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

А.М. ШИХАЛЁВ

ГРУППИРОВКА НАБЛЮДЕНИЙ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЯЕНИЙ

Учебно-методическое пособие

Казань – 2015

УДК (075.8)311

ББК 60.6я73

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

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

Шихалёв А.М.

Группировка наблюдений социально-экономических явлений / А.М. Шихалёв. – Казань: Казан. ун-т, 2015. – 38 с.

Методическая пособие предназначено для подготовки и выполнения студентами самостоятельной работы 1 для раздела «Общая теория статистики». Группировка исходной информации в виде статистической совокупности, полученной в результате наблюдения, представленной как правило в виде набора вещественных чисел (элементов совокупности) представляет собой многоэтапный процесс, результатом которого является получение вариационного ряда, отражающего структурные свойства исследуемой статистической совокупности, которые в явном виде в ней не содержатся. Полученные результаты позволяют оценить характер распределения исследуемой случайной величины в виде того или иного статистического показателя.

Ключевые слова: статистическое наблюдение, сводка данных, группировка данных, среднее взвешенное, мода, медиана, асимметрия распределения, дисперсия, среднее квадратическое отклонение, вариабельность.

СОДЕРЖАНИЕ

	Стр.
1. СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ	4
1. Содержание сводки и группировки статистической информации	4
1.2. Построение дискретных вариационных рядов	10
1.3. Особенности нахождения среднего значения вариационного ряда	13
1.4. Построение интервальных вариационных рядов	15
2. СРЕДНИЕ И СТРУКТУРНЫЕ СРЕДНИЕ ВАРИАЦИОННЫХ РЯДОВ	19
2.1. Определение моды и медианы графо-аналитическим методом	19
2.2. Определение моды и медианы аналитическим методом	21
2.3. Среднее значение вариационного ряда и степень асимметрии	23
2.4. Особый случай вычисления структурных средних	25
3. ОЦЕНКА ВАРИАБЕЛЬНОСТИ ИСХОДНОЙ ИНФОРМАЦИИ	32
4. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ЗАЩИТЕ РЕЗУЛЬТАТОВ РАБОТЫ	33
5. ВАРИАНТЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ 1	34
ЛИТЕРАТУРА	37

1. СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ

Сводка и группировка статистической информации в виде статистической совокупности (СС) $Y = \{y_j\}$, $j = 1,N$, где N – число элементов совокупности, представляет собой многоэтапный процесс, связанный с такими предварительными операциями, как выбор максимального и минимального элементов (y_{max}, y_{min}) совокупности Y, определение размаха выборки (R), оценка числа интервалов (n) будущего вариационного ряда (ВР), определение величины шага интервала h (постоянной или переменной величины).

Собственно построение ВР как средства группировки исходной статистической информации в виде СС в виде новой, более компактной совокупности $X = \{x_i\}$, $i = 1,n$, где n – число строк ВР, представляющего собой компактную таблицу (при этом $n << N$) заключается в оценке числа попаданий в тот или иной интервал с последующим подсчетом числа попаданий (f_i). Дополняется процесс построения ВР вычислением значений накопленных частот q_i и значений середин интервалов $x_i^{ср}$, если ВР интервальный.

После получения ВР можно приступить к вычислениям среднего взвешенного ($x^{ср}_{вз}$) и структурных средних – моды (M_o) и медианы M_e) и производить последующие необходимые расчеты. Однако для их расчета (графо-аналитическим или аналитическим методом) сначала необходимо определить номер модального интервала (i_m).

1.1. Содержание сводки и группировки статистической информации

Статистическое *наблюдение* – это первый этап любого статистического исследования и заключается в сборе данных о массовых явлениях путем регистрации их признаков (источник первичной информации). По документам отчетности можно провести и единовременные выборочные обследования (см. выборочные и генеральные совокупности).

На втором этапе статистического исследования на основе первичной информации выступает *сводка* – большей частью неформализованная, содержательная процедура обработки первичных материалов наблюдений с целью получения итоговых или упорядоченных числовых характеристик изучаемой статистической совокупности. Ее важным моментом является группировка.

Группировка – это объединение статистических данных в однородные по определенным признакам группы. Она помогает изучить структуру статистической совокупности и некоторые элементы взаимосвязи меду явлениями.

Распределение единиц совокупности по количественному признаку называется вариационным рядом (ВР). Ряды могут быть построены по дискретным или непрерывным признакам. Отсюда и название – дискретные ВР (задача1) и интервальные ВР (задача 2). При этом интервальные ВР могут быть как с интервалами (шагами) одинаковой величины, так и с шагами неодинакового размера. Таки образом, в достаточной степени формализованный аппарат ВР является средством группировки на уровне дискретных или непрерывных признаков.

Исходной информацией для построения интервальных ВР служит однородная последовательность N наблюдаемых элементов, объединенным общим признаком - *статистическая совокупность*. Путем группирования может быть получен ВР, имеющий форму таблицы с «n» строками (вариантами), причем всегда n < N. В результате вместо исходной неструктурированной совокупности однородных элементов (исходная СС) получаем компактную таблицу (построенный ВР), отражающую структуру исходных данных. В связи с этим полученная таблица (ВР) вместо исходной СС приобретает дополнительные свойства.

Если для исходной СС можно определить число ее элементов (N), сумму элементов, то легко найтй среднее значение СС – так называемое «механическое среднее», - поскольку каждый элемент СС из N штук встречается в ней ровно один раз ($f = 1$). По-иному выглядят варианты ВР числом «n», заменяющие N исходных элементов СС. Тогда в ВР, в отличие от исходной СС, каждый «вариант», заменяет собой как правило некоторую группу элементов (от одного и более) исходной СС, то есть появляется много раз, иначе говоря, - имеет некоторую, присущую данному интервалу, частоту своего появления (f_i). Тогда, располагая частотами появления всех интервалов сформированного ВР, легко получить «накопленную частоту» (q_i) по простому правилу:

$$q_i = q_{i-1} + f_i. \tag{1}$$

Понятно, что $q_1 = f_1$, остальные накопленные частоты (частости) вычисляются по формуле (1). Однако здесь и далее будет полезным заметить, что для вычисления первой накопленной частоты (частости) применение выражения (1) также корректно, поскольку для первой строки q_{i-1} отсутствует, то есть $q_{i-1} = 0$. Заметим также, что данное правило распространяется и на вычисление параметров структурных средних при определенных условиях (например, когда модальным интервалом является первый интервал).

Следовательно, значение каждого варианта (i-ой строки таблицы, отображающей ВР) поставлена в соответствии с определенным значением частоты f_i и накопленной частоты q_i. Отсюда и дополнительные свойства, кроме уже упомянутого среднего значения как свойства исходной СС (напомним – механического среднего, когда каждый элемент совокупности появляется строго один раз; или так: элементы исходной СС имеют одинаковую частоту появления, например, равной единице).

В целом построение вариационного ряда из исходной статистической совокупности можно представить как *отображение* τ одного (исходного множества) Y мощностью $|Y| = N$ значений (элементов СС) на новое множество X (вариационный ряд) мощностью $|X| = n$ значений (элементов ВР в виде строк таблицы):

$$\tau : Y \rightarrow X. \tag{2}$$

ВР также можно, как и СС, характеризовать средним значением ВР с тем лишь отличием, что варианты ВР (аналоги элементов СС) как правило появляются неодинаковое число раз: в один вариант попало больше значений из исходной СС, в другой меньше и т.д. Следовательно, при вычислении среднего значения по данным ВР необходимо как-то учитывать частоту их появления f: вычисляется «средневзвешенное» - с учетом значения частот появления вариантов, - вместо «среднего механического» - без учета частот появления каждого элемента исходной статистической совокупности. Однако, если быть более точным, заметим, что частота появления каждого элемента нам известна: каждый элемент исходной СС появляется в ней, напомним, ровно один раз.

Поскольку частота появления каждого варианта ВР (строки таблицы, в форме которой и существует созданный нами ВР) неодинакова, у вновь созданной нами конструкции (ВР вместо СС) появляются новые свойства, которые можно отобразить с помощью дополнительных параметров по отношению к среднему значению (в данном случае – с средневзвешенному) модной M_o. и медианой M_e, которые относятся к так называемым «структурным средним».

