

519
P. 556 B.C.
57923

311
P. 55

1932
Проф. Ю. Л. ПОМОРСКИЙ

311
П 55

ВАРИАЦИОННАЯ СТАТИСТИКА

ЭЛЕМЕНТАРНОЕ
ПРАКТИЧЕСКОЕ РУКОВОДСТВО
ДЛЯ ВРАЧЕЙ, ПЕДАГОГОВ,
ПЕДОЛОГОВ, ПСИХОТЕХНИКОВ,
РАБОТНИКОВ ФИЗКУЛЬТУРЫ
И АГРОНОМОВ

13686.
98961
ЧАСТЬ II

С 148 ЧЕРТЕЖАМИ И 3 ТАБЛИЦАМИ

Ц. Н. И. И.
Самарской губернии
БИБЛИОТЕКА
Народного
и Центросоюза

Издание Детского Обследовательного Института
им. проф. А. С. Грибоедова
ЛЕНИНГРАД
1930



11/6.

ОТ АВТОРА.

Настоящая книга, как по своему содержанию, так и по методу изложения, является прямым продолжением первой части моего руководства того же наименования, напечатанного впервые в 1927 г. и переизданного с добавлениями в 1929 г. *) Такая преемственность материала делает необходимым предварительное ознакомление читателя с содержанием предыдущей части, без чего данное руководство может оказаться все же не вполне для него понятным. При этом имеется в виду главным образом читатель не математик, для которого исключительно и предназначены обе книги.

Далее считаю своим долгом отметить, что книга эта совершенно не рассчитана на использование ее в области социально-экономических дисциплин, в условиях статистической разработки которых содержатся некоторые специфические особенности, вовсе не предусмотренные и не затронутые в содержании данного руководства. Круг читателей этой книги определяется, таким образом, лишь категориями, перечисленными в подзаголовке, а также работниками в смежных с ними областях знания по преимуществу биологического характера.

Вторая часть настоящего руководства состоит из трех отделов.

В первом из них (до XI гл. включ.) излагаются вопросы, связанные с характеристикой распределения отдельных вариантов по классам их числовых значений и, в частности, дается понятие о так называемом „нормальном распределении“ и различных от него отклонениях (асимметрия и эксцесс). Понятие о распределении вариантов позволило использовать здесь нагляд-

*) В настоящее время печатается 3-е издание этой книги.

ную графическую иллюстрацию вариационных рядов в виде соответствующих им гистограмм и полигонов, а также указать ряд дополнительных констант, характеризующих данное распределение в целом (мода, медиана, квартиль). Параллельно с описанием обычного способа моментов, при помощи которого чаще всего вычисляются показатели асимметрии и эксцесса, в этом отделе указан также и упрощенный (но вполне точный) метод определения этих показателей с помощью особой таблицы поправок.

Второй отдел (от XII до XXIV гл. включ.) включает в себе описание существующих способов оценки индивидуальных вариантов с помощью стандартов, выведенных на массовом материале. Сюда относится установление границ оценочных зон, приемы децисигмальной оценки и перцентиляжа, составление оценочных таблиц, графиков и трафаретов для вычерчивания индивидуального профиля (антропометрического, психометрического и проч.). Все эти вопросы и приемы имеют в настоящее время широкое применение и вполне актуаль-

I. Графическое изображение вариационных рядов.

В результате разбивки варьирующего признака V на те или иные классы мы получаем ряд частот p , показывающих, как распределились отдельные значения этого признака внутри установленных нами классовых границ *). Обычно в крайние классы при такой разбивке попадает сравнительно небольшое число вариантов, а в средних они накапливаются в значительно большем количестве. В связи с этим ряд частот p в большинстве случаев начинается с малых чисел, которые постепенно нарастают по мере приближения к средним классам, после чего опять заметно начинают убывать. Чтобы составить наглядное представление о характере распределения вариантов по отдельным классам, во многих случаях полезно бывает изобразить это нарастание и убывание частот p в форме „графика распределения“. Наиболее распространенные способы такого графического изображения вариационных рядов мы разберем здесь на следующем числовом примере.

Пусть имеем вариационный ряд, разбитый на 11 классов с классовым промежутком $k=2$:

в границах от 25 до 23						1	варианта
»	»	»	23	»	21	»	2
»	»	»	21	»	19	»	3
»	»	»	19	»	17	»	5
»	»	»	17	»	15	»	10
»	»	»	15	»	13	»	14
»	»	»	13	»	11	»	21
»	»	»	11	»	9	»	21
»	»	»	9	»	7	»	15
»	»	»	7	»	5	»	6
»	»	»	5	»	3	»	2

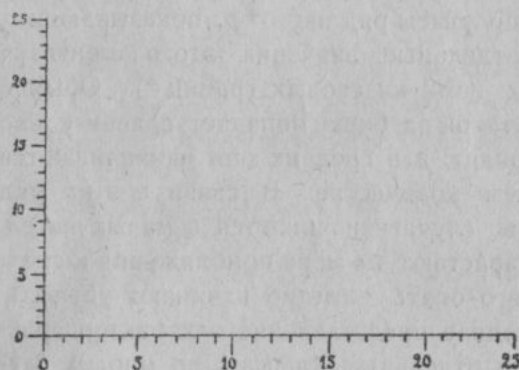
*) См. гл. XI первой части моего руководства.

Здесь разбивка на классы произведена так, что нижняя граница высшего класса (напр. 23) в то же время является и верхней границей следующего низшего класса (от 23 до 21). Сделано это в расчете на полную непрерывность варьирующего признака V , могущего принимать любые промежуточные значения, как угодно близко подходящие к отдельным классовым границам (раньше мы фиксировали эти границы например так: от 25 до 23,5 или от 25 до 23,1 в зависимости от точности самих измерений, которые, таким образом, предполагались изменяющимися скачками, а не непрерывно).

Общее число измерений (сумма всех частот p) в нашем примере = 100.

Итак:

$$n = 100.$$



Черт. 1.

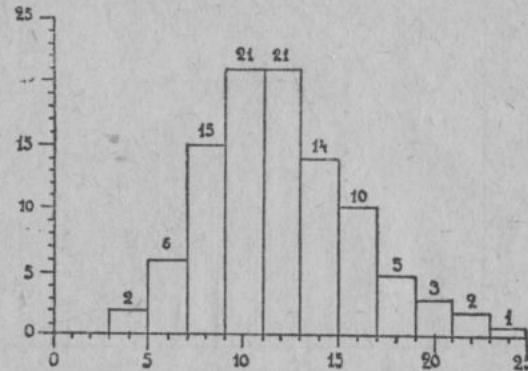
При изображении заданного ряда в форме графика условимся числовые значения самого варьирующего признака V отмечать по горизонтальному направлению, а соответствующие этим значениям частоты p — по перпендикулярному к нему вертикальному направлению.

Начертив на листе бумаги две взаимно перпендикулярные прямые (так называемые „оси координат“), пометим на них деления какого-либо произвольного масштаба, как это показано на черт. 1, и в соответствующих точках деления поставим пояснительные цифры.

Эти цифры поставлены здесь не у всех масштабных делений, а лишь у каждой пятой черточки. Отсчет ведется по горизонтальному направлению (по „оси абсцисс“) слева направо, а по вертикальному (по „оси ординат“) — снизу вверх в обоих случаях от „начала координат“ т. е. от точки пересечения этих осей, принимаемой на 0 (нуль). Масштабные деления по каждой оси сделаны равномерными (черточки находятся на одинаковом расстоянии друг от друга), но вер-

тикальный и горизонтальный масштабы не совпадают: первый из них сделан здесь несколько мельче второго.

Один из способов построения „графика распределения“ заключается в изображении отдельных частот p в виде ряда столбиков. Ширина этих столбиков делается чаще всего одинаковой, в точности равной классовому промежутку k . В нашем примере, следовательно, все столбики должны быть шириною в 2 деления горизонтального масштаба (так как $k=2$). Размещение этих столбиков по горизонтальной оси делается с тем расчетом, чтобы каждый столбик пришелся внутри границ (в интервале) соответствующих классов. У нас, слева направо, первый столбик нужно поместить в промежутке между 3 и 5 делениями горизонтальной оси, следующий между 5 и 7 делениями и т. д. включительно до последнего 11-го столбика, который окажется в промежутке между 23 и 25 делением этого масштаба.



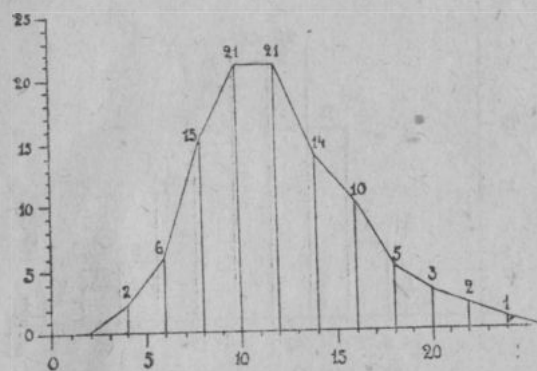
Черт. 2.

По своей высоте все эти столбики должны быть пропорциональны тем частотам p , которые они изображают. Так, крайний слева столбик, соответствующий частоте $p=2$ (для класса от 3 до 5), мы сделали высотой в 2 деления вертикального масштаба; следующий смежный с ним (справа) столбик, соответствующий частоте $p=6$ (для класса от 5 до 7) — высотой в 6 таких делений и т. д. (см. черт. 2). Для ясности сверху над всеми 11-тью столбиками поставлены цифры: 2, 6, 15, 21 и т. д., указывающие частоты p соответствующих классов.

Вариационный ряд, графически изображенный в форме таких смежных друг с другом столбиков, носит на языке „гистограммы“ распределения (или просто „гистограммы“). Здесь площадь каждого столбика изображает частоту p соответствующего класса, т. е. то количество вариантов, ко-

торое при разбивке оказалось внутри границ данного класса. Площадь же всей гистограммы (т. е. всех столбиков вместе взятых) представляет собой, очевидно, общую сумму всех частот p (Σp) т. е. число измерений n . Следовательно там, например, где заданный ряд будет иметь вдвое большее число измерений n , там и полная площадь соответствующей гистограммы распределения будет вдвое больше.

Другой способ построения „графика распределения“ заключается в изображении отдельных частот p уже не в виде широких полос (столбиков), а с помощью тонких вертикальных прямых (ординат), проведенных через точки, соответствующие серединам отдельных классов.



Черт. 3.

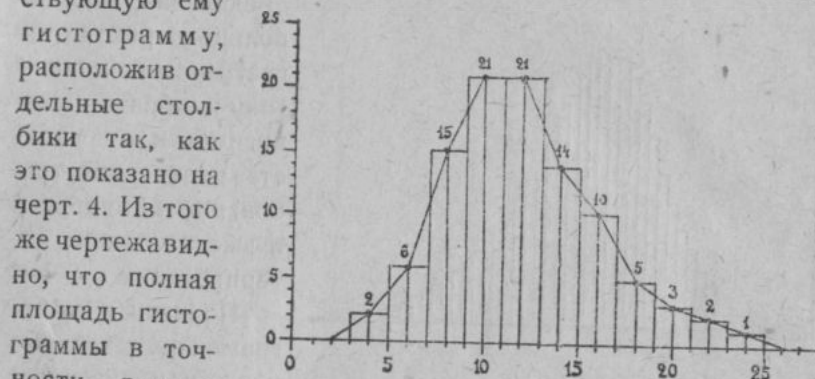
В нашем примере классы разбиты слева направо так: 3 — 5, 5 — 7, 7 — 9 и т. д. Следовательно середины этих классов приходятся в точках: 4, 6, 8 и т. д. Так вот, из этих средних точек мы и должны

теперь восставить перпендикуляры соответствующей высоты (в зависимости от величины самих частот p). Вершины всех этих перпендикуляров принято, для ясности, соединять отрезками прямых линий, а по краям графика опускать эту ломаную линию на горизонтальную ось, как показано на черт. 3. Этот прием опускания ломаной линии на горизонтальную ось основан на том, что по краям чертежа могли бы быть отмечены еще по два так называемых „пустых“ класса, в которые при разбивке не попало ни одной варианты (для этих классов $p = 0$). Середины границ таких крайних классов (1 — 3 и 25 — 27) приходятся здесь, очевидно, в точках 2 и 26. Высота соответствующих перпендикуляров в этих точках должна = 0 (так как $p = 0$), т. е. вершины таких перпендикуляров вовсе не поднимаются над осью абсцисс. Поэтому

наша ломаная линия и должна быть сведена по краям графика к этим двум точкам, лежащим на горизонтальной оси.

Вариационный ряд, графически изображенный описанным выше способом, носит название „полигона распределения“ (или просто „полигона“).

Заметим, что полигон распределения, как это видно из чертежа 4, легко может быть получен из соответствующей гистограммы. Для этого стоит лишь соединить друг с другом все средние точки верхних площадок ступенчатой фигуры этой гистограммы и свести полученную ломаную линию на ее краях к оси абсцисс. Имея, наоборот, полигон распределения, мы всегда можем построить соответствующую ему



Черт. 4.

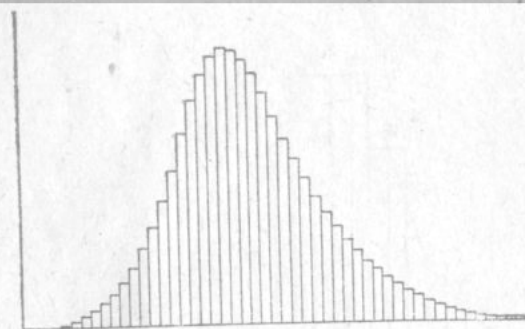
полной площади соответствующего полигона. В самом деле, если у изображенной на черт. 4 ступенчатой фигуры вырезать все выходящие за пределы полигона отдельные треугольники, то каждый из них может быть вставлен в соответствующую выемку, заполнив ее без всякого просвета, в результате чего мы и получим контур полигона, как бы сделанного из того же самого материала.

Переход от полигона к гистограмме и обратно имеет, очевидно, и логические основания. Непрерывному изменению варьирующего признака V , могущего принимать внутри каждого класса любые промежуточные значения, соответствует форма гистограммы. Если же принять, что все эти значения V в каждом классе сосредоточены в его середине (что практически и делается при вычислениях), то этому предположению будет уже отвечать форма полигона, ординаты кото-

рого как раз и приурочены к центральным точкам отдельных классовых интервалов. В случае непрерывного изменения варьирующего признака, могущего принимать только целые значения (например если V изображает число лепестков цветка или число правильно решенных задач), распределение таких вариантов V вообще говоря должно быть представлено в форме полигона. Однако и здесь мы можем принять эти целые значения V за середины отдельных классов, условно предполагая, что изменение их внутри таких классов происходит непрерывно. Этому же предположению будет отвечать уже форма гистограммы. Итак, каждый вариационный ряд мы вправе изображать любым типом графика рас-

пределения, произвольно переходя от гистограммы к полигону и обратно в зависимости от удобства построения, требований наглядности и желательной интерпретации.

Полигон и гистограмма представляют собой приближенные



Черт. 5.

графические характеристики распределения отдельных вариантов по классам, выявляя лишь общую картину грубого подсчета частот в сравнительно крупных интервалах. При достаточно большом числе измерений n мы всегда имеем возможность уменьшить классовой промежуток k и увеличить тем общее количество классов, в результате чего характеристика распределения вариантов получит большую плотность и точность. На гистограмме эти изменения скажутся очевидно в уменьшении ширины отдельных столбиков и в увеличении общего их количества при неизменном сохранении, однако, той же самой общей площади всей фигуры в целом. Если представить себе, что при чрезвычайно большом числе измерений n мы сделаем отдельные столбики гистограммы во много раз уже прежних и во столько же раз, очевидно, увеличим общее их количество, то грубая ступенчатая форма первоначального контура нашей гисто-

граммы обратится в мелко-ступенчатую, почти плавную кривую, как это схематически и показано на черт. 5.

В пределе, когда ширина отдельных столбиков делается $= 0$, а общее их количество достигнет бесконечности, сгладятся и последние зигзаги этой мелко-волнистой линии, которая обратится уже в совершенно плавную кривую, называемую „кривой распределения“.

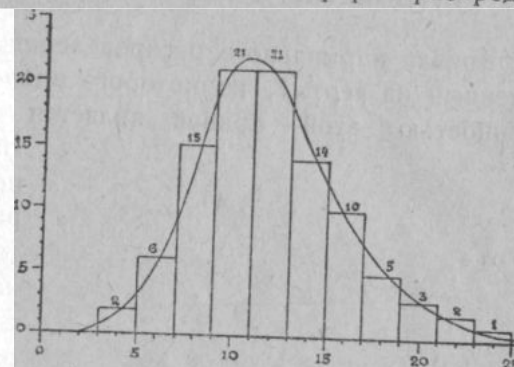
На черт. 6 эта кривая распределения изображена для ясности в сопоставлении с соответствующей ступенчатой гистограммой.

Кривая распределения, в отличие от гистограммы и полигона, дает представление об идеальной форме распре-

деления данной совокупности заданных вариантов по величине тех или иных их числовых значений. В то время как форма фактического распределения этих вариантов зависит от произвольно установленного классового промежутка k и произвольно же выбран-

ных границ разбивки, плавная теоретическая „кривая распределения“ выражает общий закон этого распределения независимо от тех или иных технических условий его выявления (или во всяком случае в минимальной от них зависимости).

Всякому эмпирически установленному распределению вариант по классам отвечает характерная для него теоретическая „кривая распределения“. В настоящее время Пирсоном установлено 8 основных типов таких кривых и указаны способы их построения в соответствии с формой эмпирического полигона или гистограммы. Однако изложение теории и техники всех этих операций не входит в задачи настоящего элементарного руководства. Читатель, заинтересовавшийся кривыми Пирсона, найдет все относящиеся сюда указания у Л. К. Лахтина („Кривые распределения числен-



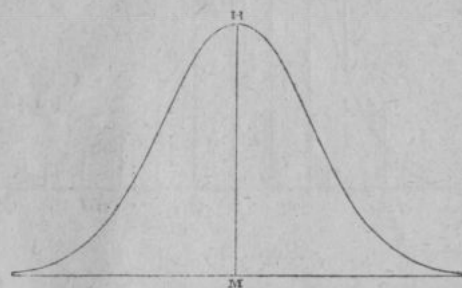
Черт. 6.

ностей и построение для них интерполяционных формул по способам Пирсона и Брунса". М. 1922), у Е. Е. Слуцкого („Теория корреляции и элементы учения о кривых распределения". Киев, 1912) и у В. П. Эльдертона („Кривые распределения численностей и корреляции". М. 1924).

Для наших целей достаточно будет рассмотреть здесь лишь один частный, но чрезвычайно важный тип распределения, известный под именем „нормального", описанию которого и будут посвящены следующие главы.

II. Понятие о нормальном распределении.

Кривая нормального распределения имеет вид, представленный на черт. 7, из которого видно, что основной особенностью этой кривой является ее симметричность.



Черт. 7.

Наибольшую высоту кривая нормального распределения имеет в точке H , находящейся как раз над точкой M , соответствующей по горизонтальному масштабу среднему арифметическому данного вариационного ряда. Так как ординаты этой кри-

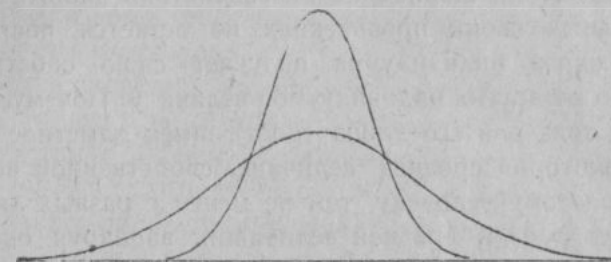
вой пропорциональны частотам p , то при нормальном распределении наибольшую частоту будет иметь следовательно тот класс, в центре которого находится среднее арифметическое M . Вправо и влево от M ординаты кривой (а стало быть и частоты p) постепенно и симметрично уменьшаются, сперва медленно, затем быстрее, а по краям кривой это ее снижение опять замедляется. Уже на небольшом сравнительно расстоянии от M обе ветви нормальной кривой довольно близко подходят к оси абсцисс, но ни здесь, ни дальше нигде с ней не совпадают вполне. Про эту кривую говорят, что она „асимптотически" приближается к оси абсцисс, нигде с ней не сливаясь. Это значит, что теоретически мы можем ожидать встретить любые отклонения от

среднего арифметического M , но что слишком большие отклонения все же будут в действительности попадаться чрезвычайно редко (так как их теоретические частоты приближаются к нулю).

Этими главными особенностями и характеризуется в общих чертах так называемое „нормальное" распределение вариант, графически представленное на черт. 7 в виде „кривой нормального распределения".

Полезно запомнить, что площадь нормальной кривой, также как и площадь всякого графика распределения, пропорциональна общему числу измерений n . Чем больше n , тем больше и площадь, ограниченная данной кривой.

При сохранении общей своей формы, соответствующей нормальному распределению, эта кривая в зависимости от



Черт. 8.

величины варьирования может иметь различную степень пологости и крутизны. На черт. 8 представлены две нормальные кривые одинаковой площади, но различного рассеяния. Сжатая по бокам высокая кривая соответствует случаю малого варьирования (т. е. малому среднему квадратическому отклонению σ), а пологая кривая — случаю большого варьирования (большой σ).

Большее или меньшее приближение к нормальному распределению можно подметить в любой области варьирования количественных признаков. Всем, например, прекрасно известно, что большинство людей обладает некоторым „средним" ростом, а „высокие" и „низкие" люди встречаются сравнительно редко, при чем „карлики" и „гиганты" попадают лишь в виде ничтожных исключений. Такое же распределение людей, надо думать, наблюдается и в отно-

шении их интеллектуального развития. Подавляющее большинство здесь также составляют несомненно люди „среднего“ умственного уровня, „талантливые“ и „отсталые“ в интеллектуальном отношении составляют меньшинство, а „гении“ и „идиоты“ представляют собой редкое исключение.

Этот известный в сущности каждому закон распределения числовых значений всякого варьирующего признака (роста, интеллектуального развития и проч.) нам остается лишь постараться объяснить здесь теоретически и придать ему по возможности более точную количественную формулировку.

При уяснении особенностей явления варьирования количественных признаков прежде всего возникает вопрос о его механизме. Под влиянием каких воздействий данный признак в отдельных своих проявлениях не остается постоянным и у каждого индивидуума получает свою собственную, отличную от других индивидуумов величину? Почему, например, вес тела или его длина (рост), имея заметное тяготение к некоторой средней величине, свойственной нормальному взрослому человеку, тем не менее у разных людей не совпадает с этой средней величиной, варьируя около нее в тех или иных пределах?

Причина этого явления заключается, очевидно, в неоднородности тех условий (внешних и внутренних), под влиянием которых данный количественный признак (например рост или вес) у отдельных индивидуумов достигает той или иной определенной величины. Так, рост человека зависит от целого ряда наследственных, социальных и прочих факторов, различным образом влияющих на величину роста у каждого данного индивидуума. Некоторые из этих факторов, взятые изолированно от остальных, содействуют увеличению роста, другие, наоборот, его уменьшают, причем влияние этих факторов может иметь различную, неодинаковую силу (одни из них влияют на интенсивность роста сильнее, другие — слабее).

Таким образом величина роста каждого отдельного индивидуума обусловлена разнообразным, часто противоположным влиянием множества таких в большинстве независимых друг от друга воздействующих факторов, в каждом отдель-

ном случае образующих в своей совокупности некоторое определенное, случайное сочетание.

Если эта комбинация факторов окажется благоприятной для увеличения данного признака — эффект получится положительный, и данный индивидуум окажется высокого роста. В обратном случае он будет низкорослым. Если же в данной случайной комбинации причинных воздействий их противоположные влияния взаимно уравновесятся — индивидуум получит некоторую среднюю величину этого признака (роста).

Такую же или подобную схематическую картину мы получим и при выяснении причин варьирования веса, окружности груди, интеллектуального развития или какого-либо иного количественного признака, определяемого в каждом своем индивидуальном проявлении влиянием совокупности многих причинных факторов.

Чем же, спрашивается, можно объяснить неодинаковое количество „средних“ и „крайних“ результатов этих случайных воздействий? Почему уравнивание противоположных влияний на деле встречается чаще, а превалирование одного какого-либо отклоняющего воздействия представляет собой явление более редкое?

Причина этого заключается в неодинакой вероятности таких сочетаний. При случайном подборе независимых друг от друга причинных факторов взаимная компенсация противоположных влияний теоретически всегда будет иметь за себя больше шансов, в то время как ожидать подбора преимущественно однородных по своему эффекту причин представляется менее вероятным.

Поясним эту мысль таким схематическим примером. Пусть 4 раза подряд бросается монета и отмечается число случаев, при которых за эти 4 бросания монета ляжет вверх гербом. Этот опыт можно повторять неограниченное число раз, причем после каждых 4-х бросаний мы будем записывать цифрами: 0, 1, 2, 3 или 4 общее количество выпадов герба, сосчитанное в данной серии 4-х подбрасываний этой монеты.

Очевидно, что при 4-х таких подбрасываниях монеты очень мало шансов получить гербы во всех четырех случаях подряд и мало же шансов не обнаружить вовсе ни

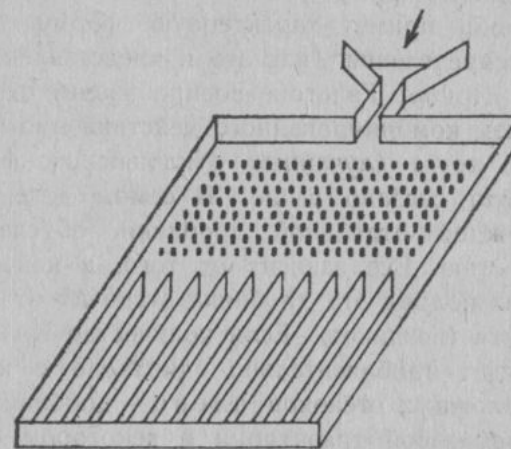
одного герба. Простой опыт с полной убедительностью покажет, что наиболее вероятной (т. е. чаще всего в действительности имеющей место) будет такая комбинация результатов, при которой число выходов герба будет равно числу выходов решетки (т. е. 2 герба и 2 решетки), менее вероятной (реже встречающимся) окажется случай выхода 3-х гербов и 1 решетки или же 1-го герба и 3 решеток, и, наконец, очень редко выпадет комбинация только 4-х гербов или только 4-х решеток.

Малая вероятность преимущественного скопления каких-либо однородных фактов (в нашем примере только гербов или только решеток) с еще большей наглядностью обнаружится при увеличении числа бросаний в каждой серии с 4-х до 10, до 50-ти или до 100. В самом деле, насколько мало, например, шансов при 100 подбрасываниях монеты все 100 раз подряд получить сверху только герб или только решетку! Такой крайний случай практически может и вовсе не встретиться на деле даже при нескольких миллионах таких стократных испытаний, в то время как приблизительное равенство числа гербов и решеток будет и в данном случае встречаться сравнительно часто.

Этот схематический пример наглядно объясняет причины указанного выше „нормального“ распределения вариант по классам. Взаимно погашающие друг друга противоположные влияния причинных факторов, приводящие в результате к „средней“ величине варьирующего признака, встречаются чаще всего, и потому количество индивидуумов, обладающих средней величиной роста или веса, будет наибольшим. Малая же или, наоборот, слишком большая величина такого признака, вызываемая подбором однородных по своему влиянию отклоняющих факторов, встречается редко, так как самая комбинация таких факторов представляется мало вероятной.

Наглядная иллюстрация схемы и действия внутреннего механизма нормального распределения может быть дана с помощью остроумного приспособления, известного под именем „прибора Гальтона“. Прибор этот состоит из наклонно поставленной плоской доски, на которую сверху через особую воронку постепенно высыпается мелкая дробь. По пути движения потока этих дробинок в доску вколочено

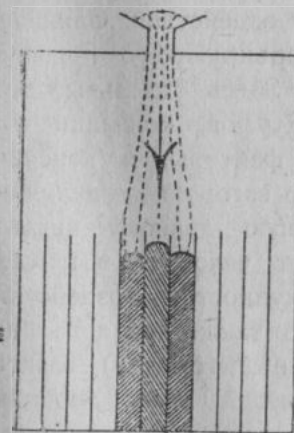
множество мелких препятствий (хотя бы в виде отдельных колышков или гвоздиков), а нижняя часть доски разгорожена несколькими параллельными друг другу переборками, в промежутках между которыми может скапливаться скатывающаяся по доске дробь (см. черт. 9).



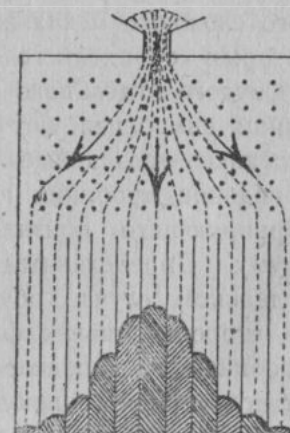
Черт. 9.

Представим себе временно, что на пути падения дробинок нет никаких препятствий (гвоздиков). Тогда вся струя дроби, скатываясь по кратчайшему направлению к середине нижней части доски, скопится только в одной или, самое большее, в нескольких центральных перегородках, вовсе не заполнив собою крайних отделений справа и слева (см. черт. 10).

Если же на пути ее движения будут расставлены пре-



Черт. 10.



Черт. 11.

пятствия, струя дроби станет рассеиваться ими в разные стороны и, скатываясь вниз все более и более расширяю-

щимся потоком, заполнит отчасти также и крайние загородки. При этом скопившаяся в нижней части прибора дробь примет характерную форму „кривой нормального распределения“, как это и представлено на черт. 11.

Прибор Гальтона воспроизводит схематически тот механизм комбинированного действия многих независимых друг от друга случайных отклоняющих факторов, о котором было сказано выше. В самом деле, характер движения каждой отдельной дробинки обусловлен лишь случайностью. Все зависит от того, в какую сторону и сколько раз подряд эта дробинка отскочит от встречающихся на ее пути гвоздиков. Если количество таких отскоков вправо будет приблизительно одинаково с количеством противоположных отскоков влево — дробинка покатится по зигзагообразной траектории в некотором среднем направлении и в результате попадет в центральную загородку. Если же большинство встречных препятствий отклонит эту дробинку в какую-либо одну сторону — она в конечном итоге упадет далеко от среднего отделения и притом тем дальше от него, чем большее количество раз пришлось ей отскочить от отдельных гвоздиков в соответствующую (одну и ту же) сторону.

Прибор Гальтона представляет собой упрощенную модель того сложного механизма случайных воздействий, влиянием которых определяется величина варьирующего признака у каждого отдельного индивидуума. Здесь отдельные дробинки символизируют этих индивидуумов, колышки или гвоздики — различные отклоняющие факторы, а конечное попадание дробинки в ту или иную загородку — величину варьирующего признака, свойственную данному индивидууму. Определенный вес или рост человека в каждом отдельном случае обусловлен совокупностью воздействий со стороны многих факторов, точно также как попадание дробинки в определенное отделение (загородку) зависит от случайной совокупности отклонений этой дробинки в результате столкновения ее со множеством расставленных по доске препятствий (гвоздиков).

Размещение по загородкам отдельных дробин, точно также как и распределение по классам роста или веса отдельных индивидуумов, представляет собою явление чисто

случайного порядка, предугадать которое мы не в состоянии, так как не в силах учесть всю сложную картину взаимодействия определяющих это явление причинных факторов. Но предсказать суммарно общий конечный результат распределения всех дробин по отдельным загородкам (что символически и схематически соответствует распределению отдельных вариантов по классам) мы можем уже довольно точно на основании учета неодинаковой вероятности различных ожидаемых комбинаций в сочетаниях этих отклоняющих факторов.

Прибор Гальтона с особой наглядностью подтверждает высказанное ранее теоретическое положение о большей вероятности „средних“ значений и меньшей вероятности „отклонений“. В результате рассеивания струи дробин при столкновении их с гвоздиками отдельные дробинки скопляются главным образом в средних загородках, в соседние с ними ближайшие отделения попадает уже значительно меньше отдельных дробин, а в самые крайние — ничтожное их количество, что и дает характерную картину „нормального“ распределения, представленного на чертежах 7, 8 и 11.

Само собою понятно, что строгое „нормальное“ распределение дробин по загородкам и вариант по классам может иметь место лишь при достаточно большом количестве распределяемых индивидуумов (дробин, вариант), а также при соблюдении указанных выше условий (множественность и независимость друг от друга всех причинных факторов, определяющих величину данного признака).

III. Построение кривой нормального распределения.

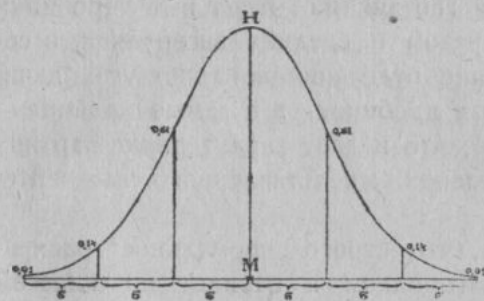
Математический анализ кривой нормального распределения, основанный на положениях теории вероятностей, дает в конечном итоге ряд практических приемов и формул, позволяющих выполнить построение этой кривой (см. черт. 12) для любой совокупности эмпирических частот.

Так как кривая нормального распределения имеет наибольшую ординату (высоту) в точке M , соответствующей по горизонтальному масштабу среднему арифметическому

заданного эмпирического ряда, то положение этой кривой на чертеже определяется положением самой точки M . Эта наибольшая высота (H) нормальной кривой может быть вычислена по формуле:

$$H = \frac{0,4 k n^*}{\sigma}$$

Вправо и влево от точки M ординаты нормальной кривой будут постепенно и симметрично уменьшаться, причем на расстоянии одной σ в обе стороны от M эти ординаты составят приблизительно 0,61 полной высоты H , на расстоянии двух σ от M они уменьшатся до 0,14 высоты H , а на расстоянии трех σ сделаются равными всего только 0,01 этой максимальной высоты H (см. черт. 12).



Черт. 12.

Обозначив эти числовые коэффициенты (0,61, 0,14 и 0,01) буквою h , мы можем написать следующую общую формулу для вычисления ординат (y) кривой нормального распределения:

$$y = h \cdot H$$

где наибольшая высота H является для данного ряда величиной постоянной, а коэффициенты h меняются в зависимости от расстояния искомой ординаты y до центральной точки M (по мере увеличения этого расстояния коэффициенты h наоборот уменьшаются).

Выше мы дали только три значения h для расстояний ординат от точки M на одну, на две и на три σ . При желании можно раз навсегда вычислить и любые промежуточные значения этих коэффициентов. Так, например, для ординат y , отстоящих в обе стороны от M на $1/2 \sigma$, коэффициент $h = 0,88$, для расстояния в $1 1/2 \sigma$ $h = 0,33$, а для $2 1/2 \sigma$ $h = 0,04$.

*) Точнее, это выражение следовало бы умножить в числителе на коэффициент 0,39894, приближенно принимаемый здесь за 0,4.

Облекая эти и промежуточные данные в форму таблицы, мы можем придать ей следующий вид:

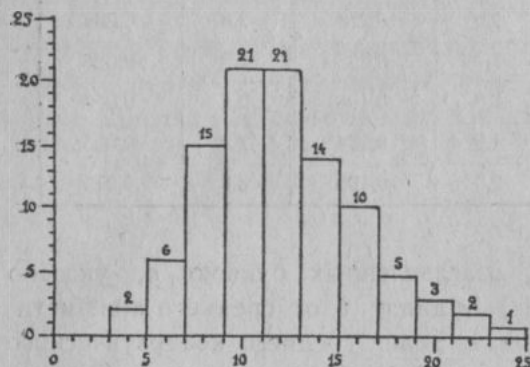
d	h	d	h	d	h
0,1	0,995	1,1	0,546	2,1	0,110
0,2	0,980	1,2	0,487	2,2	0,089
0,3	0,956	1,3	0,430	2,3	0,071
0,4	0,923	1,4	0,375	2,4	0,056
0,5	0,883	1,5	0,325	2,5	0,044
0,6	0,835	1,6	0,278	2,6	0,034
0,7	0,783	1,7	0,235	2,7	0,026
0,8	0,726	1,8	0,198	2,8	0,020
0,9	0,667	1,9	0,164	2,9	0,015
1,0	0,607	2,0	0,135	3,0	0,011

Здесь в столбцах, озаглавленных буквою d , указано отклонение (расстояние) ординат y от среднего арифметического M , выраженное в долях среднего квадратического отклонения σ , а рядом, в столбцах h , даны соответствующие этим расстояниям числовые значения искомых коэффициентов h , на которые для вычисления y следует умножить максимальную высоту H . Эту таблицу, следовательно, надо понимать так: для ординаты y , отстоящей от M на расстояние $d = 0,1$ сигмы (1-ая строчка), коэффициент $h = 0,995$; для расстояния $d = 0,2$ сигмы (2-ая строчка) $h = 0,980$ и т. д. В этой таблице расстояния d взяты через одну десятую долю сигмы, а коэффициент h вычислен с тремя знаками после запятой.

Воспользуемся данными этой таблицы для построения кривой нормального распределения в применении к разобранному уже выше примеру вариационного ряда, гистограмма которого повторена здесь на чертеже 13 (см. выше, черт. 2).

В этом распределении несколько нарушена симметрия, и весь ряд слегка скошен в левую сторону. Посмотрим, как расположится на том же самом чертеже кривая нормального распределения, соответствующая данной гистограмме.

Эту нормальную кривую мы должны вычертить так, чтобы соответствующее ей среднее арифметическое в точности совпало со средним арифметическим M заданного вариационного ряда, чтобы у обеих кривых оказались одинаковые средние квадратические отклонения σ и одинаковая площадь (соответствующая числу измерений n). Все эти условия и будут автоматически выполнены при использовании указанной выше формулы для вычисления наибольшей высоты H кривой нормального распределения и приведенной здесь таблицы значений коэффициентов h .



