И. С. КАМЕНЕЦКИЙ, А. А. УЗЯНОВ

О ПРАВИЛАХ ПОСТРОЕНИЯ ГИСТОГРАММ

Гистограммы все чаще и чаще появляются на страницах археологических статей. Еще сравнительно недавно они использовались преимущественно в работах нумизматических, исследующих разультаты различных технических анализов или обрабатывающих остеологические наблюдения. Теперь они уже проникли и в собственно археологические работы, прежде всего по археологии каменного века и, пока в меньшей мере, в другие ее разделы. При этом довольно часто допускаются ошибки в построении гистограмм, сводящие на нет труд, затраченный на их построение, или приводящие к ошибочным выводам. Это обстоятельство и побудило авторов данной статьи специально заняться этим вопросом и попытаться в предельно доступной форме изложить основные правила построения гистограмм.

Мы упомянули в заголовке статьи «гистограммы», так как этот термин уже в какой-то мере стал привычным для археологов. В действительности же речь пойдет о вариационном ряде, графическим выражением которого являются гистограмма, график и полигон частот.

Начать следует с терминологии. Существуют определенные различия при построении вариационного ряда для количественных и качественных признаков. Некоторые особенности свойственны также количественным непрерывным и количественным дискретным признакам. Вначале речь пойдет о количественных непрерывных, которые известны также под названием мерных признаков, или параметров.

Возьмем для примера диаметр венчика горшков (Дᴀ). Замерив какое-то их количество, мы получим набор значений (вариант или варианты наблюдений) данного признака, отличающихся друг от друга. Эти отличия характеризуют вариабельность признаков и образуют вариационный ряд. Если мы расположим все полученные значения от наименьшего к наибольшему (или наоборот), то получим ранжированный, или упорядоченный, ряд. Если значений немного (до 20), то все они могут быть выписаны в строчку. Если же мы имеем дело с большим числом значений, то такой способ подачи данных оказывается нерентабельным, ибо они занимают много места, а главное, плохо воспринимаются. Для того, чтобы сделать ряд «удобочитаемым» и легко обозримым, следует сгруппировать близкие значения. Допустим, мы определим границы групп через 1 *см*: 10—10,9 *см*; 11—11,9 *см* и т. д. Эти группы называются интервалами (или разрядами, классами, классовыми интервалами, класс-интервалами). Выписав в ряд интервалы, мы должны указать также число значений, входящих в каждый из интервалов, как это показано в табл. 1.

Таблица 1

Значения диаметра венчиков (D_A) краснолаковых глубоких мисок с закругленным краем (тип 1), *см*

Интервалы	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Частоты	1	1	—	3	5	11	1	16	2	3	—	—	—	1

В данной таблице 44 значения ($n=44$) разделены на 14 групп, т. е. по 14 интервалам. Причем в качестве величины интервала взят 1 *см*. Количество значений, приходящихся на каждый интервал, называется его частотой. Если те же данные выразить в процентах или долях единицы, то они будут называться частостями (или долями). В таком виде ряд называется группированным (или взвешенным) вариационным рядом. Если к этому добавим, что разность между максимальным значением признака (X_{max}) и минимальным значением (X_{min}) называется размахом колебаний, то этим исчерпаем всю необходимую нам терминологию.

Гистограмма, соответствующая вариационному ряду, данному в табл. 1, строится просто. На горизонтальной оси откладываем интервалы, на вертикальной — частоты. Частота каждого интервала изображается в виде прямоугольника, основание которого равно интервалу, а высота частоте (рис. 1-*1*). Аналогично строится полигон частот и график (рис. 1-*2,3*). Каждый из них имеет некоторые особенности, которые для нас, однако, сейчас несущественны.