Мода – это наиболее часто встречающееся значение (число в размерности СС) как ВР, так и исходной СС, и притом неважно – присутствует ли данное значение в исходной СС или нет. Важно, что оно, это значение, существует как абстракция объективно, то есть независимо от нас.

Медиана – это такое значение (число в размерности СС), которое по своей величине разделяет элементы СС пополам (отсюда и название по аналогии с геометрическим термином): половина элементов, величина которых меньше значения медианы, включая элемент, значение которого равно значению медианы, составляют одну половину исходной СС; другую половину составляют элементы, значения которых превышают величину найденной медианы. При этом нетрудно заметить, что только что сказанное может быть отнесено к такой СС, число элементов которой N – четное (а иначе – как разделить нечетное чило на два без остатка?). Если же число элементов нечетное, то поровну не получится: какая-то часть (до или после значения медианы) будет на единицу боле, какая-то на единицу меньше, что очевидно. И еще одно замечание. Если размер исходной статистической совокупности небольшой (10-20 элементов), то заявленной выше картины может не получиться даже в случае четного N: исходная СС может быть разделена на далеко не симметричные группы – до медианы (включая и ее) может быть не половина элементов исходной СС, а меньше или больше. Данное обстоятельство связано с проблемами оперирования т.н. «малыми выборками» в теории вероятностей и математической статистике, что выходит за рамки нашей темы. Для нас важен практический аспект проблемы: при увеличении мощности СС данная проблема снимается сама собой.

Здесь же будет целесообразным оценить размеры (мощность множества строк n) вариационных рядов в зависимости от размеров (мощностей множеств элементов N) исходных статистических совокупностей. На умозрительном уровне такой проблемы будто бы не существует: понятно, чем больше N, тем больше n и значит таблица, в форме которой и существует вариационный ряд как замена исходной (неудобной для вычислений и объемной) статистической совокупности. Точнее, вопрос о мере: насколько возрастет объем таблицы (ВР) в зависимости от объема исходной СС.

Для ответа на такой вопрос воспользуемся известной формулой Стерджесса, которую рекомендуют использовать для приблизительной оценки числа необходимых интервалов (вариантов ВР) m (округляется в меньшую сторону):

$$n \approx 1 + 3,22 \cdot \lg N. \qquad (3)$$

Согласно выражению (3) нетрудно провести несложные расчеты с целью оценки связи числа строк ВР n с числом группируемых элементов СС N.

Если N = 10, то $n = 1 + 3,22 \cdot \lg 10 = 1 + 3,22 \cdot 1 = 4,22 \approx 4$.

Если N = 100, то $n = 1 + 3,22 \cdot \lg 100 = 1 + 3,22 \cdot 2 = 7,44 \approx 7$.

Если N = 1000, то $n = 1 + 3,22 \cdot \lg 1000 = 1 + 3,22 \cdot 3 = 10,66 \approx 10$.

Если N = 10000, то $n = 1 + 3{,}22 \cdot \lg 10000 = 1 + 3{,}22 \cdot 4 = 13{,}88 \approx 13$.

Понятно, что N растет линейно, а n – логарифмически, поэтому со значительным ростом числа элементов исходных статистических совокупностей N размер вариационных рядов как таблиц с число строк n увеличивается несущественно: для числа элементов N 10000 и 10 различаются в тысячу раз, тогда как для n – числа строк таблиц, - 13 строк и 4 строки – всего в три раза.

Итак, в результате сравнительного анализа свойств исходных СС и получаемых из них ВР нами установлено, что ВР получаются по определенным правилам из СС, но при этом они не только наследуют их свойства (и та, и другая статистическая конструкция имеют те же размерности, допускают вычисление их средних значений - пусть и в разном виде), но и приобретают новые в виде структурных средних – моду и медиану. При этом и участки медианы (до нее и после нее) также можно разделить пополам (так получаются «квартили» Q), а то и всю СС – на десять частей (так получаются «децили» D). И те и другие находят применение в статистическом анализе и позволяют получить такие известные статистические модели, как, например, децильный коэффициент доходности (ДКД), отражающий соотношение доходов 10% наиболее обеспеченных граждан к доходам 10% наименее обеспеченных (региона, государства). ДКД при этом вычисляется как отношения девятого дециля к первому децилю ВР, отражающего доходы групп населения.

Если данное соотношение превышает 10, то объективно существует социально-политическая напряженности в обществе.

Для справки: в Скандинавии это соотношение находится в пределах 3-4; в Европе до недавних пор 5-7; для США 9-10; для РФ (2010 г.) по сообщению Президента РФ – в пределах 16,5).

Таким образом, новые параметры – структурные средние, - являются важными характеристиками вариационных рядов, определение которых можно произвести двумя способами – аналитическим (точным, с помощью специальных формул) и графо-аналитическим (приблизительным, с помощью графических средств).

1.2. Построение дискретных вариационных рядов

Задача 1. Источник статистической информации – ведомственная статистика, - отдел кадров одного из предприятий города предоставил исследователю следующие данные о тарифных разрядах (ТР) 50-ти рабочих одного из цехов завода (тот же результат можно получить и в результате собственных наблюдений, путем опроса рабочих на проходной, например).

Далее необходимо как-то *эксплицировать* (формализовать) исходную информацию в виде символов. С этой целью обозначим полученную (или созданную нами) исходную статистическую совокупность (условные номера рабочих и их разряд), которую назовем $Y = \{y_j\}$, $j = 1,N$, где $N = |Y| = 50$ – т.е. получим всего 50 штук тарифных разрядов. В данном случае для нас не представляет интерес другие свойства рабочих (их образование, возраст, семейное положение, реальный уровень квалификации, степень социальной активности, состояние здоровья, интересы и др.) – только их разряды. То есть статистическая совокупность построена нами лишь *по одному количественному признаку* с именем признака «разряд» и представлена в произвольном виде в табл. 1.

Чтобы показать распределение рабочих по ТР, построим ВР (*дискретный*, коли исходные данные носят здесь дискретный характер), то есть новую, *более компактную* статсовокупность с новым именем $X = \{x_i\}$, $i = 1,n$ $(n < N)$. Иными словами, мы хотим построить отображение (2): заменить известное нам множество Y на более компактное и неизвестное нам пока множество с именем X.

Для этого необходимо и достаточно сделать следующее.

Таблица 1

Исходная статистическая совокупность

по дискретному признаку (разрядам рабочих одного цеха)

3	5	6	3	2	4	3	5	5	6
4	3	2	3	4	5	4	2	4	6
5	3	4	5	4	3	3	6	2	3
4	6	3	4	4	5	4	5	3	4
2	6	3	4	5	3	4	4	5	4

1. Найти максимальное и минимальное значение среди элементов исходной статистической совокупности: y_{max} = 6 р., y_{min} = 2 р. Тогда размах выборки $R = y_{max} - y_{min} = 6 – 2 = 4$ (разряда), то есть установить, что в пределах размаха выборки помещаются все разряды от 2-го до 6-го (2, 3, 4, 5, 6). Однако непосредственного применения в данном случае размах выборки не находит и выступает не как этап формализации, но как сопутствующая информация. Здесь важно то, что в качестве элементов новой статистической совокупности X будут выступать разряды от минимального до максимального: $X = \{x_i\}$, $i = 1,n$. Сосчитаем их, чтобы определить мощность $|X| = n$. Тогда x_1 = 2 р.; x_2 = 3 р.; x_3 = 4 р.; x_4 = 5 р.; x_5 = 6 р. Всего насчитали пять вариантов, следовательно $n = 5$. При этом $n < N$ действительно выполняется: $5 < 50$.

2. Упорядочить разряды по возрастанию или убыванию (операция упорядочения или *рандомизации*, что и сделано в предыдущем шаге) и принять их в качестве элементов будущего, создаваемого нами более компактного по отношению к исходной информации множества с именем X. Вот эти разряды и будем именовать «вариантами» создаваемого нами вариационного ряда (см. табл. 2).

3. Сосчитать, сколько элементов исходной СС принадлежит тому или иному варианту ВР (разряду) по следующему алгоритму. Продвигаясь по табл.