Черт. 13.

Итак прежде всего определим числовые значения n , M и σ для заданного вариационного ряда. Выписав в клетках столбца p частоты этого ряда, соответствующие указанным слева классовым границам и приняв за начало отсчетов (C) середину класса от 15 до 13, т. е. положив:

$$C = 14,$$

выполним внутри расчетной решетки все обычные промежуточные вычисления по способу сумм *).

В результате получим следующие четыре итога (см. решетку на стр 25):

$$\begin{aligned} a_1 &= 42 & a_2 &= 36 \\ b_1 &= 142 & b_2 &= 122 \end{aligned}$$

Откуда определим и обе вспомогательные суммы:

$$\begin{aligned} S_1 &= a_1 - b_1 = -100 \\ S_2 &= a_1 + b_1 + 2a_2 + 2b_2 = 500 \end{aligned}$$

*) См. подробное описание этого способа в I части моего руководства.

Среднее арифметическое M найдется по обычной формуле (при $k=2$)

$$M = C + \frac{kS_1}{n} = 12$$

Сумма квадратов центральных отклонений

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 400$$

Для вычисления среднего квадратического отклонения (σ) мы здесь и в дальнейшем будем пользоваться формулой, соответствующей большому числу измерений n (т. е. деля под корнем Σx^2 не на $n-1$, а на n):

$$\sigma = \pm k \sqrt{\frac{\Sigma x^2}{n}} = \pm 4.$$

Итак, для заданного эмпирического ряда, изображенного в форме гистограммы на чертеже 13,

$$n = 100 \quad M = 12 \quad \sigma = \pm 4$$

По этим данным мы и можем теперь вычислить максимальную высоту (H) кривой нормального распределения, воспользовавшись формулой (см. стр. 22):

$$H = \frac{0,4kn}{\sigma} = \frac{0,4 \cdot 2 \cdot 100}{4} = 20.$$

Итак, в точке M , соответствующей 12-му делению горизонтального масштаба ($M=12$), наибольшая высота нормальной кривой $H=20$ делениям взятого нами вертикального масштаба.

Зададимся целью определить теперь высоты этой кривой (т. е. ее ординаты y) для нескольких точек, отстоящих друг от друга, скажем, на расстояние в $1/2\sigma$. Так как в нашем примере $M=12$, а $\sigma = \pm 4$, то $1/2\sigma = \pm 2$. Прибавляя и вычитая эту величину ($1/2\sigma = 2$) к среднему арифметическому

Границы классов

25—23
23—21
21—19
19—17
17—15
15—13
13—11
11—9
9—7
7—5
5—3

p	42	36
1	1	1
2	3	4
3	6	10
5	11	21
10	21	—
14	—	—
21	65	—
21	44	77
15	23	33
6	8	10
2	2	2
100	142	122

($M=12$) несколько раз подряд, получим следующие значения варьирующего признака V , для которых далее следует вычислить ординаты y кривой нормального распределения:

0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24.

Так как два ближайших к центру значения этого признака 10 и 14 отстоят от $M (=12)$ на $1/2 \sigma$, то при вычислении ординат кривой нормального распределения (y) по формуле:

$$y = hH$$

мы должны взять $d=0,5$, чему по приведенной выше таблице (см. стр. 23) соответствует значение $h=0,883$. Итак, для точек 10 и 14 обе ординаты кривой.

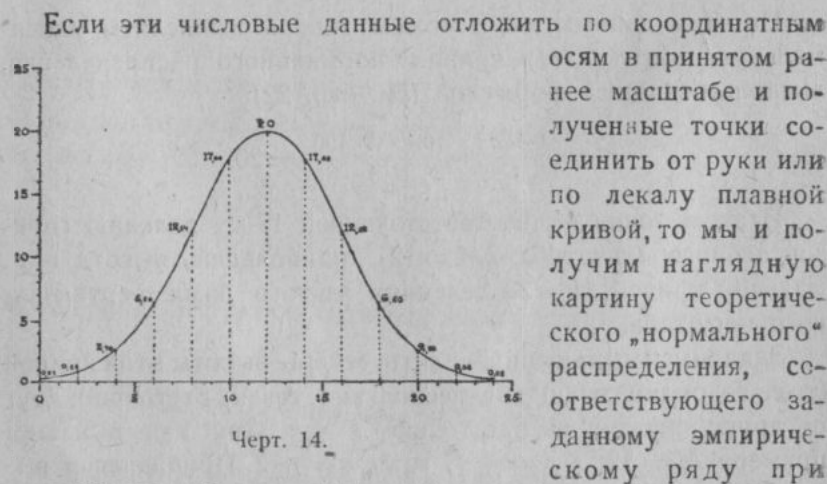
$$y = hH = 0,883 \cdot 20 = 17,6$$

Для следующей пары точек: 8 и 16, удаленных от M уже на целую σ ($d=1$), коэффициент $h=0,607$, и, следовательно, ординаты кривой в этих точках

$$y = 0,607 \cdot 20 = 12,14$$

Применяя этот же прием вычисления и дальше, получим:

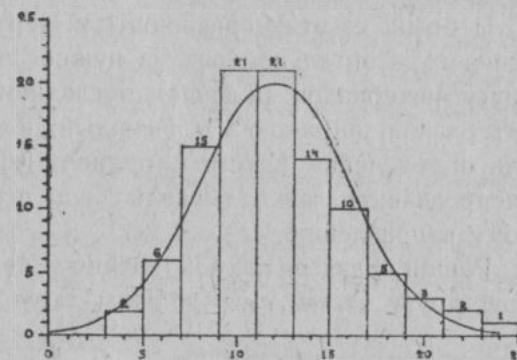
для точек: 6 и 18 (при $d=1,5$)	ординаты $y=6,50$
" " 4 и 20 (при $d=2$)	" $y=2,70$
" " 2 и 22 (при $d=2,5$)	" $y=0,88$
" " 0 и 24 (при $d=3$)	" $y=0,22$



$n=100$, $M=12$ и $\sigma=\pm 4$.

На чертеже 14 кривая нормального распределения вычерчена по точкам, сравнительно далеко отстоящим друг от друга. При желании, очевидно, можно было бы определить для этой кривой и ряд промежуточных точек.

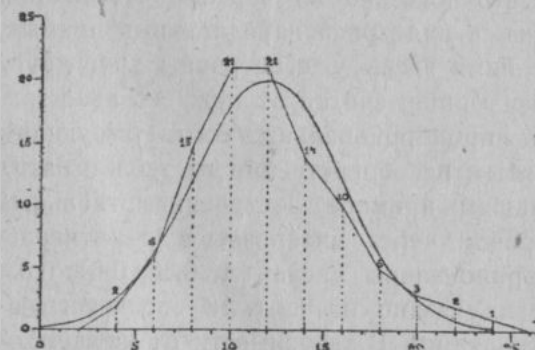
На чертеже 15 кривая нормального распределения дана в сопоставлении с гистограммой соответствующего эмпирического ряда, что



Черт. 15.

собственно и является нашей конечной задачей. Из этого сопоставления графиков видно, что заданный ряд действительно несколько не совпадает с нормальной кривой, обладая небольшой ассиметрией (сдвиг вершины влево). Ассиметрия заданного ряда выступает с еще большей наглядностью при сопоставлении кривой нормального распределения не с гистограммой, а с полигоном, как это показано на черт. 16.

Заметим, между прочим, что кривую нормального распределения можно было бы вычертить и не плавной линией, а просто соединив отдельные построенные для нее точки отрезками прямых (по линейке).



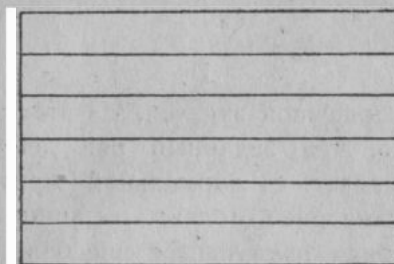
Черт. 16.

При построении нормальных кривых для большого количества эмпирических рядов полезно бывает сократить работу по вычислению ординат y и выполнить все эти построения исключительно графическим путем.

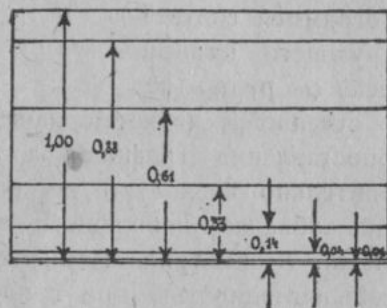
Для этого, прежде всего, необходимо раз навсегда приготовить два вспомогательные графика, образцы которых представлены на чертежах 17 и 18.

Первый из этих графиков (см. черт. 17) служит для деления горизонтальной оси на нужное число равных друг другу интервалов (в нашем последнем примере на 6 таких интервалов, шириною в $\frac{1}{2}\sigma$ каждый), а второй (см. черт. 18) — для определения искомых ординат (y) кривой нормального распределения, откладываемых на чертеже по вертикальному направлению.

Решив разделить ось абсцисс, например, на 6 равных друг другу частей, мы должны приготовить первый из этих



Черт. 17.



Черт. 18.

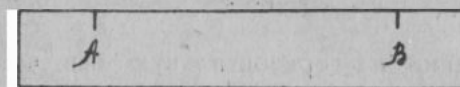
графиков в виде семи равноотстоящих друг от друга параллельных прямых, как это показано на чертеже 17. Второй график чертится также в виде ряда параллельных прямых, но расстояния между ними здесь уже не равны друг другу (см. черт. 18). Эти расстояния, считая от нижнего края графика, должны быть пропорциональны соответствующим коэффициентам h , взятым непосредственно из таблицы этих коэффициентов (в нашем примере построения отдельных точек нормальной кривой через интервалы в $\frac{1}{2}\sigma$ эти расстояния будут пропорциональны числам: 1, 0,88, 0,61, 0,33, 0,14, 0,04 и 0,01, что и указано на черт. 18 соответствующими цифрами *) и стрелками). В зависимости от желаемого размера чертежа мы можем увеличить эти табличные числа h ,

*) Здесь эти числа взяты с двумя знаками после запятой, в то время как в таблице на стр. 23 они даны с тремя знаками после запятой.

умножив их напр. на 10 или на 100. Таким образом если полную высоту этого графика (от самой нижней до самой верхней ограничивающей его прямой) сделать, например, в 10 см, то расстояние от нижней линейки до второй сверху должно равняться 8,8 см; от нижней линейки до следующей (3-ей сверху) 6,1 см и т. д. (Здесь соответствующие табличные числа увеличены в 10 раз).

Имея в своем распоряжении два таких вспомогательных графика, мы без труда сможем теперь строить с их помощью неограниченное количество кривых нормального распределения для сопоставления их с заданными эмпирическими рядами.

Пусть, например, в частности требуется построить нормальную кривую по данным последнего примера ($n=100$, $k=2$, $M=12$ и $\sigma=\pm 4$). Желая выполнить это построение для 6 точек в обе стороны от M (на расстоянии в $\frac{1}{2}\sigma$ между соседними точками), мы должны будем раздвинуть крайние из этих точек на расстояние в 3σ от M (6 интервалов в $\frac{1}{2}\sigma$ составляют в общей сложности 3σ). Так как $\sigma=4$, то $3\sigma=12$. Следовательно обе ветви нашей кривой заполнят на чертеже участок в 12 делений горизонтального масштаба с каждой стороны от M (от $M-3\sigma=0$ до $M+3\sigma=24$).

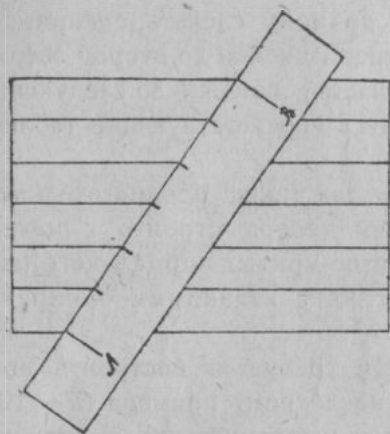


Черт. 19.

Заготовим теперь масштабную линейку в виде полоски бумаги и двумя черточками A и B помечим на ней это расстояние в 12 делений горизонтального масштаба (в данном случае возьмем для ясности этот масштаб несколько крупнее, чем раньше). Итак, расстояние от A до B равно здесь 12 делениям принятого нами горизонтального масштаба (см. черт. 19).

Чтобы разделить это расстояние на 6 равных частей, наложим нашу линейку на первый вспомогательный график так, чтобы черточки A и B точно совпали с верхним и нижним его краем, как это показано на черт. 20 (для этого линейку придется поставить наклонно). Если теперь у всех промежуточных точек (между A и B), в которых край масштабной линейки встречается с параллельными линиями

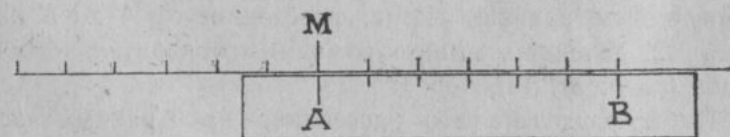
графика, сделать на линейке пометки в виде черточек, то этими черточками расстояние от A до B и разобьется на 6 равных частей.



Черт. 20.

Разделив расстояние AB с помощью первого графика на 6 равных частей, мы, далее, переносим эти деления на тот чертеж, на котором предполагаем построить кривую нормального распределения. Для этого наша линейка прикладывается к точке M краем A , как это показано на черт. 21, и все промежуточные черточки непосредственно переносятся с линейки на горизонтальную ось чертежа. После этого линейка передвигается влево так, чтобы с точкой M совпала правая черточка B , и эти же деления переносятся с линейки на горизонтальную ось влево от M (операция эта на черт. 21 не указана).

Конечно, в данном числовом примере можно было бы и не прибегать к столь сложному способу разбивки отрезка в 12 делений горизонтального масштаба на 6 равных частей, так как эту разбивку можно было бы выполнить просто путем деления 12 на 6. Но в общем случае, когда σ представляет собой дробное или многозначное число, применение указанного графического приема является вполне целесообразным.



Черт. 21.

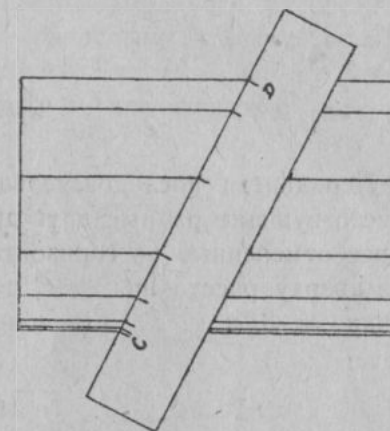
Пометив на горизонтальной оси все точки, для которых далее, следует построить ординаты (y) кривой нормального распределения, мы должны затем обычным путем вычислить

наибольшую высоту этой кривой H (в точке M) по формуле:

$$H = \frac{0,4kp}{\sigma}$$

В нашем примере $H = 20$ делениям вертикального масштаба (см. стр. 25). Пусть эти 20 делений соответствуют отрезку CD , отмеченному черточками C и D по краю бумажной масштабной линейки (см. черт. 22).

Эту линейку мы должны теперь наложить на второй вспомогательный график опять так, чтобы черточки C и D точно совпали с верхним и нижним его краем, как это показано на черт. 22. В точках пересечения края линейки со всеми параллельными прямыми, проведенными на вспомогательном графике, следует поставить черточки (на самой линейке). Этими черточками отрезок CD и разделится на части, равные искомым ординатам y кривой нормального распределения (считая от точки C).

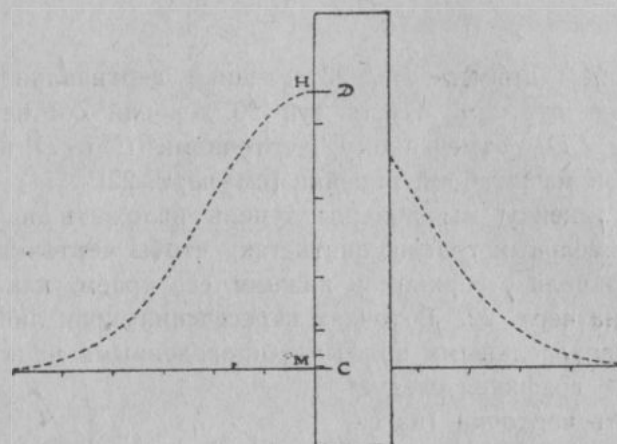


Черт. 22.

Теперь эти ординаты следует перенести на окончательный чертеж. Для этого край масштабной линейки накладывается на чертеж по перпендикулярному к оси абсцисс направлению так, чтобы черточка C на линейке пришлась точно на уровне этой горизонтальной оси в точке M (см. черт. 23). Тогда верхняя черточка линейки D и определит на чертеже положение наивысшей точки H кривой нормального распределения. На чертеже 23 пунктиром изображена для ясности также и нормальная кривая в том виде, как она в дальнейшем получится при построении всех определяющих ее точек.

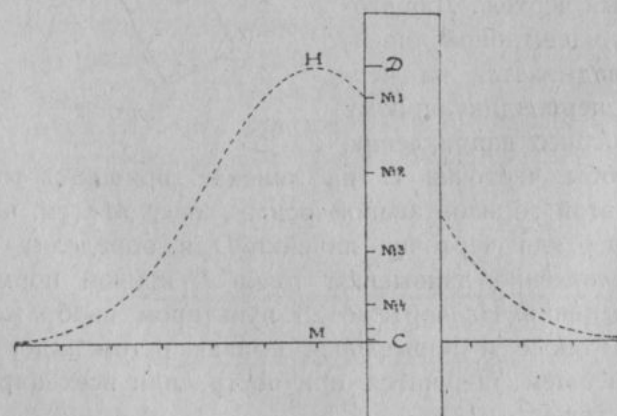
Для построения следующей ординаты этой кривой в соседней с M точке (справа от M на расстоянии $1/2 \sigma$) масштабную линейку следует передвинуть вправо до ближайшего деления горизонтальной оси и отложить кверху соот-

ветствующие расстояния (от C до точки № 1), как это показано на черт. 24.



Черт. 23.

Передвигая последовательно эту линейку вправо на соответствующие равные друг другу расстояния (в $1/2 \sigma$), заранее отмеченные по горизонтальной оси, и отмеряя каждый раз кверху расстояния от C до точки № 2, затем от C до точки № 3 и т. д., мы и получим на чертеже ряд точек,



Черт. 24.

определяющих положение нормальной кривой. Затем эти же построения следует выполнить в левую сторону от точки M .

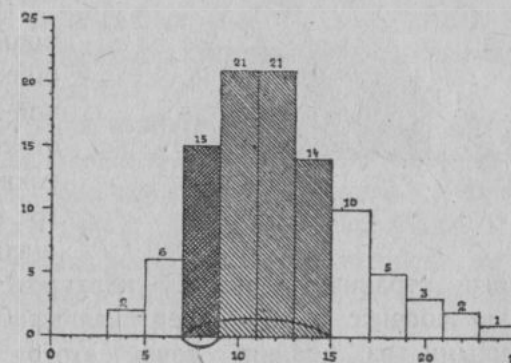
Соединив далее все эти точки плавной линией (от руки или по лекалу) или, в крайнем случае, просто прямыми отрезками, мы и получим искомую кривую нормального распределения, соответствующую заданным величинам n , M и σ .

Полезно запомнить, что при изготовлении вспомогательных графиков необходимо высоту их делать несколько меньше (а не больше) того отрезка, который разделяется этим графиком на соответствующие части (см. черт. 20 и 22).

IV. Процентное распределение вариантов.

Математический анализ кривой нормального распределения указывает также и способ вычисления площади отдельных ее участков, что во многих случаях может представить большой практический и теоретический интерес.

Вспомним, что при вычерчивании гистограммы эмпирического ряда мы строили ее столбики пропорционально частотам p соответствующих классов, в результате чего площадь каждого такого столбика величиной своей точно определяла число ва-



Черт. 25.

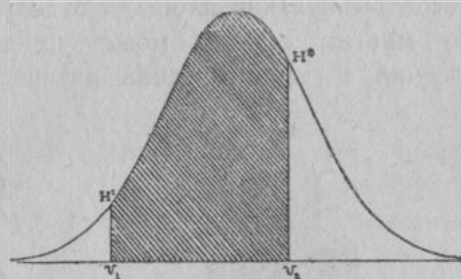
риант, попавших при их разбивке в границы данного класса. Так, например, если в границах класса от 7 до 9 оказалось 15 вариантов ($p=15$), то соответствующий столбик (заштрихованный на чертеже 25 в клетку) должен заключать в себе ровно 15 принятых за единицу квадратных мер площади. Соседний справа столбик, соответствующий $p=21$ (для класса от 9 до 11)—21 таких единиц площади и т. д. Если теперь нам потребовалось бы определить общее количество вариантов, попавших при разбивке в несколько смежных друг с другом классов одно-

временно (например в интервал от 7 до 15 включительно, заключающий в себе 4 таких класса), то соответствующий подсчет частот p этих классов даст нам в сумме

$$15 + 21 + 21 + 14 = 71.$$

Итак, общее количество вариантов, по величине своей больших 7, но меньших 15 (т. е. попавших в интервал от 7 до 15), оказалось здесь = 71.

Это количество вариантов (71) графически соответствует, очевидно, всем четырем столбикам, расположенным на гистограмме в промежутке между точками 7 и 15 (по масштабу горизонтальной оси). Сумма площадей этих четырех столбиков, т. е. общая площадь всего заштрихованного участка гистограммы и будет определять это число вариантов 71



Черт. 26.

(в этом участке гистограммы будет заключаться 71 единица площади).

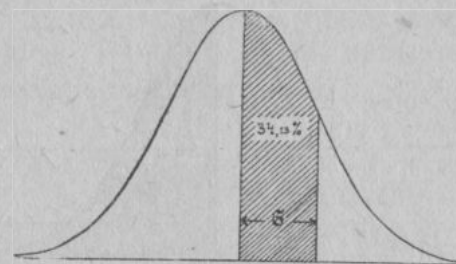
Итак, количество вариантов, попавших при разбивке на классы в какой-либо интервал, состоящий из нескольких таких классов, на гистограмме измеряется площадью того участка, который ограничен внешним контуром самой гистограммы, осью абсцисс и двумя вертикальными ординатами, проведенными из крайних точек этого интервала. Это же соображение относится, очевидно, и к полигону, поскольку его площадь всегда равновелика площади гистограммы.

Так как плавная кривая распределения представляет собой предельную форму соответствующей гистограммы или полигона (при бесконечном увеличении общего количества классов и бесконечном уменьшении классического промежутка k), то этот же принцип графического определения числа вариантов, попавших в данный интервал, может быть применим и ко всякой кривой распределения и, в частности, к нормальной кривой.

Пусть, например, требуется определить число вариантов, попавших при разбивке в интервал от V_1 до V_2 при строго

нормальном их распределении по классам (см. черт. 26). Проведя из точек V_1 и V_2 ординаты H_1 и H_2 до пересечения с кривой

нормального распределения в точках H_1 и H_2 , мы выделим заштрихованный на чертеже 26 участок $V_1 H_1 H_2 V_2$, площадь которого и даст представление об искомой численности вариантов, попавших в границы за-

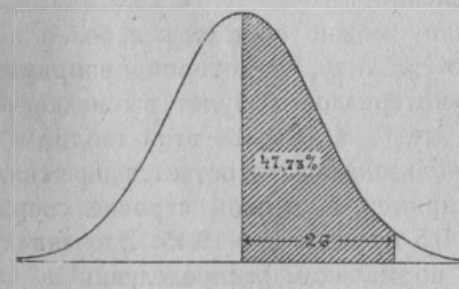


Черт. 27.

данного интервала (этих вариантов окажется столько, сколько квадратных единиц площади заключаются в данном выделенном участке кривой нормального распределения).

Во многих случаях удобнее бывает выражать эту площадь не в абсолютных единицах (напр. в квадратных миллиметрах или в квадратных сантиметрах), а в процентах от всей площади этой кривой.

Анализ геометрических свойств кривой нормального распределения дает возможность вычислять площадь любого участка этой кривой, если известны расстояние границ заданного интервала (V_1 и V_2) до среднего арифметического M , и если эти расстояния выражены в долях среднего квадратического отклонения σ . В виду полной симметрии этой кривой подобные вычисления производятся обычно только

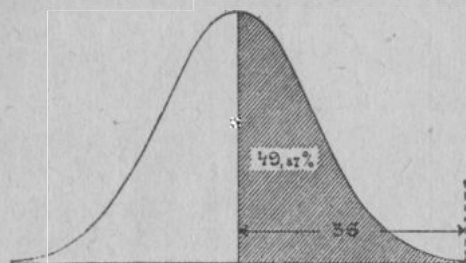


Черт. 28.

для одной какой-либо ее половины, так как справа и слева от центральной точки M площади соответствующих одинаковых участков этой кривой всегда будут в точности равны друг другу.

Соответствующие вычисления (метод и технику которых мы однако здесь приводить не будем) дают в конечном итоге следующие главные результаты:

1) Площадь участка нормальной кривой от M до границы интервала V , отстоящей от M на расстоянии одной σ , составляет



Черт. 28.

приблизительно 24,2% полной площади всей кривой нормального распределения (см. черт. 27).

2) Площадь участка от M до границы V , отстоящей на расстоянии 2σ от M , составляет 47,73% полной площади всей нормальной кривой (см. черт. 28).

3) Площадь участка от M до 3σ составляет 49,87% всей площади кривой (см. черт. 29).

Если расстояние от среднего арифметического M до этой передвижной границы V оцениваемого интервала обозначить буквой d и помнить, что расстояние d выражается всегда в долях среднего квадратического отклонения σ , то указанные выше результаты геометрического анализа нормальной кривой можно изобразить в форме приведенной здесь таблички. Эта табличка состоит из двух вертикальных столбцов d и q . В столбце d указана величина интервала от M до V (в сигмах), а в соседнем справа столбце q дан % вариант, заключающихся в этом интервале при нормальном распределении.

d	$q\%$
1	24,2
2	47,73
3	49,87

По этому же принципу можно составить и более подробную таблицу, в которой, например, ширина интервалов d будет выписана через каждые $1/2\sigma$. Смысл этой таблицы и способ пользования ею остается прежним. Так, например, в первой строчке сверху для $d=0,5$ процент $q=19,15$. Это значит, что при нормальном распределении в какую-либо одну сторону от среднего арифметического M на расстоянии от этого M на 0,5 сигмы будет заключаться 19,15% общего числа всех вариантов (всех измерений n).

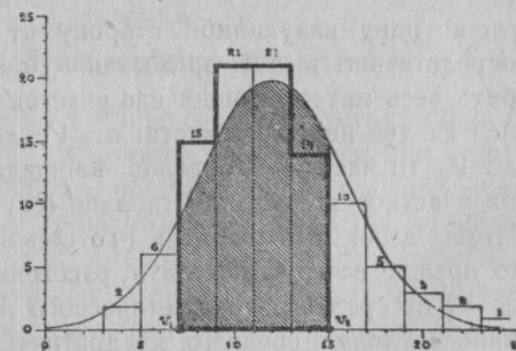
d	$q\%$
0,5	19,15
1	24,2
1,5	43,32
2	47,73
2,5	49,38
3	49,87
3,5	49,98

Ниже приведена таблица процентного распределения вариант для интервалов, увеличивающихся по ширине через каждую десятую долю сигмы (т. е. для последовательных значений $d: 0,1, 0,2, 0,3$ и т. д. до 4,0 включительно). Соответствующие же этим значениям проценты q даны в ней с тремя знаками после запятой.

d	$q\%$	d	$q\%$	d	$q\%$	d	$q\%$
0,1	3,983	1,1	36,433	2,1	48,214	3,1	49,903
0,2	7,226	1,2	37,453	2,2	48,610	3,2	49,931
0,3	11,791	1,3	40,300	2,3	48,928	3,3	49,952
0,4	15,542	1,4	41,924	2,4	49,180	3,4	49,966
0,5	19,146	1,5	43,319	2,5	49,379	3,5	49,977
0,6	22,575	1,6	44,520	2,6	49,534	3,6	49,984
0,7	25,804	1,7	45,543	2,7	49,653	3,7	49,989
0,8	28,814	1,8	46,407	2,8	49,744	3,8	49,993
0,9	31,594	1,9	47,128	2,9	49,813	3,9	49,995
1,0	34,134	2,0	47,725	3,0	49,865	4,0	49,997

Покажем, как следует пользоваться этой таблицей для решения некоторых практических задач. Пусть, например, требуется определить количество вариантов, которое заключалось бы в интервале от $V_1=7$ до $V_2=15$ при нормальном распределении заданного ряда (см. черт. 30).

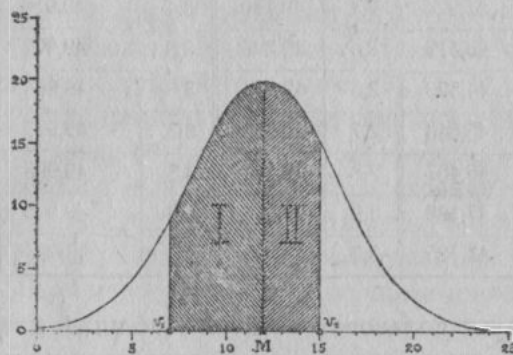
Фактически в этот интервал попало при разбивке на классы столько вариантов, сколько квадратных единиц (соответствующих одной вариантке) заключается в участке площади гистограммы, обведенном на черт. 30 жирной



Черт. 30.

линией (ранее мы высчитали уже, что площадь эта = 71 квадратной единице, и следовательно в указанном интервале заключается 71 варианта из 100, т. е. 71% общего их количества n). Если бы этот ряд имел нормальное распределение, то в том же участке (от $V_1 = 7$ до $V_2 = 15$) оказалось бы, очевидно, иное количество вариантов, именно столько, сколько единиц площади будет иметь заштрихованный на черт. 30 участок нормальной кривой, ограниченный с краев двумя ординатами (в точках V_1 и V_2). Наша задача и заключается теперь в определении площади этого участка, выраженной в процентах от всей площади кривой нормального распределения в целом.

Вспомним, что для данного ряда $M = 12$, а $\sigma = 4$. Следо-



Черт. 31.

жательно, в точке 12 наша нормальная кривая имеет наибольшую ординату, относительно которой она и располагается симметрично. Таблица процентного распределения вариант позволяет вычислять площади только таких участков кривой, которые

лежат по одну какую-либо сторону от средней ординаты, непосредственно к ней примыкая. Поэтому нам придется разбить весь интересующий нас участок этой кривой (от V_1 до V_2) на две неравные части: от V_1 до M (I часть) и от M до V_2 (II часть), определив площадь каждого из двух таких участков порознь по таблице (см. черт. 31).

Чтобы вычислить площадь I-го (левого) участка, необходимо прежде всего определить расстояние его левой границы V_1 от среднего арифметического M , выразив это расстояние в долях среднего квадратического отклонения σ . Так как граница V_1 по условию совпадает с 7 делением горизонтального масштаба, то в данном случае

$$V_1 = 7.$$

Отклонение (расстояние) этой границы до $M = 12$ выразится очевидно разностью

$$V_1 - M = 7 - 12 = -5.$$

Итак, граница V_1 отстоит от точки M влево (знак —) на расстояние = 5 делениям горизонтального масштаба. Чтобы выразить это расстояние в долях среднего квадратического отклонения, следует, очевидно, разделить его на σ (у нас $\sigma = 4$)

$$\frac{5}{4} = 1,25 \text{ или приблизительно } = 1,3.$$

(Знак — здесь всюду отброшен).

Эта величина (1,3) и даст нам искомое значение d , для которого можно определить по таблице соответствующий процент q . Итак для I части заштрихованного участка

$$d = 1,3.$$

Значению $d = 1,3$ по таблице процентного распределения вариант (см. стран. 37) соответствует $q = 40,3$ (приблизительно). Следовательно слева от M в I участке заштрихованной площади при нормальном распределении должно заключаться 40,3% общего количества всех вариантов.

Для II-го участка крайняя его граница $V_2 (= 15)$ отстоит от $M (= 12)$ на расстояние

$$V_2 - M = 15 - 12 = 3,$$

т. е. на 3 деления горизонтального масштаба. Разделив это расстояние (3) на $\sigma = 4$, получим:

$$\frac{3}{4} = 0,75 \text{ или приблизительно } = 0,8.$$

Значение $d = 0,8$ по таблице соответствует $q = 28,8$. Следовательно вправо от M должно заключаться при нормальном распределении 28,8% общего количества всех вариантов.

Итак, влево от M в заштрихованный участок нормальной кривой попадет 40,3%, а справа от M окажется 28,8%. Поэтому всего в границах от V_1 до V_2 будет заключаться при нормальном распределении:

$$40,3 + 28,8 = 69,1\% *),$$

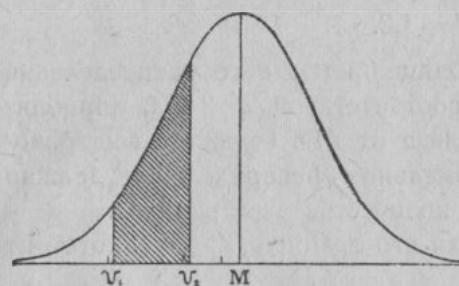
т. е. на 1,9% меньше, чем оказалось там фактически (71%).

*) При точном подсчете этого % с помощью таблицы, приложенной в конце книги, получим 66,77%.

Итак, при вычислении площади участка кривой нормального распределения в том случае, когда среднее арифметическое M расположено внутри его границ (т. е. между V_1 и V_2), следует разбить весь заданный интервал на две неравные части, расположенные слева и справа от M . Для каждого из двух полученных этим путем участков нормальной кривой следует, далее, вычислить отклонение его крайней границы V от среднего арифметического M , выразив это расстояние в долях среднего квадратического отклонения σ , т. е. по формуле:

$$d = \frac{V - M}{\sigma} *).$$

Отыскав по таблице проценты q , соответствующие полученным значениям d , мы и определим площади каждой из двух частей интересующего нас участка. Для вычисления же площади всего этого участка в целом придется оба эти числа (q) сложить.



Черт. 32.

Рассмотрим теперь второй случай вычисления площади участка нормальной кривой, расположенного по одну какую либо сторону от M (см. черт. 32). Здесь искомая (заштрихованная) площадь представляет собой разность между площадью участка от M до V_1 и площадью участка от M до V_2 .

Пусть в числах: $V_1 = 5$, $V_2 = 9$ при тех же значениях $M = 12$ и $\sigma = 4$.

Для участка, крайняя граница которого V_1 удалена от M на более далекое расстояние, величина

$$d_1 = \frac{V_1 - M}{\sigma} = \frac{5 - 12}{4} = \frac{-7}{4} = -1,75$$

или приблизительно $d_1 = 1,8$ (без знака —).

*) Для вычисления по таблице значение d всегда берется со знаком +.

Для участка, крайняя граница которого V_2 расположена ближе к M :

$$d_2 = \frac{V_2 - M}{\sigma} = \frac{9 - 12}{4} = \frac{-3}{4} = -0,75$$

или приблизительно $d_2 = 0,8$ (без знака —).

Значению $d_1 = 1,8$ соответствует процент $q_1 = 46,4 *$

„ $d_2 = 0,8$ „ „ „ $q_2 = 28,8 *$

разность этих процентов $q = q_1 - q_2 = 17,6 **$.

Следовательно в заданных границах (от 5 до 9) при нормальном распределении должно заключаться 17,6% общего количества всех вариантов n (фактически их оказалось там 21%, что легко определить непосредственно по гистограмме путем подсчета частот p в соответствующих классах: $6 + 15 = 21$).

Итак, при вычислении площади участка нормальной кривой в том случае, когда обе его границы (V_1 и V_2) располагаются по одну сторону от среднего арифметического M , следует определить по таблице площадь участков кривой от M до наиболее удаленной границы (напр. V_1), затем, независимо от этого, площадь участка от M до ближайшей границы (V_2) и вычесть из первой (большой) площади вторую (меньшую).

С помощью указанных приемов можно вычислить и площади отдельных полосок нормальной кривой, совпадающих по своей ширине и расположению со столбиками гистограммы эмпирического ряда. Этим путем мы получим так называемые „теоретические частоты“ (p_0), показывающие, сколько вариант из общего их числа n попало бы при разбивке в каждый отдельный класс, если бы данный ряд имел строго нормальное распределение. Вычислив эти теоретические частоты p_0 , мы получаем, далее, возможность сравнить их с эмпирическими частотами p , найденными путем непосредственного подсчета вариантов, заключающихся в границах каждого отдельного класса. На основании такого сопоставления этих частот (p_0 и p) можно судить о степени

*) Приблизительно.

**) Точнее 18,65, если вычисления вести по таблице, приложенной в конце книги.

соответствия или, наоборот, несоответствия заданного эмпирического ряда теоретическому „нормальному“ ряду. При такого рода вычислениях крайними границами (V_1 и V_2) измеряемых участков нормальной кривой будут очевидно служить границы соответствующих классов.

Так как в конечном итоге этих вычислений мы получаем проценты q (от общего числа измерений n), а теоретические частоты p_0 желательно иметь в виде абсолютных (а не процентных) чисел, то для окончательного определения каждой теоретической частоты p_0 следует пользоваться формулой:

$$p_0 = \frac{nq}{100}.$$

Так, например, если из общего числа 200 вариант 6% попадает при нормальном распределении в границы какого-либо класса ($n=200$, $q=6$), то теоретическая частота этого класса

$$p_0 = \frac{nq}{100} = \frac{200 \cdot 6}{100} = 12.$$

Очевидно, теоретические частоты p_0 могут получиться и в виде дробных чисел, в то время как эмпирические частоты p являются всегда числами целыми.