Построение группированного ряда может иметь две цели. Во-первых, он облегчает вычисление различных характеристик ряда, таких как арифметическая средняя, стандартное квадратическое отклонение (сигма) и пр. Например, для вычисления средней арифметической мы не должны складывать все 44 значения, а можем сложить только произведения значений интервалов на частоты, т. е. $13+14+(16\times3)+$ $+(17\times5)+... +26$. Для вычисления приближенных (а такие способы являются приближенными) характеристик ряда существуют определенные правила, которых мы здесь не будем касаться. Они дают значительный выигрыш в вычислениях при большом количестве наблюдений. Для этих целей удобнее строить цифровой ряд, подобный изображенному на табл. 1. Во-вторых, группированный ряд строится с целью выяснения характера распределения наших данных. В этих случаях удобнее избрать один из графических способов. Именно для этого обычно и строят гистограммы археологи. При этом археологи (по крайней мере,

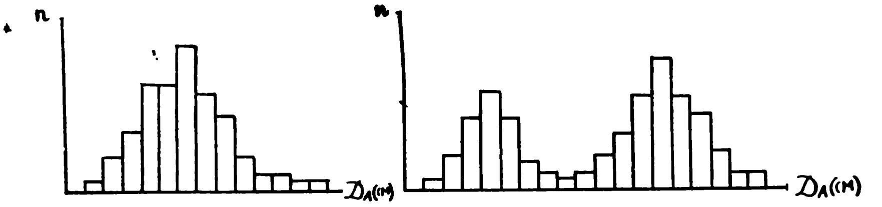
Рис. 1.

на данном этапе), говоря о типе распределения, не имеют в виду статистико-математическое значение этого термина, т. е. речь идет не о том, к какому типу — нормальному, биномиальному, пуассоновскому и т. д.— отнести полученное распределение, а хотят выяснить однородна или неоднородна исследуемая серия предметов.

Допустим, мы получили одновершинное распределение (рис. 2-*1*).

Тогда археологи говорят, что серия однородна (хотя это не совсем точно), что значения колеблются около некоей средней, которая являлась мысленным эталоном древних мастеров и отклонения от которой случайны. Если мы получим двувершинное распределение (рис. 2-*2*), то мы вправе утверждать, что замеренная нами серия неоднородна, что мы имеем, как минимум, дело с двумя различными типами, каждый из которых имел свой эталон. С этой точки зрения, распределение, представленное на рис. 1, будет многовершинным.

Ряд табл. 1 составлен так, как его очень часто составляют: за величину интервала принята точность измерения. Диаметры обломков измерены на кругах, проведенных через 0,5 *см*, и соответственно диаметр замеряемого обломка округляется с точностью до 1 *см*. Очевидно, что это величина произвольная. Если бы мы



1. Одновершинное распределение 2. Двувершинное распределение

Рис. 2.

имели дело с целыми сосудами, то могли бы измерять с точностью до 0,1 *см*. При наличии соответствующих инструментов можно измерять какие-то предметы, например бусы, с точностью до 0,01 *см* и т. д. Таким образом, точность измерения не связана прямо с величиной интервала и не может использоваться для его определения. Поэтому предлагаем такой опыт. Для того же ряда табл. *1* определим интервал в 2 *см* (табл. 2).

Таблица 2

Значения диаметра венчиков (D_A) краснолаковых мисок с закругленным краем (тип 1), *см*

Интервалы	13—14	15—16	17—18	19—20	21—22	23—24	25—26
Частоты	2	3	16	17	5	—	1

Если мы теперь сравним ряды табл. 1 и 2 или соответствующие им гистограммы (рис. 1-*1* и 3), то увидим, что вместо многовершинного первоначального распределения получилось одновершинное (одиночным значением в интервале 25—26 *см* можно пренебречь). Опираясь на первый вариант ряда, мы могли бы говорить о неоднородности измеренной серии сосудов, о том, что тут смешаны как минимум три типа (или размеренных варианта), что и соответствует трем основным вершинам на рис. 1. Опираясь на второй вариант ряда, мы этого сказать уже не можем, ибо имеем одновершинное распределение, которое не дает нам никаких указаний на разнородность материала, более того, предполагает его однородность. Совершенно очевидно, что один из этих вариантов ошибочен. А так как вывод основывается на избранной величине интервала, то очевидно, что