1 по строкам или столбцам (как кому покажется удобнее) при прохождении каждого очередного разряда делаем отметку в виде «слежа» (косой черточки) до тех пор, пока не достигнем конца табл. 1 (сумма всех косых черточек в сумме в табл. 2 должно быть равна 50-ти; это понятно). После этого подсчитаем, сколько раз тот или иной вариант (разряд) зафиксирован нами как случайное событие и запишем информацию в интервальной (привычной нам) шкале (графа f_i) в виде цифры. Это и будет частота появления варианта в новом множестве с именем X: отображение вида (2) нами построено. Заодно подсчитаем и накопленные частоты q_i путем прибавления следующей частоты к предыдущей по формуле (1). Табл. 2 как выражение дискретного вариационного ряда заполнена.

Таблица 2

Дискретный вариационный ряд. Результаты отображения вида (2)

Счетчик Интервалов	Имя и значения вариантов	Частота появления разряда в табл. 1	Значения частоты	Значения накопленной частоты
i	x_i.	счет	f_i	q_i
1	$x_1 = 2$ р.	//////	5	5
2	$x_2 = 3$ р.	/////////////	13	18
3	$x_3 = 4$ р.	////////////////	16	34
4	$x_4 = 5$ р.	//////////	10	44
n=5	$x_5 = 6$ р.	//////	6	50
		$\sum f_i = 50$		

В итоге дискретный вариационный ряд (табл. 2) из исходной СС (табл. 1) построен. Хотя и по размерам страницы табл. 1 и 2 вполне соизмеримы, табл. 2 – более информативна по отношению к табл. 1, поскольку она (табл. 2) демонстрирует *структуру* исходной СС. Мы видим, что средних разрядов (3-х и 4-х) – побольше, самых низких (2-е) и самых высоких (5-е и 6-е) – поменьше, и в чем-то напоминает (хоть и отдаленно) в графе табл. 2 «счет» гауссовский закон нормального распределения случайной величины.

С целью визуализации полученных результатов нашего отображения вида (1) – другими словами, - вариационного ряда, полученного из исходной статистической совокупности, достаточно поставить в соответствие номера вариантов на оси абцисс и значения частот на оси ординат (получим т.н. «полигон распределения» - ломаную кривую), а также номера вариантов и значения накопленных частот – соответственно (получим т.н. «кумуляту» - плавная кривая, что подробнее рассмотрим на примере интервальных ВР).

Однако в построении подобных графиков настоятельной необходимости нет, поскольку распределение разрядов среди 50-ти рабочих как случайной величины и так достаточно наглядно характеризует картинка со «слежами» в графе «счет» табл. 2. Однако попутно обратим внимание на другое, присущее в равной мере как дискретным, так и интервальным ВР – особенность нахождения среднего значения ВР.

1.3. Особенности нахождения среднего значения вариационного ряда

На вопрос, каким же будет *среднее значение* ($x_{ср}$) разряда рабочих на умозрительном уровне, ответить легко: по второй графе табл. 2 – 4-й разряд, поскольку среднее – это когда сумму всех элементов совокупности (здесь – вариантов ВР) делим на их число. Тогда сумма разрядов составит 2+3+4+5+6 = 20. Среднее как частное от деления суммы на число элементов совокупности 20 / n = 20 / 5 = 4 (р.) характеризует меру этой неодинаковости.

Однако это будет справедливым, если каждый разряд в табл. 2 появился бы равное число раз: или все по одному разу, или все по нескольку раз, но с одинаковой частотой. Но частота проявления каждого разряда в исходной СС в нашем первоначальном восприятии уже была неодинаковой, и содержание рабочей табл. 2 (так выглядит дискретный ВР) показывает количественную меру этой неодинаковости: 2-й разряд проявляется в 5 случаях из 50-ти, третий – 13-ти случаях из 50-ти и т.д. В этом случае, чтобы учесть неодинаковую частоту появления каждого варианта, вычисляют т.н. «взвешенную среднюю» $x_{ср}^{вз}$:

$$x_{ср}^{вз} = \frac{\sum x_i \cdot f_i}{\sum f_i}. \qquad (4)$$

Произведя вычисления по формуле (4), получим:

$$x_{ср}^{в} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{2\cdot 5 + 3\cdot 13 + 4\cdot 16 + 5\cdot 10 + 6\cdot 6}{50} = \frac{199}{50} = 3{,}98 \text{ (р.)}.$$

Хоть и близко значение 3,98 к 4,00, но все же они разные по своей сути. А вот если бы частота для всех вариантов (разрядов) была бы одинаковой 10; всего вариантов n=5, на каждый вариант по 10, в сумме 50; все сходится). Умножим числитель и знаменатель выражения (2) на единицу и внесем ее как постоянную величину в знак сумм. Тогда одинаковость частот можно так:

$$f_i = f_1 = f_2 = f_3 = f_4 = f_5 = f_{const} = f = 10, \qquad (5)$$

а выражение (4) с преобразованиями примет вид (постоянное значение частоты выносим за знак сумм числителя и знаменателя):

$$x_{ср}^{вз} = \frac{\sum x_i \cdot f}{\sum 1 \cdot f} = \frac{f \sum x_i}{f \sum 1} = \frac{\sum x_i}{N} = \frac{1}{N} \sum x_i = x_{ср}. \qquad (6)$$

То есть сумма единиц в знаменателе от 1 до N и есть (1+1+1+ …… +1) = N единиц. Остальное соответственно в числителе и знаменателе выражения (5) сокращается.

Таким образом при одинаковой частоте появления вариантов *среднее взвешенное* $x_{ср}^{вз}$ сводится к простой *механической средней* $x_{ср}$.

Или, другими словами, механическое среднее представляет собой средневзвешенную величину в случае, когда частота появления вариантов ВР одинакова. То есть среднее механическое – это в общем случае частный случай среднего взвешенного. Но вместе с тем это по своей сути разные параметры.

Заметим также, что формула вычисления среднего взвешенного в виде (4) будет справедлива и для интервальных вариационных рядов.

1.4. Построение интервальных вариационных рядов

Задача 2. Источник статистической информации – результаты мониторинга, проведенного самим исследователем состояния основных фондов (капитала) в млн. руб. выбранных для исследования малых предприятий (МП) города в рамках выполнения хоздоговорного заказа со стороны Совета министров РТ. В результате наблюдений составлена таблица исходных данных в виде однородной статистической совокупности, приведенной в табл. 3.

Таблица 3

Исходная статистическая совокупность по непрерывному признаку (размер ОФ выбранных для исследования МП города)

9.4	8.0	6.3	10.0	15.0	8.2	7.3	9.2	5.8	8.7
5.2	13.2	8.1	7.5	11.8	14.6	8.5	7.8	10.5	6.0
5.1	6.8	8.3	7.7	7.9	9.0	10.1	8.0	12.0	14.0
8.2	9.8	13.5	12.4	5.5	7.9	9.2	10.8	12.1	12.4
12.9	12.6	6.7	9.7	8.3	10.8	15.0	7.0	13.0	9.5

Здесь, как и в предыдущей задаче, мы хотим отобразить исходное множество Y в более компактное множество X – произвести ту же операцию вида (2). Для этого необходимо сделать что-то из того, что было раньше – еще с дискретными СС, - но и что-то еще доплнительно. Оно и понятно: там, в случае формирования СС по дискретному признаку, имеем дело с *четко фиксированными значениями*. Они, эти значения, могут повторяться, но их градация известна: в задаче 1 – исходных значений целых 50, а вся информация «зашифрована» всего лишь пятью значениями в виде разрядов 2, 3 4 5 и 6-го. Потом подсчитали, как часто эти разряды себя проявили в общей совокупности - и задача практически решена: дискретный вариационный ряд построен. Структура исходной статистической совокупности выявлена: средних разрядов (3,4) больше, слишком малых и слишком больших (2, 5 и 6) – меньше. Это с одной стороны. С другой стороны, сама последовательность построения ВР как наследует все свои этапы, так и предусматривает новые, дополнительные операции.

Итак, исходная информация в виде СС представлена в табл. 3.Для нас – это СС, то есть множество значений с именем Y, числом N = 50 (шт.).

1. Сначала, как и прежде для дискретного ВР, необходимо найти максимальное и минимальное значение среди элементов исходной статистической совокупности: y_{max} = 15,0 млн. руб., y_{min} = 5,1 млн. руб. Тогда размах выборки

$$R = y_{max} - y_{min} . \qquad (7)$$

По формуле (7) получим:

R = 15,0 – 5,1 = 9,9 (млн. руб.).