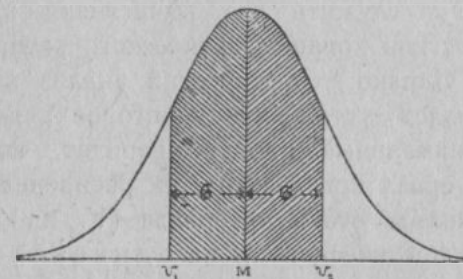
Вычисление теоретических частот p_0 удобнее производить с помощью более подробной таблицы процентного распределения вариант, приложенной в конце книги (см. стр. 421). В этой таблице значения d выписаны через каждую сотую долю сигмы, а соответствующие им проценты q даны с четырьмя знаками после запятой.

Таблицы подобного типа и предназначения в руководствах по статистике обычно называются „таблицами значений нормального интеграла вероятностей“, что объясняется способом их составления, основанным на положениях теории вероятностей и правилах интегрального исчисления.

Воспользуемся этой таблицей для определения площади некоторых участков нормальной кривой, для которых границы V_1 и V_2 отстоят от M на одинаковое расстояние. Пусть, например (черт. 33) эти границы раздвинуты равно на одну σ от M . В этом случае для обеих границ $d=1$ и, стало быть, $q=34.1345$. Полная же площадь обеих половин

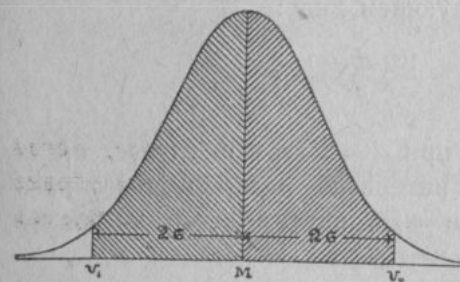
этого участка будет вдвое больше q , т. е. окажется равной 68,2690% всей площади нормальной кривой в целом.

Если же эти границы (V_1 и V_2) раздвинуть на расстояние вдвое большее, т. е. на 2σ от M (см. черт. 34), то для каждой половины этой площади $d=2$ и $q=47,7250$, а вся площадь участка составит 95,4500%.



Черт. 33.

Для случая: $d=3$, представленного на чертеже 35, процент $q=49,8650$, а обе половины этого участка, вместе взятые, составят 99,7300%.



Черт. 34.

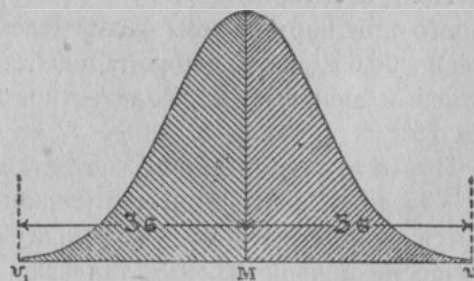
Этот процент (99,73%) является уже настолько близким к 100, что существующей между ними разницей в 0,27% практически можно совершенно пренебречь и считать, что внутри границ,

удаленных от M на 3σ , заключаются все варианты данного ряда. На этом основании отклонение отдельных вариант от M на три σ может считаться максимальным.

По отношению к вариантам, отклоняющимся от M на большую величину, может даже возникнуть вопрос о принадлежности их к какому-либо иному ряду, так

как такие отклонения не являются уже типичными.

Таблица процентного распределения вариант составлена в применении только к нормальной кривой. В тех же слу-



Черт. 35.

чаях, когда изучаемый ряд значительно отличается от нормального, приведенные в этой таблице проценты q уже не могут служить для вычисления площади тех или иных участков кривой фактического распределения.

Однако, теоретический анализ вариационных рядов позволяет установить некоторое неравенство, определяющее минимальный процент вариант, попадающих в заданный интервал при любом их распределении, если известно, что границы этого интервала (V_1 и V_2) отстоят от среднего арифметического M на одинаковое расстояние d (выраженное, как всегда, в долях σ).

Обозначив процентное количество вариант, попадающих в обе симметрично расположенные части такого интервала (от V_1 до M и от M до V_2) через q_1 и q_2 , мы можем записать это неравенство в следующем виде:

$$q_1 + q_2 > \frac{100(d^2 - 1)}{d^2}.$$

В частности, например, при $d=2$ (т. е. в случае, когда обе границы исследуемого интервала расположены справа и слева от M на расстоянии $=2\sigma$) внутри этого интервала всегда окажется не меньше, чем

$$\frac{100(d^2 - 1)}{d^2} = \frac{100(4 - 1)}{4} = 75\%$$

общего количества всех вариант n .

Итак, в интервале от $M - 2\sigma$ до $M + 2\sigma$, в границы которого при нормальном распределении попадает приблизительно 95,65% всех вариант, в действительности может оказаться и меньшее их количество, однако все же больше, чем 75%.

При $d=3$ в границах соответствующего интервала (от $M - 3\sigma$ до $M + 3\sigma$) при нормальном распределении заключается приблизительно 99,73% всех вариант и во всяком случае не меньше 88,89%, так как

$$\frac{100(d^2 - 1)}{d^2} = \frac{100(9 - 1)}{9} = 88,89$$

и следовательно

$$q_1 + q_2 > 88,89\%.$$

Этим же способом можно вычислить наименьший процент вариант и для каких угодно промежуточных (дробных) значений d .

Полезно запомнить, что при $d=1$

$$\frac{100(d^2 - 1)}{d^2} = \frac{100(1 - 1)}{1} = 0$$

и следовательно для интервала от $M - \sigma$ до $M + \sigma$

$$q_1 + q_2 > 0.$$

Это значит, что по отношению к интервалу $M \pm \sigma$ мы можем ручаться лишь за то, что какое-либо количество вариант в него непременно попадет ($q_1 + q_2$ не $= 0$, а больше 0), но и только.

Для интервалов же, границы которых раздвинуты от M на расстояние, меньшее σ , например при $d=0,9$

$$\frac{100(d^2 - 1)}{d^2} = \frac{100(0,81 - 1)}{0,81} = -23,5$$

и следовательно

$$q_1 + q_2 > -23,5.$$

Стало быть в этом случае мы не можем уже ручаться

даже за попадание в этот интервал хотя бы и одной варианты, т. к. в таких суженных границах, при некоторых особенно неблагоприятных условиях распределения может вовсе не оказаться ни одной варианты (например при провале кривой распределения в ее средней части).

Выше приведена сравнительная таблица процентного распределения вариант при нормальном и при каком угодно их распределении для некоторых интервалов.

d	Границы интервалов		Минимальный процент вариант при нормальном распределении	Минимальный процент вариант при любом распределении
	от	до		
1	$M - \sigma$	$M + \sigma$	68,3	0
1,5	$M - 1\frac{1}{2}\sigma$	$M + 1\frac{1}{2}\sigma$	86,6	55,5
2	$M - 2\sigma$	$M + 2\sigma$	95,5	75,0
2,5	$M - 2\frac{1}{2}\sigma$	$M + 2\frac{1}{2}\sigma$	98,8	84,0
3	$M - 3\sigma$	$M + 3\sigma$	99,7	88,9
3,5	$M - 3\frac{1}{2}\sigma$	$M + 3\frac{1}{2}\sigma$	99,9	91,8

V. Показатель асимметрии (А) и показатель эксцесса (Е).

Графическое сопоставление гистограммы или полигона с кривой нормального распределения, так же как и сравнение эмпирических частот с теоретическими, дает возможность судить о соответствии данного фактического распределения вариант с тем образцовым их распределением, которое носит название „нормального“. Однако полного совпадения эмпирического (полученного из опыта) ряда с рядом, имеющим строго нормальное распределение, в действительности не наблюдается никогда. Пусть, например, в каком-либо классе теоретически должно заключаться, скажем, 3,2 варианты, а фактически там оказалось только 3 варианты ($p_0 = 3,2$, а $p = 3$). Спрашивается, является ли эта разница в 0,2 варианты показателем существенного расхождения сравниваемых распределений, или такое малое различие, да к тому же еще и дробное, объясняется здесь простой случайностью или тем обстоятельством, что эмпирическая частота $p = 3$ в качестве целого числа и не может совпадать с дробным значением теоретической частоты $p_0 = 3,2$? Такие же вопросы возникают и по отношению ко всем остальным парам сопоставляемых друг с другом частот p_0 и p или, вернее, по отношению ко всей их совокупности в целом.

Если бы даже нам и удалось установить такой объективный критерий оценки для каждой отдельной пары сравниваемых частот p_0 и p , то все они, вместе взятые, все же продолжали бы в некоторых случаях возбуждать наши сомнения. В самом деле, пусть, например, в главной своей части данный ряд на основании такого критерия вполне слился с нормальной кривой, но кое-где все же отошел от нее на заметную величину. Как в этом случае мы будем квалифицировать заданное распределение? Посчитаем ли этот ряд все же нормальным или признаем его уклоняющимся от нормального и в какой степени?

Эти соображения вынуждают нас установить такие общие числовые характеристики совпадения заданного ряда с кривой нормального распределения, которые своей величиной и знаком (+ или —) точно определяли бы принадлежность эмпирического распределения вариант к типу „нормального“

или, в случае несоответствия, указывали бы на его степень, характер и форму.

Для уяснения принципа составления таких числовых показателей рассмотрим следующий схематический пример:

Пусть имеем два ряда, каждый из которых состоит для простоты всего лишь из 10 вариантов:

I ряд: 2, 2, 2, 2, 3, 3, 3, 4, 4 и 5

II ряд: 1, 2, 2, 3, 3, 3, 4, 4, 4 и 4.

Вычислим прежде всего для каждого из этих рядов среднее арифметическое M и среднее квадратическое отклонение σ .

Сумма всех вариант I ряда:

$$\Sigma V = 30.$$

Так как общее число этих вариант $n = 10$, то среднее арифметическое для I ряда

$$M_1 = \frac{\Sigma V}{n} = \frac{30}{10} = 3.$$

Вычислив отклонения x отдельных вариант V этого ряда от найденного значения $M_1 = 3$ (по формуле $x = V - M_1$), получим:

$$-1, -1, -1, -1, 0, 0, 0, +1, +1, +2.$$

Возвысив все эти отклонения в квадрат, будем иметь (x^2):

$$1, 1, 1, 1, 0, 0, 0, 1, 1, 4.$$

Сумма квадратов всех отклонений

$$\Sigma x^2 = 10.$$

Следовательно, среднее квадратическое отклонение для I ряда

$$\sigma_1 = \pm \sqrt{\frac{\Sigma x^2}{n}} = \pm \sqrt{\frac{10}{10}} = \pm \sqrt{1} = \pm 1.$$

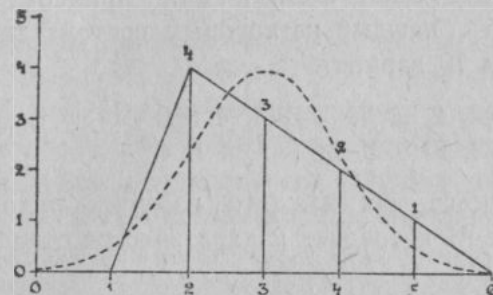
Итак для I ряда $M_1 = 3$, а $\sigma_1 = \pm 1$.

Проделав эти вычисления и для II ряда, получим совершенно те же числовые значения M и σ (в чем нетрудно убедиться, выполнив соответствующие вычисления в действительности). Итак

$$M_2 = 3 \text{ и } \sigma_2 = \pm 1.$$

Чем же, спрашивается, отличаются заданные ряды друг от друга?

Для ответа на этот вопрос выясним повторяемость в этих рядах отдельных числовых значений вариант V (т. е. подсчитаем для них частоты p).



Черт. 36.

В I ряду самая малая варианта $V=2$ повторяется 4 раза ($p=4$), следующая варианта $V=3$ повторяется 3 раза ($p=3$), варианта

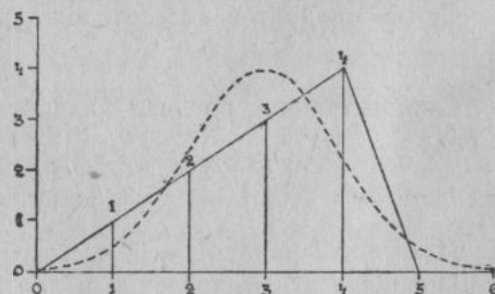
$V=4$ повторяется 2 раза ($p=2$), а самая большая варианта $V=5$ всего только 1 раз ($p=1$).

Вычертив этот ряд в форме полигона, получим график распределения, представленный на чертеже 36. Выполнив аналогичный подсчет частот p и соответствующее построение полигона для II ряда, получим график распределения, представленный на чертеже 37.

Сравнивая эти полигоны друг с другом и с нормальной кривой, мы видим, что I ряд имеет ясно выраженную асимметрию со сдвигом вершины полигона влево (левосторонняя асимметрия), а II ряд является асимметричным в противоположном направлении (правосторонняя асимметрия).

Эти заметные на графике нарушения симметрии в нашем примере вовсе не отразились на величине средних арифметических M и средних квадратических отклонений σ (оказавшихся в обоих случаях совершенно

одинаковыми), и стало быть помимо этих основных числовых характеристик мы должны сконструировать какой-то



Черт. 37.

новый показатель, который улавливал бы наличие и степень ε -ой асимметрии.

Подобно тому, как σ вычисляется из квадратов центральных отклонений (x^2), такой показатель асимметрии (A) удобно составить с помощью их третьих степеней (кубов), сумма которых (Σx^3) и явится, как увидим дальше, существенной частью этого показателя.

Выпишем в строчку найденные уже ранее отклонения (x) всех 10-ти вариантов I ряда, под ними во вторую строчку запишем их квадраты (x^2), а ниже поместим кубы этих отклонений (x^3):

$$\begin{array}{rcccccccccc} x = & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & +1 & +1 & +2 \\ x^2 = & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 4 \\ x^3 = & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & +1 & +1 & +8 \end{array}$$

Сумма чисел первой строчки (Σx), в качестве суммы центральных отклонений, здесь, как и всегда, равна 0; сумма квадратов этих отклонений $\Sigma x^2 = 10$, а сумма кубов $\Sigma x^3 = +6$.

Прodelав аналогичные вычисления для II ряда, будем иметь:

$$\begin{array}{rcccccccccc} x = & -2 & -1 & -1 & 0 & 0 & 0 & +1 & +1 & +1 & +1 \\ x^2 = & 4 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ x^3 = & -8 & -1 & -1 & 0 & 0 & 0 & +1 & +1 & +1 & +1 \end{array}$$

Суммы чисел в этих строчках соответственно равны: 0, 10 и — 6.

Сравним теперь эти результаты по рядам:

$$\begin{array}{ll} \text{Для I ряда:} & \text{Для II ряда:} \\ \Sigma x = & 0 \\ \Sigma x^2 = & 10 \\ \Sigma x^3 = & +6 \end{array}$$

Из этого сопоставления видно, что различие в асимметрии заданных рядов отразилось здесь лишь на сумме третьих степеней центральных отклонений, причем для случая левосторонней асимметрии (I ряд) эта Σx^3 оказалась величиной положительной (+6), а для правосторонней (II ряд) — отрицательной (— 6).

Это и понятно. При левосторонней, например, асимметрии вершина кривой, а стало быть и среднее арифметическое M отодвигается влево, а расположенные справа от M отдельные варианты V получают, вследствие этого, большие по своей абсолютной величине отклонения. Так как они имеют знак $+$, то при возвышении в 3-тью степень эти большие положительные отклонения возрастут особенно сильно и при суммировании дадут заметный перевес над кубами сравнительно малых отрицательных отклонений. Поэтому сумма Σx^3 и оказывается здесь величиной положительной.

В случае правосторонней асимметрии картина получается обратная, и в сумме Σx^3 начинают уже превалировать отрицательные слагаемые.

Таким образом знак суммы Σx^3 ($+$ или $-$) прямо указывает на характер имеющейся асимметрии, а значение суммы $\Sigma x^3 = 0$, как это нетрудно догадаться — на отсутствие асимметрии, т. е. на строго симметричный характер заданного ряда. В самом деле, при симметричном распределении вариант относительно среднего арифметического M каждое положительное отклонение ($+x$) в правой части ряда имеет равное себе по абсолютной величине отрицательное отклонение ($-x$) в левой части данного ряда. Третьи степени этих отклонений ($+x^3$ и $-x^3$) также, очевидно, сохраняют это равенство, в силу чего они и погашают друг друга при подсчете общей суммы.

Но этого мало. Величина суммы Σx^3 дает представление не только о характере, но также и о величине асимметрии. При сравнительно небольшом сдвиге вершины кривой описанное выше превалирование положительных или отрицательных отклонений будет, очевидно, также незначительным в то время как при сильных степенях асимметрии этот перевес отклонений одного знака над отклонениями противоположного знака будет также усилен. Стало быть величина суммы Σx^3 определяет собой не только характер, но и степень асимметрии заданного ряда.

Однако непосредственное использование суммы Σx^3 в качестве показателя асимметрии все же представляет некоторое неудобство. В нашем числовом примере для I ряда $\Sigma x^3 = +6$, а для II ряда $\Sigma x^3 = -6$. Нетрудно убедиться

в том, что суммы Σx^3 в обоих случаях зависят не только от характера и степени асимметрии, но также и от общего числа вариант n . В этом примере $n = 10$. Если бы при совершенно такой же асимметрии число вариант n было бы вдвое большим (например каждая варианта встречалась бы дважды), то и сумма Σx^3 увеличилась бы, очевидно, также в 2 раза (так как каждый куб отклонения в этом случае повторялся бы дважды).

Чтобы избавиться от этой нежелательной зависимости суммы Σx^3 от числа измерений n , можно сконструировать показатель асимметрии в виде отношения суммы Σx^3 к числу измерений n (т. е. разделить Σx^3 на n). Тогда увеличение или уменьшение числителя этой дроби (т. е. суммы Σx^3), зависящее от числа измерений n , будет автоматически компенсировано точно таким же увеличением или уменьшением знаменателя (n), и значение всей дроби в целом останется неизменным (т. е. не будет уже зависеть от числа измерений n).

Но и этого мало. В нашем примере как для I, так и для II ряда среднее квадратическое отклонение $\sigma = \pm 1$. Нетрудно догадаться, что при неодинаковом варьировании этих рядов (т. е. при различных значениях σ) обе суммы Σx^3 оказались бы также несравнимыми друг с другом. Там, где ряд более разбросан относительно M (т. е. где σ больше), там и все отклонения x окажутся большими, и притом во столько же раз. Кубы же этих отклонений, а стало быть и вся сумма Σx^3 возрастет, очевидно, пропорционально третьей степени, соответствующей σ (т. е. пропорционально σ^3). Чтобы уравнивать эти результаты и в данном отношении, следует разделить сумму Σx^3 не только на n , но и на σ^3 . Тогда полученный этим путем показатель асимметрии окажется независимым также и от величины рассеяния заданного ряда (т. е. от σ) и будет определять только характер и степень асимметрии заданного ряда.

Итак, для вычисления показателя асимметрии (обозначим его буквой A) может служить следующая окончательная формула:

$$A + \frac{\Sigma x^3}{n\sigma^3} *)$$

*) Здесь σ берется всегда со знаком $+$.

В нашем примере (при $n = 10$ и $\sigma = \pm 1$) для I ряда:

$$A_1 = \frac{+6}{10 \cdot 1} = +0,6$$

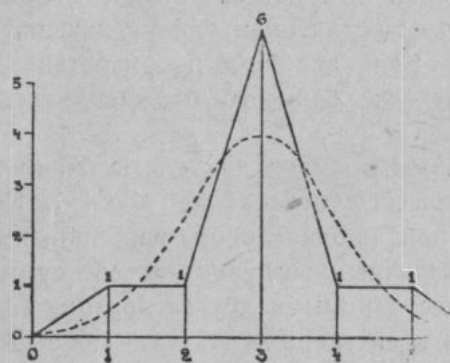
а для II ряда:

$$A_2 = \frac{-6}{10 \cdot 1} = -0,6$$

Понятие об асимметрии не охватывает, однако, всех типовых случаев нарушения нормального распределения. Пусть, например, имеем следующие два ряда (III и IV), как и ранее состоящие для простоты всего лишь из 10 вариантов:

III ряд: 1 2 3 3 3 3 3 3 4 и 5

IV ряд: 2 2 2 2 2 4 4 4 4 и 4

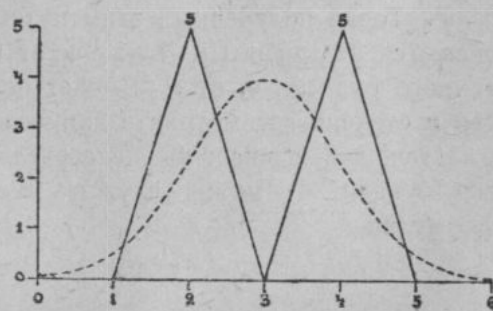


Черт. 38.

Для этих рядов по-прежнему: $M = 3$ и $\sigma = \pm 1$. Соответствующие полигоны в сопоставлении с нормальной кривой приведены на черт. 38 и 39.

Здесь в III ряду вершина полигона заметно выступает за пределы нормальной кривой (вытянута вверх), а в IV ряду наблюдается,

наоборот, провал середины полигона. Иначе говоря, в первом случае подавляющее большинство вариантов (6 из 10) сосредоточилось главным образом в центре ряда, но зато отдельные варианты оказались отброшенными в обе стороны от M на сравнительно большое расстояние, а во втором большинстве вариантов, наоборот, равномерно распределилось по обе стороны от M (сравнительно



Черт. 39.

недалеко от него) засчет ущерба в центральной части ряда. Такое нарушение нормального распределения ряда при сохранении полной его симметрии носит название „эксцесса“. Эксцесс, выражающийся высоковершинностью полигона, мы будем в дальнейшем называть прямым, а в случае низковершинности, плосковершинности или провала ряда в середине — говорить об эксцессе обратном.

Каким же показателем можно охарактеризовать направление и величину эксцесса? В целях решения этого вопроса выпишем для каждого из двух заданных выше рядов первые 4 степени их центральных отклонений.

Для III ряда:

$x =$	-2	-1	0	0	0	0	0	0	+1	+2
$x^2 =$	4	1	0	0	0	0	0	0	1	4
$x^3 =$	-8	-1	0	0	0	0	0	0	+1	+8
$x^4 =$	16	1	0	0	0	0	0	0	1	16

Для IV ряда:

$x =$	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1
$x^2 =$	1	1	1	1	1	1	1	1	1	1
$x^3 =$	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1
$x^4 =$	1	1	1	1	1	1	1	1	1	1

Подсчитав для каждого из этих рядов суммы первых, вторых, третьих и четвертых степеней всех отклонений x , получим:

Для III ряда:

$$\begin{aligned}\Sigma x &= 0 \\ \Sigma x^2 &= 10 \\ \Sigma x^3 &= 0 \\ \Sigma x^4 &= 34\end{aligned}$$

Для IV ряда:

$$\begin{aligned}\Sigma x &= 0 \\ \Sigma x^2 &= 10 \\ \Sigma x^3 &= 0 \\ \Sigma x^4 &= 10\end{aligned}$$

Отсюда видно, что разница в характере распределения этих рядов сказалась лишь на величине суммы 4-х степеней центральных отклонений (Σx^4), которая для случая прямого эксцесса (III ряд) получилась значительно больше, чем для случая обратного эксцесса (IV ряд).

Однако непосредственное использование этой суммы (Σx^4) в качестве показателя эксцесса (E) представляется не-

удобным, поскольку эта сумма Σx^4 , подобно сумме Σx^3 , зависит кроме того от числа измерений n и от среднего квадратического отклонения σ (вернее от его 4-й степени: σ^4). Чтобы избавиться от этих привносящих влияния, следует, как и раньше, разделить сумму Σx^4 на n и на σ^4 .

В результате получим выражение вида:

$$\frac{\Sigma x^4}{n\sigma^4}$$

по величине которого и можно судить о характере и степени эксцесса.

В нашем примере (при $n = 10$ и $\sigma = \pm 1$) для III ряда выражение

$$\frac{\Sigma x^4}{n\sigma^4} = \frac{34}{10 \cdot 1} = 3,4$$

и для IV ряда:

$$\frac{\Sigma x^4}{n\sigma^4} = \frac{10}{10 \cdot 1} = 1$$

Эти числовые характеристики эксцесса (3,4 и 1) интересно, далее, сравнить с тем значением выражения

$$\frac{\Sigma x^4}{n\sigma^4}$$

которое соответствует случаю строго нормального распределения (т. е. отсутствию какого бы то ни было эксцесса).

Математический анализ кривой нормального распределения показывает, что для этой кривой предельная величина выражения

$$\frac{\Sigma x^4}{n\sigma^4} = 3$$

Сравнивая наши данные (3,4 и 1) с этим постоянным числом 3, мы замечаем, что числовая характеристика прямого эксцесса (3,4) в нашем примере (III ряд) превышает число 3 на величину $= 0,4$, а в примере обратного эксцесса (для IV ряда) оказывается меньше 3 на величину $= 2$. Эти окончательные результаты сопоставления числовых значений выражения

$$\frac{\Sigma x^4}{n\sigma^4}$$

с постоянным значением этого же выражения для случая нормального распределения (т. е. числом 3) и принимаются обычно за показатель эксцесса E .

Таким образом показатель эксцесса E вычисляется по формуле:

$$E = \frac{\Sigma x^4}{n\sigma^4} - 3.$$

В наших примерах при $n = 10$ и $\sigma = \pm 1$ для случая прямого эксцесса (III ряд)

$$E_3 = +0,4$$

а для обратного эксцесса (IV ряд)

$$E_4 = -2.$$

Прямой эксцесс, дающий всегда показатель E со знаком $+$, называется вследствие этого положительным, а обратный (дающий у E знак $-$) отрицательным.

Само собой разумеется, что наши числовые примеры асимметрии и эксцесса, искусственно составленные всего лишь из 10 вариантов различного распределения, ни в коем случае не могут считаться достаточными в количественном отношении. Мы взяли их здесь в этом виде исключительно лишь с целью большей наглядности описания метода вычисления показателей A и E , дабы иметь возможность одновременно обозреть всю совокупность этих вариантов в целом.

Вычислив показатели A и E по формулам:

$$A = \frac{\Sigma x^3}{n\sigma^3} \quad \text{и} \quad E = \frac{\Sigma x^4}{n\sigma^4} - 3$$

необходимо определить также и их средние ошибки m_A и m_E , пользуясь формулами:

$$m_A = \pm \sqrt{\frac{6}{n}} \quad \text{и} \quad m_E = \pm 2 \sqrt{\frac{6}{n}}.$$

Из этих формул видно, что ошибки m_A и m_E не зависят от величины самих показателей A и E и определяются только числом измерений n . Для вычисления же ошибки m_E можно взамен определения ее по вышеуказанной формуле просто умножить ошибку m_A на 2

$$m_E = 2m_A.$$

чения показателя A и положительные значения E не ограничены никакими теоретическими пределами (т. е. могут при некоторых условиях достигать любой абсолютной величины), в то время как отрицательные значения показателя E колеблются лишь в пределах от 0 до -2 . Значение $E = -2$ является уже предельным, соответствующим наибольшей низковершинности ряда, выражающейся провалом его по середине, доходящим до самой оси абсцисс. В этом случае, как видно из черт. 39, данный ряд распадается на 2 отдельные части, являющиеся как бы двумя самостоятельными рядами.

VI. Вычисление показателей A и E при большом числе измерений n .

На приведенных в прошлой главе упрощенных примерах мы разобрали общий способ определения показателей A и E , сводящийся в главной своей части к вычислению суммы третьих и четвертых степеней центральных отклонений (Σx^3 и Σx^4). При наличии в заданных рядах всего лишь по 10 отдельных вариантов в каждом, эти вычисления можно было выполнить там непосредственно для каждого центрального отклонения (x) отдельно. Однако при большем числе этих вариантов (а стало быть и отклонений) такой способ вычислений является чрезвычайно обременительным. Выход из создавшегося здесь затруднения может быть найден в использовании следующего приема.

Пусть имеем по прежнему 10 центральных отклонений (x), полученных ранее для I-го из четырех разобранных рядов:

$$-1 \quad -1 \quad -1 \quad -1 \quad 0 \quad 0 \quad 0 \quad +1 \quad +1 \quad +2$$

Третьи степени (кубы) этих отклонений будут равны:

$$-1 \quad -1 \quad -1 \quad -1 \quad 0 \quad 0 \quad 0 \quad +1 \quad +1 \quad +8$$

Взамен непосредственного подсчета всех этих чисел, мы имеем возможность сгруппировать их предварительно по классам и, перемножив отдельные повторяющиеся значения x^3 на число их повторений p , сложить все полученные этим

путем произведения вида $p x^3$, в результате чего и получим очевидно то же самое значение искомой суммы Σx^3 *). Эта группировка дает следующие результаты:

-1	-1	-1	-1	0	0	0	+1	+1	+8
4 раза				3 раза			2 раза		1 раз
$p=4$				$p=3$			$p=2$		$p=1$

Для подсчета общей суммы Σx^3 этим способом следует значение $x^3 = +8$ помножить на его повторяемость (частоту) $p=1$, значение $x^3 = +1$ помножить на $p=2$, $x^3 = 0$ — на $p=3$, а $x^3 = -1$ — на $p=4$ (идя в направлении справа налево)

$$\begin{aligned} (+8) \cdot 1 &= +8 \\ (+1) \cdot 2 &= +2 \\ (0) \cdot 3 &= 0 \\ (-1) \cdot 4 &= -4 \end{aligned}$$

Сумма этих произведений, как и следовало ожидать (см стр. 49), по прежнему

$$\Sigma x^3 = +6.$$

Точно также можно поступить и при подсчете суммы четвертых степеней этих отклонений. Для III-го ряда мы имеем например следующие значения x :

$$-2 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad +1 \quad +2$$

и, стало быть, следующий ряд их четвертых степеней (x^4):

$$16 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 16$$

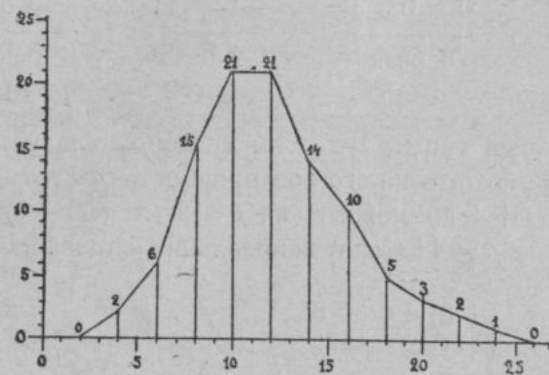
Перемножив все эти значения x^4 на соответствующие им частоты p , получим:

$$\begin{aligned} 16 \cdot 1 &= 16 \\ 1 \cdot 1 &= 1 \\ + \quad 0 \cdot 6 &= 0 \\ 1 \cdot 1 &= 1 \\ 16 \cdot 1 &= 16 \\ \hline \text{сумма } \Sigma x^4 &= 34 \end{aligned}$$

*) Эту сумму правильнее было бы в данном случае обозначить через $\Sigma p x^3$, чего однако мы делать здесь не будем, чтобы не вводить лишних символов.

что мы имели и раньше, непосредственно складывая отдельные значения x^4 (см. стр. 53).

Воспользуемся теперь этим приемом при вычислении показателей A и E для ряда, состоящего из сравнительно



Черт. 40.

большого числа вариантов. Пусть имеем например ряд, уже разобранный нами раньше (см. его полигон, представленный здесь на чер. 40).

Вспомним, что в этом ряду число измерений $n=100$, среднее арифметическое $M=12$,

а среднее квадратическое отклонение $\sigma = \pm 4$ (см. стр. 25).

Для производства предстоящих нам вычислений заготовим расчетную решетку из 9 вертикальных столбцов и такого количества горизонтальных строчек, которое соответствует количеству имеющихся классов (в данном случае следовательно из 11 строчек). Кроме того сверху и снизу добавим еще по одной строчке для заголовков и для итогов.

V	p	x	x^2	x^3	x^4	px^2	px^3	px^4
24	1							
22	2							
20	3							
18	5							
16	10							
14	14							
12	21							
10	21							
8	15							
6	6							
4	2							
—	100							

В крайний слева столбец, озаглавленный буквою V , мы занесем все 11 числовых значений наших вариантов (большие наверху, малые внизу), а рядом в столбце p запишем соответствующие частоты. Сумму этих частот ($n=100$) пометим снизу.

Вычитая из каждого значения V общее для всего ряда среднее арифметическое $M=12$, получим отклонения x . Эти отклонения, далее, нужно последовательно возвысить во 2-ю, в 3-тью и в 4-ю степень.

После записи этих результатов в соответствующие столбцы x , x^2 , x^3 и x^4 наша расчетная решетка примет следующий вид:

V	p	x	x^2	x^3	x^4	px^2	px^3	px^4
24	1	+12	144	+1728	20736			
22	2	+10	100	+1000	10000			
20	3	+8	64	+512	4096			
18	5	+6	36	+216	1296			
16	10	+4	16	+64	256			
14	14	+2	4	+8	16			
12	21	0	0	0	0			
10	21	-2	4	-8	16			
8	15	-4	16	-64	256			
6	6	-6	36	-216	1296			
4	2	-8	64	-512	4096			
—	100	—	—	—	—			

После этого все x^2 , x^3 и x^4 следует помножить на соответствующие частоты p , полученные произведения записать в решетку и сложить по столбцам (см. таблицу ниже).

Сложив числа последних трех столбцов (px^2 , px^3 и px^4), получим в суммах: 1600, +3648 и 82432. Эти суммы мы запишем под соответствующими столбцами расчетной решетки в самой нижней ее строчке, заготовленной для итогов.

Итак, в результате этих вычислений мы получим:

$$\Sigma x^2 = 1600 \quad \Sigma x^3 = +3648 \quad \Sigma x^4 = 82432.$$

V	p	x	x ²	x ³	x ⁴	px ²	px ³	px ⁴
24	1	+ 12	144	+ 1728	20736	144	+ 1728	20736
22	2	+ 10	100	+ 1000	10000	200	+ 2000	20000
20	3	+ 8	64	+ 512	4096	192	+ 1536	12288
18	5	+ 6	36	+ 216	1296	180	+ 1080	6480
16	10	+ 4	16	+ 64	256	160	+ 640	2560
14	14	+ 2	4	+ 8	16	56	+ 112	224
12	21	0	0	0	0	0	0	0
10	21	- 2	4	- 8	16	84	- 168	336
8	15	- 4	16	- 64	256	240	- 960	3840
6	6	- 6	36	- 216	1296	216	- 1296	7776
4	2	- 8	64	- 512	4096	128	- 1024	8192
—	100	—	—	—	—	1600	+ 3648	82432

С помощью этих сумм представляется теперь возможным вычислить σ , A и E по формулам:

$$\sigma = \pm \sqrt{\frac{\Sigma x^2}{n}} \quad A = \frac{\Sigma x^3}{n\sigma^3} \quad E = \frac{\Sigma x^4}{n\sigma^4} - 3$$

В нашем случае для $n = 100$

$$\sigma = \pm \sqrt{\frac{1600}{100}} = \pm \sqrt{16} = \pm 4$$

что мы имели уже и раньше (см. стран. 25).

$$A = \frac{+3648}{100 \cdot 4^3} = \frac{+3648}{6400} = +0,57$$

$$E = \frac{82432}{100 \cdot 4^4} - 3 = \frac{82432}{25600} - 3 = 3,22 - 3 = +0,22$$

Итак, в заданном ряду мы имеем некоторую левостороннюю асимметрию ($A = +0,57$) и слабый прямой эксцесс ($E = +0,22$).

В разобранный выше примере классовый промежуток $k = 2$. Мы можем значительно упростить все наши вычисления, производимые в клетках расчетной решетки, если условно примем $k = 1$. Это допущение, как увидим дальше, совершенно не отразится на величине показателей A и E , если только с соответствующим изменением будет вычислено и среднее квадратическое отклонение σ .

Выше при $k = 2$, смежные значения центральных отклонений (x) отличались друг от друга на величину этого k (т. е. на 2 единицы). При $k = 1$ абсолютные значения x -ов вверх и вниз от 0 будут, очевидно, возрастать с каждой клеткой ровно на единицу, т. е. в данном случае окажутся вдвое меньше прежних. Соответственно изменятся и высшие степени центральных отклонений (x^2 , x^3 и x^4), уменьшившись в число раз, пропорциональное тем же степеням k (т. е. в 4, в 8 и в 16 раз). Вследствие этого все числа, записанные в клетках расчетной решетки, а стало быть и суммы их Σx^2 , Σx^3 и Σx^4 окажутся уменьшенными по сравнению с ранее полученными их значениями, что и явится моментом, упрощающим технику вычисления.

При этом способе вычисления расчетная решетка будет иметь следующий вид:

p	x	x ²	x ³	x ⁴	px ²	px ³	px ⁴
1	+ 6	36	+ 216	1296	36	+ 216	1296
2	+ 5	25	+ 125	625	50	+ 250	1250
3	+ 4	16	+ 64	256	48	+ 192	768
5	+ 3	9	+ 27	81	45	+ 135	405
10	+ 2	4	+ 8	16	40	+ 80	160
14	+ 1	1	+ 1	1	14	+ 14	14
21	0	0	0	0	0	0	0
21	- 1	1	- 1	1	21	- 21	21
15	- 2	4	- 8	16	60	- 120	240
6	- 3	9	- 27	81	54	- 162	486
2	- 4	16	- 64	256	32	- 128	512
100	—	—	—	—	400	+ 456	5152

Итак (при $k = 1$) $\Sigma x^2 = 400$, $\Sigma x^3 = +456$ и $\Sigma x^4 = 5152$ *).

Приняв эти заведомо ошибочные значения полученных сумм за истинные, подставим их в обычные формулы для вычисления σ , A и E .

$$\sigma = \pm \sqrt{\frac{\Sigma x^2}{n}} = \pm \sqrt{\frac{400}{100}} = \pm \sqrt{4} = \pm 2$$

(ранее при $k = 2$ мы имели $\sigma = \pm 4$).