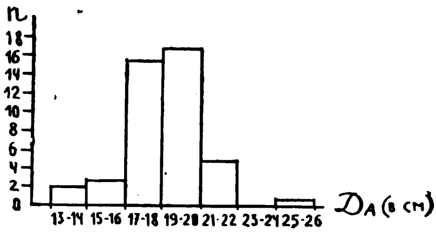


Рис. 3.

в одном (как минимум) случае эта величина избрана неверно. Очевидно также, что, произвольно увеличивая величину интервала, мы можем любое распределение свести к одновершинному, так как мы имеем возможность вместить все имеющиеся значения в два или даже в один интервал. Если объединить все значения гистограммы, изображенной на рис. *2—2*, в три интервала, то получим частоты 53—58—2, т. е. одновершинное распределение. Если бы мы пошли в противоположном направлении и начали уменьшать величину интервала, то опять бы пришли в конце концов к абсурду, так как на каждый интервал у нас приходилось бы одно-два значения, которые в большинстве случаев образовывали бы отдельные вершины. Уже из приведенного ясно, что ряд не должен быть и чрезмерно сжат, ибо в этом случае сглаживаются возможные особенности распределения, и чрезмерно растянут, ибо, приближаясь к негруппированному ряду, он теряет наглядность, читаемость. Значит, суть вопроса в том, чтобы определить оптимальный вариант группировки. Для этого следует уяснить, с какими величинами связана группировка ряда.

В первую очередь надо указать на размах колебаний ($R = X_{max} - X_{min}$). Связь эта не требует доказательства и определяется формулой $C = \frac{R}{k}$ (1), где C — величина интервала, а k — число интервалов. Используется также вариант этой формулы $C = \frac{R}{k-1}$ (2), что ведет к некоторому укрупнению интервалов. Таким образом, чем шире размах колебаний, тем больше величина интервала. В этом рассуждении число интервалов выступает как величина заданная. Действительно, оно не зависит от размаха колебаний. Но оно прямо связано с другой величиной — с числом наблюдений, с числом значений. Эта связь не кажется столь очевидной, но, несомненно, ее легко показать при помощи простых логических рассуждений. Мы говорили, что если уменьшать величину интервалов и, следовательно, увеличивать их число, то картина распределения становится ненаглядной, так как в каждый интервал попадает слишком малое число значений. Однако если одновременно увеличивать и число значений, то при известном условии этого можно избежать. Становится ясно, что чем больше наблюдений, тем на большее количество интервалов мы можем и должны их разделить. Или, что то же самое, чем больше число наблюдений, тем меньше может быть величина интервала. Таким образом, величина интервала зависит от двух величин: размаха колебаний и числа наблюдений.

Все сказанное, как мы надеемся, позволяет уяснить суть проблемы, однако не дает никаких конкретных указаний о способе определения величины интервала. Для этого обратимся к специальной литературе. В части руководств не содержится никаких конкретных указаний, и авторы ограничиваются только общими замечаниями типа приведенных выше[1]. В ряде работ указывается «оптимальное» число интервалов, иногда в зависимости от числа наблюдений. Называются такие числа: от 10 до 20 интервалов[2] при числе наблюдений порядка 200—300[3]; не более 20[4]; от 7 до 15 интервалов[5]; от 8 до 20 интервалов при сотне наблюдений и больше[6]; от 15 до 25 интервалов[7]; 6—9 (5—9) интервалов при числе наблюдений менее 100 и 9—12 (или 9—13) интервалов при числе наблюдений более 100[8]. Эти рекомендации иногда сопровождаются указанием на их необязательность и даже невозможность каких-либо рецептов[9]. Авторы приведенных рекомендаций обычно исходят из того, что группировка ряда преследует только одну цель — удобство вычисления характеристик ряда. В других случаях имеются в виду социологические или демографические исследования, где материал чаще всего сам диктует число интервалов. Наконец, исходят из теории нормального распределения. В последнем случае рекомендуется брать 12 интервалов[10] или дается указание, что частота интервала не должна превышать 15—20% от общего числа значений[11]. Задача выяснения характера распределения, основная для археологов, в этих случаях не ставится совсем или является второстепенной.