Значит все значения элементов множества Y расположены в данном диапазоне, который необходимо разделить на необходимое количество интервалов, что и составляет сущность *формирования вариантов* будущего ВР.

2. Определение приблизительного числа интервалов (вариантов ВР) по формуле Стерджесса (3):

$$n = 1 + 3{,}22 \cdot \lg N = 1 + 3{,}22 \cdot \lg 50 = 1 + 3{,}22 \cdot 1{,}69 \approx 6{,}44 \approx 6 \text{ (интервалов).}$$

Для построения ВР примем решение ограничить число интервалов пятью интервалами (n=5). Поскольку практика построения интервальных ВР показывает, что с увеличением числа интервалов (особенно при незначительных N, как данной задаче) могут быть сформированы такие отдельные интервалы, в которые, что легко представить, не попадет ни одно элемента из исходных значений.

Тогда график ВР в виде гистограммы (о чем далее) будет иметь т.н. «провал», что в известной степени может затруднить процесс получения графо-аналитической (да и аналитической) оценок значений структурных средних – моды и медианы. В таком случае рекомендуется уменьшить число строк таблицы как формы существования ВР «n» и повторить вычисления по последующим пунктам 3 и 4.

3. Оценим величину шага h, разделив размах выборки на число шагов:

$$h = R / n. \qquad (8)$$

По выражению (8) получим:

h = 9,9 млн. руб. / 5 = 1,98 млн. руб.

Это при условии, что формировать интервалы (варианты интервального ВР) начнем с минимального значения 5,1 млн. руб. Однако возможен более удобный вариант, если начнем формировать интервалы не с минимального значения 5,1 млн. руб, а со значения 5,0 млн. руб. Тогда R = 15,0 – 5,0 = 10,0 (млн. руб.) и h = R / n = 10.0 млн. руб./ 5 = 2 млн. руб., что при проведении практических расчетов гораздо удобнее. Заметим только, что подобное округление оправдано, если при этом «не потеряем» крайние значения исходной совокупности – минимальное и максимальное. В данном случае не теряется ни 5,1 млн. руб., ни 15,0 млн. руб. Значит шаг по выражению (8) с округлениями вполне оправдан: 5,1 млн. руб. будет принадлежать первому интервалу с номером i = 1; элемент совокупности 15,0 млн. руб. будет принадлежать интервалу с номером i = n = 5.

4. Заполняем рабочую таблицу (см. табл. 4) – будущий интервальный ВР с постоянным шагом h = 2 млн. руб.. Также введем дополнительно по отношению к табл. 2 дискретного вариационного ряда графу для фиксирования середин $x^{ср}_i$ сформированных интервалов.

Замечание. При такой записи непрерывного признака, когда одну и ту же величину можно отнести как к верхней границе, так и к нижней (например, 7), начала интервалов значений следует воспринимать (за исключением первого интервала) как открытые границы интервалов, тогда как концы интервалов как закрытые границы. Иначе говоря, если при заполнении табл. 4 в исходной СС встретится, например, число 7, то его справедливым будет отнести к первому интервалу. А число, большее 7 (например 7,01), уже ко второму интервалу.

5. Определяем номер модального интервала и значение модальной (максимальной) частоты f_m. Модальным интервалом будет тот, которому принадлежит наибольшее число попаданий исследуемой случайной величины (размер капитала того или иного обследованного МП города). В нашем случае в качестве *модального* будет выступать второй интервал, частота попаданий в котором максимальная. Следовательно $f_m = f_2$,= 16 (шт.), и модальный интервал имеет номер i=2.

Интервальный вариационный ряд с постоянным шагом построен: исходная СС в виде множества Y, представленное данными табл. 3, отображено в более компактное множество X в виде табл. 4. Это и есть искомый интервальный ВР с постоянным шагом h = 2 млн. руб.

Далее заполним графу «Средние интервальные значения» и «Значения накопленной частоты» по формуле (1). Интервальный вариационный ряд построен.

Таблица 4

Интервальный вариационный ряд. Результаты отображения вида (1)

Номера Интервалов	Начало – конец интервалов	Среднее интервальное значение	Частота попаданий в интервалы	Значение Частоты	Значения накопленной частоты
I	$x_i^н$ - $x_i^в$	$x_i^{ср}$	счет	f_i	q_i
1	5 - 7	6	/////////	9	9
2	7 - 9	8	////////////////	16	25
3	9 - 11	10	///////////	11	36
4	11 - 13	12	////////	8	44
5 = n	13 - 15	14	//////	6	50
			$\sum f_i$= 50	$\sum f_i$= 50	

2. ОПРЕДЕЛЕНИЕ СТРУКТУРНЫХ СРЕДНИХ ВАРИАЦИОННЫХ РЯДОВ

Определение структурных средних (моды и медианы) можно осуществить двумя способами: графо-аналитическим и аналитическим.

2.1. Определение моды и медианы графо-аналитическим методом

Визуализируем интервальный ВР с постоянным шагом в виде двух графиков, один из которых построим в координатах (x, f) – гистограмму (см. рис. 1), а другой в координатах (x, q) – кумуляту (см. рис. 2).

Сначала построим гистограмму. Студентам рекомендуется при этом использовать место в конспектах так, чтобы его использовать с максимальной наглядностью. Для этого необходимо определиться с ценой деления (графики при этом рекомендуется делать без линейки, от руки). Так, в случае построения гистограммы по горизонтальной оси абсцисс следует сделать так, чтобы весь

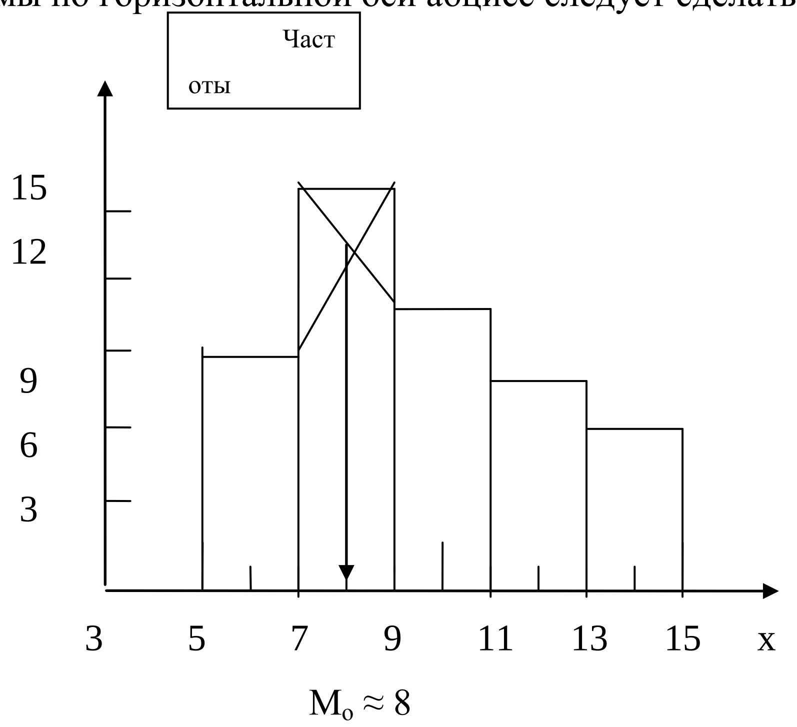

Рис. 1. Гистограмма, отражающая зависимость частот от вариантов

график со всеми вариантами уместился от 5 до 15-ти (млн. руб.), тогда как для отображения частоты попадания в тот или иной вариант на вертикальтной оси ординат достаточно ограничиться числом 15; чуть выше – и 16 будет (см. пятую графу табл. 4).

Значение наиболее часто встречающегося значения в исходной СС, то есть мода M_o, графо-аналитически путем графически в координатах (x, f) определяется так, как это показано на рис. 1 сплошными линиями внутри столбика гистограммы, принадлежащему модальному интервалу (перпендикуляр к оси абцисс строится из пересечения прямых). Здесь мода получается равной где-то 8,0 млн. руб. (как на графике), а на самом деле она должна быть где-то между 8,1 – 8,3 млн. руб. (с учетом специфики рисования графиков в window); в среднем 8,2 млн. руб. как наиболее часто встречающееся значение в исходной статистической совокупности Y. При этом неважно, имеется ли данное значение в исходной совокупности Y или не имеется (здесь – имеется: см. табл. 3), важно, что такая характеристика полученного ВР является *объективной величиной*, пусть даже в исходной совокупности и отсутствующей. В итоге приближенное значение моды $M_o \approx 8{,}1 – 8{,}3$ млн. руб. (в среднем 8,2 млн. руб.)