*) Здесь обозначения Σx^2 , Σx^3 и Σx^4 оставлены лишь условно, чтобы не вводить новых символов.

$$A = \frac{\Sigma x^3}{n^3} = \frac{+456}{100 \cdot 2^3} = \frac{+456}{800} = +0,57$$

$$E = \frac{\Sigma x^4}{n^4} - 3 = \frac{5152}{100 \cdot 2^4} - 3 = \frac{5152}{1600} - 3 = 3,22 - 3 = +0,22$$

Итак показатель $A = +0,57$, а показатель $E = +0,22$, что мы имели и раньше при $k=2$ и $\sigma = \pm 4$.

Заметим, кстати, что показатель E можно было бы вычислить и по другой формуле, в которую вовсе не входит σ :

$$E = \frac{n \Sigma x^4}{(\Sigma x^2)^2} - 3$$

В нашем последнем примере $n = 100$, $\Sigma x^2 = 400$, а $\Sigma x^4 = 5152$. Следовательно

$$E = \frac{100 \cdot 5152}{(400)^2} - 3 = +0,22$$

Пользоваться этой формулой удобно в тех случаях, когда вычисляется только один показатель E , а надобности в A и σ нет.

V	$a_1 \quad a_2$			
	p	77	78	
24	1	1	1	
22	2	3	4	
20	3	6	10	
18	5	11	21	
16	10	21	42	
14	14	35	—	
C=12	21	—	—	
10	21	44	—	
8	15	23	33	
6	6	8	10	
4	2	2	2	
	100	77	45	
	n	b ₁	b ₂	

Чтобы не смешивать σ , вычисленную с учетом данного значения классового промежутка k (у нас $k=2$), с тем вспомогательным значением σ , которое получается в условном предположении $k=1$, обозначим эту вспомогательную σ буквою ς (сигма концевая). Итак, у нас $\sigma=4$, а $\varsigma=2$. Вполне понятно, что $\sigma=k\varsigma$ и обратно

$$\varsigma = \frac{\sigma}{k}$$

Рассмотренный выше прием вычисления сумм Σx^2 , Σx^3 и Σx^4 может быть назван „способом произведений“. Покажем теперь, как выполняются эти вычисления по „способу сумм“.

Составив расчетную решетку обычного типа, но с большим числом вертикальных столбцов, запишем в нее все наши частоты p .

Далее, против частоты $p=21$, соответствующей началу отсчетов $C=12$, проведем четыре черточки и выполним все обычные суммирования чисел в клетках второго и третьего столбца. Подсчитав затем верхние и нижние итоги, получим:

$$a_1 = 77 \quad a_2 = 78$$

$$b_1 = 77 \quad b_2 = 45$$

Откуда суммы вспомогательных отклонений:

$$S_1 = a_1 - b_1 = 77 - 77 = 0$$

$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 = 77 + 77 + 2 \cdot 78 + 2 \cdot 45 = 400$$

В нашем примере начало отсчетов $C=12$ точно совпало со средним арифметическим $M=12$. Поэтому обе суммы $S_1=0$ и $S_2=400$ в то же самое время являются здесь и суммами центральных отклонений и их квадратов (Σx и Σx^2). Итак, в данном случае

$$S_1 = \Sigma x = 0$$

$$S_2 = \Sigma x^2 = 400$$

что мы имели уже и раньше.

Продолжим теперь эти вычисления, выполнив их и для последних двух столбцов расчетной решетки.

Проставив черточки, как это показано на приведенной здесь решетке, и заполнив все клетки последних двух столбцов обычным способом, подсчитаем суммы чисел в этих столбцах сверху и книзу отдельно. В результате такого подсчета получим итоги: a_3 , a_4 , b_3 и b_4 , которые

и запишем сверху и снизу рядом с найденными ранее итогами a_1 , a_2 , b_1 и b_2 *). В нашем примере

$$a_3 = 57 \quad a_4 = 28$$

$$b_3 = 14 \quad b_4 = 2$$

*) Заметим, кстати, что взамен действительного сложения всех чисел какого либо столбца, для получения этих итогов можно складывать лишь два нижних числа, непосредственно примыкающих к черточкам в данном и в соседнем справа столбце. Напр. $a_3 = 36 + 21 = 57$ или $b_2 = 33 + 12 = 45$.

Теперь нам предстоит вычислить с помощью этих итогов сумму третьих степеней вспомогательных отклонений S_3 и сумму четвертых их степеней S_4 (в данном частном случае при $C=M$ эти суммы S_3 и S_4 очевидно точно совпадают с суммами соответствующих степеней центральных отклонений Σx^3 и Σx^4).

Для вычисления вспомогательных сумм S_3 и S_4 служат следующие общие формулы:

$$S_3 = (a_1 - b_1) + 6(a_2 - b_2) + 6(a_3 - b_3) \\ S_4 = (a_1 + b_1) + 14(a_2 + b_2) + 36(a_3 + b_3) + 24(a_4 + b_4)$$

Подставляя в эти формулы полученные из решетки значения итогов:

$$\begin{array}{cccc} a_1 = 77 & a_2 = 78 & a_3 = 57 & a_4 = 28 \\ b_1 = 77 & b_2 = 45 & b_3 = 14 & b_4 = 2 \end{array}$$

будем иметь:

$$S_3 = (77 - 77) + 6(78 - 45) + 6(57 - 14) = +456 \\ S_4 = (77 + 77) + 14(78 + 45) + 36(57 + 14) + \\ + 24(28 + 2) = 5152.$$

Итак, в данном случае:

$$S_3 = \Sigma x^3 = +456 \\ S_4 = \Sigma x^4 = 5152$$

что мы имели и раньше, выполняя вычисления по способу произведений.

Из рассмотрения формул для вычисления вспомогательных сумм S_1, S_2, S_3 и S_4 видно, что суммы нечетных степеней (S_1 и S_3) получаются из разностей соответствующих итогов ($a - b$), а в формулы для вычисления сумм четных степеней (S_2 и S_4) входят суммы этих же итогов ($a + b$).

Эти суммы или разности должны быть умножены, далее на постоянные числовые коэффициенты, помещенные в соответствующих местах указанных выше формул перед скобками. Чтобы облегчить запоминание этих формул и входящих в их состав коэффициентов, полезно иметь перед собой изображенную здесь ориентировочную табличку. Слева на ее полях выписаны обозначения сумм: S_1, S_2, S_3 и S_4 , а справа против соответствующих строчек указан спо-

соб их вычисления: либо из разностей ($-$), либо из сумм ($+$) соответствующих итогов a и b . Вертикальные столбцы таблички соответствуют отдельным парам этих итогов: крайний слева столбец (1) — итогам a_1 и b_1 следующий (2) — итогам a_2 и b_2 и т. д. В отдельных клетках внутри самой таблички записаны те числовые коэффициенты, на которые при вычислении сумм S следует умножать отдельные разности ($a - b$) или суммы ($a + b$) соответствующих итогов.

Пользование этой табличкой крайне просто. Пусть, например, требуется вычислить сумму S_1 . Буква S_1 записана против первой строки, на уровне которой справа стоит знак $-$. Следовательно сумма S_1 вычисляется из разностей итогов $a_1 - b_1$. Просматривая цифры в клетках этой строки, мы находим там лишь 1, записанную в крайнем слева столбце. Т. к. этот столбец относится к первой паре итогов (a_1 и b_1), то сумма S_1 будет состоять из одной (1) разности первых итогов a_1 и b_1 т. е.

$$S_1 = a_1 - b_1$$

Для вычисления S_2 мы должны оперировать уже с суммами этих итогов (справа у второй строчки стоит знак $+$). Так как в первом столбце стоит 1, а во втором 2, то к одной (1) сумме первых итогов ($a_1 + b_1$) следует прибавить удвоенную (2) сумму вторых итогов ($a_2 + b_2$). Итак

$$S_2 = (a_1 + b_1) + 2(a_2 + b_2).$$

Из той же таблички видно, что при вычислении S_3 следует первую разность ($a_1 - b_1$) сложить с умноженной на 6 второй разностью ($a_2 - b_2$) и с умноженной на 6 же третьей разностью ($a_3 - b_3$).

$$S_3 = (a_1 - b_1) + 6(a_2 - b_2) + 6(a_3 - b_3)$$

Наконец, при вычислении S_4 первая сумма ($a_1 + b_1$) умножается на 1 (т. е. остается без изменения), вторая сумма ($a_2 + b_2$) — на 14, третья ($a_3 + b_3$) — на 36, а четвертая

$(a_4 + b_4)$ — на 24, после чего все эти произведения складываются.

$$S_4 = (a_1 + b_1) + 14 (a_2 + b_2) + 36 (a_3 + b_3) + 24 (a_4 + b_4)$$

p	77	78	57	28
1				
2				
3				
5				
10				
14				
21				
21				
15				
6				
2				
100	77	45	14	2

$$\begin{array}{l} a - b \quad 0 \quad +33 \quad +43 \quad +26 \\ a + b \quad 154 \quad 123 \quad 71 \quad 30 \end{array}$$

При пользовании вместо формул этой табличкой коэффициентов удобно внизу самой расчетной решетки выписать под соответствующими столбиками значения всех разностей и сумм итогов a и b , взятых попарно. Слева приведен образец расчетной решетки с такими дополнительными записями. Для ясности средние клетки этой решетки оставлены незаполненными. Умножая эти разности или суммы итогов на определенные числа, взятые из ориентировочной таблички, и складывая эти произведения (с учетом знаков), мы и получаем искомые суммы: S_1 , S_2 , S_3 и S_4 :

$$\begin{aligned} S_1 &= 0 \\ S_2 &= 154 + 2 \cdot 123 = 400 \\ S_3 &= 0 + 6 (+33) + 6 (+43) = 456 \\ S_4 &= 154 + 14 \cdot 123 + 36 \cdot 71 + 24 \cdot 30 = 5152 \end{aligned}$$

VII. Вычисление показателей A и E по способу моментов.

В разобранный выше примере вычисления показателей A и E начало отсчетов C точно совпадало со средним арифметическим M , вследствие чего суммы различных степеней вспомогательных отклонений S_1 , S_2 , S_3 и S_4 в то же самое время являлись и суммами соответствующих степеней центральных отклонений Σx , Σx^2 , Σx^3 и Σx^4 .

Покажем, как следует поступать в том общем случае, когда начало отсчетов C не совпадает со средним арифметическим M .

Пусть в том же самом числовом примере за начало отсчетов C принято значение варианты $V = 14$, частота которой $p = 14$. В этом случае, следовательно, начало отсчетов

$$C = 14$$

в то время как среднее арифметическое $M = 12$.

Приняв условно $k = 1$ и выполнив обычные промежуточные вычисления по способу сумм, получим следующие итоги:

$$a_1 = 42 \quad a_2 = 36$$

$$a_3 = 21 \quad a_4 = 7$$

$$b_1 = 142 \quad b_2 = 122$$

$$b_3 = 59 \quad b_4 = 16$$

Разности и суммы этих итогов записаны здесь внизу расчетной решетки. Пользуясь ориентировочной табличкой, описанной в прошлой главе (см. стр. 67), найдем значения всех вспомогательных сумм:

$$\begin{aligned} S_1 &= -100 \\ S_2 &= 184 + 2 \cdot 158 = 500 \\ S_3 &= -100 + 6 (-86) + \\ &\quad + 6 (-38) = -844 \\ S_4 &= 184 + 14 \cdot 158 + 36 \cdot 80 + \\ &\quad + 24 \cdot 23 = 5828 \end{aligned}$$

V	p	42	36	21	7
24	1	1	1	1	1
22	2	3	4	5	6
20	3	6	10	15	—
18	5	11	21	—	—
16	10	21	—	—	—
14	14	—	—	—	—
12	21	65	—	—	—
10	21	44	77	—	—
8	15	23	33	45	—
6	6	8	10	12	14
4	2	2	2	2	2
	100	142	122	59	16

$$\begin{array}{l} a - b \quad -100 \quad -86 \quad -38 \quad -9 \\ a + b \quad 184 \quad 158 \quad 80 \quad 23 \end{array}$$

Чтобы от этих вспомогательных сумм перейти к суммам 2-х, 3-х и 4-х степеней центральных отклонений (т. е. отклонений, отсчитанных от точного значения среднего арифметического $M = 12$, а не от произвольного начала отсчетов $C = 14$), можно было бы воспользоваться формулами, аналогичными знакомой уже нам формуле для вычисления Σx^2 с помощью сумм S_1 и S_2 :

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n}$$

Эти формулы для Σx^3 и Σx^4 будут иметь следующий вид:

$$\Sigma x^3 = S_3 - \frac{3 S_1 S_2}{n} + \frac{2 S_1^3}{n^2}$$

$$\Sigma x^4 = S_4 - \frac{4 S_1 S_3}{n} + \frac{6 S_1^2 S_2}{n^2} - \frac{3 S_1^4}{n^3}$$

Подставляя сюда найденные выше значения сумм S_1, S_2, S_3 и S_4 , получим:

$$\Sigma x^2 = -844 - \frac{3(-100) \cdot 500}{100} + \frac{2(-100)^3}{100^2} = +456$$

$$\Sigma x^4 = 5828 - \frac{4(-100) \cdot (-844)}{100} + \frac{6(-100)^2 \cdot 500}{100^2} - \frac{3(-100)^4}{100^3} = 5152$$

Итак, $\Sigma x^3 = +456$, а $\Sigma x^4 = 5152$, что мы имели и раньше при совпадении начала отсчетов C со средним арифметическим $M = 12$ (см. стр. 66).

В приведенные выше формулы для вычисления Σx^3 и Σx^4 входят высшие степени числа измерений n (n^2 и n^3). В данном примере при $n = 100$ это возвышение 100 в квадрат и в куб не представляло особых затруднений, но в общем случае, когда n является числом неокругленным, эти действия с ним и последующее деление числителей на такие многозначные знаменатели представляют уже большое техническое осложнение.

В целях упрощения этих формул введем некоторые дополнительные обозначения.

Разделив все четыре суммы S_1, S_2, S_3 и S_4 на число измерений n , будем иметь следующие четыре отношения (дроби):

$$\frac{S_1}{n} \quad \frac{S_2}{n} \quad \frac{S_3}{n} \quad \frac{S_4}{n}$$

называемые вспомогательными „моментами распределения“ (или просто „моментами“) и обозначаемые обычно греческой буквой v (ню).

Итак, вспомогательный момент I-й степени (или просто первый вспомогательный момент):

$$v_1 = \frac{S_1}{n}$$

второй вспомогательный момент:

$$v_2 = \frac{S_2}{n}$$

и т. д.

В нашем примере при $n = 100$, $S_1 = -100$, $S_2 = 500$, $S_3 = -844$ и $S_4 = 5828$ будем иметь:

$$v_1 = \frac{S_1}{n} = \frac{-100}{100} = -1$$

$$v_2 = \frac{S_2}{n} = \frac{500}{100} = 5$$

$$v_3 = \frac{S_3}{n} = \frac{-844}{100} = -8,44$$

$$v_4 = \frac{S_4}{n} = \frac{5828}{100} = 58,28$$

Аналогично вспомогательным моментам v_1, v_2, v_3 и v_4 вычисляются и моменты центральные, получаемые посредством деления сумм $\Sigma x, \Sigma x^2, \Sigma x^3$ и Σx^4 на число измерений n . Эти центральные моменты обозначаются обычно греческой буквой μ (мю).

В нашем примере при $\Sigma x = 0$, $\Sigma x^2 = 400$, $\Sigma x^3 = +456$ и $\Sigma x^4 = 5152$ получим:

$$\mu_1 = \frac{\Sigma x}{n} = \frac{0}{100} = 0$$

$$\mu_2 = \frac{\Sigma x^2}{n} = \frac{400}{100} = 4$$

$$\mu_3 = \frac{\Sigma x^3}{n} = \frac{+456}{100} = +4,56$$

$$\mu_4 = \frac{\Sigma x^4}{n} = \frac{5152}{100} = 51,52$$

Заметим, кстати, что первый центральный момент μ_1 всегда $= 0$, так как и $\Sigma x = 0$.

Итак в нашем примере:

$$\begin{array}{cccc} v_1 = -1 & v_2 = 5 & v_3 = -8,44 & v_4 = 58,28 \\ \mu_1 = 0 & \mu_2 = 4 & \mu_3 = +4,56 & \mu_4 = 51,52 \end{array}$$

Центральные моменты μ находятся в определенной зависимости от вспомогательных моментов v , причем зависимость эта определяется следующими формулами:

$$\mu_2 = v_2 - v_1^2$$

$$\mu_3 = v_3 - 3v_1 v_2 + 2v_1^3$$

$$\mu_4 = v_4 - 4v_1 v_3 + 6v_1^2 v_2 - 3v_1^4$$

Подставив в эти формулы числовые данные нашего примера, получим:

$$\begin{aligned}\mu_2 &= 5 - (-1)^2 = 5 - 1 = 4 \\ \mu_3 &= -8,44 - 3(-1) \cdot 5 + 2(-1)^3 = +4,56 \\ \mu_4 &= 58,28 - 4(-1) \cdot (-8,44) + 6(-1)^2 \cdot 5 - 3(-1)^4 = 51,52\end{aligned}$$

Итак, указанные формулы приводят нас к совершенно тем же числовым результатам, которые мы имели и раньше.

Определив этим путем центральные моменты μ_3 и μ_4 , можно вычислить с их помощью и оба показателя A и E по формулам:

$$A = \frac{\mu_3}{\sigma^3}$$

$$E = \frac{\mu_4}{\sigma^4} - 3,$$

где через σ обозначено значение σ , вычисленное в условном предположении, что классовый промежуток $k = 1$.

В данном случае при $\mu_3 = +4,56$, $\mu_4 = 51,52$ и $\sigma = 2$ будем иметь:

$$\begin{aligned}A &= \frac{+4,56}{2^3} = \frac{+4,56}{8} = +0,57 \\ E &= \frac{51,52}{2^4} - 3 = \frac{51,52}{16} - 3 = 3,22 - 3 = +0,22\end{aligned}$$

т. е. те же самые результаты, которые мы получили и раньше.

С помощью моментов представляется возможным вычислять также среднее арифметическое M и среднее квадратическое отклонение σ или ς по формулам:

$$M = C + k v_1 \quad \sigma = \pm k \sqrt{\mu_2} \quad \varsigma = \pm \sqrt{\mu_2}$$

В нашем примере $C = 14$, $k = 2$, $v_1 = -1$, а $\mu_2 = 4$. Подставляя эти данные в указанные выше формулы, получим:

$$\begin{aligned}M &= 14 + 2(-1) = 14 - 2 = 12. \\ \sigma &= \pm 2 \sqrt{4} = \pm 2 \cdot 2 = \pm 4 \quad \varsigma = \pm \sqrt{4} = \pm 2\end{aligned}$$

что мы имели и раньше.

Чтобы иллюстрировать процесс вычисления M , σ , A и E по способу моментов, рассмотрим еще следующий числовой пример. Пусть имеем вариационный ряд, в котором

в границах от 9 до 8,5 заключается 3 варианты

"	"	"	8,5	"	8	"	9	"
"	"	"	8	"	7,5	"	16	"
"	"	"	7,5	"	7	"	28	"
"	"	"	7	"	6,5	"	35	"
"	"	"	6,5	"	6	"	38	"
"	"	"	6	"	5,5	"	36	"
"	"	"	5,5	"	5	"	31	"
"	"	"	5	"	4,5	"	25	"
"	"	"	4,5	"	4	"	14	"
"	"	"	4	"	3,5	"	8	"
"	"	"	3,5	"	3	"	4	"
"	"	"	3	"	2,5	"	2	"
"	"	"	2,5	"	2	"	1	"

Этот ряд следовательно состоит из $n = 250$ вариантов, разбитых на 14 классов с классовым промежутком $k = 0,5$. Его гистограмма представлена на чертеже 41. Частоты p ука-

заны цифрами у вершин соответствующих столбиков (см. стр. 74).

Запишем эти частоты в расчетную решетку против границ соответствующих классов (проставленных на полях) и возьмем за начало отсчетов C середину границ класса от 6 до 5,5, т. е. примем:

$$\begin{aligned}C &= \frac{6 + 5,5}{2} = \\ &= \frac{11,5}{2} = 5,75\end{aligned}$$

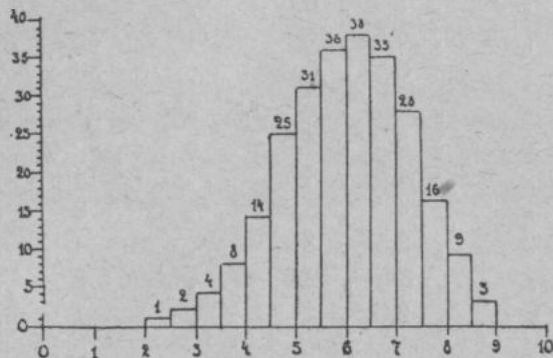
Далее проведем черточки, как показано на расчетной решетке, и выполним все промежуточные вычисле-

Границы классов	p	319	350	242	106
9-8,5	3	3	3	3	3
8,5-8	9	12	15	18	21
8-7,5	16	28	43	61	82
7,5-7	28	56	99	160	—
7-6,5	35	91	190	—	—
6,5-6	38	129	—	—	—
6-5,5	36	—	—	—	—
5,5-5	31	85	—	—	—
5-4,5	25	54	109	—	—
4,5-4	14	29	55	97	—
4-3,5	8	15	26	42	64
3,5-3	4	7	11	16	22
3-2,5	2	3	4	5	6
2,5-2	1	1	1	1	1
	250	194	206	161	93

$$\begin{aligned}a - b &= 125 & 144 & 81 & 13 \\ a + b &= 513 & 556 & 403 & 199\end{aligned}$$

ния внутри клеток этой решетки. В результате получим следующие значения итогов a и b .

$$\begin{array}{llll} a_1 = 319 & a_2 = 350 & a_3 = 242 & a_4 = 106 \\ b_1 = 194 & b_2 = 206 & b_3 = 161 & b_4 = 93 \end{array}$$



Черт. 41.

Разности и суммы этих итогов записаны внизу решетки в две строчки рядом с соответствующими заголовками: $a - b$ и $a + b$. Отсюда:

$$\begin{aligned} S_1 &= +125 \\ S_2 &= 513 + 2.556 = 1625 \\ S_3 &= +125 + 6(+144) + 6(+81) = +1475 \\ S_4 &= 513 + 14.556 + 36.403 + 24.199 = 27581 \end{aligned}$$

Разделив эти суммы на $n = 250$, получим четыре вспомогательных момента v_1, v_2, v_3 и v_4

$$\begin{aligned} v_1 &= \frac{S_1}{n} = \frac{+125}{250} = +0,50 \\ v_2 &= \frac{S_2}{n} = \frac{1625}{250} = 6,50 \\ v_3 &= \frac{S_3}{n} = \frac{+1475}{250} = +5,90 \\ v_4 &= \frac{S_4}{n} = \frac{27581}{250} = 110,32 \end{aligned}$$

Подставляя эти значения вспомогательных моментов в формулы для вычисления центральных моментов μ_2, μ_3 и μ_4 будем иметь:

$$\begin{aligned} \mu_2 &= v_2 - v_1^2 = 6,5 - (+0,5)^2 = 6,25 \\ \mu_3 &= v_3 - 3v_1v_2 + 2v_1^3 = 5,9 - 3(+0,5) \cdot 6,5 + 2(+0,5)^3 = -3,6 \\ \mu_4 &= v_4 - 4v_1v_3 + 6v_1^2v_2 - 3v_1^4 = 110,32 - 4(+0,5)(+5,9) + \\ &\quad + 6(+0,5)^2 \cdot 6,5 - 3(+0,5)^4 = 108,08. \end{aligned}$$

Итак в нашем числовом примере

$$\begin{array}{ll} v_1 = +0,50 & \\ v_2 = 6,50 & \mu_2 = 6,25 \\ v_3 = +5,90 & \mu_3 = -3,60 \\ v_4 = 110,32 & \mu_4 = 108,08 \end{array}$$

По этим данным можно определить M, σ, ς, A и E

$$M = C + kv_1 = 5,75 + 0,5(+0,5) = 6$$

$$\sigma = \pm k \sqrt{\mu_2} = \pm 0,5 \sqrt{6,25} = \pm 1,25$$

Нужное для вычисления показателей A и E вспомогательное значение ς найдем по формуле:

$$\varsigma = \frac{\sigma}{k} = \frac{\pm 1,25}{0,5} = \pm 2,5$$

Подставляя это значение ς в формулы для вычисления A и E , получим:

$$A = \frac{\mu_3}{\varsigma^3} = \frac{-3,60}{(2,5)^3} = \frac{-3,60}{15,63} = -0,23$$

$$E = \frac{\mu_4}{\varsigma^4} - 3 = \frac{108,08}{(2,5)^4} - 3 = \frac{108,08}{39,06} - 3 = 2,77 - 3 = -0,23$$

Итак в нашем примере

$$\begin{array}{ll} M = 6 & A = -0,23 \\ \sigma = \pm 1,25 & E = -0,23 \end{array}$$

Чтобы судить о надежности этих данных, необходимо вычислить их средние ошибки по формулам:

$$m = \frac{\pm \sigma}{\sqrt{n}} = \frac{\pm 1,25}{\sqrt{250}} = \pm 0,08$$

$$m_{\varsigma} = \frac{\sigma}{\sqrt{2n}} = \frac{\pm 1,25}{\sqrt{500}} = \pm 0,06$$

$$m_A = \pm \sqrt{\frac{6}{n}} = \pm \sqrt{\frac{6}{250}} = \pm 0,15$$

$$m_E = \pm 2 \sqrt{\frac{6}{n}} = \pm 2 \sqrt{\frac{6}{250}} = \pm 0,31$$

Учитывая эту среднюю величину возможного варьирования всех числовых характеристик заданного ряда, мы можем теперь записать их в такой символической форме:

$$\begin{array}{ll} M = 6 \pm 0,08 & A = -0,23 \pm 0,15 \\ \sigma = 1,25 \pm 0,06 & E = -0,23 \pm 0,31 \end{array}$$

В частности, в применении к показателям A и E может возникнуть вопрос о достоверности обнаруженных ими отрицательной асимметрии и отрицательного эксцесса. Применяя обычный критерий

$$\frac{A}{m_A} > 3 \text{ и } \frac{E}{m_E} > 3$$

получим:

$$\frac{A}{m_A} = \frac{0,23}{0,15} = 1,53$$

$$\frac{E}{m_E} = \frac{0,23}{0,31} = 0,74$$

Следовательно наличие в заданном ряду асимметрии и эксцесса является не доказанным, и при данных условиях мы можем считать это распределение отличающимся от нормального лишь в пределах случайных ошибок.

К этому же выводу можно прийти и другим путем, вычислив наименьшее число измерений n , при котором данные показатели A и E могли бы считаться достоверными.

Для показателя асимметрии

$$\min n = \frac{54}{A^2} = \frac{54}{(-0,23)^2} = 1021$$

Для показателя эксцесса

$$\min n = \frac{216}{E^2} = \frac{216}{(-0,23)^2} = 4083$$

Так как в действительности мы имели значительно меньшее количество измерений (у нас $n = 250$), то оба эти показателя следует признать недостоверными.

При сравнении друг с другом двух одноименных показателей A или E применяется обычный критерий оценки:

$$\frac{A_1 - A_2}{\sqrt{m_1^2 + m_2^2}} > 3 \text{ и } \frac{E_1 - E_2}{\sqrt{m_1^2 + m_2^2}} > 3$$

где через m_1 и m_2 обозначены средние ошибки соответствующих показателей A_1 и A_2 или E_1 и E_2 .

Для облегчения вычислительной техники можно пользоваться также следующими рабочими формулами:

$$\frac{(A_1 - A_2)^2}{m_1^2 + m_2^2} > 9 \quad \frac{(E_1 - E_2)^2}{m_1^2 + m_2^2} > 9.$$

Пусть, например, в одном случае показатель эксцесса $E_1 = +0,89$, а в другом $E_2 = +0,73$, причем средние ошибки их соответственно равны $m_1 = \pm 0,03$ и $m_2 = \pm 0,04$. Подставляя эти данные в последнюю формулу, получим:

$$\frac{(0,89 - 0,73)^2}{(0,03)^2 + (0,04)^2} = \frac{(0,16)^2}{0,0009 + 0,0016} = \frac{0,0256}{0,0025} = 10,24.$$

Так как это отношение получилось больше 9, то разница между заданными показателями эксцесса E_1 и E_2 может считаться доказанной. Это значит, что I-ый ряд, для которого $E_1 = +0,89$, действительно (а не случайно) является более высоковершинным, нежели II-й ряд, для которого $E_2 = +0,73$. В данном случае, следовательно, больший эксцесс I-го ряда свойственен ему по природе самого материала.

Если средние ошибки m_A и m_E показателей A и E вычислены не были, оценку достоверности различий между A_1 и A_2 или E_1 и E_2 можно выполнить по следующему критерию:

$$\frac{(A_1 - A_2)^2 n_1 n_2}{6(n_1 + n_2)} > 9 \text{ и } \frac{(E_1 - E_2)^2 n_1 n_2}{24(n_1 + n_2)} > 9$$

где через n_1 и n_2 обозначено число измерений, из которых эти показатели были вычислены.

Пусть, например, для I-го ряда:

$$A_1 = -0,3 \quad E_1 = +1,2 \quad n_1 = 200$$

а для II-го ряда:

$$A_2 = +0,4 \quad E_2 = +0,4 \quad n_2 = 300$$

Вычислим два выражения:

$$\frac{(A_1 - A_2)^2 n_1 n_2}{6(n_1 + n_2)} = \frac{(-0,3 - 0,4)^2 \cdot 200 \cdot 300}{6(200 + 300)} = \frac{(-0,7)^2 \cdot 60.000}{3.000} = 9,8,$$

$$\frac{(E_1 - E_2)^2 n_1 n_2}{24(n_1 + n_2)} = \frac{(+1,2 - 0,4)^2 \cdot 200 \cdot 300}{24(200 + 300)} = \frac{(+0,8)^2 \cdot 60.000}{12.000} = 3,2.$$

Следовательно различие между показателями A_1 и A_2 можно считать бесспорным, в то время как различие между E_1 и E_2 не является еще доказанным.

VIII. Упрощенный способ вычисления показателей A и E .

При определении показателей A и E по способу моментов некоторое техническое затруднение представляет процесс вычисления центральных моментов μ_3 и μ_4 по формулам:

$$\begin{aligned}\mu_3 &= v_3 - 3v_1v_2 + 2v_1^3 \\ \mu_4 &= v_4 - 4v_1v_3 + 6v_1^2v_2 - 3v_1^4.\end{aligned}$$

Это затруднение можно устранить путем применения особого приема, заключающегося в вычислении сперва заведомо неправильных значений показателей асимметрии и эксцесса непосредственно из сумм S_1 , S_2 , S_3 и S_4 с последующим исправлением этих результатов по специальной таблице. Вспомним, что показатели A и E при помощи сумм высших степеней центральных отклонений Σx^3 и Σx^4 мы определяли по формулам:

$$\begin{aligned}A &= \frac{\Sigma x^3}{n\sigma^3} \\ E &= \frac{\Sigma x^4}{n\sigma^4} - 3.\end{aligned}$$

В тех случаях, когда среднее арифметическое M точно совпадает с принятым началом отсчетов C , обе суммы высших степеней вспомогательных отклонений S_3 и S_4 будут в свою очередь совершенно точно совпадать с суммами соответствующих степеней центральных отклонений Σx^3 и Σx^4 , вследствие чего при $M=C$ показатели A и E могут быть вычислены по формулам:

$$\begin{aligned}A &= \frac{S_3}{n\sigma^3} \\ E &= \frac{S_4}{n\sigma^4} - 3.\end{aligned}$$

Но это только в том частном случае, когда $M=C$. По мере увеличения разницы между средним арифметическим M и началом отсчетов C эти последние формулы будут давать все большую и большую неточность, находящуюся в определенной зависимости от разности $M-C$.

Обозначим такие заведомо ошибочные значения показателей A и E через A_0 и E_0 и будем помнить, что

$$A_0 = \frac{S_3}{n\sigma^3} \quad E_0 = \frac{S_4}{n\sigma^4} - 3$$

в то время, как ранее мы имели:

$$A = \frac{\Sigma x^3}{n\sigma^3} \quad E = \frac{\Sigma x^4}{n\sigma^4} - 3.$$

Величина неточности этих ошибочных показателей A_0 и E_0 зависит, как было уже сказано выше, от разности $M-C$. Чем больше абсолютная величина этой разности, тем сильнее отличается A_0 от A и E_0 от E . В частности при $M-C=0$ (т. е. при $M=C$) эти вспомогательные и истинные значения обоих показателей совершенно точно совпадают друг с другом (т. е. $A_0=A$ и $E_0=E$).

Выразим величину разности между M и C в долях среднего квадратического отклонения σ . Обозначив результат этого вычисления через d , будем иметь следующую формулу:

$$d = \frac{M-C}{\sigma}.$$

Величина d определяет, следовательно, то количество средних квадратических отклонений σ (или ту долю σ), на которую отстоит (отклоняется) точное значение среднего арифметического M от произвольно выбранного начала отсчетов C .

В зависимости от разности d можно определить по особой таблице две „поправки“: a и e , с помощью которых представляется возможным вычислить истинные значения показателей A и E по величине вспомогательных (неправильных) их значений A_0 и E_0 .

Для этого служат следующие формулы:

$$\begin{aligned}A &= A_0 + a \\ E &= E_0 - 4dA_0 + e.\end{aligned}$$

Покажем применение этого способа на разобранном в конце предыдущей главы числовом примере. Там за начало отсчетов было принято значение $C=5,75$, в то время, как истинная величина среднего арифметического $M=6$. Следовательно разность $M-C=0,25$. Эта разность выражена здесь в обычных единицах измерения данного варьирующего признака. Чтобы отклонение M от C выразить в долях самой σ , следует разделить эту разность $M-C=$

$= 0,25$ на среднее квадратическое отклонение σ . В нашем примере $\sigma = \pm 1,25$. Следовательно

$$d = \frac{M - C}{\sigma} = \frac{+0,25}{1,25} = +0,2.$$

Итак, среднее арифметическое M в нашем примере отклоняется от начала отсчетов C на $+0,2$ сигмы (больше C на $\frac{2}{10} \sigma$). Эта величина

$$d = +0,2$$

и определяет степень той неточности, которую мы можем получить в наших вычислениях, принимая за Σx^3 и Σx^4 вспомогательные суммы S_3 и S_4 .

Заметим, что величину d можно было бы определить и не вычисляя предварительно точного значения M . Для этого служит следующая формула, которую в дальнейшем мы и будем всегда применять:

$$d = \frac{S_1}{n\sigma}.$$

Так как в данном случае (см. стр. 74 и 75) $S_1 = +125$, $n = 250$, а $\sigma = 2,5$ то

$$d = \frac{+125}{250 \cdot 2,5} = +0,2$$

$\pm d$	$\mp a$	$+e$
0,1	0,301	0,0603
0,2	0,608	0,2448
0,3	0,927	0,5643
0,4	1,264	1,0368
0,5	1,625	1,6875
0,6	2,016	2,5488
0,7	2,443	3,6603
0,8	2,912	5,0688
0,9	3,429	6,8283
1,0	4,000	9,0000
1,1	4,631	11,6523
1,2	5,328	14,8608

что мы и получили выше при пользовании другой формулой.

Вычислив d , определим две поправки a и e по специальной таблице, приложенной в конце этой книги (см. стран. 424). Здесь приведена эта таблица лишь в кратком виде. В ней значения d выписаны через 1 десятую долю сигмы, а поправки a и e даны с точностью до 3-го и 4-го знака после запятой. В заголовках этой таблицы у d поставлены два знака $+$ и $-$, а у a наоборот $-$ и $+$. Этим указывается, что для положительных значений d поправку a следует брать со знаком $-$, а для отрицательных d — со зна-

ком $+$. Поправка e берется всегда со знаком $+$, что и указано в соответствующем заголовке третьего столбца ($+e$).

ком $+$. Поправка e берется всегда со знаком $+$, что и указано в соответствующем заголовке третьего столбца ($+e$).

Во второй строчке сверху этой таблицы для $d = 0,2$ находим: $a = 0,608$ и $e = 0,2448$. Так как в нашем примере d имеет знак $+$, то поправку a следует взять с минусом. Итак:

$$a = -0,608$$

$$e = +0,2448.$$

Это и есть те поправки, с помощью которых мы можем получить истинные значения A и E по величине вспомогательных (заведомо неправильных) показателей A_0 и E_0 . Подставляя в формулы для вычисления этих вспомогательных показателей числовые данные нашего примера, получим:

$$A_0 = \frac{S_3}{n\sigma^3} = \frac{+1475}{250(2,5)^3} = +0,378.$$

$$E_0 = \frac{S_4}{n\sigma^4} - 3 = \frac{27581}{250(2,5)^4} - 3 = 2,8243 - 3 = -0,1757.$$

При помощи этих ошибочных показателей $A_0 = +0,378$ и $E_0 = -0,1757$ находим истинные их значения A и E по формулам

$$A = A_0 + a$$

$$E = E_0 - 4d A_0 + e$$

где $d = +0,2$, $a = -0,608$ и $e = +0,2448$.

Вычисления дают здесь следующие результаты:

$$A = +0,378 - 0,608 = -0,23$$

$$E = -0,1757 - 4(+0,2)(+0,378) + 0,2448 = -0,23.$$

Итак, в данном случае, так же как и раньше, показатель асимметрии

$$A = -0,23,$$

и показатель эксцесса

$$E = -0,23.$$

Приложенная в конце книги полная таблица поправок a и e составлена лишь до значения $d = 2$. Если при каких-либо вычислениях понадобилось бы продолжить эту таблицу для значений d , превышающих 2, или вычислить поправки a и e для более мелких долей d (например с точностью до

0,001 разности d), можно воспользоваться следующими формулами зависимости поправок a и e от величины d :

$$a = -d(d^2 + 3)$$

$$e = 3d^2(d^2 + 2).$$

Приведем еще один числовой пример определения показателей A и E по этому упрощенному способу, указав попутно также и приемы проверки всех промежуточных вычислений внутри расчетной решетки при заполнении ее клеток по способу сумм.