Задача эта является актуальной и для биологов. Поэтому в различных руководствах по биометрии мы находим более точные указания в виде таблиц зависимости числа интервалов от числа значений[12]. Эти таблицы дают опять же приблизительные указания и весьма сходны между собой (см. табл. 3, где n — число наблюдений, а k — число интервалов).

Таблица 3

Соотношение числа значений и числа интервалов для составления группированного вариационного ряда (у разных авторов)

Е. К. Меркурьева и Г. Н. Зайцев		В. Г. Вольф		П. Ф. Рокицкий	
n	k	n	k	n	k
—	—	—	—	25—40	5—6
40—60	6—8	30—60	6—7	40—60	6—8
60—100	7—10	60—100	7—8	60—100	7—10
100—200	9—12	100—200	8—15	100—200	8—12
200—500 и более	12—17	200—500	15—25	более 200	10—15

Таблицы Е. К. Меркурьевой и П. Ф. Рокицкого практически одинаковы (следует думать, что в основе приведенных таблиц лежат какие-то расчеты, но они, к сожалению, не приведены). Наконец, в литературе имеются формулы для точного определения числа интервалов в зависимости от числа наблюдений, что представляет для нас наибольший интерес. Правда, формулы эти предназначены для определения величины интервалов. Нам известны четыре формулы для определения величины интервала. Первая предложена Г. А. Стерджесом[13] в 1926 г.:

$$C = \frac{X_{max} - X_{min}}{1 + 3{,}322 \lg n} \qquad (3)$$

Вторая формула предложена в 1962 г. К. Бруксом и Н. Карузерсом[14]:

$$C = \frac{X_{max} - X_{min}}{5 \lg n} \qquad (4)$$

Три следующие формулы заимствованы нами из работы Г. Н. Зайцева [15]:

$$C = \frac{(X_{max} - X_{min}) \lg 2}{\lg n}, \tag{5}$$

где $\lg 2 = 0,30102$ (напомним, что lg — это десятичный логарифм).

$$C = \frac{(X_{max} - X_{min}) \ln 2}{\ln n}, \tag{6}$$

где $\ln 2 = 0,69314$ (ln — это натуральный логарифм).

$$C = \frac{X_{max} - X_{min}}{\sqrt{n} - 1} \tag{7}$$

Формулы (5) и (6) совершенно тождественны, и выбор одной из них зависит только от того, какими таблицами располагает исследователь (или какими логарифмами ему удобно пользоваться).

