008. 256 с.
2. Ионин В.Г., Щеглова Т.Л. Многомерная статистическая оценка показателей инвестиционной привлекательности фирм // Вестник НГУЭУ. 2010. № 1. С. 148–155.
3. Ионин В.Г. Структуризация населения по уровню денежных доходов (на примере Новосибирской области) // Научные записки НГАЭУ. 2001. С. 70–76.
4. Ионин В.Г., Ярославцева Л.П. Стратификация населения Новосибирской области по уровню денежных доходов // Вестник НГУЭУ. 2010. № 1. С. 116–130.
5. Ионин В.Г., Ярославцева Л.П. К проблеме выделения социально-экономических типов населения // Вестник НГУЭУ. 2012. № 4. Т. 2. С. 74–86.
6. Ионин В.Г., Ярославцева Л.П. Структура населения по социально-экономическим типам: временной разрез // Вызовы современного мира. Материалы международной научно-практической конференции (Новосибирск, 24–25 сентября 2013 г.). Новосибирск, 2013. С. 111–112.
7. Ионин В.Г. Статистический анализ типовых экономических и социальных процессов на ПЭВМ: учеб. пособие. Новосибирск: НГУЭУ, 2005. 170 с.

Bibliography

1. Glinskij V.V. Statisticheskie metody podderzhki upravlencheskih reshenij. Novosibirsk: NGUJeU, 2008. 256 p.
2. Ionin V.G., Shheglova T.L. Mnogomernaja statisticheskaja ocenka pokazatelej investicionnoj privlekatel'nosti firm // Vestnik NGUJeU. 2010. № 1. P. 148–155.
3. Ionin V.G. Strukturizacija naselenija po urovnju denezhnyh dohodov (na primere Novosibirskoj oblasti) // Nauchnye zapiski NGAJeU. 2001. P. 70–76.

4. *Ionin V.G., Jaroslavceva L.P.* Stratifikacija naselenija Novosibirskoj oblasti po urovnju denezhnyh dohodov // Vestnik NGUJeU. 2010. № 1. P. 116–130.
5. *Ionin V.G., Jaroslavceva L.P.* K probleme vydelenija social'no-jekonomicheskikh tipov naselenija // Vestnik NGUJeU. 2012. № 4. T. 2. P. 74–86.
6. *Ionin V.G., Jaroslavceva L.P.* Struktura naselenija po social'no-jekonomicheskim tipam: vremennoj razrez // Vyzovy sovremennogo mira. Materialy mezhdunarodnoj nauchno-prakticheskoi konferencii (Novosibirsk, 24–25 sentjabrja 2013 g.). Novosibirsk, 2013. P. 111–112.
7. *Ionin V.G.* Statisticheskij analiz tipovyh jekonomicheskikh i social'nyh processov na PJeVM: ucheb. posobie. Novosibirsk: NGUJeU, 2005. 170 p.