Пусть имеем два следующие вариационные ряда:

I ряд:

в границах от 33 до 30	заключается 1	варианта
" " " 30 " 27	" 2	"
" " " 27 " 24	" 3	"
" " " 24 " 21	" 6	"
" " " 21 " 18	" 14	"
" " " 18 " 15	" 29	"
" " " 15 " 12	" 12	"
" " " 12 " 9	" 4	"
" " " 9 " 6	" 1	"

II ряд:

в границах от 55 до 50	заключается 6	вариант
" " " 50 " 45	" 10	"
" " " 45 " 40	" 12	"
" " " 40 " 35	" 11	"
" " " 35 " 30	" 10	"
" " " 30 " 25	" 12	"
" " " 25 " 20	" 11	"
" " " 20 " 15	" 9	"
" " " 15 " 10	" 4	"

Здесь в первом ряду классовой промежуток $k=3$, а во втором $k=5$.

Занесем все эти данные в расчетную решетку, состоящую из 3-х частей (для двух заданных рядов и для их контрольной сводки), причем черточки в этих частях проставим всюду на одном и том же уровне, хотя бы например на уровне 5-той строчки сверху (против частот 14 и 10).

I ряд:					II ряд:					Контр. сводка:				
p					p					p				
1					6					7				
2				—	10				—	12				—
3			—	—	12			—	—	15			—	—
6		—	—	—	11		—	—	—	17		—	—	—
14	—	—	—	—	10	—	—	—	—	24	—	—	—	—
29		—	—	—	12		—	—	—	41		—	—	—
12			—	—	11			—	—	23			—	—
4				—	9				—	13				—
1					4					5				

Таким образом начало отсчетов в первом ряду:

$$C = \frac{21 + 18}{2} = \frac{39}{2} = 19,5,$$

а во втором ряду:

$$C = \frac{35 + 30}{2} = \frac{65}{2} = 32,5.$$

Просуммировав частоты обоих заданных рядов по горизонтальным строчкам, получим частоты сводки. Здесь середина черточек пришлась на уровне частоты $p=24$.

В данном случае такая сводка частот составлена лишь для контроля вычислений и не имеет, очевидно, никакого реального смысла.

Выполнив во всех 3-х частях расчетной решетки обычные промежуточные операции, получим следующие итоги a и b (см. таблицу ниже).

Эти итоги следует теперь проверить. Для этого одноименные итоги во всех частях решетки складываются, и сумма их сравнивается с соответствующим итогом контрольной сводки. Если эта сумма оказывается в точности равной данному итогу сводки, вычисления могут быть признаны правильными, в противном случае следует искать ошибку где либо в промежуточных операциях.

I ряд:					II ряд:					Сводка:				
p	22	15	6	1	p	89	78	34	6	p	111	93	40	7
1	1	1	1	1	6	6	6	6	6	7	7	7	7	7
2	3	4	5	—	10	16	22	28	—	12	19	26	33	—
3	6	10	—	—	12	28	50	—	—	15	34	60	—	—
6	12	—	—	—	11	39	—	—	—	17	51	—	—	—
14	—	—	—	—	10	—	—	—	—	24	—	—	—	—
29	46	—	—	—	12	36	—	—	—	41	82	—	—	—
12	17	23	—	—	11	24	41	—	—	23	41	64	—	—
4	5	6	7	—	9	13	17	21	—	13	18	23	28	—
1	1	1	1	1	4	4	4	4	4	5	5	5	5	5
72	69	30	8	1	85	77	62	25	4	157	146	92	33	5

В нашем примере для I ряда итог $a_1 = 22$, а для II ряда $a_1 = 89$. Сумма этих итогов $22 + 89 = 111$ точно совпала с итогом a_1 в контрольной сводке (там $a_1 = 111$).

Для итогов a_2 имеем следующие результаты их проверки:

в I ряду	$a_2 = 15$	В контрольной сводке
во II "	$a_2 = 78$	$a_2 = 93$
сумма их = 93		

Проверяя этим же способом итоги a_3 , получим:

$$6 + 34 = 40.$$

Для итогов a_4 будем иметь:

$$1 + 6 = 7.$$

Аналогично проверяются и все нижние итоги b_1, b_2, b_3 и b_4 .

$$69 + 77 = 146$$

$$30 + 62 = 92$$

$$8 + 25 = 33$$

$$1 + 4 = 5.$$

Вычислив и проверив все 8 итогов, определим для каждого из двух заданных рядов и для их контрольной сводки отдельные пары разностей и сумм этих итогов: $a - b$

и $a + b$. Записав эти разности и суммы в соответствующие части расчетной решетки, получим внизу этой решетки следующие две дополнительные строчки *):

	Для I ряда:				Для II ряда:				Для сводки:			
$a - b$	-47	-15	-2	0	+12	+16	+9	+2	-35	+1	+7	+2
$a + b$	91	45	14	2	166	140	59	10	257	185	73	12

Все эти суммы и разности итогов следует проверить указанным выше способом. Так, например, в частности:

$$\text{в I ряду разность} \quad a_1 - b_1 = -47$$

$$\text{во II " " "} \quad a_1 - b_1 = +12$$

$$\text{сумма этих чисел} \quad = -35$$

что мы имеем и в сводке, для которой также

$$a_1 - b_1 = -35.$$

Аналогично проверяются и все остальные разности этих итогов:

$$-15 + 16 = +1$$

$$-2 + 9 = +7$$

$$0 + 2 = +2.$$

Проверка сумм ($a + b$) дает здесь следующие результаты:

$$91 + 166 = 257$$

$$45 + 140 = 185$$

$$14 + 59 = 73$$

$$2 + 10 = 12.$$

Проверив эти вычисления, следует помножить отдельные разности или суммы итогов на определенные числовые коэффициенты, указанные в ориентировочной таблице (см. стран. 67). После сложения таких произведений получим для каждого ряда и для сводки все 4 вспомогательные суммы S_1, S_2, S_3 и S_4 .

*) В целях экономии места здесь воспроизведена лишь нижняя часть расчетной решетки. Верхняя же остается без изменения.

Для I ряда:	Для II ряда:	Для сводки:
$S_1 = -47$	$S_1 = +12$	$S_1 = -35$
$S_2 = 181$	$S_2 = 446$	$S_2 = 627$
$S_3 = -149$	$S_3 = +162$	$S_3 = +13$
$S_4 = 1273$	$S_4 = 4490$	$S_4 = 5763$

Эти суммы необходимо также проверить, складывая их в отдельных рядах и сличая результаты со сводкой:

Для S_1 получим	$-47 + 12 = -35$
" S_2 "	$181 + 446 = 627$
" S_3 "	$-149 + 162 = +13$
" S_4 "	$1273 + 4490 = 5763$

На этом проверка кончается, и в дальнейшем контрольная сводка уже не принимается во внимание.

Определяя для обоих заданных рядов M , Σx^2 , σ и вспомогательное значение сигмы ς , получим:

Для I ряда (при $n=72$, $C=19,5$ и $k=3$)

$$M = 17,54 \quad \sigma = \pm 4,33$$

$$\Sigma x^2 = 150,32 \quad \varsigma = \pm 1,44$$

Для II ряда (при $n=85$, $C=32,5$ и $k=5$)

$$M = 33,21 \quad \sigma = \pm 11,43$$

$$\Sigma x^2 = 444,31 \quad \varsigma = \pm 2,29$$

Для вычисления показателей асимметрии и эксцесса определим оба значения d по формуле:

$$d = \frac{S_1}{n\varsigma}$$

Для I ряда будем иметь:

$$d = \frac{S_1}{n\varsigma} = \frac{-47}{72 \cdot 1,44} = -0,45$$

Для II ряда получим:

$$d = \frac{S_1}{n\varsigma} = \frac{+12}{85 \cdot 2,29} = +0,06$$

Так как первое из этих значений d отрицательное, то соответствующую ему поправку a следует брать со знаком

плюс. Для второго d (положительного) поправку a следует взять со знаком минус. По таблице, приложенной в конце книги (см. стран. 424), находим:

$$\begin{array}{lll} \text{для } d = -0,45 & a = +1,4411 & e = 1,338019 \\ \text{" } d = +0,06 & a = -0,1802 & e = 0,021639 \end{array}$$

Теперь вычислим вспомогательные (ошибочные) значения показателей A_0 и E_0 по формулам:

$$A_0 = \frac{S_3}{n\varsigma^3}$$

$$E_0 = \frac{S_4}{n\varsigma^4} - 3$$

В данном случае, для I ряда:

$$A_0 = \frac{-149}{72 \cdot (1,44)^3} = -0,69$$

$$E_0 = \frac{1273}{72 \cdot (1,44)^4} - 3 = +1,06$$

а для II ряда:

$$A_0 = \frac{+162}{85 \cdot (2,29)^3} = +0,16$$

$$E_0 = \frac{4490}{85 \cdot (2,29)^4} - 3 = -1,07$$

Имея в своем распоряжении все числовые значения A_0 , E_0 , d , a и e , мы можем определить и точные значения показателей A и E по формулам:

$$A = A_0 + a$$

$$E = E_0 - 4dA_0 + e$$

В нашем примере для I ряда: *)

$$A = -0,69 + 1,44 = +0,75$$

$$E = +1,06 - 4(-0,45)(-0,69) + 1,34 = +1,16$$

а для II ряда:

$$A = +0,16 - 0,18 = -0,02$$

$$E = -1,07 - 4(+0,06)(+0,16) + 0,02 = -1,09$$

*) Так как оба показателя A_0 и E_0 вычислены здесь с двумя знаками после запятой, то и поправки a и e достаточно взять с тем же количеством знаков.

Записав эти данные в расчетную решетку, получим:

I ряд					II ряд					Сводка				
<i>p</i>	22	15	6	1	<i>p</i>	89	78	34	6	<i>p</i>	111	93	40	7
1	1	1	1	1	6	6	6	6	6	7	7	7	7	7
2	3	4	5	—	10	16	22	28	—	12	19	26	33	—
3	6	10	—	—	12	28	50	—	—	15	34	60	—	—
6	12	—	—	—	11	39	—	—	—	17	51	—	—	—
14	—	—	—	—	10	—	—	—	—	24	—	—	—	—
29	46	—	—	—	12	36	—	—	—	41	82	—	—	—
12	17	23	—	—	11	24	41	—	—	23	41	64	—	—
4	5	6	7	—	9	13	17	21	—	13	18	23	28	—
1	1	1	1	1	4	4	4	4	4	5	5	5	5	5
72	69	30	8	1	85	77	62	25	4	157	146	92	33	5
$a-b$	-47	-15	-2	0	$a-b$	+12	+16	+9	+2	$a-b$	-35	+1	+7	+2
$a+b$	91	45	14	2	$a+b$	166	140	59	10	$a+b$	257	185	73	12
$S_1 = -47$ $S_3 = -149$					$S_1 = +12$ $S_3 = +162$					$S_1 = -35$ $S_3 = +13$				
$S_2 = 181$ $S_4 = 1273$					$S_2 = 446$ $S_4 = 4490$					$S_2 = 627$ $S_4 = 5763$				
$C = 19,5$ $k = 3$					$C = 32,5$ $k = 5$					— —				
$M = 17,54$ $A_0 = -0,69$					$M = 33,21$ $A_0 = +1,16$									
$\sigma = \pm 4,33$ $E_0 = +1,06$					$\sigma = \pm 11,43$ $E_0 = -1,07$									
$\epsilon = \pm 1,44$ $d = -0,45$					$\epsilon = \pm 2,29$ $d = +0,06$									
$\lg \epsilon = 0,15984$					$\lg \epsilon = 0,35913$									
$a = +1,44$ $e = 1,34$					$a = -0,18$ $e = 0,02$									
$A = +0,75$ $E = +1,16$					$A = -0,02$ $E = -1,09$									

Здесь выписаны также и логарифмы вспомогательных сигм ($\lg \epsilon$), так как вычисление показателей A_0 и E_0 удобнее производить с помощью логарифмических таблиц.

IX. Некоторые соображения по поводу образования асимметрии и эксцесса.

Явления асимметрии и эксцесса графически выражаются в тех или иных перемещениях вершины соответствующей кривой распределения либо в горизонтальном направлении при асимметрии, либо по вертикали для случаев эксцесса. Однако, то или иное положение вершины этой кривой само по себе далеко еще не исчерпывает всех характерных особенностей асимметричного и эксцессивного распределения и не может поэтому рассматриваться в качестве единственного и основного признака данных явлений. Дело в том, что положение вершины кривой распределения тесно связано также и с величиной крайних отклонений, в чем собственно говоря и заключается ключ к интерпретации интересующих нас случаев распределения.

Анализ этих случаев показывает, что при положительной асимметрии ($A > 0$), графически выражающейся сдвигом вершины кривой влево, некоторое сравнительно небольшое число положительных отклонений ($+x$), помечаемых на оси абсцисс справа от M , достигают очень большой величины, в то время как слева от M значительное количество отрицательных отклонений ($-x$) имеют лишь сравнительно малую абсолютную величину.

При отрицательной асимметрии ($A < 0$), т. е. при сдвиге вершины кривой вправо, наблюдается обратное соотношение: малое количество больших отрицательных отклонений ($-x$) в левой части кривой уравнивается множеством мелких положительных отклонений ($+x$) справа от M .

При положительном эксцессе ($E > 0$), когда кривая распределения имеет чрезмерно высокий подъем в середине ряда, большинство положительных и отрицательных отклонений оказываются по своей абсолютной величине сравнительно небольшими, но зато отдельные отклонения достигают чрезвычайно больших значений, увеличивая собой общий размах колебаний варьирующего признака в обе стороны от M (например в пределах до $\pm 4 \sigma$ от M , в то время как при нормальном распределении эти пределы ограничены приблизительно 3σ).

При отрицательном же эксцессе ($E < 0$), т. е. при низковоершинности, плосковершинности или провале кривой распределения в ее середине вовсе не наблюдается особенно больших положительных и отрицательных отклонений, вся масса которых таким образом варьирует в сравнительно узких пределах (напр. в пределах всего лишь $\pm 2 \sigma$ от M).

Присутствием малого количества этих больших по своей абсолютной величине отклонений и определяется характер распределения ряда. Если в общей их массе содержится несколько сравнительно очень больших положительных отклонений, ряд обнаружит положительную асимметрию; при наличии в нем неболь-

Подобные случаи наблюдаются иногда в рядах распределения линейных размеров тех или иных частей организма животных и растений, естественно ограниченных минимальным своим пределом $V = 0$. Впрочем при сравнительно небольшом варьировании этих размеров такой минимальный их предел (0) чаще всего находится на относительно далеком от M расстоянии, вследствие чего, как общее правило, у большинства морфологических признаков (а в частности и у признаков антропометрических) явление асимметрии наблюдается редко, и если наблюдается, то в очень слабых степенях.

Двухстороннее ограничение ряда по тем же соображе-

Аналогичным ходом рассуждений представляется иногда возможным проанализировать и другие случаи проявления асимметрии и эксцесса, для чего необходимо, очевидно, учесть все особенности данного явления. Чисто формальный анализ кривых распределения, основанный только на их внешнем виде и геометрических свойствах, сам по себе сможет дать очень немного, если у исследователя нет полной ориентировки в той области специальных знаний, к которой относятся анализируемые им факты.

В рамках элементарного и к тому же не специального руководства совершенно не представляется возможным сколько-нибудь подробно изложить причины образования асимметрии и эксцесса в различных случаях, встречающихся на практике, тем более, что этот вопрос в настоящее время далеко еще не может считаться разрешенным с исчерпывающей полнотой. Поэтому мы ограничимся здесь лишь приведенными выше общими соображениями о приемах интерпретации этих явлений, отослав читателей, интересующихся детальным их анализом, к тем специальным сочинениям, в которых излагается статистический метод в тех или иных конкретных его приложениях.

Для нас же представляют здесь особый интерес главным образом те случаи образования асимметрии и эксцесса, при которых эти явления данному материалу вообще говоря несвойственны и обнаруживаются в нем лишь под влиянием привходящих второстепенных причин.

Одна из этих причин может заключаться в смешении неоднородного или недостаточно однородного материала.

Пусть, например, имеем два следующие ряда:

I ряд:		II ряд:	
Варианта 8 встретиась . . .	1 раз.	Варианта 10 встретиась . . .	2 раза.
" 7 " . . .	9 "	" 9 " . . .	18 "
" 6 " . . .	45 "	" 8 " . . .	90 "
" 5 " . . .	70 "	" 7 " . . .	140 "
" 4 " . . .	45 "	" 6 " . . .	90 "
" 3 " . . .	9 "	" 5 " . . .	18 "
" 2 " . . .	1 "	" 4 " . . .	2 "

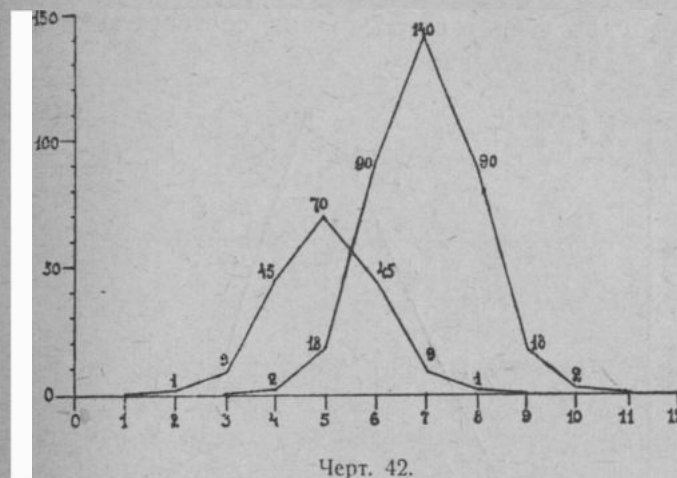
Подвергнув эти данные обычной обработке, получим для I ряда: ($n = 180$):

$$M = 5 \quad A = 0 \\ \sigma = \pm 1 \quad E = 0$$

а для II ряда ($n = 360$):

$$M = 7 \quad A = 0 \\ \sigma = \pm 1 \quad E = 0$$

в чем нетрудно убедиться, выполнив соответствующие вычисления в действительности.



Черт. 42.

Отсюда видно, что при совершенно одинаковом варьировании заданных рядов (в обоих случаях $\sigma = 1$) и при отсутствии в них как асимметрии, так и эксцесса ($A = 0$ и $E = 0$) I ряд имеет лишь несколько меньшее среднее по сравнению со II рядом и включает в себе вдвое меньшее количество вариантов.

Полигоны этих рядов представлены на одном общем чертеже 42. Из этого чертежа видно, что полигоны здесь как бы заходят друг на друга, причем площадь полигона II ряда вдвое больше площади I полигона, расположенного слева.

Соединим теперь оба ряда в один, подсчитав для каждого отдельного значения варианты V ее повторяемость

в обоих рядах. Так, например, варианта $V=4$ в I ряду встречается 45 раз, а во II ряду только 2 раза. Следовательно в смешанном ряду это значение варианты $V=4$ будет встречаться уже $45+2=47$ раз. Аналогично выясним и все остальные частоты p этого смешанного ряда. В результате получим новый ряд, полигон которого представлен на черт. 43 (частотами этого ряда служат следующие числа: 1, 9, 47, 88, 135, 149, 91, 18 и 2).

Для этого ряда:

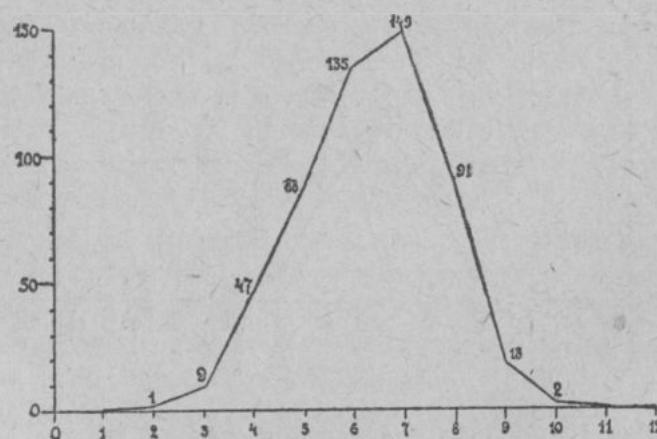
$$M = 6,33$$

$$\sigma = \pm 1,37$$

$$A = -0,23$$

$$E = -0,33$$

в чем нетрудно убедиться, выполнив соответствующие вычисления в действительности.



Черт. 43.

Итак, асимметрия и эксцесс в некоторых случаях могут появиться в результате смешения неоднородного материала, части которого, взятые изолированно, имеют строго нормальное распределение.

Подобный случай может встретиться на практике при соединении, например, антропометрических, психометрических или иных данных, относящихся к детям различного возраста или пола.

Если соединяемые ряды имеют одинаковое число измерений n и одинаковую σ , но сильно различаются по вели-

чине средних арифметических M , такое их объединение может дать в смешанном ряду отрицательный эксцесс. На черт. 44 представлен пример подобного случая в форме соответствующих полигонов.

Здесь для смешанного ряда:

$$M = 6,5$$

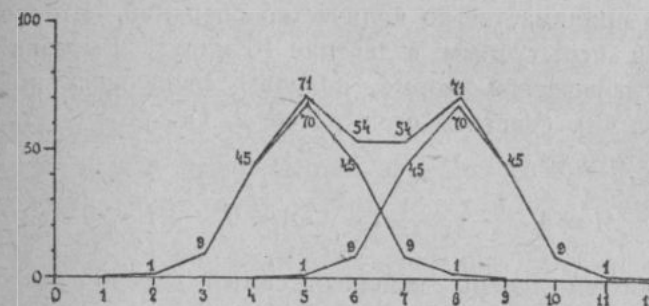
$$\sigma = \pm 1,80$$

$$A = 0$$

$$E = -0,96^*)$$

в то время, как в каждом из двух заданных рядов мы имели бы:

$$A = 0 \quad \text{и} \quad E = 0.$$



Черт. 44.

Итак, в данном случае, при объединении двух рядов, имеющих нормальное распределение, мы получили новый смешанный ряд, хотя и симметричный, но с заметным отрицательным эксцессом.

При подобной очень часто возможной неоднородности материала всегда представляется ценным своевременно ее обнаружить и устранить. В этих случаях наличие асимметрии и эксцесса в каком либо фактически полученном распределении может рассматриваться, как один из весьма показательных симптомов смешения двух или нескольких разнородных рядов, если для интерпретации этих явлений нет иных обоснованных предположений.

Образование асимметрии и эксцесса в некоторых случаях может вызываться также и выбором тех единиц изме-

*) Показатель E здесь удобно вычислять по формуле (см. стр. 64):

$$E = \frac{n \sum x^4}{(\sum x^2)^2} - 3$$

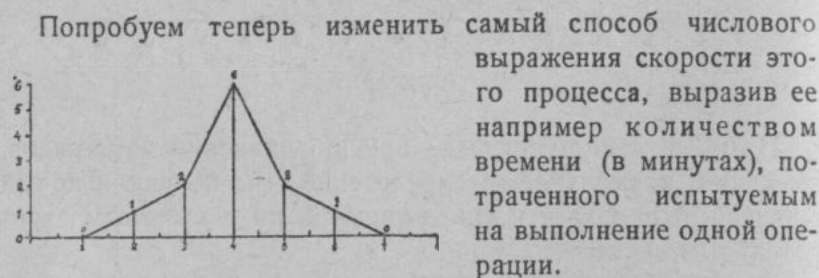
рения, в которых данный признак выражен количественно. Пусть, например, в результате испытания скорости какого либо процесса, слагающегося из однотипных операций, мы получим следующие данные:

1 человек в 10 минут выполнил 6 операций	
2 " " 10 " " 5 "	
6 " " 10 " " 4 "	
2 " " 10 " " 3 "	
1 " " 10 " " 2 "	

Здесь за единицу измерений проявленной скорости этого процесса принимается то количество операций, которое выполняется испытуемым в течение 10 минут. Выразив этот ряд в графической форме, получим следующий полигон (см. черт. 45). (Частоты его: 1, 2, 6, 2, 1).

Для этого ряда:

$$M = 4 \quad \sigma = \pm 1 \quad A = 0 \quad E = 0.$$



Черт. 45.

Попробуем теперь изменить самый способ числового выражения скорости этого процесса, выразив ее например количеством времени (в минутах), потраченного испытуемым на выполнение одной операции. Тот испытуемый, который в 10 минут выполнил 6 операций, на каждую из них потратил, очевидно, уже не 10 минут, а время, в 6 раз меньшее, т. е.

$$\frac{10}{6} = 1,666 \dots \text{ или приблиз. } = 1,7 \text{ мин.}$$

Выполнение 5 операций в тот же промежуток времени (10 минут) будет при этих условиях соответствовать времени в 5 раз меньшему, т. е.

$$\frac{10}{5} = 2 \text{ минуты.}$$

Выполнив это перечисление по отношению ко всем значениям скорости, встречающимся в заданном ряду, получим:

6 операций в 10 минут соответствуют	$\frac{10}{6} = 1,7$ мин. на 1 опер.
5 " " 10 " " "	$\frac{10}{5} = 2$ " " 1 "
4 " " 10 " " "	$\frac{10}{4} = 2,5$ " " 1 "
3 " " 10 " " "	$\frac{10}{3} = 3,3$ " " 1 "
2 " " 10 " " "	$\frac{10}{2} = 5$ " " 1 "

Так как мы изменили здесь лишь единицу измерения скорости, число же испытуемых, проявивших каждую данную скорость, осталось неизменным, то и частоты p нашего ряда сохранят здесь свою первоначальную величину. Переписав новые значения этих вариантов в обычном порядке (сверху большие, снизу малые), получим следующее их распределение:

Скорость в 5 мин. проявил 1 испытуемый	
" " 3,3 " " 2 "	
" " 2,5 " " 6 "	
" " 2 " " 2 "	
" " 1,7 " " 1 "	

Здесь с увеличением скорости числовые результаты ее измерения уменьшаются и наоборот, в то время как раньше наши варианты изменялись прямо пропорционально этим изменениям скорости.

Полигон этого видоизмененного ряда представлен на черт. 46, в котором обе крайние ветви ломаной линии опущены на ось абсцисс в тех точках, которые в прежних единицах скорости соответствовали значению крайних вариантов 1 и 7 (для которых частоты $p = 0$).

Для новых единиц измерения скорости будем иметь:

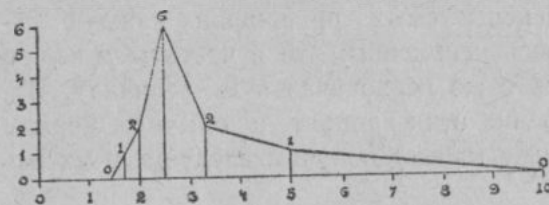
1 операция в 10 минут соответствует	$\frac{10}{1} = 10$ мин. на 1 опер.
7 " " 10 " " "	$\frac{10}{7} = 1,4$ " " 1 "

Эти значения вариантов: 1,4 и 10 и взяты на черт. 46 за крайние границы полигона.

Новый полигон обнаруживает явную асимметрию, что легко объясняется сжатием его по абсциссе с левой стороны.

Вычислим оба показателя A и E , соответствующие этим данным. В виду неравномерности интервалов между отдельными вариантами этого полигона обработку такого материала удобно выполнить по способу произведений (способ сумм применим лишь при разбивке ряда на равные друг другу классы).

Прежде всего вычислим среднее арифметическое M этого ряда. Перемножив все новые значения наших вариант V на их частоты p и сложив все эти произведения, получим сумму всех 12 вариант (ΣV):



Черт. 46.

$$\begin{aligned} 5 \cdot 1 &= 5 \\ 3,3 \cdot 2 &= 6,6 \\ 2,5 \cdot 6 &= 15 \\ 2 \cdot 2 &= 4 \\ 1,7 \cdot 1 &= 1,7 \\ \hline \Sigma V &= 32,3 \end{aligned}$$

Разделив эту сумму (32,3) на число измерений n (12) найдем среднее арифметическое

Итак для нового ряда:

$$\Sigma x^2 = 8,23 \quad \Sigma x^3 = +10,865 \quad \text{и} \quad \Sigma x^4 = 29,7331.$$

Откуда среднее квадратическое отклонение:

$$\sigma = \varsigma = \pm \sqrt{\frac{\Sigma x^2}{n}} = \pm \sqrt{\frac{8,23}{12}} = \pm \sqrt{0,79} = \pm 0,89.$$

Здесь σ совпадает с ς , так как никаких искусственных изменений вариант пропорционально k мы не производили. Все варианты и их отклонения сохранили при вычислениях свою первоначальную величину, что и соответствует случаю $k=1$, при котором $\sigma = \varsigma$.

Подставляя найденные значения всех сумм Σ и ς в формулы для A и E , получим:

$$\begin{aligned} A &= \frac{\Sigma x^3}{n \varsigma^3} = \frac{+10,865}{12 \cdot (0,89)^3} = +1,29 \\ E &= \frac{\Sigma x^4}{n \varsigma^4} - 3 = \frac{29,7331}{12 \cdot (0,89)^4} - 3 = -0,37. \end{aligned}$$

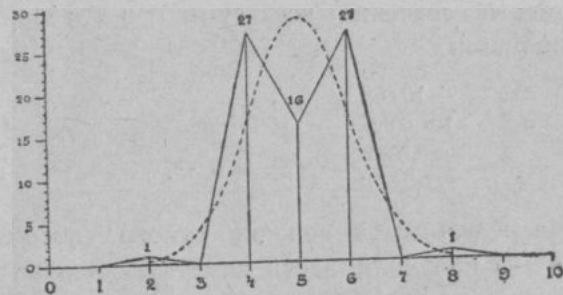
Таким образом в результате изменения самого способа измерения скорости мы получили здесь заметную положительную асимметрию ряда и некоторый его отрицательный эксцесс.

С подобными явлениями мы всегда можем встретиться при пользовании сложными единицами измерения, допускающими различные способы их конструирования. Элемен-

мым условием принадлежности ряда к категории „нормальных“, тем не менее в некоторых случаях может оказаться все же не вполне еще достаточным.

Пусть, например, имеем ряд, полигон которого в сопоставлении с нормальной кривой представлен на черт. 47. (Частотами p этого ряда служат числа: 1, 0, 27, 16, 27, 0, и 1).

Несмотря на строгую симметричность этого ряда, признать его „нормальным“, очевидно, нельзя, так как здесь наблюдается во первых провал в середине полигона, а во вторых некоторое расхождение крайних его ветвей, особенно заметное по сравнению с центральной частью полигона.



Черт. 47.

Однако вычисления дают здесь совершенно неожиданный результат:

$$E = 0,$$

указывающий, казалось бы, на строго нормаль-

ное распределение этих 72 вариант по 7-ми классам их числовых значений.

Этот парадоксальный на первый взгляд вывод следует объяснить здесь взаимной компенсацией влияний со стороны двух противоположных особенностей заданного полигона. Провал его в середине, определяющий отрицательный эксцесс, в сочетании с большими отклонениями крайних вариантов, характерными для положительного эксцесса, друг друга уравновесили и при вычислении показателя E обратили его в 0.

Таким образом значение $E = 0$, всегда и без исключения свойственное нормальному распределению, может быть в то же время констатировано иногда и в эксцессивных рядах. Поэтому, получив $E = 0$, мы не можем еще утверждать, что заданный ряд непременно является нормальным, но должны будем учесть также и геометрическую форму данного распределения. Только соединение этих двух симп-

томов может дать достаточные основания для суждения о принадлежности фактически полученного распределения к категории нормальных, в то время как значение $A = 0$ прямо уже указывает на симметрию ряда без необходимости обращаться к рассмотрению соответствующего графика.

Х. Медиана (Med) и мода (Mo).

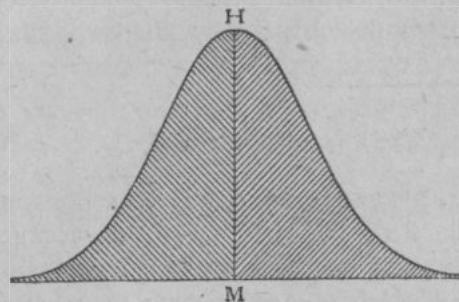
В первой части этого руководства и во всех предшествующих главах настоящей (II-й) его части излагались методы статистического анализа вариационных рядов, основанные на положениях так называемого способа „наименьших квадратов“. При этом способе обработки массовых материалов, приводящем в конечном итоге к определению для заданного ряда четырех основных его характеристик: M , σ , A и E , мы оперировали исключительно с „центральными отклонениями“ (x), сумма квадратов которых (Σx^2), как известно, является наименьшей по сравнению с суммами квадратов „вспомогательных отклонений“, отсчитанных от произвольной величины C , не совпадающей со средним арифметическим M . Это условие (предельной малости суммы Σx^2) определяет здесь как самую величину M , от которой отсчитываются центральные отклонения, так и все остальные сводные характеристики ряда: σ , A и E , в формулы для вычисления которых также входит эта минимальная сумма (Σx^2) вместе с суммами высших степеней тех же самых „центральных“ отклонений (Σx^3 и Σx^4).

Но на ряду с этой системой характеристики средней величины варьирующего признака (M) и средних пределов его колебаний (σ) мыслимы также и другие способы количественной оценки различных свойств и особенностей заданного вариационного ряда, в которых вместо M и σ применяются аналогичные сводные характеристики распределения, но вычисляемые уже по совершенно иному принципу.

Рассмотрим, прежде всего, существующие способы такой замены по отношению к среднему арифметическому M . Как известно, при нормальном распределении ряда (см. черт. 48) его среднее арифметическое M обладает следующими двумя основными свойствами:

1) В точке M кривая распределения имеет наибольшую высоту (ординату).

2) Эта наибольшая ордината делит всю площадь кривой на две равные части.



Черт. 48.

Вспомним, что высота кривой распределения указывает на частоту (повторяемость) того или иного значения варьирующего признака, а площадь отдельных участков этой кривой — на количество вариантов, заключенных в границах соответствующего интервала. Отсюда следует, что при строго нормальном распределении ряда среднее арифметическое M представляет собой в то же самое время и наиболее часто встречающуюся (типичную) величину данного варьирующего признака, и что количество вариантов меньших M в точности равно количеству вариантов больших M .

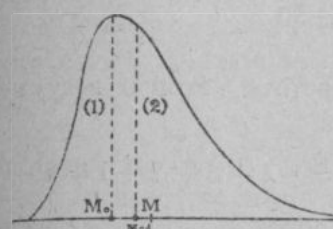
Но эти свойства присущи среднему арифметическому M только лишь при наличии нормального (или вообще симметричного) распределения. При малейшей асимметрии ряда наиболее часто встречающееся значение варьирующего признака (так называемая „мода“) уже не совпадает с M , точно так же как и граница, разделяющая все варианты на части, равные друг другу по численности (так называемая „медиана“), окажется расположенной где-либо слева или справа от M .

Но эти свойства присущи среднему арифметическому M только лишь при наличии нормального (или вообще симметричного) распределения. При малейшей асимметрии ряда наиболее часто встречающееся значение варьирующего признака (так называемая „мода“) уже не совпадает с M , точно так же как и граница, разделяющая все варианты на части, равные друг другу по численности (так называемая „медиана“), окажется расположенной где-либо слева или справа от M .

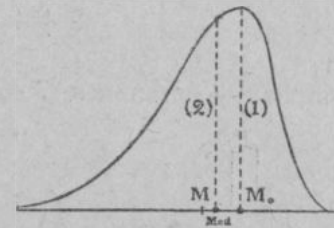
На чертежах 49 и 50 представлены две асимметричные кривые распределения с двумя ординатами в каждой. Одна из этих ординат (1) является здесь наибольшей (по высоте), а другая (2) делит площадь кривой на две равные части.

У основания этих ординат на оси абсцисс проставлены символические обозначения: Mo (мода) у основания наибольшей ординаты и Med (медиана) у основания ординаты, разделяющей площадь кривой пополам. Среднее арифметическое помечено на оси абсцисс черточкой с поставленной около нее буквой M .

Из этих чертежей видно, что при положительной асимметрии (черт. 49) обе ординаты оказались сдвинутыми влево от M , а при отрицательной (черт. 50) — наоборот вправо, причем мода (Mo) в обоих случаях отстоит от среднего арифметического (M) дальше, чем медиана (Med), и, стало



Черт. 49.



Черт. 50.

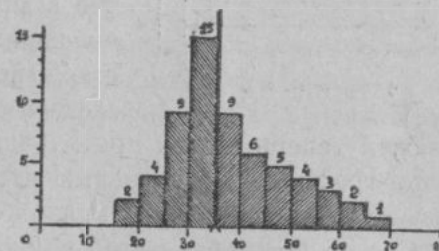
быть, медиана (Med) заключается всегда в промежутке между модой (Mo) и средним арифметическим (M).

Математический анализ кривых распределения показывает, что для случаев умеренной (не слишком сильной) асимметрии расстояние между точками Mo и Med оказывается приблизительно вдвое больше расстояния между точками Med и M , на основании чего между модой, медианой и средним арифметическим можно установить следующие приближенные соотношения:

$$1) Mo = 3 Med - 2M$$

$$2) Med = \frac{Mo + 2M}{3}$$

$$3) M = \frac{3 Med - Mo}{2}$$



Черт. 51.