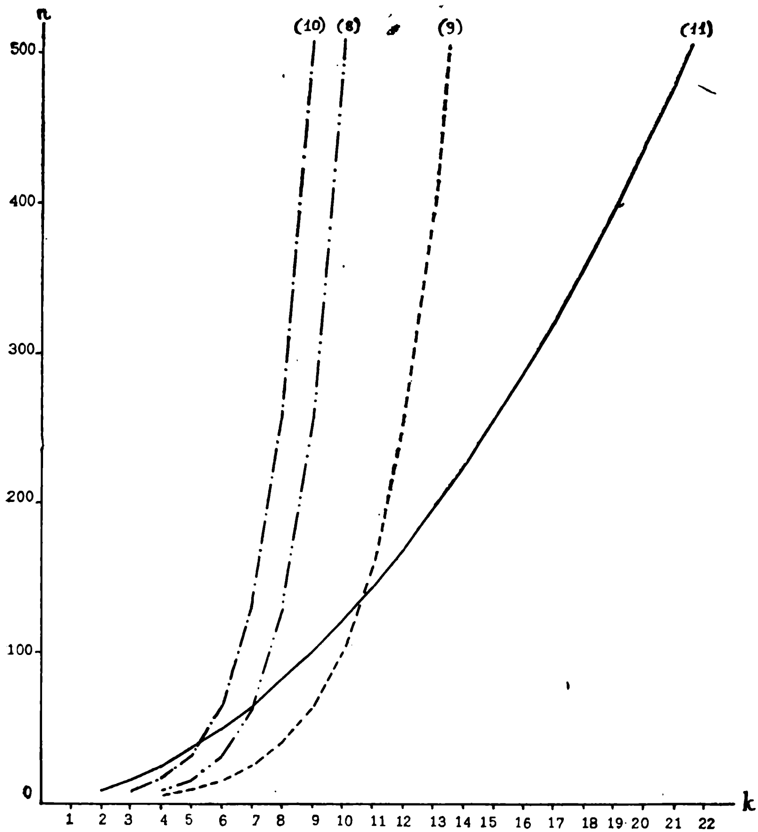


Рис. 4.

Используя формулу (1) легко получить, подставляя в нее значения С, формулы для определения числа интервалов (k). Напомним при этом, что $R = X_{max} - X_{min}$. Соответственно получаем четыре формулы:

$$k = 1 + 3{,}322 \cdot \lg n \quad (8)$$

$$k = 5 \lg n \quad (9)$$

$$k = \frac{\ln n}{\ln 2} = \frac{\lg n}{\lg 2} \quad (10)$$

$$k = \sqrt{n} - 1 \quad (11)$$

Формулы эти имеют разное обоснование и дают соответственно отличающиеся результаты. В табл. 4 и на рис. 4 эти результаты представлены.

По формулам (8), (9) и (10) расчеты произведены для числа наблюдений до 10 000, что, как нам представляется, значительно перекрывает потребности археологов. По формуле (11) расчеты даны до

Таблица 4

Соотношение числа значений и числа интервалов для составления группированного вариационного ряда (по разным формулам)

К	Число наблюдений по формулам			
	(8)	(9)	(10)	(11)
4	6—11	6—7	12—22	21—30
5	12—22	8—12	23—45	31—42
6	23—45	13—19	46—90	43—56
7	46—90	20—31	91—180	57—72
8	91—180	32—50	181—361	73—90
9	181—361	51—79	362—723	91—110
10	362—724	80—126	724—1447	111—132
11	725—1447	127—199	1448—2895	133—156
12	1448—2895	200—316	2896—5791	157—183
13	2896—5796	317—501	5792—11585	184—211
14	5797—11585	502—794		212—241
15		795—1259		242—273
16		1260—1995		274—307
17		1996—3162		308—343
18		3163—5012		344—382
19		5013—7944		383—422
20		7945—12590		423—464
21				465—507
22				508—554
23				555—602
24				603—652
25				653—704
26				705—758
27				759—815
28				816—873
29				874—933
30				934—995
31				996—1060

1000 наблюдений, что в большинстве случаев также вполне достаточно. Отметим, что при 10 000 наблюдений мы имели бы по формуле (11) количество интервалов 99. Из приведенных данных очевидно сходство результатов, получаемых по формулам (8) и (10). Они дают одинаковые рекомендации, но число интервалов отличается на 1. Формулы (8) и (10) можно признать тождественными, так как $3{,}322 \lg n = \lg n : \lg 2$.

При небольшом числе наблюдений ($n < 100$) результаты, полученные по всем формулам, отличаются не очень значительно. С увеличением же числа наблюдений расхождение быстро увеличивается, достигая уже при $n \approx 500$ различия в два с лишним раза (рис. 4).