Значение медианы M_e (млн. руб.) также можно определить графо-аналитическим путем по графику кумуляты в координатах (x, q) – см. рис. 2. График накопленных частот строится так. В начале первого интервала [5 – 7] еще не накоплено ничего (в точке 5 млн. руб. на оси абцисс должен быть нуль), зато к концу первого интервала накопилось 9 попаданий случайной величины – объемов ОФ попавших в этот интервал МП и т.д.

Поскольку из определения медианы следует, что она делит исходную совокупность на две равные части, разделим всю сумму частот $\sum f_i$. Пополам и получим: ($\sum f_i$./ 2) = 50 / 2 = 25 (штук, величина безразмерная). Затем – спроецируем это значение на график кумуляты (по стрелке, параллельной оси абцисс, с точки пересечения проекции и кумуляты опустим к горизонтальной оси перпендикуляр) – и получим приближенное значение медианы так, как это показано на рис. 2. Приближенные значения медианы $M_e \approx 9$ млн. руб.

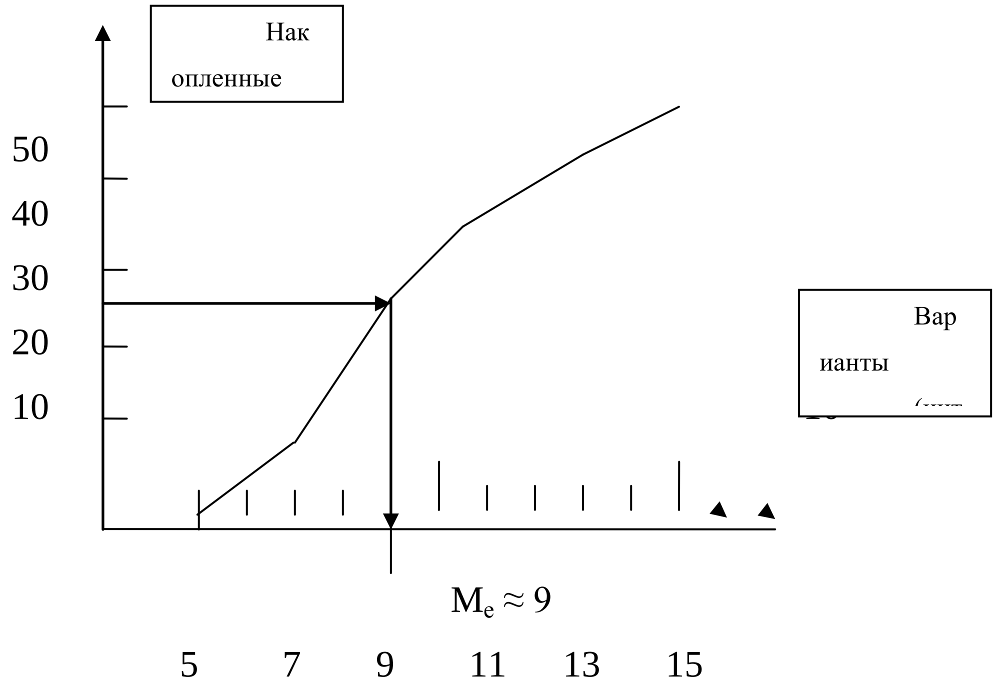

Рис. 2. Кумулята, отражающая зависимость накопленных частот от вариантов

2.2. Определение моды и медианы аналитическим методом

Таким образом, графо-аналитическим путем были определены приближенные значения новых структурных средних построенного нами вариационного ряда, значение которых может быть уточнено строго аналитическим способом по формулам (9) и (10) - соответственно. Начало модального интервала = 7, размер шага = 2, модальная частота f_m= 16. Тогда частота интервала, предшествующему модальному, f_{m-1} = 9, а частота, последующего за модальным интервалом f_{m+1} = 11, q_{m-1} = 9, как и следует из табл. 4.

$$M_o = x_m^{н} + h \cdot \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})}. \qquad (9)$$

По формуле (9) получим:

$$M_o = 7 + 2 \cdot \frac{16 - 9}{(16 - 9) + (16 - 11)} = 8{,}16 \approx 8{,}2 \text{ (млн. руб.).}$$

Значение медианы вычислим по формуле (10):

$$M_e = x_m^{н} + h \cdot \frac{(\sum f_i / 2) - q_{m-1}}{f_m} = \qquad (10)$$

$$= 7 + 2 \cdot \frac{(50 / 2) - 9}{16} = 7 + 2 \cdot \frac{25 - 9}{16} = 9{,}0 \text{ (млн. руб.).}$$

Если сравнить результаты, полученные графо-аналитическим и аналитическим способами, то по части величины моды вместо 8,1 – 8,3 (в среднем 8,2) получили точно 8,2 млн. руб., по части медианы – также точно 9,0 млн. руб. хоть тем, хоть другим способом.

Чтобы лучше уяснить физический смысл медианы, приведем здесь еще раз содержание табл. 3 в виде табл. 5 и выделим шрифтом те значения элементов исходной СС, значения которых будут меньше или равны величине медианы = 9,0 млн. руб.

Таблица 5

Исходная статистическая совокупность Y

9.4	**8.0**	**6.3**	10.0	15.0	**8.2**	**7.3**	9.2	**5.8**	**8.7**
5.2	13.2	**8.1**	**7.5**	11.8	14.6	**8.5**	**7.8**	10.5	**6.0**
5.1	**6.8**	**8.3**	**7.7**	**7.9**	**9.0**	10.1	**8.0**	12.0	14.0
8.2	9.8	13.5	12.4	**5.5**	**7.9**	9.2	10.8	12.1	12.4
12.9	12.6	**6.7**	9.7	**8.3**	10.8	15.0	**7.0**	13.0	9.5

Выделенных шрифтом значений получилось из 50-ти ровно половина – 25 штук, как и следовало ожидать. В этом и есть смысл *медианы*, делящей всю исходную совокупность пополам: половина обследованных МП города имеют капитал до 9 млн. руб. включительно, вторая половина (оставшиеся 25 МП) имеют размер основных фондов более 9 млн. руб.

Попутно заметим, что такую структурную среднюю как медиана, можно так или иначе *увидеть*, то есть убедиться, что она реально существует, тогда как величину моды (наиболее часто встречающегося значения в исходной СС) увидеть можно далеко не всегда. Да, в исходной СС значения отдельных ее элементов 8,2 млн. руб. встречаются даже дважды. Но это – округленное до десятых долей, тогда как количественный показатель 8,16 млн. руб., полученный теоретически, не встречается в исходной таблице ни разу. Однако он объективно существует, независимо от нас, как некоторое обобщение, абстракция. Хотя, конечно, капиталы в 8,2 млн. руб. весьма близки к нему.

После того, как структурные средние – мода и медиана определены нами даже альтернативными методами, необходимо оценить среднее значение полученного ВР.

2.3. Среднее значение вариационного ряда и степень асимметрии

Поскольку каждый вариант согласно табл. 4 проявляет себя с неодинаковыми частотами, применим формулу для вычисления среднего взвешенного вида (4). В качестве вариантов будем использовать их средние значения $x_i^{ср}$:

$$x_{ср}^{в} = \frac{\sum x_i^{ср} \cdot f_i}{\sum f_i} = \frac{6 \cdot 9 + 8 \cdot 16 + 10 \cdot 11 + 12 \cdot 8 + 14 \cdot 6}{50} = \frac{472}{50} = 9{,}44 \text{ (млн. руб.)}.$$

Располагая значениями моды, медианы и среднего взвешенного, можно аналитически оценить степень асимметрии распределения случайной величины – объема основных фондов выбранных для анализа малых предприятий города.