Для иллюстрации понятия о медиане рассмотрим следующий числовой пример. Пусть имеем вариационный ряд, гистограмма которого представлена на черт. 51. (Частотами его служат числа: 2, 4, 9, 15, 9, 6, 5, 4, 3, 2, 1). Чтобы определить для этого ряда медиану (Med), найдем на оси абсцисс такую точку (отмеченную на чертеже крестиком), через которую проходит ордината, разделяющая всю площадь гистограммы на две равные части. В нашем примере эта

точка, очевидно, придется на 35-ом делении горизонтального масштаба, так как с обеих сторон от этой границы расположено по одинаковому количеству вариантов. В самом деле: слева от точки, помеченной крестиком, находятся 4 класса, общая сумма частот которых:

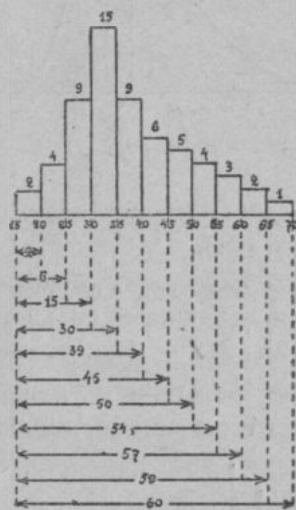
$$2 + 4 + 9 + 15 = 30$$

а справа — остальные 7 классов, частоты которых в сумме дают:

$$9 + 6 + 5 + 4 + 3 + 2 + 1 = 30$$

Таким образом 35-ое деление горизонтального масштаба и разделяет в нашем примере площадь всей гистограммы на две равные части (по 30 вариант с каждой стороны от *Med* из общего их числа $n = 60$).

В приведенном здесь примере медиана (*Med*) пришлась как раз на границе между двумя классами, что в значительной степени облегчило нам подсчет частот p для участков гистограммы, расположенных слева и справа от этой медианы.



Черт. 52.

Укажем теперь общий прием решения этой задачи при помощи графика „накопленных частот“ или, как его часто называют, „интегральной кривой“.

Способ построения этой кривой мы разясним на примере гистограммы уже разо ранного выше ряда (см. черт. 52). Поставим себе целью по отношению к каждой отдельной границе между двумя соседними классами выяснить то количество вариантов, которое лежит слева от этой границы (т. е. количество вариантов, меньших числового значения данной границы, так как по оси абсцисс слева отменяются малые значения вариант, а справа — большие).

Наш ряд слева ограничен значением признака $V = 15$, и вариант меньших этой крайней границы очевидно здесь вовсе не встречается (количество их = 0).

Слева от следующей по порядку границы (20) расположен один класс (от 15 до 20), частота которого $p = 2$. Следовательно слева от границы 20 расположены две варианты. Это количество вариант указано на черт. 52 соответствующей цифрой 2, поставленной внизу гистограммы (у стрелки).

Слева от границы 25 лежат уже два класса, частоты которых составляют в сумме

$$2 + 4 = 6,$$

что и указано внизу гистограммы соответствующей цифрой.

Слева от границы 30 к 6 прежним вариантам присоединяются еще 9 вариант следующего класса (частота которого $p = 9$), что дает в сумме

$$6 + 9 = 15.$$

Для границы 35 будем иметь очевидно

$$15 + 15 = 30 \text{ и т. д.}$$

Результаты этих подсчетов записаны на чертеже 52 соответствующими цифрами внизу гистограммы. Ясно, что эти „накопленные частоты“ получаются здесь посредством последовательного прибавления обычных классовых частот p к сумме этих же частот, вычисленных для ближайшей слева границы. В конечном результате получим следующий ряд накопленных частот для всех 12 границ наших 11-ти классов:

Границы классов: 15 20 25 30 35 40 45 50 55 60 65 70
Накоплен. частоты: 0 2 6 15 30 39 45 50 54 57 59 60

Здесь, например, число 6 (третье слева во второй строчке), подписанное под числом 25 (в первой строчке), указывает на то, что слева от границы 25 в нашем ряду заключается 6 вариант; последнее же справа число 60, подписанное под числом 70, указывает на то, что слева от крайней границы 70 располагаются все 60 вариант, составляющих данный ряд.

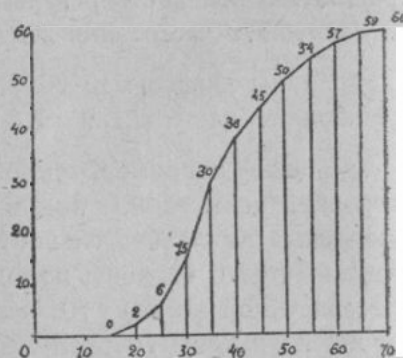
Этот же ряд „накопленных частот“ (обозначим их символом Σp) мы могли бы получить и привычным для нас способом сумм в соответствующей расчетной решетке. Произведя обычные суммирования частот p в клетках столбца

Σp в направлении снизу вверх (указанном стрелкой), мы и получим, очевидно, те же самые накопленные частоты Σp .

V	p	Σp
70 — 65	1	60
65 — 60	2	59
60 — 55	3	57
55 — 50	4	54
50 — 45	5	50
45 — 40	6	45
40 — 35	9	39
35 — 30	15	30
30 — 25	9	15
25 — 20	4	6
20 — 15	2	2

Каждая из накопленных частот при этом способе вычисления будет относиться к верхней (большей) границе соответствующего класса (выписанной в столбце V слева). Так, например, накопленная частота 57 (третья сверху в столбце Σp), относится к верхней границе этого класса 60 (а не к нижней — 55). Для нижней границы самого малого класса (в нашем примере для 15) накопленная частота очевидно всегда будет = 0 (в расчетной решетке она не помечена вовсе), а для верхней границы самого большого класса (у нас для 70) она будет = общему числу всех вариантов n (в нашем примере 60).

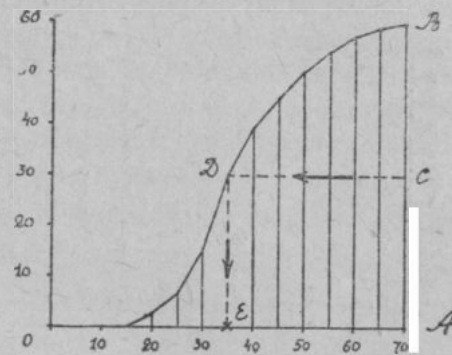
Вычислив эти накопленные частоты и приняв их за ординаты, мы можем построить особый „график накопленных частот“ или „интегральную кривую“ (см. черт. 53). Здесь вертикальный масштаб (для накопленных частот) взят несколько мельче прежнего масштаба обычных классовых частот p , а горизонтальный масштаб (для классовых границ) оставлен без изменений. У вершин всех ординат цифрами помечены вычисленные ранее накопленные частоты. В этом графике следовательно высота ординат определяет количество вариантов, расположенных на гистограмме слева от каждой границы. Крайняя же справа наибольшая накопленная частота 60 (у границы 70) равна по вертикальному масштабу общему числу всех измерений n .



Черт. 53.

Интегральную кривую удобно вычерчивать на миллиметровой бумаге.

Так как медиана (Med) представляет собой такую границу, слева и справа от которой заключается ровно половина всех вариантов заданного ряда, то наша задача теперь сводится к определению на интегральной кривой Медианной точки, для которой ордината графика накопленных частот = половине максимальной его высоты 60, т. е. = 30 делениям вертикального масштаба.



Черт. 54.

Разделив наибольшую ординату AB пополам в точке C , мы и получим высоту ее половины AC (см. черт. 54). Чтобы найти на оси абсцисс точку E , для которой ордината ED также равнялась бы половине полной высоты AB

(т. е. = AC), следует из точки C провести прямую, параллельную оси абсцисс до пересечения с интегральной кривой в точке D , а из этой точки пересечения (D) опустить перпендикуляр на ось абсцисс. Основание этого перпендикуляра E по масштабу горизонтальной оси и определит значение медианы Med , которая в данном случае, так же как и раньше, оказалась = 35 (пришлась на 35-м делении горизонтального масштаба).

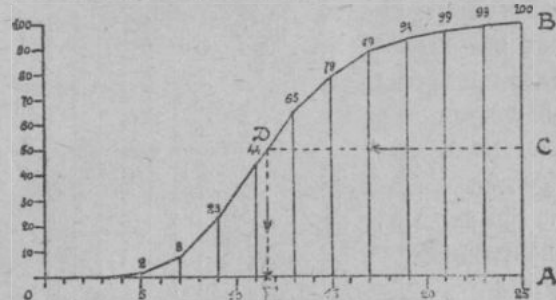
Этот графический прием определения медианы особенно удобен в тех случаях, когда

значение медианы не совпадает ни с одной из классовых границ. Зададимся, например, целью вычислить медиану для неоднократно уже приводимого ряда, состоящего из 100 ва-

V	p	Σp
25 — 23	1	100
23 — 21	2	99
21 — 19	3	97
19 — 17	5	94
17 — 15	10	89
15 — 13	14	79
13 — 11	21	65
11 — 9	21	44
9 — 7	15	23
7 — 5	6	8
5 — 3	2	2

риант и разбитого на 11 классов с классовым промежутком $k=2$ (см. стр. 7).

Вычислим накопленные частоты (Σp) и построим по ним интегральную кривую, представленную на чертеже 55.



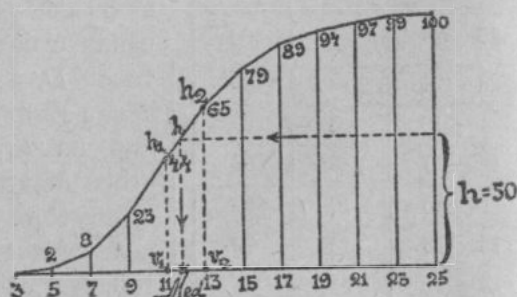
Черт. 55.

Разделив наибольшую ординату (высоту) этой кривой AB пополам в точке C , проведем через C горизонтальную прямую CD (влево) и из точки D ее пересечения с интегральной кривой опустим перпендикуляр DE . В данном случае основание этого перпендикуляра E пришлось между 11 и 12 делением горизонтального масштаба, несколько ближе к делению 12-му, приблизительно на расстоянии 0,6 от левой границы 11. Следовательно для заданного ряда его медиана:

$$Med = 11,6 \text{ (приблизительно).}$$

Точность определения медианы графическим способом всецело зависит от тщательности построения и от размера чертежа: чем крупнее масштаб, тем больше и точность. Читателю рекомендуется самому проделать эти (и все последующие) построения на большом листе миллиметровой бумаги.

Кроме указанного выше графического способа медиана может быть вычислена также и аналитическим путем (по формуле). Обозначим границы того класса, внутри которого оказалась при построении медиана Med , через V_1 (левая



Черт. 56.

граница) и V_2 (правая граница), а накопленные частоты в этих границах — через h_1 (для V_1) и h_2 (для V_2).

В нашем примере искомая медиана попала в границы класса от 11 до 13 (см. черт. 56). Следовательно у нас левая граница $V_1=11$, а правая $V_2=13$. Накопленная частота (ордината) h_1 для границы V_1 в нашем примере $=44$, а для границы V_2 ее $h_2=65$. Кроме того, через h мы обозначим ту высоту (ординату), которая при построении отсекается от интегральной кривой (она равна половине ее максимальной высоты). В нашем примере при числе измерений $n=100$ отсекается высота $h = \frac{n}{2} = 50$. Вспомним, далее, что классовой промежуток k у нас $=2$. Итак для заданного примера:

$$\begin{aligned} V_1 &= 11 & h_1 &= 44 & h &= 50 \\ V_2 &= 13 & h_2 &= 65 & k &= 2 \end{aligned}$$

Окончательная формула для вычисления медианы (Med) имеет следующий вид:

$$Med = V_1 + k \left(\frac{h - h_1}{h_2 - h_1} \right)$$

Подставляя сюда наши числовые данные, получим:

$$Med = 11 + 2 \left(\frac{50 - 44}{65 - 44} \right) = 11,57$$

Итак, медиана заданного ряда:

$$Med = 11,57.$$

Покажем теперь, как вычисляется мода (Mo), т. е. то значение варьирующего признака, которое в заданном ряду встречается чаще всего и для которого следовательно кривая распределения имеет наибольшую высоту.

Самый грубый способ приближенного вычисления моды заключается в определении середины того интервала, внутри которого значения варьирующего признака встречаются наиболее часто.

В нашем последнем примере (см. гистограмму на черт. 57) этому условию удовлетворяют два смежных класса: от 9 до 11 и от 11 до 13, частоты которых (21 и 21) являются наибольшими по сравнению со всеми остальными частотами этого ряда. Следовательно в данном случае границами интервала, заключающего в себе наиболее часто встречающиеся варианты, служат значения признака 9 и 13. Полу-

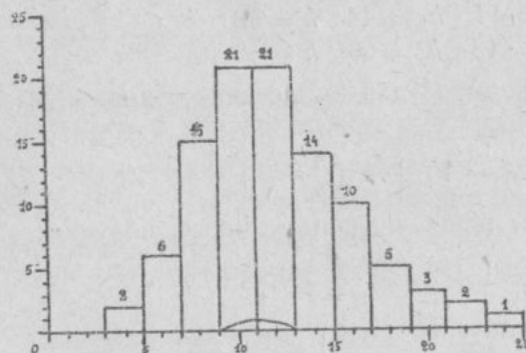
сумма этих значений (середина интервала) и может быть принята здесь за моду. Итак

$$Mo = \frac{9 + 13}{2} = \frac{22}{2} = 11.$$

Если бы у нас был всего один класс с наибольшей частотой p , то за моду (Mo) следовало бы принять, очевидно, середину этого единственного класса, вычисленную, как полусумму его границ.

Второй несколько более точный, но все же еще приближенный способ вычисления моды заключается в применении формулы, указанной в начале этой главы (см. стран. 103):

$$Mo = 3 Med - 2 M.$$



Черт. 57.

В нашем последнем примере медиана

$$Med = 11,57,$$

а среднее арифметическое

$$M = 12.$$

Подставляя эти данные в формулу для вычисления мо-

ды, получим:

$$Mo = 3 \cdot 11,57 - 2 \cdot 12 = 10,71.$$

Третий наиболее точный, но в то же время и самый сложный способ вычисления моды заключается в использовании с этой целью показателей асимметрии (A) и эксцесса (E), а также M и σ , между которыми с одной стороны и точным значением моды с другой существует следующая зависимость:

$$Mo = M - \frac{A \sigma (E + 6)}{2(5E - 6A^2 + 6)}.$$

В нашем примере (см. стр. 62)

$$M = 12$$

$$\sigma = 4$$

$$A = +0,57$$

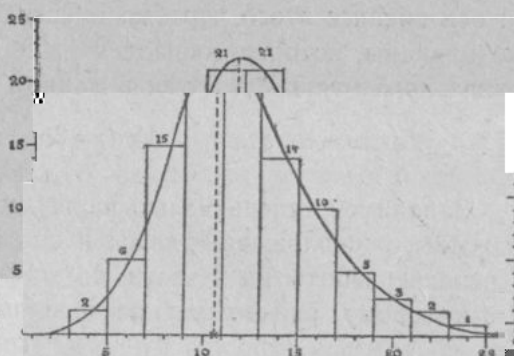
$$E = +0,22.$$

Подставляя эти данные в указанную выше формулу, получим:

$$Mo = 12 - \frac{0,57 \cdot 4 (0,22 + 6)}{2(5 \cdot 0,22 - 6 \cdot 0,57^2 + 6)} = 10,62.$$

При этом способе вычисления моды мы определяем то значение варьирующего признака, для которого ордината кривой распределения имеет наибольшую высоту (см. черт. 58).

Итак, применяя последовательно все три способа вычисления моды (с возрастающей точностью), мы получили:



Черт. 58.

- 1) $Mo = 11$
- 2) $Mo = 10,71$
- 3) $Mo = 10,62$.

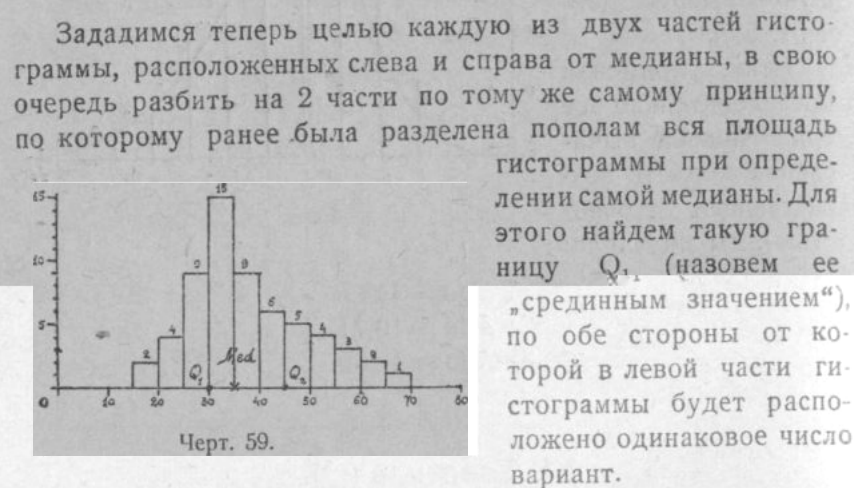
XI. Квартиль (Q).

В предыдущей главе мы познакомились с двумя видами сводных характеристик распределения: Med и Mo , принимаемых иногда за типичные значения варьирующего признака. Эти два возможных способа общей числовой характеристики вариационно о ряда могут в некоторых случаях заменять лишь его среднее арифметическое M , но совершенно не определяют еще средних размахов варьирования данного признака, как впрочем не отражает эту сторону явления и само среднее арифметическое M . Та или иная степень разбросанности (рассеяния) отдельных вариантов должна иметь свою особую меру, независимую от медианы и моды, подобно тому, как характеризующее это рассеяние среднее квадратическое отклонение σ является независимым от среднего арифметического M . Такой мерой рассеяния, аналогичной σ , но связанной уже не с M , а с медианой Med , служит так

называемый „квартиль“ Q . Принцип определения квартиля заключается в следующем.

Пусть имеем вариационный ряд, для которого в начале предыдущей главы (см. стр. 103) мы определили медиану Med . Гистограмма этого ряда представлена здесь на черт. 59. На оси абсцисс этого чертежа крестиком помечено положение медианы, которая совпадает здесь с 35-м делением горизонтального масштаба. Итак в данном случае:

$$Med = 35.$$



Зададимся теперь целью каждую из двух частей гистограммы, расположенных слева и справа от медианы, в свою очередь разбить на 2 части по тому же самому принципу, по которому ранее была разделена пополам вся площадь гистограммы при определении самой медианы. Для этого найдем такую границу Q_1 (назовем ее „срединным значением“), по обе стороны от которой в левой части гистограммы будет расположено одинаковое число вариант.

$$2 + 4 + 9 = 15,$$

а справа (до медианы) — только 1 класс, частота которого p также = 15.

Аналогичным способом найдем и „срединное значение“ Q_2 для правой части гистограммы, совпадающее в нашем случае с 45-м делением горизонтального масштаба. Здесь

два класса расположенные слева от границы Q_2 (до медианы), дают сумму частот:

$$9 + 6 = 15$$

и справа от Q_2 сумму частот 5-ти расположенных там классов:

$$5 + 4 + 3 + 2 + 1 = 15.$$

Итак тремя точками: Q_1 , Med и Q_2 мы разбили всю площадь гистограммы на 4 равные части, в каждую из которых вошло следовательно по 15 вариант из общего их числа $n = 60$.

Расстояния от Q_1 до Med и от Med до Q_2 по величине своей могут служить мерой варьирования заданного ряда. В самом деле, при сильной разбросанности отдельных вариантов в обе стороны от медианы границы Q_1 и Q_2 в свою очередь далеко отойдут от средней точки Med , при малом же рассеянии ряда обе они также приблизятся друг к другу *и к медиане*.

В нашем примере граница $Q_1 = 30$ удалена от $Med = 35$ на расстояние:

$$35 - 30 = 5,$$

а граница $Q_2 = 45$ от той же медианы $Med = 35$ на расстояние:

$$45 - 35 = 10.$$

Эти два неодинаковые расстояния от Q_1 и Q_2 до Med (в нашем примере 5 и 10) и могут характеризовать среднюю величину рассеяния заданного ряда. Так как неравенство этих расстояний, зависящее, очевидно, от асимметрии ряда, в данном случае нас само по себе не интересует, то в качестве искомой меры рассеяния в обе стороны от Med мы можем взять среднее арифметическое этих расстояний (5 и 10), т. е. попросту их полусумму:

$$\frac{5 + 10}{2} = \frac{15}{2} = 7,5.$$

Это среднее расстояние обеих границ Q_1 и Q_2 до медианы Med (т. е. 7,5) и может быть принято за окончательную характеристику средней величины варьирования ряда,

называемую „квартилем“ и обозначаемую обычно буквою Q . Итак, в нашем примере квартиль:

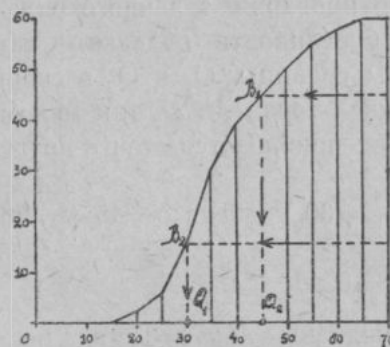
$$Q = 7,5.$$

Заметим, что то же самое значение квартиля Q мы могли бы получить и проще, вычислив его, как полуразность обоих „серединных“ значений левой и правой половины ряда Q_1 и Q_2 , т. е. по формуле:

$$Q = \frac{Q_2 - Q_1}{2}.$$

В нашем примере $Q_1 = 30$, а $Q_2 = 45$. Подставляя эти данные в указанную выше формулу, получим:

$$Q = \frac{45 - 30}{2} = 7,5$$



Черт. 60.

т. е. совершенно тот же окончательный результат.

В дальнейшем мы будем вычислять квартиль Q всегда по этой последней формуле.

Определение квартиля Q удобно выполнять также графическим способом при помощи интегральной кривой

(см. черт. 60, являющийся повторением черт. 53). Разделив наибольшую высоту этой кривой на 4 равные части, следует из точек A_1 и A_2 провести две горизонтальные прямые $A_1 B_1$ и $A_2 B_2$ до пересечения с интегральной кривой в точках B_1 и B_2 , а из этих точек опустить на ось абсцисс перпендикуляры $B_1 Q_1$ и $B_2 Q_2$. Основания этих перпендикуляров Q_1 и Q_2 , очевидно, и будут служить „серединными“ значениями левой и правой половины ряда. В данном случае, также как и раньше, $Q_1 = 30$, а $Q_2 = 45$, откуда искомый квартиль:

$$Q = \frac{45 - 30}{2} = 7,5.$$

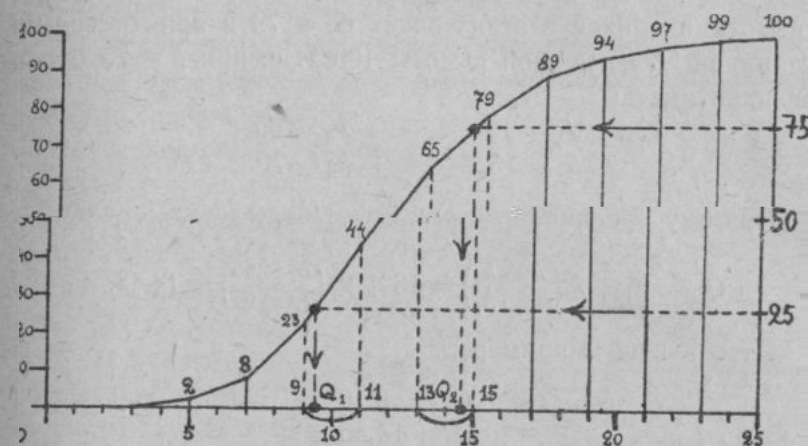
В том общем случае, когда „серединные“ значения придутся где-либо внутри классовых промежутков (а не точно на их границах, как это имело место в последнем примере),

оба значения Q_1 и Q_2 удобно вычислить по формуле, аналогичной той, которая была указана в прошлой главе для определения медианы Med (см. стран. 109):

$$Med = V_1 + k \left(\frac{h - h_1}{h_2 - h_1} \right).$$

В этой формуле следует заменить лишь медиану (Med) величинами Q_1 или Q_2 и взять по чертежу соответствующие значения V_1 , h_1 , h_2 и h .

Пусть имеем интегральную кривую, вычерченную по данным второго примера, разобранный в конце предыдущей главы (см. стран. 108).*)



Черт. 61.

Разделив наибольшую высоту этой кривой на 4 равные части и проведя две горизонтальные прямые до пересечения с интегральной кривой (см. черт. 61), опустим из этих точек два перпендикуляра и найдем их приблизительное положение на оси абсцисс.

В данном случае точка Q_1 легла в промежутке между границами класса от 9 до 11, а точка Q_2 оказалась в границах класса от 13 до 15.

Начнем вычисление с первой точки Q_1 . Для нее левая граница $V_1 = 9$. Высота интегральной кривой в этой точке $h_1 = 23$, а в соседней справа границе (11) высота $h_2 = 44$

*) Полезно выполнить это построение на большом листе миллиметровой бумаги.

Отсекаемая на интегральной кривой высота $h = 25$ ($1/4$ наибольшей высоты этой кривой 100).

Итак для точки Q_1 мы имеем:

$$\begin{aligned} V_1 &= 9 & h_1 &= 23 \\ h &= 25 & h_2 &= 44. \end{aligned}$$

Подставляя эти данные в указанную выше формулу и помня, что классовый промежуток $k = 2$, получим:

$$Q_1 = V_1 + k \left(\frac{h - h_1}{h_2 - h_1} \right) = 9 + 2 \left(\frac{25 - 23}{44 - 23} \right) = 9,19.$$

Для точки Q_2 , попавшей в границы класса от 13 до 15, с высотами кривой в этих точках 65 и 79 и при отсечении от кривой $3/4$ ее полной высоты 100, т. е. при $h = 75$ будем очевидно иметь:

$$\begin{aligned} V_1 &= 13 & h_1 &= 65 \\ h &= 75 & h_2 &= 79. \end{aligned}$$

Поэтому срединное значение правой половины ряда:

$$Q_2 = V_1 + k \left(\frac{h - h_1}{h_2 - h_1} \right) = 13 + 2 \cdot \left(\frac{75 - 65}{79 - 65} \right) = 14,43.$$

Итак в нашем примере:

$$\begin{aligned} Q_1 &= 9,19 \\ Q_2 &= 14,43, \end{aligned}$$

откуда искомый квартиль:

$$Q = \frac{Q_2 - Q_1}{2} = \frac{14,43 - 9,19}{2} = 2,62.$$

Вспомним, что для этого ряда мы определили уже среднее арифметическое M и среднее квадратическое отклонение σ . Сопоставим эти данные с вычисленными здесь значениями медианы Med и квартиля Q :

$$\begin{aligned} M &= 12 & Med &= 11,57 \\ \sigma &= 4 & Q &= 2,62. \end{aligned}$$

Итак медиана (Med) оказалась здесь несколько меньше среднего арифметического M , что легко находит себе объяснение в положительной асимметрии заданного ряда (при отрицательной асимметрии она получилась бы, наоборот,

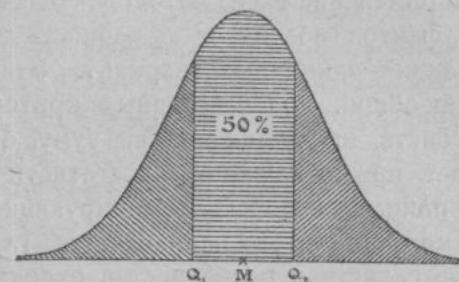
больше M), и квартиль Q не совпал со средним квадратическим отклонением σ . Это различие между σ и Q будет обнаруживаться при всяком (в том числе и при нормальном распределении, так как квартиль, как общее правило, оказывается всегда меньше сигмы.

При строго нормальном распределении между σ и Q существует следующее постоянное соотношение:

$$Q = \frac{2}{3} \sigma \text{ или точнее } Q = 0,6745 \sigma$$

$$\sigma = \frac{3}{2} Q \text{ или точнее } \sigma = 1,4826 Q.$$

Следовательно, при разбивке площади кривой нормального распределения на 4 равные части оба „срединные“ значения правой и левой ее половины Q_1 и Q_2 окажутся раздвинутыми от медианы Med (совпадающей здесь со средним арифметическим M) на одинаковые расстояния, приблизительно равные $2/3$ среднего квадратического отклонения σ



Черт. 62.

(см. черт. 62). Поэтому внутри интервала от $M - 2/3 \sigma$ до $M + 2/3 \sigma$ при нормальном распределении будет заключаться половина всех вариантов данного ряда (т. е. 50%).

Полезно запомнить, что медиана и квартиль, подобно M и σ , представляют собой именованные числа, выраженные в тех же самых единицах (см., килгр. и т. д.), в которых были измерены и отдельные варианты V . Однако при сравнении средних размахов колебаний у различных признаков (напр. у веса и у роста) удобнее бывает располагать мерами их варьирования в виде отвлеченных, а не именованных чисел, выразив для этого квартили Q в процентах от соответствующей медианы Med (подобно тому, как это делается при вычислении вариационного коэффициента v , представляющего собою процентное отношение σ к M).

Такой процентный квартиль Q_0 в применении к нашему последнему примеру (при $Q = 2,62$ и $Med = 11,57$) окажется равным:

$$Q_0 = \frac{100 Q}{Med} = \frac{100 \cdot 2,62}{11,57} = 22,64\%$$

аналогичный же ему вариационный коэффициент v в данном случае (при $\sigma = 4$ и $M = 12$):

$$v = \frac{100 \sigma}{M} = \frac{100 \cdot 4}{12} = 33,33\%$$

ХII. Оценочные зоны.

В некоторых отраслях приложения статистики на ряду с применением ее методов к изучению массовых явлений и фактов возникает надобность также и в объективной оценке единичных вариантов. Эта потребность в применении общих количественных критериев для интерпретации свойств, присущих индивидууму, особенно рельефно выступает при обслуживании статистическим методом тех дисциплин, в которых доминирующий интерес исследователя в конечном итоге направляется главным образом на изучение свойств и особенностей отдельной человеческой личности и в которых изучение окружающей социальной среды и биологических моментов, свойственных остальной массе этих объектов исследования в целом, служит лишь материалом для выяснения индивидуальных особенностей отдельного человека. Подобную задачу мы имеем, например, в педологии и в психотехнике. Здесь статистическое изучение человеческой массы главным образом и имеет своей конечной целью установку объективных критериев, с помощью которых представляется возможность оценить индивидуальные свойства и особенности отдельного человека или, в частности, отдельного ребенка.

Принцип и техника такой оценки индивидуальных вариантов по стандартным данным, выведенным на массовом материале, в общих чертах сводится к следующему.

Пусть, например, в результате безукоризненно точного измерения мы определили полную длину тела (рост) отдель-

ного ребенка, скажем мальчика в возрасте $11\frac{1}{2}$ лет, и нашли что рост его оказался равным 132 см . Спрашивается, должны ли мы на основании этих данных считать измеренного ребенка обладающим „средним“ ростом или же его следует отнести к группе „высоких“ или „низких“ детей? Не имея соответствующих сведений о росте других детей того же самого пола и возраста, мы, очевидно, сделать такую оценку единичного измерения вовсе не сможем. Для точной и объективной квалификации роста отдельного ребенка необходимо во-первых знать величину типичного роста соответствующей возрастно-половой группы детей, а во-вторых — учесть при этом также и среднюю величину варьирования данного количественного признака у всей совокупности остальных объектов исследования в целом. Сопоставив индивидуальный рост этого ребенка (132 см) с типичным ростом его сверстников, мы найдем абсолютную величину имеющегося здесь отклонения, а сравнив это отклонение со средней величиной отклонений роста других детей — определим относительную (оценочную) величину данного индивидуального различия.

Пусть типичный рост, свойственный мальчикам от 11 до 12 лет, составляет 134 см , а средний размах колебаний отдельных измерений роста у этих детей (в обе стороны от типичного роста 134 см) равен $\pm 5\text{ см}$ *).

Сравнивая рост данного ребенка (132 см) с этим типичным ростом (134 см), мы замечаем, что здесь обнаруживается некоторая отсталость в росте, количественно определяемая разностью:

$$132 - 134 = -2\text{ см}.$$

Естественно возникает вопрос: велика ли эта отсталость (на 2 см)? Так как рост других детей той же самой возрастно-половой группы варьировал в средних пределах $\pm 5\text{ см}$, то такая разница (только в 2 см), очевидно, не представляется еще существенной, и мы смело можем считать данного ребенка обладающим все же „средним“ (а не „низким“) ростом.

*) Приблизительно.

Аналогичные рассуждения могут иметь место и при оценке результатов какого-либо психотехнического испытания...

Пусть, например, группе испытуемых было предложено 10 задач одинаковой трудности, из которых один испытуемый решил, скажем, 7 таких задач. Как следует оценить индивидуальный результат этого теста?

В такой постановке этот вопрос является неопределенным. Нам необходимо знать среднее количество решений у всей массы испытуемых и средние пределы колебаний этих индивидуальных результатов испытания. Пусть типичное (среднее) число решений = 4 задачам с колебанием в обе стороны от него в среднем на величину $\pm 1,5$ задачи.

При этих условиях указанный выше единичный результат испытания (7 задач) получает совершенно точную и объективную оценку. В данном случае имеется проявленный успех, количественно определяемый разностью:

$$7 - 4 = +3 \text{ задачи,}$$

а так как средние пределы колебаний числа решений у других испытуемых $= \pm 1,5$ задачи, то это абсолютное превышение среднего результата на 3 задачи оказывается здесь вдвое больше среднего превышения (1,5), т. е. во всяком случае довольно значительным.

Таким образом для производства объективной количественной оценки какой-либо индивидуальной варианты (роста, веса и т. д.) нам необходимо предварительно определить типичную величину данного признака и средний размах его колебаний на достаточно большом и однородном материале, состоящем из аналогичных вариантов. При этом за типичную величину данного признака мы можем принять или среднее арифметическое M , или медиану Med , или, наконец, моду Mo , а средний размах колебаний этого признака численно охарактеризовать или средним квадратическим отклонением σ или квартилем Q . В качестве стандартных величин с этой целью чаще всего применяются M и σ , часто Med и Q (или другая сводная характеристика, аналогичная квартилю Q), а иногда также и константы Mo и σ .

При оценке индивидуальных вариантов с помощью стандартных M и σ за „средние“ значения данного количественного признака (за средний рост, средний вес, средний результат какого-нибудь психотехнического испытания и т. д.) чаще всего принимается интервал его значений, отличающихся от M (в обе стороны) не более, чем на одну σ . Так, если в каком-нибудь частном случае:

$$M = 10, \quad \text{а} \quad \sigma = \pm 3,$$

то „средними“ значениями варьирующего признака можно признать все колебания его в пределах:

$$\begin{aligned} \text{от } M - \sigma &= 10 - 3 = 7 \\ \text{до } M + \sigma &= 10 + 3 = 13. \end{aligned}$$

В этом интервале (от 7 до 13) числовые значения данного признака могут очевидно отличаться от среднего M не более, чем на одну σ , т. е. будут иметь исключительно малые по своей абсолютной величине отклонения. Это понятие о больших и малых отклонениях устанавливается здесь по отношению к сигме, характеризующей некоторую среднюю величину такого отклонения. Отклонения, меньшие сигмы, естественно поэтому принять за малые, а превышающие сигму — за большие.

Установленная этим путем область „средних“ значений варьирующего признака и принимается условно за „норму“.

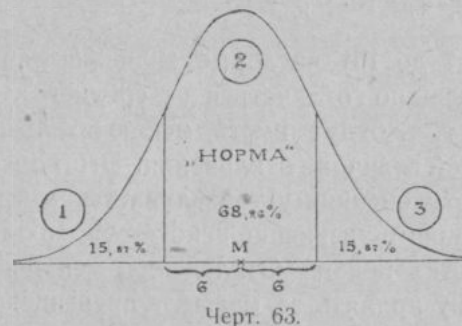
Такое вполне произвольное, но весьма удобное и распространенное понятие о „норме“, может быть сформулировано и иначе.

Вспомним, что при нормальном распределении *) в центральном интервале, границы которого симметрично удалены от M на расстояние = одной σ , заключается приблизительно 68,26% всех вариантов, а за пределами этого интервала, стало быть, 31,74% (по 15,87% с каждой стороны). Поэтому за „норму“ в указанном смысле слова принимаются здесь те условия, которым при нормальном распределении удовлетворяют приблизительно 68% всех объектов исследования (всех вариантов).

*) См. гл. IV, стр. 43, черт. 33.

Словесные наименования области „нормы“ и примыкающих к ней смежных участков меньших и больших значений данного признака („низкий“, „средний“ и „высокий“ рост; „плохие“, „средние“ и „хорошие“ результаты испытания и т. д.) для краткости можно иногда заменить какими-либо условными цифровыми или буквенными обозначениями, например тремя баллами: 1, 2 и 3 (1 — низкий, 2 — средний, 3 — высокий рост или, по желанию, наоборот).

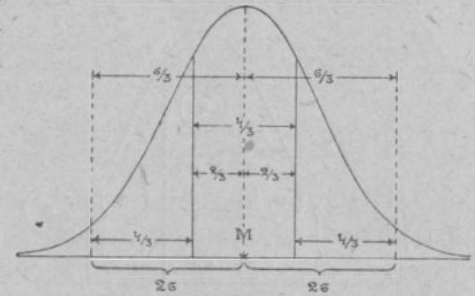
На черт. 63 представлена кривая распределения с выделенной в ее центре областью „нормы“. Цифрами в кружках здесь помечены условные баллы, а внутри этой кривой указано процентное распределение вариантов по трем ее участкам. Эти участки в дальнейшем мы будем называть „оценочными зонами“.



Способ подразделения всех возможных значений варьирующего признака всего лишь на 3 оценочные зоны конечно не является единственно возможным. Если по условиям вопроса столь широкое понимание „нормы“ (включающее в себе 68% всех вариантов) нас не удовлетворяет, мы можем как угодно сгустить обе ее границы, приняв, например, за „норму“ те условия, которым при нормальном распределении удовлетворяют всего только 50% всех объектов исследования. В этом случае границы центральной оценочной зоны (т. е. границы „нормы“) по закону нормального распределения должны быть отставлены в обе стороны от M на $\frac{2}{3}$ сигмы (или точнее на 0,6745 сигмы *). При этом мы можем разбить весь диапазон числовых значений варьирующего признака не только грубо на 3 участка, но ввести сюда и более мелкие подразделения, установив, например, для оценки роста пять последовательных и равномерных градаций: „очень низкий“, „низкий“, „средний“, „высокий“

*) См. таблицу процентного распределения вариант в конце книги на стр. 421.