По графику (рис. 4) хорошо видно, что кривая, построенная по формуле (9), имеет тот же характер, что и кривые, построенные по формулам (8) и (10), но сдвинута вправо. В основе этих формул лежит утверждение[16], что $k \leqslant 5 \lg n$. В формуле (9) взят крайний случай, когда $k = 5 \lg n$. В формулах же (8) и (10) взят случай, когда $k < 5 \lg n$, а именно $k = 3{,}322 \lg n$. Таким образом, формула (9) дает предельное число интервалов.

Если мы сравним теперь результаты, полученные по формулам (табл. 4) с рекомендациями таблиц (табл. 3), то легко увидим, что таблицы Е. К. Меркурьевой и П. Ф. Рокицкого при числе наблюдений менее 100 близки[17] с данными формулы (8), а при $n \geqslant 100$ с данными формулы (9). Таблица же В. Г. Вольфа довольно искусственно сочетает данные формулы (8) при $n \leqslant 100$ и формулы (11) при $n > 100$. Говоря об искусственности, мы имеем в виду, что если данные двух первых таблиц на графике дают плавную кривую того же характера, что и кривые формул (8), (9) и (10), то данные таблицы В. Г. Вольфа такой кривой не образуют, а дают скачкообразный переход между двумя отрезками кривой.

Мы думаем, что появление таблиц и формулы (11), а частично и формулы (9) связано с тем, что исследователи часто интуитивно и эмпирически ощущали неудовлетворительность формул (8) и (10) при значительном числе наблюдений. Действительно, для того чтобы получить один дополнительный интервал, приходится увеличивать число наблюдений в два раза. Это становится уже затруднительным, часто невозможным после 9—10 интервалов, а при использовании формулы (9) примерно после 12 интервала. По этим формулам удобно проводить вычисление характеристик ряда. Проверять же по ним однородность выборок, а тем более устанавливать границу между различными ее составными (если мы имеем неоднородную выборку, представленную двувершинным или многовершинным распределением) затруднительно. Причем всегда желательное увеличение числа наблюдений здесь только мешает, ибо сглаживает кривую распределения.

Мы не знаем, на каком допущении основана формула (11), но даваемые ею рекомендации по соотношению n и k представляются наиболее удобными при проверке однородности выборки, так как в этом случае мы при увеличении наблюдений получаем ощутимый результат в виде увеличения числа интервалов, что дает возможность проведения весьма тонких наблюдений.

Установление зависимости между числом и величиной интервалов, размахом колебаний и числом наблюдений подводит нас и к вопросу о мере точности измерения. По существу, величина интервала — это и есть та точность, которая необходима при обработке данной конкретной коллекции, минимально допустимая точность. Естественно, что точность измерения желательно иметь несколько более высокую, чем минимальная. Допустим, что интервалы у нас равны 2 *см*. В этом случае измерение с точностью до 1 *см* вполне достаточно. Однако это рассуждение действительно только в том случае, когда мы уверены, что никогда в будущем объем нашей коллекции не увеличится. Практически эта ситуация, по-видимому, нереальна. При увеличении коллекции (допустим, числа отщепов определенного типа) величина интервалов, как мы уже знаем, уменьшится, следовательно, должна быть повышена точность замеров. Чтобы избежать повторения замеров всей серии, надо с самого начала предусмотреть такую точность измерений, чтобы она оказалась достаточной и на случай увеличения материала. Это уже отдельная тема. Приведенными рассуждениями мы только хотели показать связь этих вопросов.

Но один аспект этой темы необходимо рассмотреть. Иногда возникает ситуация, когда степень точности наших замеров оказывается ниже, чем величина интервала. Допустим, при замерах диаметров венчиков на «кругах», проведенных через 1 *см*, мы получаем замеры диаметра [18] с точностью до 2 *см*. В то же время число наблюдений позволяет нам установить величину интервала в 1 *см*. Очевидно, что, построив гистограмму с такими интервалами, мы получим вершину на каждом втором интервале, так как среди полученных нами значений будут отсутствовать все четные (или нечетные). В таких случаях точность измерений диктует нам величину интервала, ставит ограничения для ее уменьшения. Величина интервала не может быть меньше, чем принятая точность измерения. Но так как величина интервала связана с числом наблюдений, то мы, пользуясь приведенными формулами, можем определить максимальное число наблюдений, необходимое для построения правильной гистограммы.