Оценка степени асимметрии распределения определяется по нестрогому неравенству следующего вида:

$$\left| M_o - x_{ср}^{вз} \right| \leq 3 \cdot M_e - x_{ср}^{вз}. \qquad (11)$$

Если нестрогое неравенство вида (11) выполняется, распределение основных фондов (ОФ) МП как случайной величины является умеренно асимметричным, если не выполняется, то асимметрия распределения считается повышенной. Тога, согласно условию (11),

$$\left| 8{,}17 - 9{,}44 \right| \leq 3 \cdot \left| 9{,}00 - 9{,}44 \right|,$$

$$1{,}27 \leq 1{,}32:$$

Нестрогое неравенство вида (11) выполняется. Следовательно. Распределение ОФ МП (в млн. руб.) в исходной статистической совокупности является умеренно-асимметричным.

Попутно заметим, что если бы гистограмма на рис. 1 была бы строго симметричной по отношению к максимальной величины среднему столбцу, то значения моды, медианы и среднего взвешенного совпали бы: $M_o = M_e = x_{ср}^{вз}$. Тогда неравенство (11) выглядело бы так:

$$\left| M_o - x_{ср}^{вз} \right| \leq 3 \cdot M_e - x_{ср}^{вз} \left|. \right.$$

$$0 \leq 0.$$

Выражение (11) выполняется: абсолютная симметрия распределения относительно среднего столбца (не как на рис. 1).

Замечание. Выражение (11) свидетельствует о степени асимметрии как степени «скошенности» распределения вправо или влево от середины графика, вариант которого приведен на рис. 1, однако не дает ответа в какую именно сторону. Но нам на рис. 1 видно, что та самая «скошенность» распределения размеров капитала у обследованных МП города скошена именно влево (см. рис.

1), то есть МП с меньшими объемами ОФ в млн. руб. среди обследованных МП преобладают.

2.4. Особый случай вычисления структурных средних

При группировке исходной информации в виде данной нам статистической совокупности Y мощностью (числом вещественных элементов совокупности) N в более компактную совокупность X мощностью n, в которой n < N, строим обычный вариационный ряд, иначе говоря, отображение Y на X вида (2). Однако может случиться так, что при выполнении очередного этапа – определения модального интервала, - в процессе анализа уже построенного нами ВР может оказаться так, что на роль модального интервала (где частоты максимальны) будут претендовать не один, а, например, два интервала. Возникает естественный со стороны исследователя (студента) вопрос: какому из интервалов отдать предпочтение, какой из одинаковых интервалов выбрать в качестве модального? Продемонстрируем выбор модального интервала (а значит и значение структурных средних – моды и медианы) на следующем примере.

В качестве исходных данных для такого примера пусть будем располагать некоторой статистической совокупностью $Y = \{y_j\}$, $j = 1,N$, где $N = |Y|$ – мощность множества Y. В качестве исходных данных примем объем реализованной продукции десяти малых предприятий за отчетный период в млн. руб. (задача 3):

$$Y=\{5,2; 13,2; 8,1; 7,5; 11,8; 14,6; 8,5; 7,8; 10,5; 6,0). \qquad (12)$$

Оценку структуры исходных данных проведем в терминах вариационных рядов, когда мы хотим получить отображение исходного множества Y мощностью N = 10 шт. единиц совокупности в некоторое более компактное множество X мощностью n. То есть наша цель – построить отображение исходного множества Y на более компактное множество X вида (2): $\tau : Y \rightarrow X$.

Для реализации операции (2) необходимо и достаточно сгруппировать элементы статистической совокупности (1) по известным правилам.

Итак, мощность исходной совокупности N=10; максимальный элемент совокупности y_{max} = 14,6 млн. руб.; y_{min} = 5,2 млн. руб.

Размах выборки согласно (7) $R = y_{max} - y_{min} = 14{,}6 – 5{,}2 = 9{,}4$ (млн. руб.)

Чтобы определить шаг выборки h, согласно выражению (8), необходимо размах выборки R поделить на число интервалов «n»: h = R / n.

Число интервалов n, в свою очередь, можно приближенно найти по формуле Стерджесса (3):

$$n = 1 + 3{,}22\cdot\lg N = 1 + 3{,}22\cdot\lg 10 = 1 + 3{,}22\cdot 1 = 4{,}22 \text{ (шт.)} \approx 4 \text{ (шт.)}.$$

Тогда согласно выражению (8) шаг $h = 9{,}4 / 4 = 2{,}35 \approx 2{,}40 = 2{,}4$ (млн. руб.).

Таким образом, мы должны составить рабочую таблицу (вариационный ряд) для четырех вариантов (интервалов); вместо 10-ти исходных элементов получили 4 варианта. Далее только останется подсчитать, сколько элементов исходной совокупности Y вида (12) попадет в тот или иной интервал, что представлено в табл. 6. Также заполним графы значений середин полученных интервалов и вычислим значения накопленной частоты по выражению (1).

Таблица 6

Интервальный вариационный ряд для исходной совокупности

Y вида (12)

Номера Интервалов	Начало – конец интервалов	Среднее интервальное значение	Частота попаданий в интервалы	Значение частоты	Значения накопленной частоты
I	$x_i^н$ - $x_i^в$	$x_i^{ср}$	счет	f_i	q_i
1	5,2 - 7,6	6,4	///	3	3
2	7,6 - 10,0	8,8	///	3	6
3	10,0 - 12,4	11,2	//	2	8
4 = n	12,4 - 14,8	13,6	//	2	10
			$\sum f_i = 10$	$\sum f_i = 10$	

Вариационный ряд X по исходной информации Y в виде табл. 6 построен. В первую очередь необходимо определить номер модального интервала, на который могут претендовать не один, а два интервала с номерами i = 1 и i = 2: частоты попадания в них элементов исходной совокупности Y максимальны и одинаковы: с $f_{max} = f_1 = 3$ шт. и $f_{max} = f_2 = 3$ шт. На данном этапе оценим точные (аналитические) значения структурных средних (моду и медиану) по выражениям (9) и (10) - соответственно. Вычислим последовательно значения моды и медианы для обоих модальных интервалов.

Итак, модальная частота $f_m = f_2 = 3$. Сначала возьмем в качестве модального второй интервал: $i_m = 2$. Начало модального интервала $x_i^н = 7{,}6$ млн. руб. (см. табл. 6), шаг (размер интервала) h = 2,4 млн. руб. Значение моды по выражению (9):

$$M_o = 7{,}6 + 2{,}4 \cdot \frac{3 - 3}{(3 - 3) + (3 - 2)} = 7{,}6 \text{ (млн. руб.)}.$$

Значение медианы вычислим по формуле (10):

$$M_e = 7{,}6 + 2{,}4 \cdot \frac{(10 / 2) - 3}{3} = 7{,}6 + 2{,}4 \cdot \frac{5 - 3}{3} = 9{,}2 \text{ (млн. руб.)}.$$

Теперь в качестве модального интервала примем первый интервал $i_m = 1$. Начало модального интервала $x_i^н = 5{,}2$ млн. руб. (см. табл. 6), шаг (размер интервала) – тот же h = 2,4 млн. руб.; значение $f_m - f_{m-1} = f_1 - f_{1-1} = f_1 - f_0$, а значения f_0 в табл. 6 нет; значение накопленной частоты $q_{m-1} = q_{1-1} = q_0 = 0$, а такого значения в табл. 6 также нет.

Новое значение моды по формуле (9):

$$M_o = 5{,}2 + 2{,}4 \cdot \frac{3 - 0}{(3 - 0) + (3 - 3)} = 7{,}6 \text{ (млн. руб.) – прежнее значение.}$$

Значение медианы вычислим по формуле (10):

$$M_e = 5,2 + 2,4 \cdot \frac{(10 / 2) - 0}{3} = 5,2 + 2,4 \cdot \frac{5}{3} = 9,2$$ (млн. руб.) – прежнее значение.

Таким образом, на формальном уровне в случае одинаковых значений максимальных частот, соответствующих соседним интервалам, в качестве модального можем выбрать любой из них. Однако тот факт, что значение медианы = 9,2 млн. руб. находится во втором интервале (7,6 – 10.0), то в качестве модального следует считать все-таки второй интервал, а не первый.