и „очень высокий“ рост или в соответствующих условных баллах: 1, 2, 3, 4 и 5 (что удобно например при оценке результатов психо-технических испытаний). Здесь возникает лишь вопрос о том, где следует установить границы, отделяющие просто малые (или большие) значения варьирующего признака от его значений очень малых (или очень больших). Из практических соображений (по аналогии с классовым промежутком) удобно выбирать эти границы с тем расчетом, чтобы все зоны, кроме крайних, имели одинаковую ширину. Крайние же зоны (1 и 5) очевидно не должны быть ограничены, так как оцениваемый признак, вообще говоря, может изменяться в обе стороны от M беспредельно.

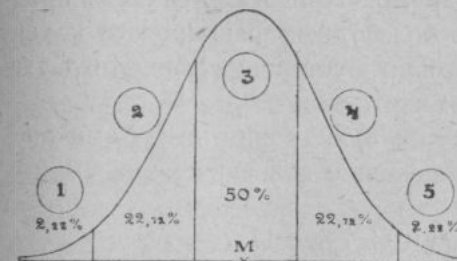


В нашем случае (при 5-ти оценочных зонах) границы „нормы“ отстоят от M на $\frac{2}{3}\sigma$ (см. черт. 64), а полная ширина центральной оценочной зоны равна следовательно:

$$\frac{2}{3} + \frac{2}{3} = \frac{4}{3} \text{ сигмы.}$$

Желая сохранить ту же самую ширину (т. е. $\frac{4}{3}\sigma$) и для соседних зон, мы должны отодвинуть их границы на расстояние $\frac{4}{3}\sigma$ от предыдущих, в свою очередь удаленных от M на $\frac{2}{3}\sigma$, т. е. в общем на расстояние от M , равное:

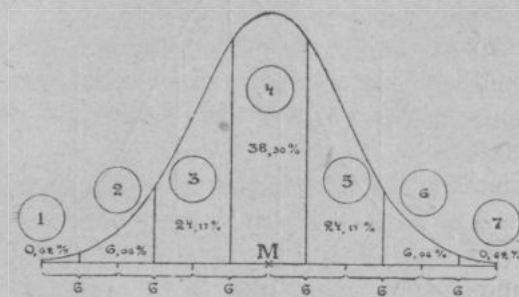
$$\frac{4}{3} + \frac{2}{3} = \frac{6}{3} = 2 \text{ сигмы.}$$



Этот способ разбивки всего диапазона числовых значений варьирующего признака на 5 оценочных зон может быть иллюстрирован черт. 65, на котором по прежнему цифрами в кружках

обозначены баллы, а в соответствующих участках кривой выписан процент вариантов, попадающих в эти участки при нормальном распределении.

Если в частности $M = 10$, а $\sigma = \pm 3$, то 4 границы пяти



Черт. 66.

таких оценочных зон могут быть вычислены по формулам:

$$M + 2\sigma = 10 + 6 = 16$$

$$M + \frac{2}{3}\sigma = 10 + 2 = 12$$

$$M - \frac{2}{3}\sigma = 10 - 2 = 8$$

$$M - 2\sigma = 10 - 6 = 4$$

В данном случае, следовательно, оценку баллом:

5	получат варианты, большие 16
4	" в пределах от 16 до 12
3	" " " " 12 " 8
2	" " " " 8 " 4
1	" " " " меньшие 4.

Если принять за „норму“ те условия, которым удовлетворяет всего лишь 38,30% общего количества всех объектов исследования, то, в предположении нормального распределения этих вариантов, границы центральной оценочной зоны окажутся на расстоянии $\frac{1}{2}$ сигмы от M (или на расстоянии одной σ друг от друга). При этих условиях представляется возможным еще более дифференцировать конечную оценку отдельных вариантов, установив для этого уже 7 соответствующих зон (см. черт. 66).

Здесь (при $M = 10$ и $\sigma = \pm 3$) 6 границ этих семи оценочных зон будут иметь следующие значения:

$$M + 2,5\sigma = 10 + 7,5 = 17,5$$

$$M + 1,5\sigma = 10 + 4,5 = 14,5$$

$$M + 0,5\sigma = 10 + 1,5 = 11,5$$

$$M - 0,5\sigma = 10 - 1,5 = 8,5$$

$$M - 1,5\sigma = 10 - 4,5 = 5,5$$

$$M - 2,5\sigma = 10 - 7,5 = 2,5$$

Поэтому оценку баллом:

7	получают варианты, большие 17,5
6	" " от 17,5 до 14,5
5	" " " 14,5 " 11,5
4	" " " 11,5 " 8,5
3	" " " 8,5 " 5,5
2	" " " 5,5 " 2,5
1	" " меньшие 2,5.

Сужение области, условно принимаемой за „норму“, и увеличение, в связи с этим, общего количества оценочных зон можно было бы, вообще говоря, продолжать и далее, но в этом чаще всего практически не встречается уже надобности.

Основной особенностью описанного здесь способа разбивки варьирующего признака V на оценочные зоны является нечетное количество (3, 5, 7...) таких зон (что вызывается необходимостью иметь центрально расположенную область „нормы“) и одинаковая ширина всех зон, кроме двух крайних, вовсе не ограниченных с одного конца и распространяющихся в соответствующую сторону бесконечно.

Выбор общего количества оценочных зон и установление той или иной их ширины, вообще говоря, совершенно произволен. Здесь все зависит от того, что условно принимается за „норму“, и насколько подробно желательно получать конечную оценку индивидуальных вариантов.

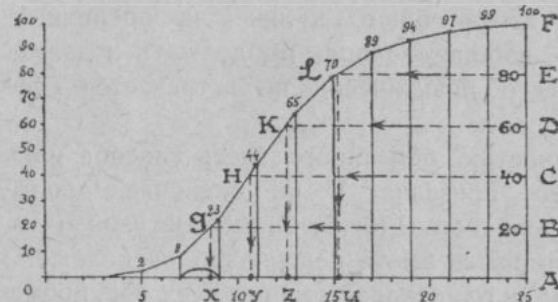
Наряду с описанным выше способом разбивки материала на равномерные по своей ширине участки иногда с этой же целью применяются и зоны неравномерные, но заключающие внутри своих границ одинаковое число вариантов (назовем их „равночисленными“). При определении границ таких равночисленных оценочных зон отсчет ведется уже не от среднего арифметического M , а от медианы (Med *), которая попадает здесь в центральную зону (и является ее „срединным“ значением), подобно тому как при разбивке материала на равномерные зоны в центр этой разбивки попадало среднее арифметическое M . Самая же

*) См. гл. X, стр. 101.

разбивка на равночисленные зоны выполняется по принципу, аналогичному разбивке ряда на 4 части при определении квартиля Q .

Пусть, например, требуется разбить на 5 равночисленных оценочных зон следующий вариационный ряд, интегральная кривая которого представлена на черт. 67. *)

Разделив наибольшую высоту (AF) интегральной кривой на 5 равных частей (AB , BC , CD , DE и EF), проведем через точки деления (B , C , D и E) влево ряд прямых, парал-



Черт. 67.

лельных горизонтальной оси, до пересечения с интегральной кривой в точках H , K и L . Опустив из этих точек перпендикуляры на горизонтальную ось, мы и получим при основании этих перпендикуляров (в точках x , y , z и u) четыре границы, разделяющие пять „равночисленных“ оценочных зон. В нашем примере искомые границы (помеченные на черт. 67 жирными точками) пришлись между 8 и 9 делением горизонтального масштаба, между 10 и 11 делением, между 12 и 13 и между 15 и 16. Если данное построение выполнить на листе миллиметровой бумаги в достаточно крупном масштабе (что всегда и рекомендуется делать), то можно определить и более точное положение всех этих границ. Так, нетрудно выяснить, что крайняя слева граница (в точке x) по горизонтальному масштабу соответствует числу 8,6, следующая граница — числу 10,6, соседняя с ней — числу 12,6 и, наконец, последняя — числу 15,2.

Таким образом при данном способе разбивки всего диапазона варьирования (от 5 до 25) на 5 неравномерных по

*) Гистограмма этого ряда уже неоднократно приводилась выше (см., напр., черт. 2, 13, 57 и др.). Интегральная кривая, аналогичная представленной здесь (на черт. 67), была ранее приведена на черт. 55, 56 и 61, а расчет ее ординат дан на стр. 107 (в отдельной таблице).

величине, но одинаковых по численности зон оценку высшим баллом:

5	получат все варианты, большие	15,2
4	„ „ „ „	в пределах от 15,2 до 12,5
3	„ „ „ „	12,5 „ 10,6
2	„ „ „ „	10,6 „ 8,6
1	„ „ „ „	меньшие 8,6.

При этом в каждую такую зону попадет при разбивке по 20% всех вариантов (поровну).

Для определения еще более точного положения этих равночисленных оценочных зон можно выполнить соответствующие вычисления по обычным формулам, указанным для определения квартиля Q *), помня, что отсекаемая от интегральной кривой высота h в отдельных случаях равна здесь 20, 40, 60 и 80 делениям вертикального масштаба а классовый промежуток k во всех случаях = 2.

Для крайней слева границы, разделяющей зону оценки баллом 1 от смежной с нею зоны балла 2 (обозначим ее символом g_{12}), левая (меньшая) граница того класса, в котором заключается искомая граница зоны $V_1 = 7$, правая (большая) граница этого же класса $V_2 = 9$, высота интегральной кривой в точке V_1 , т. е. $h_1 = 8$, высота ее в точке V_2 , т. е. $h_2 = 23$, отсекаемая высота $h = 20$, а классовый промежуток $k = 2$. Следовательно:

$$g_{12} = V_1 + k \left(\frac{h - h_1}{h_2 - h_1} \right) = 7 + 2 \left(\frac{20 - 8}{23 - 8} \right) = 8,60.$$

Для следующей границы, разделяющей зоны баллов 2 и 3, получим $V_1 = 9$, $h_1 = 23$, $h_2 = 44$, $h = 40$ и попрежнему $k = 2$. Следовательно:

$$g_{23} = 9 + 2 \left(\frac{40 - 23}{44 - 23} \right) = 10,62.$$

Аналогично для границы зон балла 3 и 4:

$$g_{34} = 11 + 2 \left(\frac{60 - 44}{65 - 44} \right) = 12,52.$$

Наконец, для крайней справа границы:

*) См. стр. 116.

$$g_{45} = 15 + 2 \left(\frac{80 - 79}{89 - 79} \right) = 15,20.$$

Аналогично можно выполнить разбивку этого же ряда и на другое нечетное число неравномерных оценочных зон, напр. на 3 или на 7.

Наряду с описанными выше двумя основными способами разбивки материала на равномерные и неравномерные по своей ширине оценочные зоны иногда применяется также особый вариант первого из этих способов, при котором отсчет границ оценочных зон ведется не от среднего арифметического M , а от моды Mo , но ширина всех зон по-прежнему делается одинаковой и находится в определенной зависимости от среднего квадратического отклонения σ .

В нашем последнем примере, на котором была произведена разбивка материала на 5 равномерных оценочных зон (см. стр. 111 и 62):

$$Mo = 10,62 \text{ и } \sigma = \pm 4.$$

При разбивке этого же материала на 5 равномерных зон с отсчетом границ от моды мы должны, следовательно, вычислить эти границы по формулам:

$$Mo + 2\sigma = 10,62 + 8 = 18,62$$

$$Mo + \frac{2}{3}\sigma = 10,62 + 2,67 = 13,29$$

$$Mo - \frac{2}{3}\sigma = 10,62 - 2,67 = 7,95$$

$$Mo - 2\sigma = 10,62 - 8 = 2,62.$$

При этом способе разбивки оценку баллом:

5	получат варианты, большие	18,62
4	" "	от 18,62 до 13,29
3	" "	" 13,29 " 7,95
2	" "	" 7,95 " 2,62
1	" "	меньшие 2,62.

При производстве оценки большого количества отдельных вариантов все эти критерии разбивки на зоны полезно бывает облечь в форму особой „оценочной таблицы“. В применении, например, к последнему способу разбивки на 5 зон (с отсче-

том границ от моды Mo) эта таблица будет иметь следующий вид:

1	2	3	4	5
$Mo - 2\sigma$	$Mo - \frac{2}{3}\sigma$	$Mo + \frac{2}{3}\sigma$	$Mo + 2\sigma$	
2,62	7,95	13,29	18,62	

Здесь в верхней строчке даны в виде заголовков формулы для вычисления соответствующих границ, под ними выписаны числовые значения этих границ, а сверху над линиями вертикального графления указаны условные (балловые) обозначения всех пяти оценочных зон.

Способ пользования такой таблицей крайне прост. Пусть, например, требуется оценить вариант $V=6$. Среди четырех выписанных здесь пограничных значений этого признака находим два числа, ближайšie к 6. Одно из них (слева) будет несколько меньше 6 (2,62), другое (справа), наоборот, больше 6 (7,95). Как только такая пара оценочных чисел найдена, следует заметить разделяющую их в таблице вертикальную линию и посмотреть на цифровой заголовок, стоящий над этой линией вверху таблицы. Эта цифра (в данном случае балл 2) и определит искомую оценку нашей варианты $V=6$.

Если бы оказалось, что оцениваемая варианта будет меньше самого малого из пограничных чисел (2,62) или, наоборот, больше самого большого из них (18,62), то оценочный балл следует считать у соответствующего края этой таблицы (в первом случае 1, во втором 5).

Для случая разбивки этого же материала на 5 равномерных оценочных зон с отсчетом их границ от среднего арифметического $M=12$ при той же величине $\sigma=\pm 4$ получим следующую оценочную таблицу:

1	2	3	4	5
$M - 2\sigma$	$M - \frac{2}{3}\sigma$	$M + \frac{2}{3}\sigma$	$M + 2\sigma$	
4,00	9,33	14,67	20,00	

Способ пользования этой таблицей остается прежним. При оценке, например, варианты $V=15$ мы находим два ближайших к ней числа 14,67 и 20 и над линией, разделяющей в таблице эти числа, считываем конечную балловую оценку, в данном случае 4.

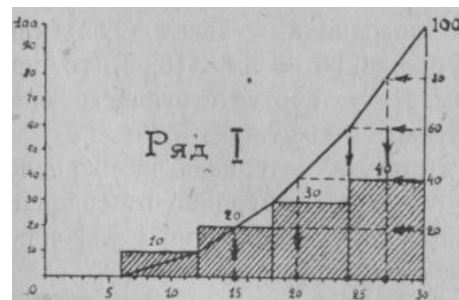
При разбивке этого материала на 5 неравномерных оценочных зон (произведенной с помощью интегральной кривой) получим следующую оценочную таблицу:

1	2	3	4	5
8,60	20,62	12,52	15,20	

В данном случае в заголовках не указаны расчетные формулы, так как определение этих границ было произведено здесь графическим путем (окончательные же формулы, уточняющие этот расчет, не обладают необходимой для заголовков наглядностью).

ХIII. Сравнительный обзор основных способов зональной разбивки.

Поставим себе целью выяснить некоторые особенности трех указанных в прошлой главе основных способов разбивки варьирующего материала на оценочные зоны, определив попутно также и условия применимости этих способов в некоторых случаях, могущих встретиться на практике.

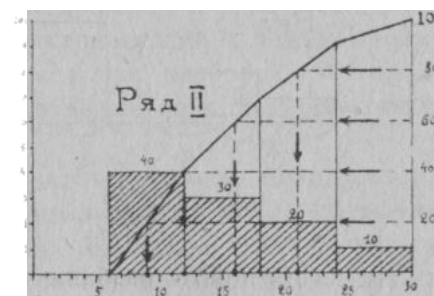


Черт. 68.

Пусть, например, имеем два вариационные ряда, представляющие собой результаты какого-либо психотехнического испытания. Гистограммы этих рядов вместе с соответствующими интегральными кривыми представлены на черт. 68 и 69.

Здесь в обоих случаях классовый промежуток $k=6$, общее количество измерений $n=100$, числовые значения варьирующего признака колеблются в максимальных пределах от 6 до 30, а весь эмпирический интервал варьирования разбит на 4 класса.

Очевидно I-ый из этих рядов (черт. 68) представляет собой результаты легкого испытания (так как большинство испытуемых проявило в нем хорошие успехи), а II-ой ряд (черт. 69), наоборот, свидетельствует о трудности данного испытания



Черт. 69.

(большинство испытуемых обнаружило плохие результаты).

Здесь для I-го ряда (легкий тест):

$$M = 21 \quad \sigma = \pm 6 \quad Med = 22,$$

а для второго ряда (трудный тест):

$$M = 15 \quad \sigma = \pm 6 \quad Med = 14,$$

в чем нетрудно убедиться путем обычных вычислений.

Разобьем каждый из этих рядов на 5 оценочных зон двумя способами: на равномерные и неравномерные (равночисленные) зоны.

После определения числовых значений всех границ этих зон получим следующие две оценочные таблицы:

	1	2	3	4	5		1	2	3	4	5
Для I ряда:	9	17	25	33		Для II ряда:	3	11	19	27	
	15	20	24	27			9	12	16	21	

В первой строчке этих таблиц (сверху) выписаны числовые значения четырех границ равномерных оценочных зон, а во второй строчке (под ними) — границы зон равночисленных.

Представим эти конечные результаты разбивки в графической форме, пометив оценочные зоны дугами у оси абсцисс соответствующих гистограмм и обозначив условные баллы

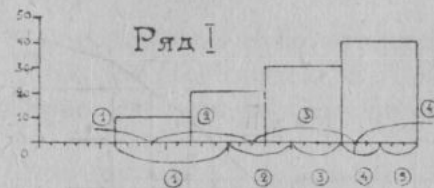
цифрами в кружках: сверху для равномерной, а снизу для равночисленной разбивки (см. черт. 70 и 71).

Отсюда видно, что при разбивке на равномерные зоны в I ряду ни одна варианта (в пределах от 6 до 30) не получает оценки баллом 5, а во II ряду — баллом 1, в то время как при разбивке на равночисленные зоны обе эти крайние оценки (1 и 5) встречаются как в I, так и во II ряду.

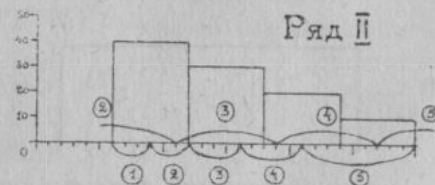
Эта особенность разбивки на равномерные зоны легко объясняется здесь наличием довольно значительной асимметрии заданных рядов (для I-го ряда показатель асимметрии $A_1 = -0,6$, а для II-го ряда $A_2 = +0,6$). В результате сдвига вершины гистограммы I ряда вправо, в том же направлении естественно переместились и все 4 границы оценочных зон, причем крайняя из этих границ (между зонами баллов 4 и 5) при таком перемещении вышла даже за пределы фактического варьирования данного признака (эта граница оказалась $= 33$, в то время как значения отдельных вариантов колеблются в этом ряду лишь в пределах от 6 до 30).

Во II ряду картина получилась обратная: среднее арифметическое M , а с ним и все 4 границы оценочных зон передвинулись влево, в результате чего граница, разделяющая зоны баллов 1 и 2 (в данном случае $= 3$), оказалась меньше самого малого предела фактического варьирования заданного признака (т. е. меньше 6).

Итак, при разбивке материала на 5 равночисленных оценочных зон (с помощью интегральной кривой) отдельные варианты непременно получают оценку всеми 5 баллами, и притом оценки эти встретятся в заданном материале приблизительно в одинаковом количестве случаев (теоретически одинаково часто), в то время как при разбивке этого же материала на зоны равномерные (в долях σ с отсчетом



Черт. 70.



Черт. 71.

от M) крайние оценки (1 или 5) могут и вовсе не встретиться. То же самое, очевидно, может случиться и при ином количестве оценочных баллов (напр. при 3-х, 7-ми и т. д.).

Учитывая это существенное различие в способах разбивки материала на зоны, мы должны прежде всего выяснить, является ли такое отсутствие крайних оценок преимуществом равномерных зон или же, наоборот, в этом их свойстве следует усматривать недостаток данного способа зональной разбивки.

Тут все зависит, очевидно, от конечной цели, которая преследуется разбивкой данного материала на оценочные зоны. Если при этой оценке имелось в виду выявить на ряду со „средним“ результатом какого-либо измерения или испытания также и резкие отклонения от этой „нормы“, то разбивка на равночисленные зоны окажется нецелесообразной. В самом деле, при таком способе оценки в обе крайние (и во все средние) зоны всегда войдет по одинаковому количеству вариантов, так что на этом основании мы должны будем в любом обследованном коллективе констатировать наличие некоторого постоянного процента индивидуумов исключительно выдающихся и исключительно же отсталых, что, конечно, противоречит здравому смыслу. Нельзя же в самом деле утверждать, что из 100 испытуемых при разбивке их на 5 оценочных зон 20 человек непременно окажутся талантливыми, а 20 — бездарными, если под этими крайними градациями оценки мы будем понимать исключительно резкие отклонения от „нормы“, а не просто категорию сравнительно лучших и сравнительно худших. Вполне понятно, что талантливых и бездарных в указанном смысле слова может и вовсе не оказаться среди данной группы объектов исследования, в то время как некоторое произвольно установленное количество просто „самых хороших“, „самых плохих“ результатов испытания всегда может быть выделено из общей их массы.

В одном из двух рассмотренных выше рядов мы имели пример легкого теста, так как большинство испытуемых обнаружило там сравнительно хорошие результаты. Поэтому здесь будет вполне уместно и справедливо квалифицировать отдельные плохие результаты этого испытания самым низким баллом 1 и вовсе воздержаться от чрезмерно высо-

кой оценки какого-либо из испытуемых самым высоким баллом 5. В самом деле, ведь хорошие результаты были проявлены большинством испытуемых, и в них нельзя еще усматривать выявление исключительных способностей (балл 5). Но если при таком легком задании отдельный испытуемый все же справился с ним неудовлетворительно, то такой столь необычный для остальной их массы результат несомненно должен уже считаться исключительно плохим (балл 1).

В применении ко II ряду наши рассуждения будут носить обратный характер. Мы имеем здесь тест трудный, с которым большинство испытуемых справилось неудовлетворительно. Поэтому отдельные хорошие результаты этого испытания могут заслуженно получить здесь высшую оценку баллом 5, в то время как типичное для остальных испытуемых неудовлетворительное решение данной задачи не может еще считаться исключительно плохим, и никто из них низшего балла 1 здесь не получит.

Если бы целью этой разбивки являлось выделение из общей массы объектов исследования некоторого заранее установленного процента сравнительно лучших или сравнительно плохих испытуемых (например для приема на какую-либо работу или, наоборот, для увольнения), то разбивка на неравномерные оценочные зоны оказалась бы здесь очевидно вполне целесообразной.

В тех случаях, когда границы зон устанавливаются на обширном материале, а самая оценка производится по отношению лишь к тем или иным мелким его частям, оба способа зональной разбивки одинаково могут выявить крайние (исключительные) отклонения отдельных вариантов либо вовсе не обнаружить этих крайних оценок. В самом деле, пусть на материале, скажем, в 10.000 человек мы установили границы 5-ти равночисленных оценочных зон. В каждую из них войдет, очевидно, по 2.000 индивидуумов из общего их числа 10.000. Но, подвергая такой оценке лишь малую часть этого коллектива, например только 100 человек, мы можем встретиться с любым характером распределения членов этого коллектива по отдельным зонам и в частности натолкнуться, например, даже на такой крайний случай, когда все 100 человек попадут только в одну зону

оценки баллом 5 или только в противоположную зону балла 1.

Итак, для оценки результатов отдельных измерений или испытаний могут с успехом служить как равномерные так и неравномерные оценочные зоны, причем применение тех или других в каждом конкретном случае зависит от конечной цели производимой оценки и от того значения, которое мы придаем крайним баллам.

Посмотрим теперь, что может нам дать разбивка материала на равномерные зоны, границы которых отсчитываются не от среднего арифметического M , а от моды Mo . Пользуясь приближенной формулой для вычисления моды:

$$Mo = 3 Med - 2 M$$

и помня, что для I-го из двух разобранных выше рядов:

$$M = 21 \quad Med = 22,$$

а для II-го ряда:

$$M = 15 \quad Med = 14,$$

получим для I-го ряда:

$$Mo = 3 \cdot 22 - 2 \cdot 21 = 66 - 42 = 24,$$

а для II-го ряда:

$$Mo = 3 \cdot 14 - 2 \cdot 15 = 42 - 30 = 12.$$

Определив по этим данным числовые значения четырех границ пяти наших зон (при $\sigma = \pm 6$), получим следующие 2 оценочные таблицы:

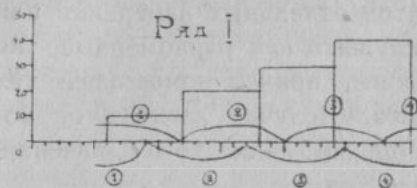
Для I-го ряда.

$Mo - 2\sigma$	$Mo - \frac{2}{3}\sigma$	$Mo + \frac{2}{3}\sigma$	$Mo + 2\sigma$
12	20	28	36

Для II-го ряда.

$Mo - 2\sigma$	$Mo - \frac{2}{3}\sigma$	$Mo + \frac{2}{3}\sigma$	$Mo + 2\sigma$
0	8	16	24

Представим эти конечные результаты зональной разбивки в обычной графической форме (см. черт. 72 и 73).



Черт. 72.

Здесь выше оси абсцисс дугами обозначены только что установленные зоны, границы которых отсчитаны от моды M_0 , а внизу для сравнения — прежние зоны, зависящие от среднего арифметического M . Цифрами в кружках, как всегда, помечены соответствующие этим зонам условные баллы.

Из этого сопоставления видно, что при отсчете границ не от среднего арифметического M , а от моды M_0 наши оценочные зоны особенно сильно перемещаются в направлении сдвига вершины соответствующей гистограммы, и, стало быть, в данном случае мы также будем иметь отсутствие оценок крайними баллами 1 и 5, причем это свойство разбивки здесь выступает особенно резко.

В двух приведенных выше примерах отсутствие крайних оценок было вызвано асимметрией заданных рядов. Подобное же явление может иметь место также и при наличии в этих рядах обратного (отрицательного) эксцесса.

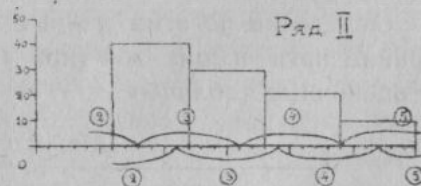
Пусть, например, имеем вариационный ряд, гистограмма которого представлена на черт. 74.

Для этого ряда:

$$M = 45 \quad \sigma = \pm 20.$$

Выполнив разбивку этих данных на 5 равномерных и равночисленных зон, получим следующую оценочную таблицу:

	1	2	3	4	5
Для равномерных зон	5,0	31,7	58,3	85,0	
Для равночисленных зон	24,2	35,6	54,4	65,8	



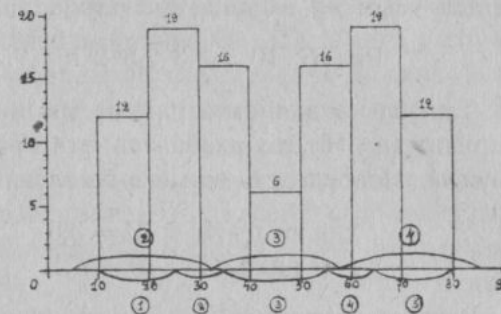
Черт. 73.

Эти зоны помечены на чертеже 74 дугами на оси абсцисс: сверху — для равномерной, а снизу — для равночисленной разбивки.

Здесь при пользовании равномерными зонами мы обнаруживаем отсутствие обеих крайних оценок (1 и 5), в то время как при разбивке материала на равночисленные зоны все пять оценок, как и раньше, встречаются одинаково часто.

Эта особенность разбивки на равномерные зоны станет понятной, если вспомнить, что основным признаком отрицательного эксцесса (а здесь показатель $E = -1,39$) является ограниченность пределов варьирования соответствующего материала.

Присущая этому ряду малая абсолютная величина отклонений и выразилась здесь в отсутствии крайних оценок 1 и 5, которые в нашем понимании предназначены как раз для характеристики чрезвычайно резких отличий отдельных вариантов от „нормы“.



Черт. 74.

При пользовании же равночисленными зонами эта специфическая особенность обратно-эксцессивных рядов вовсе не получает надлежащего отражения в балловой оценке крайних отклонений, которые при обратном эксцессе попадают в зоны баллов 1 и 5 совершенно так же, как попали бы они туда и при наличии в данном ряду прямого эксцесса (и связанной с ним сильной разбросанности небольшого числа отдельных вариантов). Поэтому равночисленные зоны в данном случае не выявляют особенностей распределения оцениваемых вариантов, в то время как оценка по зонам равномерным находится в определенной зависимости как от асимметрии ряда, так и от его эксцесса, т. е. оказывается в этом отношении более гибкой.

Однако разбивка материала на равночисленные зоны имеет все же одно крупное преимущество, заставляющее признать этот способ оценки наиболее целесообразным в том

случае, когда отдельные варианты измерены в сложных единицах, допускающих различные адекватные друг другу формы их количественного выражения. Пусть, например, мы имеем ряд *A*, состоящий из 6 вариантов, представляющих собой результат измерения скорости какого-либо процесса, причем скорость эта выражена здесь числом единичных операций, выполняемых испытуемым в течение десятиминутного периода времени.

Ряд *A*: 1, 2, 3, 4, 5 и 6.

Выразив эти же данные количеством минут, затраченных испытуемым на выполнение одной операции, получим:

Ряд *B*: 10 5 3,3 2,5 2 и 1,7.

Так как в данном примере мы имеем ничтожное число вариантов ($n = 6$), то разбивать этот ряд на классы не имеет смысла. Непосредственные вычисления здесь дают:

$$\text{для ряда } A: M = 3,5 \quad \sigma = \pm 1,8$$

$$\text{для ряда } B: M = 4,1 \quad \sigma = \pm 3,1.$$

Разобьем каждый из этих рядов для простоты всего лишь на 3 равномерные зоны, соответствующие условным баллам 1, 2 и 3, взяв за разделяющие их границы числовые значения: $M + \sigma$ и $M - \sigma$.

Для ряда *A* этими границами будут служить:

$$M + \sigma = 3,5 + 1,8 = 5,3$$

$$M - \sigma = 3,5 - 1,8 = 1,7,$$

а для ряда *B*:

$$M + \sigma = 4,1 + 3,1 = 7,2$$

$$M - \sigma = 4,1 - 3,1 = 1,0.$$

Выпишем теперь заново оба заданные ряда вариант друг под другом и каждый из них разобьем на 3 зоны в местах, соответствующих только что вычисленным границам:

	балл 1	балл 2				балл 3
Ряд А:	1	2	3	4	5	6
Ряд Б:	10	5	3,3	2,5	2	1,7
	балл 1	балл 2				

Из этого сопоставления видно, что оценочные зоны здесь не вполне совпадают друг с другом. Так, например, варианта 6 в ряду *A* получила оценку баллом 3, в то время как соответствующая ей (в сущности та же самая) варианта 1,7 в ряду *B* попала в зону балла 2 (зоны высшего балла 3 в ряду *B* не имеется вовсе).

Отсюда парадоксальный вывод: если скорость изучаемого процесса выражать числом операций в 10-ти минутный промежуток времени, то значение этой скорости, равное например 6 операциям в минуту, оказывается высоким (балл 3), но если эту же скорость перечислить в другие единицы, выразив ее количеством минут, потраченных на выполнение одной операции (получим 1,7 минут), то по сравнению с теми же данными эта варианта должна быть отнесена уже к категории средних значений скорости (балл 2). Таким образом в данном случае оценка скорости находится как бы в зависимости от выбора единиц ее измерения, что конечно и практически и принципиально является недопустимым.

Технически этот парадокс объясняется здесь чрезвычайно просто. Вспомним (см. стр. 96—99), что при подобном перечислении единиц измерения мы всегда нарушаем симметрию ряда, чем и вызывается сдвиг всей системы зональных границ в соответствующем направлении.

Ничего подобного не произойдет (и не может произойти) при разбивке этого материала по принципу равночисленных зон, которые в данном частном случае легко могут быть установлены просто путем деления каждого из двух заданных рядов на 3 части по 2 варианты в каждой.

Здесь каждая варианта сохраняет свою балловую оценку независимо от способа ее измерения: количеством операций в 10 минут или количеством минут, затраченных на выполнение одной операции, что, очевидно, в большей степени отвечает требованию здравого смысла *).

*) Заметим кстати, что и среднее арифметическое M при таком перечислении единиц измерения теряет свое значение в качестве показателя некоторой типичной для данного ряда неизменной величины. В самом деле, для ряда *A* среднее $M = 3,5$ (операций в 10 минут). Перечислив эту среднюю скорость в другие единицы, получим $M = 2,9$ (минут на одну операцию), в то время как для ряда *B* мы получили $M = 4,1$. Обратно: при перечислении $M = 4,1$ в другие единицы скорости получим $M = 2,44$, в то

Итак, применение того или иного способа зональной разбивки всецело диктуется соображениями практического удобства, теми целями, которые преследуются данной оценкой, и нашей трактовкой значения крайних зон.

В задачи настоящего руководства не входит рассмотрение тех или иных специфических особенностей метода зональной разбивки в различных его конкретных приложениях. Достаточно будет лишь указать на то, что приемы разбивки оцениваемого материала по принципу „равномерных“ зон находят широкое применение в антропометрии, в то время как „равночисленные“ зоны (содержащие в своих границах заранее фиксированный % вариант) часто используются при оценке результатов психотехнических испытаний.

Последнее станет понятным, если принять во внимание различную значимость тех чисел, которыми характеризуется количественная сторона варьирующих признаков в этих дисциплинах.

XIV. Некоторые особые случаи зональной разбивки.

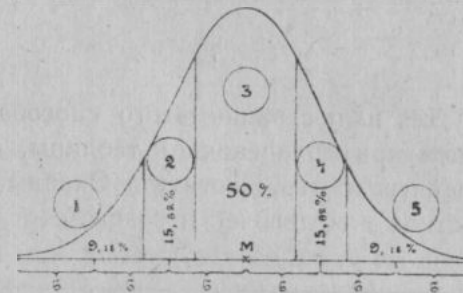
В двух предыдущих главах под именем „равномерных“ и „неравномерных“ („равночисленных“) оценочных зон мы понимали в первом случае зоны одинаковой ширины, (по оси абсцисс), а во втором — зоны одинаковой площади. Эти условные термины были введены там лишь для удобства и краткости наименования соответствующих способов зональной разбивки, но сам по себе признак равномерности или неравномерности расстояний между соседними зональными границами не является еще, очевидно, главной характерной особенностью этих способов разбивки.

Очень часто, например, оценочные зоны, отсчитанные в тех или иных долях σ от среднего арифметического M (т. е. по прежней нашей терминологии „равномерные“ зоны), фактически делаются и неодинаковой ширины.

время как для ряда A мы имели $M = 3,5$. Медиана же (Med) в обоих случаях сохраняет здесь свою неизменную величину, разделяя ряд на две равные по количеству вариант половины.

На черт. 75 представлена такая разбивка ряда на 5 зон с выделением в центре распределения области „нормы“, границы которой удалены в обе стороны от M на расстояние $= \frac{2}{3}\sigma$. Эта средняя зона, заключающая в себе при нормальном распределении 50% всех вариант заданного ряда, имеет полную ширину в $\frac{4}{3}\sigma$. Две соседние зоны, соответствующие баллам 2 и 4, сделаны здесь вдвое уже (шириною только в $\frac{2}{3}\sigma$) и заключают в своих границах, следовательно, по 15,82% вариант *). Крайние зоны (1 и 5), как всегда, не ограничены. На каждую из них приходится по 9,18% всех вариант.

Такая разбивка имеет в виду наиболее точно дифференцировать сравнительно небольшие отклонения от нормы, соответствующие баллам 2 и 4, за счет менее четкой квалификации крайних значений признака, попадающих в несколько расширенную здесь область оценки баллами 1 и 5, что во многих случаях практически представляется весьма целесообразным.



Черт. 75.

Пример аналогичного сужения двух ближайших к „норме“ участков по сравнению со всеми остальными „равномерными“ зонами мы имеем в часто применяемой разбивке антропометрических признаков на 7 оценочных зон по способу Мартина.

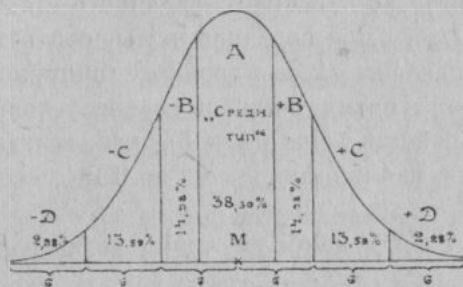
Способ этот может быть иллюстрирован черт. 76, из которого видно, что все зоны имеют здесь одинаковую ширину в одну целую σ кроме двух ближайших к центру участков, сделанных шириною всего лишь в $\frac{1}{2}\sigma$.

Центральная зона по Мартину определяет так называемый „средний тип“, условно обозначенный на черт. 76 буквою A . **)

*) См. таблицу процентного распределения вариант в гл. IV.

**) Мартин обозначает „средний тип“ символом T , указанные же здесь обозначения взяты несколько иными, так, как ими пользуются многие антропометрические кабинеты у нас в СССР.

Слева и справа проставлены буквы A , B и C со знаками $+$ и $-$. Символом $+B$ и $-B$ определяются здесь сравнительно незначительные отклонения от среднего типа A , буквами $+C$ и $-C$ отклонения уже существенные, но еще



Черт. 76.