В начале статьи мы упоминали о том, что признаки количественные, дискретные и признаки качественные имеют свои особенности, сказывающиеся на построении группированного ряда. Суть в том, что эти признаки не могут быть группированы произвольно, т. е. границы групп в них заданы и не могут быть изменены. Например, если мы строим гистограмму для какого-то могильника по признаку возраста и имеем варианты этого признака: дети, подростки, юноши и т. д.,— то число групп у нас уже задано (5 или 6). Тут мы уже ничего не можем изменить, и число интервалов у нас остается постоянным, вне зависимости от количества наблюдений. Встречаются случаи, когда некоторая группировка возможна и даже необходима и здесь. Допустим, строится гистограмма, отражающая набор кремневых изделий (вариант — коммулятивные графики). При небольшом числе находок следует объединить их в более крупные группы. Например, рассматривать типы резцов не отдельно, а объединить в один интервал все резцы и т. д. Когда речь идет о дискретных количественных признаках, то при небольшом числе вариантов признака число интервалов определяется числом этих вариантов, как и в случае с качественными признаками. Когда же вариантов много, то с ними поступают, как с мерными признаками, группируя их по изложенным правилам. Например, если речь идет о количестве ручек у сосудов, то число интервалов будет соответствовать максимальному числу ручек, встречающихся на одном сосуде. Если же речь будет идти о количестве бус в погребении, то мы будем группировать полученные данные так же, как в случае с мерными признаками. Одним словом, при работе с качественными и количественными признаками никогда не следует забывать о смысловой нагрузке, которую несут рассматриваемые варианты этих признаков. Описанные выше правила в этом случае играют подсобную, ориентировочную роль.

На этом мы считаем возможным закончить изложение правил составления группированного ряда или соответствующей ему гистограммы. Считаем необходимым в заключение дать алгоритм построения такого ряда. Он выглядит так.

1. Производим подсчет числа значений (наблюдений) — n.

2. Находим максимальное (X_{max}) и минимальное (X_{min}) значения и определяем размах колебаний ($R = X_{max} - X_{min}$).

3. По табл. 4 определяем число интервалов k, необходимое для данного числа наблюдений. При этом, если мы проводим группировку для выяснения однородности выборки, то пользуемся данными, полученными по формуле (11), если же имеем дело с однородной выборкой и проводим группировку для вычислений характеристик ряда, то пользуемся данными одной из трех формул (8), (9) или (10).

4. Определяем величину интервала. При этом если мы для определения числа интервалов пользовались формулой (8), то теперь используем формулу (2); если формулой (11) — обращаемся к формуле (1). При использовании на предыдущем этапе формул (9) и (10) мы можем в равной мере выбирать формулу (1) или (2).

5. Если интервал не представлен круглой цифрой, то полученное значение округляется, желательно в большую сторону. Например, если мы получаем $C = 1{,}76$ *см*, то округляем до 2 *см* (или до 1,8 *см*, в зависимости от условий задачи).

6. Существует два способа совмещения крайнего левого интервала с минимальным значением (X_{min}): когда X_{min} приходится на середину этого интервала и когда X_{min} определяет начало этого интервала. Этим, собственно, и объясняется наличие двух формул — (1) и (2). Если мы ранее воспользовались формулой (1), то следует X_{min} совместить с краем интервала, если формулой (2), то следует X_{min} поместить в середину первого интервала. Если мы группируем ряд для вычисления его характеристик, то желательно, чтобы середины интервалов приходились на круглые цифры. Это значительно облегчает последующие вычисления.