Средневзвешенное значение исходной статистической совокупности (12) находится по той же формуле (4):

$$x_в^{ср} = \frac{6,4 \cdot 3 + 8,8 \cdot 3 + 11,2 \cdot 2 + 13,6 \cdot 2}{10} = 9,50 \text{ (млн. руб.)}.$$

Располагая значениями моды, медианы и средневзвешенного, можно оценить степень асимметричности распределения. Если нестрогое неравенство (11) выполняется, распределение является «умеренно асимметричным», ели нет, то «асимметрия повышенная». Для наших расчетов:

$|\ 7,6 - 9,5\ | \le 3 \cdot |\ 9,2 - 9,5\ |$ или

$1,9 \le 3 \cdot 0,3 = 0,9$ – нестрогое неравенство не соблюдается: 0,9 не больше, чем 1,9; все наоборот. Следовательно, структура данных такова, что распределение их по вариантам имеет повышенную асимметрию.

Таким образом, в случае наличия рядом расположенных интервалов с максимальными частотами, для расчета структурных средних в качестве модального можно использовать любой из них. Тогда графо-аналитический метод оценки тех же величин можно проиллюстрировать на рис. 1 и рис. 2.

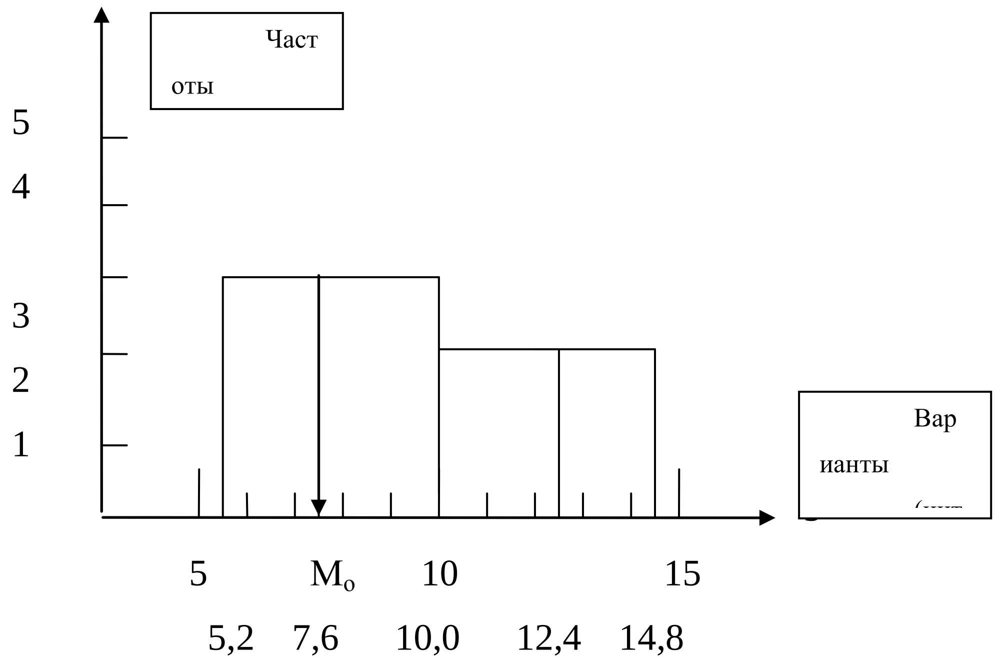

Рис. 3. Гистограмма, отражающая зависимость частот от вариантов

При исходных данных табл. 6 максимальные частоты просматриваются на интервалах с номерами 1 и 2. Значит оба эти варианта можно считать модальными. Тогда медиана займет место строго между ними, что и показано на рис. 4. Здесь значение моды - $M_o \approx 7{,}5$ млн. руб.

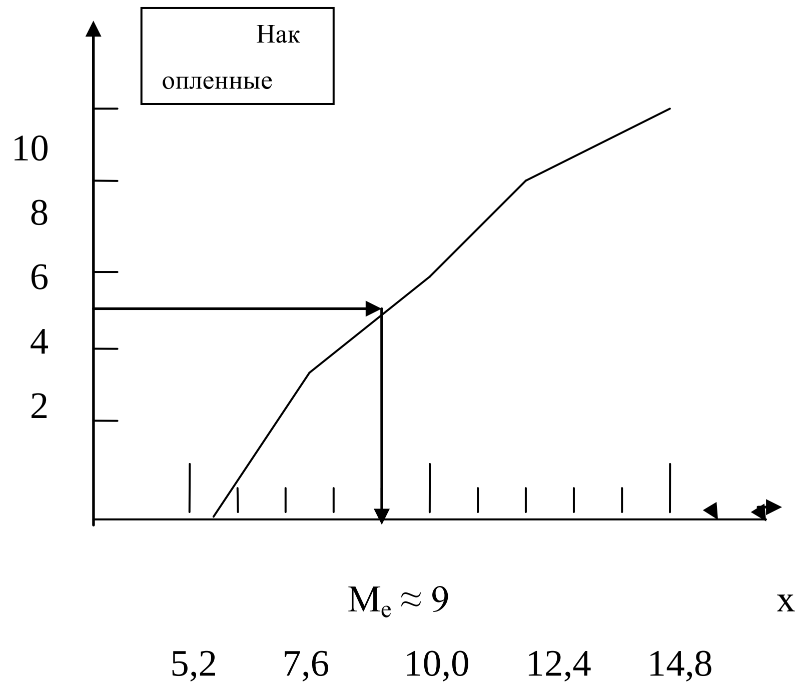

Рис. 4. Кумулята, отражающая зависимость накопленных частот от вариантов

Теперь определим приближенное значение медианы M_e также графо-аналитическим способом (см. рис. 4). Кумулята (накопленная на интервалах частота), в отличие от гистограммы, строится следующим образом. В начале первого интервала ВР (от 5,2 до 7,6 млн. руб.) накопленная частота отсутствует, однако к концу его достигает своего значения $q_1 = 3$ (см. табл. 1); к концу второго интервала $q_2 = 6$; к концу третьего - $q_3 = 8$; к концу четвертого $q_4 = 10$, что и представим на рис. 4.

Поскольку из определения медианы следует то, что она делит исходную совокупность на две равные части, разделим всю сумму частот $\sum f_i$. пополам и получим: $(\sum f_i./ 2) = 10 / 2 = 5$ (штук, величина безразмерная). Затем – спроецируем это значение на график кумуляты (по стрелке, параллельной оси абцисс, с точки пересечения проекции и кумуляты опустим к горизонтальной оси перпендикуляр – и получим приближенное значение медианы так, как это показано на рис. 4.

Значения двух новых параметров – структурных средних, - моды и медианы графо-аналитическим методом нами определены. Точность полученных результатов определяется точностью произведенных графических работ.

Однако здесь необходимо сделать следующее замечание. Если одинаковые максимальные частоты следуют друг за другом на протяжении, скажем 3 и более интервалов, значит расчеты, связанные с определением числа интервалов по выражению (3) и величиной шага (8) необходимо уточнить (в сторону уменьшения числа интервалов n) и составить новую рабочую таблицу (вариационный ряд). Такой итеративный процесс является естественным при работе с созданием вариационных рядов как средства группировки исходной статистической информации.

Заметим также, что если мощность (число элементов) исходной статистической совокупности Y невелико (например, как в последнем примере, всего 10), то построение ВР является средством структурирования исходных данных, которое предоставляет в распоряжение исследователя дополнительную информацию об исследуемом объекте. Если же численность элементов исходной совокупности достаточно велико (несколько десятков и более), то группирование средствами ВР является, пожалуй, единственным средством скомпрессировать (сжать) исходные данные, привести к виду, удобному для последующего анализа плюс дополнительная информация о структуре данных. Поэтому вследствие отдельных особенностей окончательная группировка может быть осуществлена не сразу, но в результате последовательного выполнения нескольких шагов до тех пор, пока полученные результаты будут удовлетворять исследователя. В частности, варьирование величиной шага h, может случиться так, что его постоянной величины придерживаться нецелесообразно. Тогда выражение (8) может быть использовано как начальное, приблизительное.

Однако в любом случае может возникнуть вполне справедливый вопрос относительно элементов исходной статистической совокупности Y: достаточно ли она статистически однородна? Актуальность такой постановки вопроса связана с тем, что если для анализа мы выберем более или менее однородные элементы (в задаче 2 размер основных фондов МП варьировался в пределах от 5 до 15 млн. руб.), то такая статистическая совокупность будет характеризовать определенную группу МП. Если же выберем МП с капиталом 20 – 50 млн. руб. – это будет уже другая группа, качественно отличающаяся от первой. Если же среди МП, например, из размера основных фондов сформировать выборку, в основном, от 5 до 15 млн. руб. и внести в нее, скажем, два МП с капиталом 50 и 70 млн. руб., то впоследствии будем дело иметь явно с неоднородной выборкой, и полученные результаты анализа вряд ли смогут претендовать на типичные для случайно сформированной выборки.

Поэтому при анализе полученного вариационного ряда целесообразно рассчитать коэффициент вариации и сравнения его с пороговым значением.

3. ОЦЕНКА ВАРИАБЕЛЬНОСТИ ИСХОДНОЙ СТАТИСТИЧЕСКОЙ СОВОКУПНОСТИ

Степень вариабельности оценивается в статистике с помощью коэффициента вариации v как отношение среднего квадратического отклонения элементов статсовокупности σ из выражения (14) к их среднему значению $x_{ср}^{вз}$ из выражения (4). Так как мощность исходной СС в задаче 2 достаточно велик (50 единиц), воспользуемся возможностями созданного нами интервального вариационного ряда (вместо 50-ти – всего 5 единиц вариантов), приведенного в виде табл. 4.

Коэффициент вариации вычисляется так (D – дисперсия анализируемого признака):

$$v = \frac{\sigma}{x_{ср}^{вз}}, \qquad (13)$$

$$\sigma = (D)^{½}, \qquad (14)$$

$$D = \frac{\sum (x_i^{ср} - x_{ср}^{вз})^2 \cdot f_i}{\sum f_i}. \qquad (15)$$

Вычисляем дисперсию D из выражения (15) и среднее квадратическое отклонение σ из выражения (14):

$$D = \frac{\sum(x_i^{ср} - x_{ср}^{в})^2 \cdot f_i}{\sum f_i} = \frac{(6–9,44)^2 9+(8-9,44)^2 16+\ldots+(12-9,44)^2 8+(14-9,44)^2 6}{50} =$$

$$= \frac{320,32}{50} = 6,406 \text{ (млн. руб.}^2\text{)}.$$

$$\sigma = (6,406)^{1/2} = 2,53 \text{ (млн. руб.)}.$$

Тогда коэффициент вариации из выражения (13):

$$v = \frac{\sigma}{x_{ср}^{в}} = \frac{2,53}{9,44} = 0,268$$ или 26,8%, что меньше порогового значения, равного 30%.

Следовательно, степень вариабельности исходных элементов в первичной статистической совокупности является весьма умеренной, и подбор для анализа МП был достаточно корректен с позиции их статистической соизмеримости. Поставленная задача решена.

Примечание: если бы значение взвешенного коэффициента превышали бы пороговое значение 30% (в иных источниках 33%), то мы бы сократили объем исходной выборки, выведя из рассмотрения слишком маленькие и слишком большие значения, например, 5,1 млн. руб. и 15,0 млн. руб. Тогда и N уменьшился бы на две единицы: вместо 50-ти работали бы с 48-ю элементами исходной статистической совокупности. Тогда, что очевидно, сократился бы и размах выборки R.

4. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ЗАЩИТЕ РЕЗУЛЬТАТОВ РАБОТЫ

После ознакомления с приведенным здесь материалам предполагается выполнение самостоятельной работы 1, в рамках которой необходимо из 15-ти элементов исходной статистической совокупности построить интервальный вариационный ряд с равными интервалами и определить его параметры, что в основном и составляет содержание защиты результатов работы. Студент должен сообщить:

1) содержание, цели задачи;

2) число интервалов, размах выборки, величину шага;

3) номер модального интервала;

4) значение модальной частоты (частот);

5) точное значение моды и ее значение, полученное графо-аналитическим путем;

6) точное значение медианы и ее значение, полученное графо-аналитическим путем;

7) значение среднего взвешенного;

8) оценить степень асимметрии распределения (умеренная, повышенная асимметрия);

9) оценить степень вариабельности элементов исходной статистической совокупности (сравнить с ее пороговым значением 0,30 или 30%; если меньше – вариабельность умеренная, если превосходит пороговое значение – вариабельность повышенная).

5. ВАРИАНТЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ 1

Вариант 1

Для построения интервального вариационного ряда с равными интервалами воспользуемся следующим данными о стоимости основных фондов (непрерывный признак) у 15-ти малых предприятий (млн. руб.).

9.4 8.0 6.3 10.0 15.0 8.2 7.3 9.2 5.8 8.7 5.0 13.2 8.1 7.5 11.8

Необходимо сгруппировать исходные данные методом построения вариационного ряда и оценить структуру данных по следующим параметрам: найти моду и медиану графо-аналитическим и аналитическим путем (сравнить степень точности определения), среднее взвешенное, степень асимметрии распределения, среднее квадратическое отклонение (с.к.о.) и степень вариабельности исходных данных.

(*Задание для всех вариантов общее; далее приводятся исходные статистические совокупности*)

Вариант 2

11.4 10.0 6.3 10.0 15.0 8.2 7.3 9.2 5.8 8.7 5.0 13.2 8.1 7.5 11.8

Вариант 3

11.4 10.0 7.3 12.0 15.0 8.2 7.3 9.2 5.8 8.7 5.0 13.2 8.1 7.5 11.8

Вариант 4

11.4 10.0 7.3 12.0 15.0 10.2 9.3 9.2 5.8 8.7 5.0 13.2 8.1 7.5 11.8

Вариант 5

11.4 10.0 7.3 12.0 15.0 10.2 9.3 11.2 7.8 8.7 5.0 13.2 8.1 7.5 11.8

Вариант 6

11.4 10.0 7.3 12.0 15.0 10.2 9.3 11.2 7.8 10.7 5.0 10.2 8.1 7.5 11.8

Вариант 7

11.4 10.0 7.3 12.0 15.0 10.2 9.3 11.2 7.8 10.7 5.0 10.2 10.1 9.5 11.8

Вариант 8

7.4 7.0 7.3 12.0 15.0 10.2 9.3 11.2 7.8 10.7 5.0 10.2 10.1 9.5 11.8

Вариант 9

7.4 7.0 5.3 10.0 15.0 10.2 9.3 11.2 7.8 10.7 5.0 10.2 10.1 9.5 11.8

Вариант 10

7.4 7.0 5.3 10.0 15.0 8.2 7.3 11.2 7.8 10.7 5.0 10.2 10.1 9.5 11.8

Вариант 11

7.4 7.0 5.3 10.0 15.0 8.2 7.3 13.2 9.8 10.7 5.0 10.2 10.1 9.5 11.8

Вариант 12

7.4 7.0 5.3 10.0 15.0 8.2 7.3 13.2 9.8 8.7 5.0 8.2 10.1 9.5 11.8

Вариант 13

7.4 7.0 5.3 10.0 15.0 8.2 7.3 13.2 9.8 8.7 5.0 8.2 13.1 12.5 11.8

Вариант 14

5.4 5.0 5.3 10.0 15.0 8.2 7.3 13.2 9.8 8.7 5.0 8.2 13.1 12.5 11.8

Вариант 15

5.4 5.0 7.3 12.0 15.0 8.2 7.3 13.2 9.8 8.7 5.0 8.2 13.1 12.5 11.8

Вариант 16

5.4 5.0 7.3 12.0 15.0 9.2 7.3 13.2 9.8 8.7 5.0 8.2 13.1 12.5 11.8

Вариант 17

5.4 5.0 7.3 12.0 15.0 9.2 7.3 12.2 7.8 8.7 5.0 8.2 13.1 12.5 11.8

Вариант 18

5.4 5.0 9.3 10.0 15.0 9.2 7.3 12.2 7.8 8.7 5.0 8.2 13.1 12.5 11.8

Вариант 19

5.4 5.0 5.3 10.0 15.0 9.2 7.3 12.2 7.8 8.7 5.0 8.2 13.1 12.5 11.8

Вариант 20

5.4 5.0 5.3 8.0 15.0 9.2 7.3 12.2 7.8 8.7 5.0 8.2 13.1 12.5 11.8

ЛИТЕРАТУРА

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

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

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

Учебное издание

Шихалёв Анатолий Михайллович

ГРУППИРОВКА НАБЛЮДЕНИЙ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