Теперь, когда изложены все правила, вернемся еще раз к рядам, представленным в табл. 1 и 2, и определим, какой из них составлен верно. Мы знаем, что $n = 44$, а $R = 13$. По формуле (11) определяем, что k=6. По формуле (1) определяем, что C=2,16...≈2 *см*. Следовательно, группировка табл. 2 правильна или близка к правильной (если округлять в большую сторону). Так как группировка табл. 2 (рис. 3) дает одновершинное распределение, то при C=3 мы тем более будем иметь таковое. Таким образом, предположение о неоднородности выборки, которое могло бы возникнуть на основании группировки ряда в табл. 1 (рис. 1), не подтверждается.

ПРИМЕЧАНИЯ

[1] Р. А. Фишер. Статистический метод исследований. М., 1934, с. 29; Б. А. Ван дер Варден. Математическая статистика. М., 1960, с. 85; Д. Д. Финни. Применение статистики в опытном деле (сельское хозяйство). М., 1957, с. 42; Д. У. Снедекор. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. М., 1961, с. 40; Л. З. Румшиский. Элементы теории вероятностей. М., 1970, с. 56; Я. П. Черчук. Графики в математико-статистическом анализе. М., 1972, с. 21 и т. п.

[2] S. Szulc. Metody statystyczne. Warczawa, 1968, p. 153.

[3] Н. В. Смирнов, И. В. Дунин-Барковский. Курс теории вероятностей и математической статистики для технических приложений. М., 1969, с. 126.

[4] Методика и техника статистической обработки первичной социологической информации. М., 1968, с. 31.

[5] В. И. Василевич. Статистические методы в геоботанике. Л., 1969, с. 22.

[6] Е. И. Пустыльник. Статистические методы анализа и обработки наблюдений. М., 1968, с. 156.

[7] Д. Э. Юл, М. Дж. Кендэл. Теория статистики. М., 1960, с. 104.

[8] В. Ю. Урбах. Математическая статистика для биологов и медиков. М., 1963, с. 15; Он же. Биометрические методы. М., 1964, с. 15.

[9] S. Szulc. Metody...

[10] Н. В. Смирнов, И. В. Дунин-Барковский. Курс теории..., с. 126; А. К. Митропольский. Техника статистических исчислений. М., 1931, с. 29; Он же. Техника статистических вычислений. М., 1961, с. 20.

[11] Методика и техника статистической обработки первичной социологической информации, с. 31.

[12] Е. К. Меркурьева. Основы биометрии. М., 1963, с. 34; Она же. Биометрия в селекции и генетике сельскохозяйственных животных. М., 1970, с. 21; В. Г. Вольф. Статистическая обработка опытных данных. М., 1966, с. 13; П. Ф. Рокицкий. Биологическая статистика. Минск, 1973, с. 19; Г. Н. Зайцев. Методика биометрических расчетов. М., 1973, с. 213, табл. 3.

[13] H. A. Sturges. The Choiee of class Interval.— Journal of the American Stastical Association, 1926, March.; Цит. по: Ф. Миллс. Статистические методы. М., 1958, с. 49; См. также: Г. Ф. Лакин. Биометрия. М., 1973, с. 19. Имеется публикация этой формулы в несколько измененном виде (возможна, опечатка).— См.: И. Г. Венецкий, Г. Г. Кильдишев. Основы теории вероятностей и математической статистики. М., 1968, с. 55. Различие в коэффициенте, учитывая последующее округление, на результат не влияет.

[14] Цит. по: Г. Ф. Лакин. Биометрия, с. 19.

[15] Г. Н. Зайцев. Методика биометрических расчетов, с. 13.

[16] Г. Ф. Лакин. Биометрия, с. 19.

[17] Указанные таблицы дают примерное соотношение $k=(2{,}7+0{,}2k)\lg n$.

[18] Мы оставляем здесь в стороне вопрос о достоверности замеров фрагментов лепной керамики с такой точностью.

4 Заказ 293