не очень большие и, наконец, $+D$ и $-D$ — явно большие отклонения. Все буквы со знаком $+$ указывают на превышение „среднего типа“, а со знаком $-$, наоборот, на отсталость данного индивидуума по соответствующему признаку.

Для иллюстрации этого способа разбивки на зоны приведем пример оценочной таблицы, составленной по данным Ленинградского Института Охраны Здоровья Детей и Подростков для девочек в возрасте от 13 до 14 лет *).

	$-D$	$-C$	$-B$	A	$+B$	$+C$	$+D$
Наименование признака:	$M-2\sigma$	$M-\sigma$	$M-1/2\sigma$	$M+1/2\sigma$	$M+\sigma$	$M+2\sigma$	
Рост стоя . . .	131,52	138,47	141,95	148,90	152,37	159,32	
Рост сидя . . .	70,16	73,69	75,46	78,99	80,75	84,28	
Длина ног . . .	59,54	63,87	66,04	70,37	72,53	76,86	
Окружн. груди	58,53	62,35	64,26	68,08	69,99	73,81	
Окружн. головы	49,57	50,99	51,70	53,12	53,83	55,25	

Способ пользования этой таблицей остается прежним. Пусть, например, в каком либо частном случае мы получили

*) См. д-р Н. П. Шеповальников и проф. Ю. Л. Поморский: „Возрастно-половые эволютивные особенности школьника и методика обработки стандартов антропометрических величин“. Ленинград, 1927 г., издание Института О.З.Д.П.

следующие результаты измерения, относящиеся к ребенку соответствующего возраста и пола:

Рост стоя	150 см
Рост сидя	76 „
Длина ног	74 „
Окружн. груди	63 „
Окружн. головы	52 „

Сопоставляя эти абсолютные измерения с пограничными числами, выписанными в оценочной таблице, мы видим, что для данного ребенка:

его рост стоя	(150 см)	попадает в зону $+B$
„ „ сидя	(76 „)	„ „ „ A
„ длина ног	(74 „)	„ „ „ $+C$
„ окружн. груди	(63 „)	„ „ „ $-B$
„ „ головы	(52 „)	„ „ „ A .

Следовательно этот ребенок при общем росте несколько выше среднего уровня ($+B$) имеет длинные ноги ($+C$) и немного узкую грудь ($-B$).

Совершенно также и при разбивке материала по принципу равночисленных зон (с помощью интегральной кривой) площади отдельных участков соответствующей гистограммы делаются иногда не точно равными друг другу, а лишь пропорциональными заранее установленным числам, часто, например, в пропорции:

$$1 : 2 : 4 : 2 : 1.$$

При этих условиях в центре разбивки выделяется зона „нормы“, заключающая в себе при данном распределении ровно 40% всех вариантов заданного ряда; справа и слева от „нормы“ в зоны небольших отклонений попадает по 20% этих вариантов, а в самые крайние зоны по 10%.

Пусть, например, имеем ряд, гистограмма которого представлена на черт. 77. С помощью обычного построения на интегральной кривой (здесь она не указана) можно разбить этот ряд на 5 оценочных зон, заключающих в себе заданное выше количество вариантов (см. главу XII). Границами этих зон будут здесь служить следующие 4 значения варьирующего признака:

10, 13, 19 и 28,

причем, как нетрудно видеть из чертежа, внутри этих границ оказывается как раз заданное выше количество вариантов:

$$10, \quad 20, \quad 40, \quad 20, \quad 10.$$

В самом деле, в границу зоны от 4 до 10 попадают два класса, сумма частот которых:

$$2 + 8 = 10.$$

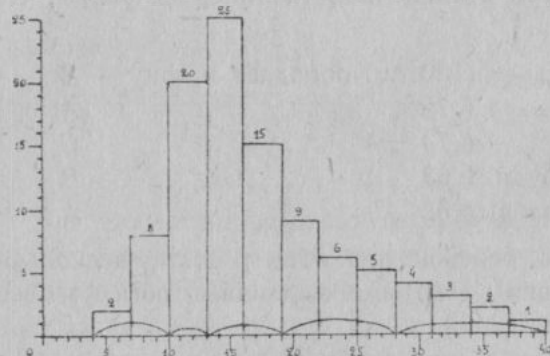
В следующей зоне от 10 до 13 заключается всего только один класс с частотой:

$$p = 20.$$

В зону от 13 до 19 попадают два класса, сумма частот которых:

$$25 + 15 = 40.$$

В границах зоны от 19 до 28 имеются три клас-



Черт. 77.

са, частоты которых дают в сумме:

$$9 + 6 + 5 = 20.$$

Наконец, в крайнюю справа зону от 28 до 40 входят остальные 4 класса, причем сумма их частот:

$$4 + 3 + 2 + 1 = 10.$$

Наряду с описанным здесь приемом часто применяется также разбивка на 5 оценочных зон, заключающих внутри своих границ количество вариантов, пропорциональное числам:

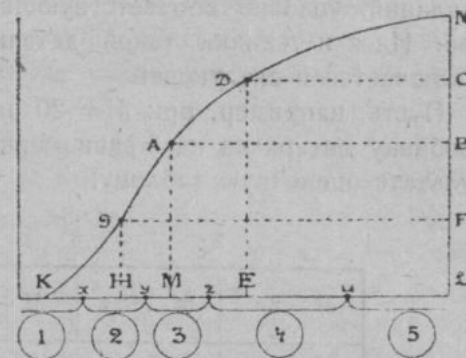
$$10, \quad 15, \quad 50, \quad 15 \text{ и } 10.$$

Иногда в целях использования особенностей и преимуществ обоих основных способов зональной разбивки применяются различные комбинированные приемы подразделения материала на оценочные участки. В качестве примера попытки подобного типа приведем краткое описание способа определения 4-х границ пяти оценочных зон, применяемого в одном

из научно-исследовательских учреждений г. Москвы. Здесь зоны эти устанавливаются с помощью интегральной кривой по обычному принципу равенства площадей, но центральная область „нормы“ определяется положением на гистограмме не медианы *Med*, а величиной среднего арифметического *M*.

Схема построения здесь такова (см. черт. 78).

По масштабу горизонтальной оси находят точку *M*, соответствующую среднему арифметическому заданного ряда. Из этой точки *M* восстанавливают перпендикуляр *MA* до пересечения с интегральной кривой в точке *A*. Далее, через точку *A* проводят горизонтальную прямую *AB* вправо до пересечения с максимальной ординатой этой кривой (*LN*) в точке *B*. Полученная точка (*B*) делит весь отрезок *LN* на две неравные части: *NB* и *BL*, причем, очевидно, величина этих отрезков указывает (по вертикальному масштабу) на количество вариантов, больших и меньших среднего арифметического *M* (т. е. лежащих справа и слева от точки *M* в участках *KM* и *ML* на оси абсцисс).



Черт. 78.

Для каждого из двух таких неравных интервалов (т. е. для отрезков *KM* и *ML*) находят затем их „срединные“ значения *H* и *E* путем обычного построения: т. е. делят отрезки *NB* и *BL* пополам в соответствующих точках *C* и *F*, из которых проводят горизонтальные прямые (влево) до пересечения с интегральной кривой в точках *D* и *G*, а из этих точек опять опускают перпендикуляры *DE* и *GH*. Основания этих перпендикуляров *H* и *E* по положению своему на оси абсцисс и укажут искомые „срединные“ значения участков *KM* и *ML*.

Тремя найденными точками *H*, *M* и *E* весь интервал фактического варьирования *KL* разобьется на 4 неравные части: *KH*, *HM*, *ME* и *EL*. Разделив все эти отрезки пополам,

получим на горизонтальной оси точки x , y , z и u , которые и принимаются здесь за 4 границы пяти оценочных зон.

На черт. 78 эти зоны, как всегда, указаны дугами у оси абсцисс.

Наряду с описанным способом возможны и другие комбинированные приемы разбивки материала на оценочные зоны, причем принцип и техника этих методов всецело зависят от особенностей подлежащего оценке материала и от конечной цели самой разбивки.

В тех случаях, когда оценка индивидуальных измерений V производится в той или иной системе балловых обозначений, иногда полезно бывает ввести сюда ряд промежуточных градаций, условно соответствующих половинам целых баллов. Идея и техника такой детализации конечной оценки заключается в следующем:

Пусть, например, при $M=20$ и $\sigma=\pm 6$ мы произвели разбивку материала на 5 равномерных зон и получили в результате оценочную таблицу:

1	2	3	4	5
$M - 2\sigma$	$M - \frac{2}{3}\sigma$	$M + \frac{2}{3}\sigma$	$M + 2\sigma$	
8	16	24	32	

Представим себе, что нашей оценке подлежат две отдельные варианты: $V_1=23,9$ и $V_2=24,1$. Первая из них, попадая в интервал от 16 до 24, оказывается, очевидно, в области „нормы“ (балл 3), а вторая, заключенная в интервале от 24 до 32, должна быть отнесена уже к зоне балла 4.

Так как обе эти варианты (23,9 и 24,1) оказываются чрезвычайно близкими к пограничному значению признака $g=24$ (разделяющему зоны баллов 3 и 4), то по отношению к ним невольно возникает вопрос о целесообразности столь резкого различия в оценке двух таких в сущности почти одинаковых вариантов. Эти сомнения усиливаются еще благодаря тому обстоятельству, что сама-то граница g не является здесь абсолютно устойчивой, так как при повторных ее вычислениях на другом аналогичном материале она неизбежно оказалась бы либо немного меньше, либо, наоборот, больше

данного, по существу все же случайного своего значения $g=24$. В самом деле, эта граница была здесь вычислена путем прибавления к M двух третей σ :

$$g = M + \frac{2}{3}\sigma = 20 + \frac{2}{3} \cdot 6 = 20 + 4 = 24.$$

Так как и M и σ при повторных вычислениях на другом материале могут варьировать в границах своих средних ошибок (m и m_σ), то и сумма их $M + \frac{2}{3}\sigma$ также будет колебаться в некоторых, вообще говоря, не очень широких пределах. Следовательно в других случаях эта граница $g=24$ могла бы или передвинуться несколько влево (т. е. сделаться меньше) или, наоборот, вправо (т. е. сделаться больше), в каком-то смысле выше мы и называли ее неустойчивой.

В силу этих соображений оценку отдельных вариантов, лежащих в непосредственной близости от такой неустойчивой границы, необходимо производить с особой осторожностью. Кто поручится, например, за то, что варианта $V_1=23,9$ действительно лежит слева от границы $g=24$ (т. е. оказывается меньше ее), принимая во внимание возможность перемещения влево же и самой этой границы? Ведь при таких условиях наша варианта V_1 могла бы оказаться, наоборот, справа от этой передвинувшейся границы g , и мы должны были бы отнести ее уже не к „норме“, а к зоне, оцениваемой баллом 4.

Правда, значения V , близкие к максимальным пределам ожидаемых колебаний самой границы g , являются уже достаточно определенными в смысле принадлежности их к той или иной оценочной зоне, потому что такие резкие перемещения этой границы сами по себе чрезвычайно мало вероятны. Но, менее удаленные от g , значения вариант V все же имеют еще много шансов оказаться при иных условиях и по другую ее сторону, что заставляет нас воздержаться здесь от точной оценки этих вариантов баллами 3 или 4, приняв для такого неопределенного случая некоторое промежуточное обозначение, напр. $3\frac{1}{2}$.

Как же определить ширину такой промежуточной зоны? Теоретические рассуждения, на которых однако мы не имеем

возможности здесь останавливаться, приводят к следующей окончательной формуле:

$$\delta = \pm \frac{2}{3} \sqrt{\frac{d^2 + 2}{2}} \cdot m$$

или точнее:

$$\delta = \pm 0,6745 \sqrt{\frac{d^2 + 2}{2}} \cdot m,$$

где через δ обозначено расстояние границ этой промежуточной зоны от основной границы g (разделяющей зоны обозначений целыми баллами); через d — расстояние этой основной границы g от среднего арифметического M , выраженное в долях среднего квадратического отклонения σ , а через m — средняя ошибка самого среднего арифметического M *).

Определим по этой формуле границы промежуточной зоны, примыкающей к основной границе $g = 24$ в нашем последнем примере (при $M = 20$ и $\sigma = \pm 6$).

Здесь заданная граница $g = 24$ отстоит от среднего арифметического $M = 20$ на расстояние:

$$g - M = 24 - 20 = 4.$$

Так как в нашем примере $\sigma = 6$, то это расстояние (4) составляет 0,67 сигмы:

$$\frac{4}{6} = 0,67.$$

Итак, в данном случае величина:

$$d = \frac{g - M}{\sigma} = 0,67.$$

Предположив, что M и σ были вычислены из 100 измерений ($n = 100$), найдем среднюю ошибку:

$$m = \pm \frac{\sigma}{\sqrt{n}} = \pm \frac{6}{\sqrt{100}} = \pm 0,6.$$

Подставляя эти значения $d = 0,67$ и $m = 0,6$ в указанную выше формулу для вычисления δ , получим (приблизительно):

$$\delta = \pm \frac{2}{3} \sqrt{\frac{d^2 + 2}{2}} \cdot m = \pm 0,67 \sqrt{\frac{(0,67)^2 + 2}{2}} \cdot 0,6 = \pm 0,44.$$

*) См. I часть этого руководства.

Итак, искомые границы промежуточной зоны должны быть проведены здесь на расстоянии $\delta = 0,44$ от основной границы $g = 24$. Поэтому числовые значения двух границ этой зоны окажутся равными:

$$24 + 0,44 = 24,44$$

$$24 - 0,44 = 23,56.$$

На этом основании все варианты, заключающиеся в пределах от 23,56 до 24,44, в нашем примере должны быть отнесены к промежуточной зоне балла $3\frac{1}{2}$, так как вероятность их попадания в ту или иную определенную зону оказывается меньше вероятности невыясненной (неопределенной) оценки. Это значит, что мы имеем меньше шансов за попадание данной варианты в зону балла 3, чем за неопределенное ее попадание либо в эту, либо в соседнюю зону (балла 4), принимая во внимание возможность перемещения вправо или влево также и самой разделяющей эти зоны основной границы g .

Для границы $g = 32$, разделяющей в нашем примере зоны баллов 4 и 5, величина:

$$d = \frac{g - M}{\sigma} = \frac{32 - 20}{6} = 2,$$

а средняя ошибка m по прежнему $= \pm 0,6$. Следовательно:

$$d = \pm \frac{2}{3} \sqrt{\frac{d^2 + 2}{2}} \cdot m = \pm 0,67 \sqrt{\frac{2^2 + 2}{2}} \cdot 0,6 = \pm 0,69.$$

После вычисления ширины промежуточных зон (δ) для всех четырех основных границ (g) получим следующие окончательные результаты:

для границы	$g = 32$	ширина промежуточной зоны	$\delta = \pm 0,69$
"	"	$g = 24$	" $\delta = \pm 0,44$
"	"	$g = 16$	" $\delta = \pm 0,44$
"	"	$g = 8$	" $\delta = \pm 0,69$

Полезно запомнить, что ширина δ всегда увеличивается по мере удаления g от M , а для границ g , удаленных от M на одинаковые расстояния, и промежуточные зоны оказываются также одинаковой ширины.

Прибавив к числовому значению каждой основной границы g вычисленную ширину соответствующей промежу-

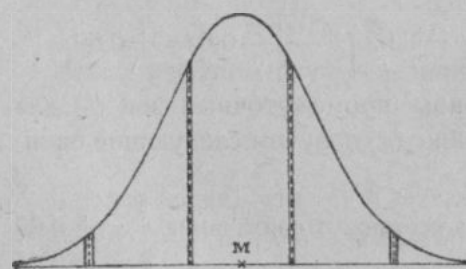
точной зоны δ и вычитая эту ширину δ из g , найдем границы всех 4-х промежуточных зон:

$$\begin{aligned} \text{Для оценки баллом 4: } & \begin{cases} g + \delta = 32 + 0,69 = 32,69 \\ g - \delta = 32 - 0,69 = 31,31 \end{cases} \\ \text{Для оценки баллом 3: } & \begin{cases} g + \delta = 24 + 0,44 = 24,44 \\ g - \delta = 24 - 0,44 = 23,56 \end{cases} \\ \text{Для оценки баллом 2: } & \begin{cases} g + \delta = 16 + 0,44 = 16,44 \\ g - \delta = 16 - 0,44 = 15,56 \end{cases} \\ \text{Для оценки баллом 1: } & \begin{cases} g + \delta = 8 + 0,69 = 8,69 \\ g - \delta = 8 - 0,69 = 7,31 \end{cases} \end{aligned}$$

Облекая эти результаты в форму оценочной таблицы обычного типа, но с более подробным подразделением всего диапазона фактического варьирования на 5 основных и 4 промежуточные зоны, получим:

1	1 1/2	2	2 1/2	3	3 1/2	4	4 1/2	5
7,31	8,69	15,56	16,44	23,56	24,44	31,31	32,69	

Ниже для ясности изображена также и кривая нормального распределения с подразделением ее на основные и промежуточные зоны в соответствии с данными нашего последнего примера (см. черт. 79).



Черт. 79.

В целях облегчения самой техники вычисления ширины промежуточных зон (δ) по формуле:

$$\delta = \pm \frac{2}{3} \sqrt{\frac{d^2 + 2}{2}} \cdot m$$

можно заранее составить таблицу значений величины

$$\frac{2}{3} \sqrt{\frac{d^2 + 2}{2}}$$

на которую приходится умножать среднюю ошибку m при определении ширины δ .

Ниже приведена такая вспомогательная таблица для отдельных значений d в пределах от 0 до 3,9 через каждую десятую долю d . В этой таблице символом δ_0 обозначена искомая величина:

$$\frac{2}{3} \sqrt{\frac{d^2 + 2}{2}}$$

d	δ_0	d	δ_0	d	δ_0	d	δ_0
0	0,67	1,0	0,83	2,0	1,17	3,0	1,58
0,1	0,68	1,1	0,86	2,1	1,21	3,1	1,63
0,2	0,68	1,2	0,89	2,2	1,25	3,2	1,67
0,3	0,69	1,3	0,92	2,3	1,29	3,3	1,71
0,4	0,70	1,4	0,94	2,4	1,33	3,4	1,76
0,5	0,72	1,5	0,99	2,5	1,37	3,5	1,80
0,6	0,73	1,6	1,02	2,6	1,41	3,6	1,85
0,7	0,75	1,7	1,06	2,7	1,45	3,7	1,89
0,8	0,78	1,8	1,09	2,8	1,50	3,8	1,93
0,9	0,80	1,9	1,13	2,9	1,53	3,9	1,98

Таким образом, при пользовании этой таблицей ширина промежуточной зоны δ вычисляется по формуле:

$$\delta = \delta_0 \cdot m,$$

где коэффициент δ_0 определяется непосредственно по таблице (для данного значения d).

Пусть, например, требуется определить ширину промежуточной зоны, примыкающей к основной границе $g=10$, удаленной от M на расстояние $d=2,5$ сигмы, причем нам известно, что средняя ошибка самого M , т. е. $m = \pm 0,8$.

Из таблицы видно, что значению $d=2,5$ соответствует величина $\delta_0=1,37$. Подставляя ее в указанную выше формулу, получим:

$$\delta = \delta_0 \cdot m = 1,37 \cdot 0,8 = 1,10.$$

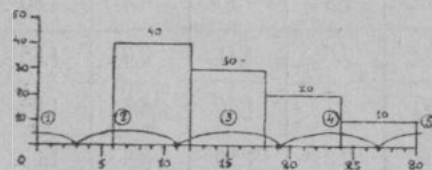
Поэтому границами этой промежуточной зоны будут служить следующие значения варьирующего признака: —

$$g + \delta = 10 + 1,10 = 11,10$$

$$g - \delta = 10 - 1,10 = 8,90.$$

XV. Распределение вариантов по оценочным зонам.

Если заданный вариационный ряд был разбит на части по принципу равночисленных зон, и оценке были подвергнуты все без исключения варианты этого ряда, то количество индивидуумов, попавших в границы каждой отдельной зоны нам очевидно всегда будет известно заранее. Но при разбивке этого же ряда по принципу равномерных зон или при оценке равночисленными зонами только какой-либо составной части всего обследованного коллектива



Черт. 80.

мы не сможем уже точно предугадать характер распределения конечных результатов такой оценки по отдельным зонам и вынуждены будем прибегнуть для этого к дополнительным вычислениям.

Пусть, например, в результате какого-либо психотехнического испытания мы получили вариационный ряд, гистограмма которого представлена на черт. 80. Для данного ряда $M = 15$, а $\sigma = \pm 6$ (этими данными мы уже пользовались однажды в гл. XIII на стр. 131). При разбивке этого ряда на 5 равномерных оценочных зон получим 4 их границы в точках:

3, 11, 19 и 27,

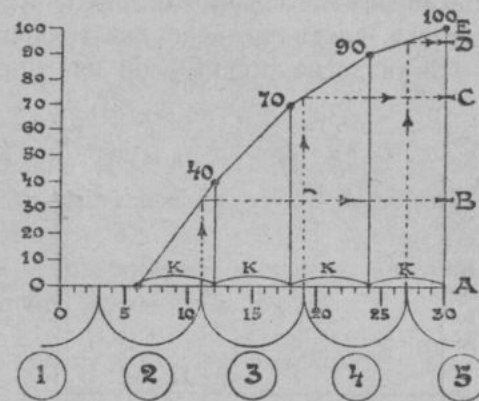
из которых первая (3) выходит здесь за пределы фактического варьирования (от 6 до 30).

В виду явной асимметрии заданного ряда, в центральную зону балла 3 здесь уже не попадет 50% всех вариантов, как это наблюдалось бы при строго нормальном их распределении, точно также, как и в соседних зонах 2 и 4 не ока-

жется уже по 22,73%, а в самых крайних (1 и 5) — по 2,27% *). В частности, например, в крайнюю слева зону балла 1 здесь вовсе не попадет ни одной варианты.

Принимая во внимание эти своеобразные (характерные только для данного теста) условия распределения, представляется весьма интересным определить фактическое количество вариантов, заключающихся внутри границ каждой отдельной оценочной зоны. Сделать это можно либо только графическим способом, либо путем дополнительных (уточняющих) вычислений по специальной формуле.

Вычертив интегральную кривую заданного ряда, пометим на оси абсцисс четыре границы наших 5-ти оценочных зон (снизу), а сверху, для ясности, обозначим дугами на той же оси (буквой k) также и все классовые границы первоначально заданного ряда (см. черт. 81).



Черт. 81.

В интервал фактического варьирования (от 6 до 30) здесь попали только 3 зональные границы, соответствующие по горизонтальному масштабу точкам 11, 19 и 27.

Для определения процентного количества вариантов, попавших в каждую отдельную зону, из этих точек следует восстановить перпендикуляры до интегральной кривой и через соответствующие точки пересечения провести три горизонтальные прямые вправо до пересечения с максимальной ординатой (AE) этой кривой в точках B, C и D **). Тогда величина полученных здесь отрезков этой максимальной

*) См. гл. XII, стр. 123.

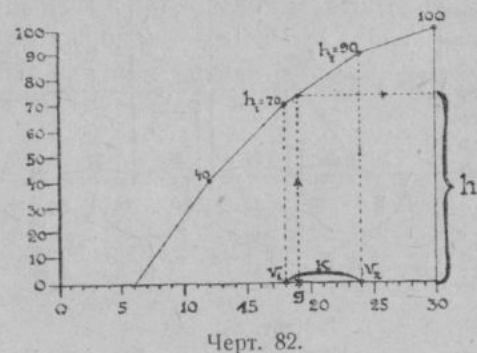
**) Т. е. выполнить построение как раз обратное тому, которым мы пользовались при разбивке ряда на равночисленные оценочные зоны, с той лишь разницей, что теперь на равные части у нас разбита ось абсцисс, а ранее мы разделяли на равные же части максимальную ординату интегральной кривой (см. стр. 126).

ординаты AB , BC , CD и DE в единицах того же самого вертикального масштаба и укажет нам количество вариантов, попавших в каждую отдельную зону. Измерив эти отрезки, мы без труда убедимся в том, что в зону балла:

5 (отрезок ED)	попадет приблизительно	5	вариант
4 (" DC)	"	"	22 "
3 (" CB)	"	"	40 "
2 (" BA)	"	"	33 "

Описанный здесь графический способ решения этой задачи является, очевидно, неполно точным.

По аналогии с формулой для определения точного положения зональной границы (см. стр. 127):



Черт. 82.

$$g = V_1 + k \left(\frac{h - h_1}{h_2 - h_1} \right)^*$$

мы можем воспользоваться здесь обратной формулой, позволяющей вычислять отсекаемую от интегральной кривой высоту h , соответствующую данной границе g :

$$h = h_1 + \frac{(g - V_1)(h_2 - h_1)}{k}$$

Эта величина h и будет в наших условиях обозначать количество вариантов, расположенных в заданном ряду слева от границы g (т. е. по величине своей меньших числового значения этой границы).

Здесь, как и раньше, через V_1 обозначена ближайшая к g левая классовая граница, через h_1 высота интегральной кривой в точке V_1 , через h_2 — высота этой кривой в соседней справа классовой границе V_2 (она в формулу не входит), а через k — как всегда, классовый промежуток.

Пусть, например, мы пожелаем вычислить по этой формуле количество вариантов, расположенных в заданном ряду слева от границы $g = 19$ (см. черт. 82).

*) См. гл. X и XI.

Здесь ближайшая слева от g классовая граница V_1 соответствует по горизонтальному масштабу точке 18, а ближайшая справа V_2 — точке 24. Высота интегральной кривой в точке V_1 , т. е. $h_1 = 70$, а в точке V_2 , т. е. $h_2 = 90$. Классовый промежуток k (расстояние от V_1 до V_2) здесь = 6. Итак, для нашего примера:

$$\begin{array}{lll} g = 19 & h_1 = 70 & \\ V_1 = 18 & h_2 = 90 & k = 6. \end{array}$$

Подставляя эти данные в указанную выше формулу для вычисления высоты интегральной кривой в точке g , получим:

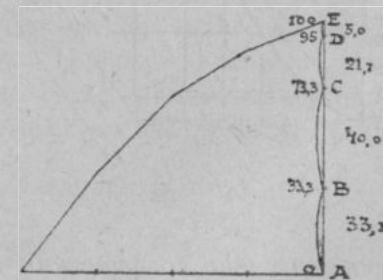
$$h = h_1 + \frac{(g - V_1)(h_2 - h_1)}{k} = 70 + \frac{(19 - 18)(90 - 70)}{6} = 73,3.$$

Следовательно зональная граница $g = 19$ слева от себя отсекает в заданном ряду приблизительно 73 варианты из общего их числа $n = 100$ (т. е. 73,3%).

Выполнив аналогичные вычисления для всех зональных границ, получим следующие результаты:

для границы $g = 27$	высота $h = 95,0$
" " $g = 19$	" $h = 73,3$
" " $g = 11$	" $h = 33,3$ *).

Полученные здесь высоты h определяют лишь расстояние соответствующих точек D , C и B от нижней точки A (см. черт. 83). Размер же интересующих нас отдельных участков ED , DC , CB и BA (пропорциональных количеству вариантов, получивших оценку баллами 5, 4, 3 и 2) может быть найден для каждого отрезка путем вычитания. Так, например, в отрезке CB верхняя его точка C имеет высоту $h = 73,3$, а для нижней точки B ее высота $h = 33,3$. Общую же длину этого



Черт. 83.

*) Очевидно в применении к границе $g = 11$ величина $h_1 = 0$, а для границы $g = 27$ величина $h_2 = 100$.

отрезка CB мы и получим здесь очевидно путем вычитания 33,3 из 73,3:

$$73,3 - 33,3 = 40.$$

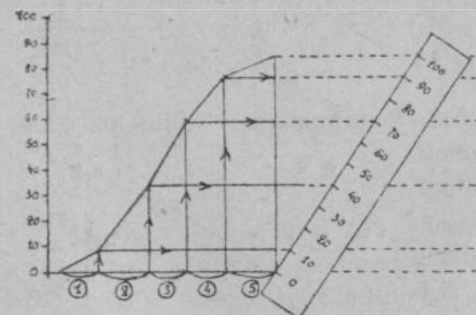
Выполнив аналогичные вычисления для всех 4-х отрезков, найдем, что:

длина отрезка ED	равна	$100,0 - 95,0 = 5,0$
" "	DC	" $95,0 - 73,3 = 21,7$
" "	CB	" $73,3 - 33,3 = 40,0$
" "	BA	" $33,3 - 0 = 33,3$

Отсюда видно, что в заданном ряду оценку:

баллом 5	получат	5 вариант
" 4	"	22 "
" 3	"	40 "
" 2	"	33 "
" 1	"	0 "

В данном случае (при $n=100$) эти абсолютные числа очевидно являются вместе с тем также и процентными. При ином же количестве измерений каждое из этих чисел можно выразить в $\%$ от n посредством обычных вычислений (умножить на 100 и разделить на n).



Черт. 84.

Если выяснение количества вариантов, попадающих в отдельные зоны, производилось исключительно графическим способом, то это конечное перечисление в $\%$ можно выполнить также путем очень несложного дополнительного построения.

Пусть, например, границы 5 равночисленных оценочных зон были установлены на обширном материале, скажем в 2.000 человек, а оценке подвергается лишь ничтожная часть этого материала, например только 85 человек. Построив интегральную кривую только для этих 85 оцениваемых

вариант и пометив на оси абсцисс границы всех зон, установленных на полном материале в 2.000 человек, выполним обычные только что описанные построения, указанные сплошными линиями на черт. 84. Далее продолжим вправо (см. пунктир) все горизонтальные линии, которыми мы разделили на части наибольшую ординату интегральной кривой. Приложив наклонно к этим прямым заранее заготовленную линейку, разделенную на 100 каких-либо равных частей, мы и получим возможность измерить все 5 интересующих нас отрезков в процентах от $n=85$. В данном случае, следовательно, в зону балла:

5	попало	10%	вариант
4	"	30%	"
3	"	30%	"
2	"	20%	"
1	"	10%	"

Определив тем или иным способом абсолютное или процентное количество вариантов, попавших в каждую отдельную оценочную зону, мы можем, далее, выразить результат такого подсчета в обычной графической форме, изобразив это распределение вариантов по зонам в виде соответствующей гистограммы (подобно гистограмме распределения вариантов по классам их первичной разбивки).

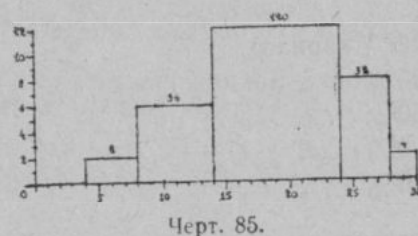
Если все оценочные зоны имели одинаковую ширину, то вычерчивание такой гистограммы очевидно не может встретить никаких затруднений. Покажем, как выполнить это построение в применении к зонам различной ширины.

Пусть, например, нам стало известно, что в границы зоны:

от 30 до 28	попало	4 варианты	(2%)
" 28 "	24 "	32 "	(16%)
" 24 "	14 "	120 "	(60%)
" 14 "	8 "	36 "	(18%)
" 8 "	4 "	8 "	(4%)

Здесь эти зоны имеют различную ширину (4, 6, 10, 4, 2), что уже лишает нас права строить соответствующие им отдельные столбики гистограммы, просто откладывая их высоту пропорционально частотам: 8, 36, 120, 32 и 4.

В самом деле, при построении гистограммы каждая частота p изображается на чертеже площадью соответствующего столбика, зависящей, очевидно, как от его высоты, так и от ширины. Ранее, когда ширина всех столбиков была у нас совершенно одинаковой (равной классовой промежутку k), эта площадь зависела только от высоты столбика, которую мы и откладывали на гистограмме, руководствуясь величиною соответствующей частоты p . Теперь этого делать уже нельзя, и при построении каждого отдельного столбика нам придется наряду с изображаемой им частотой p учесть также и переменную ширину соответствующей оценочной зоны L . Таким образом в этом общем случае высоту каждого столбика H следует вычислять по формуле:



Черт. 85.

$$H = \frac{p}{L}$$

В применении, например, к центральной оценочной зоне (от 14 до 24), в границы которой по условию попало 120 вариантов, мы будем иметь:

$$L = 24 - 14 = 10 \quad p = 120.$$

Следовательно высота соответствующего столбика:

$$H = \frac{p}{L} = \frac{120}{10} = 12 \text{ единиц условного вертикального масштаба.}$$

Выполнив аналогичные вычисления для всех 5-ти оценочных зон, получим:

для зоны от 30 до 28	высота столбика	$H = 2$
" " " 28 " 24	" "	$H = 8$
" " " 24 " 14	" "	$H = 12$
" " " 14 " 8	" "	$H = 6$
" " " 8 " 4	" "	$H = 2$

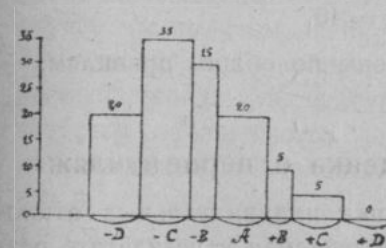
Построенная по этим данным гистограмма представлена на черт. 85. Здесь слева указан условный вертикальный масштаб, а цифрами над столбиками помечено количество вариантов, попавших в отдельные оценочные зоны.

Непосредственным вычислением легко проверить, что площади этих столбиков пропорциональны здесь частотам — 8, 36, 120, 32 и 4, что и соответствует нашему заданию.

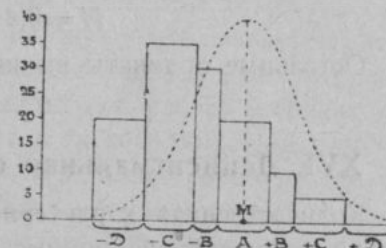
Подобные вычисления значительно упрощаются в том частном случае, когда отдельные зоны оказываются вдвое шире или уже остальных. Так, например, при вычерчивании гистограммы распределения отдельных оценок по 7-ми Мартиновским зонам*) можно пользоваться обычным вертикальным масштабом и лишь при построении столбиков для зон $-B$ и $+B$ (вдвое более узких) просто откладывать вверх соответствующую им высоту дважды.

Пусть, например, при обследовании какой-либо группы детей оказалось, что в зону:

$+D$	попало	0% всех вариантов
$+C$	"	5% " "
$+B$	"	5% " "
A	"	20% " "
$-B$	"	15% " "
$-C$	"	35% " "
$-D$	"	20% " "



Черт. 86.



Черт. 87.

При построении соответствующей этому случаю гистограммы (см. черт. 86) мы должны все столбики, включая и крайние, сделать одинаковой ширины, кроме двух полосок для зон $-B$ и $+B$, которые будут здесь вдвое уже. Высоту всех этих столбиков мы откладываем по вертикальному масштабу, указанному на чертеже слева, но столбики $-B$ и $+B$ делаем вдвое выше (вместо 15 строим 30, а вместо 5 берем 10) так как по своей ширине они вдвое уже

*) См. гл. XIV

остальных. При этих условиях площади всех наших столбиков и окажутся пропорциональными заданным числам:

$$20, 35, 15, 20, 5, 5 \text{ и } 0,$$

что нетрудно проверить, подсчитав число клеток внутри каждого отдельного столбика на черт. 86.

Приведенная здесь гистограмма наглядно показывает, что обследованный коллектив представляет собой отсталую группу (по сравнению с той массой, на которой был выработан самый критерий этой оценки).

На черт. 87 эта гистограмма дана в сопоставлении с кривой нормального распределения (для всех 2.000 вариантов), в которой центральная точка M расположена в середине зоны A , за σ принята ширина этой зоны A (или зон C и D , что все равно), а n взято равным фактическому числу вариантов (100), составляющих заданный ряд. Таким образом при вычислении наибольшей высоты кривой нормального распределения по формуле:

$$H = \frac{0,4 \cdot kn}{\sigma} *$$

мы здесь приняли $k = \sigma$, а $n = 100$, откуда:

$$H = 0,4 \cdot n = 40.$$

Остальные ординаты вычислены по общим правилам.

XVI. Децисигмальная оценка и перцентильяж.

Количественная характеристика индивидуальных отклонений от условной „нормы“, получаемая в результате разбивки материала на те или иные зоны, как показывает опыт, обычно обеспечивает уже вполне достаточную в практическом отношении точность такой оценки. Однако для одновременной характеристики индивидуума по совокупности нескольких признаков (например, по ряду антропометрических признаков или по нескольким психотехническим испытаниям одновременно) такая приближенная оценка этих признаков, в виду необходимости дальнейшей их суммации, оказывается все же не достаточно еще дифференцированной. В самом деле, вычисляя, например, средний балл по резуль-

*) См. стр. 22.

татам разбивки одного признака на те или иные оценочные зоны, мы смело можем пренебречь той неизбежной неточностью, которая вкрадывается в наши вычисления в силу неопределенности этой оценки внутри отдельных зон (от — до), так как такая неточность всегда может быть (и бывает) компенсирована достаточно большим числом измерений или наблюдений. Но при выводе средней оценки для каждого отдельного индивидуума по нескольким характеризующим его признакам одновременно (см. дальше гл. XIX) мы рискуем получить в результате неопределенно крупную ошибку вследствие того, что общее количество совместно учитываемых при этом моментов чаще всего бывает весьма ограниченным и во всяком случае не поддается произвольному увеличению.

Эти соображения побуждают нас в некоторых случаях уточнить зональную разбивку путем увеличения общего количества оценочных зон и соответственного уменьшения расстояния между их границами. При этом, как показывает практика, равномерные (сигмальные) зоны достаточно бывает сузить до $1/10 \sigma$, а при разбивке материала по принципу равночисленных зон — принять их количество = 100. В первом случае мы получим оценку, могущую быть названной „децисигмальной“, а во втором — способ разбивки на зоны, называемый иногда „перцентильяжем“ (от слова „перцентиль“, которым обозначаются участки гистограммы при разбивке ее площади на 100 равных частей). Само собой разумеется, что все соображения, высказанные о равномерных и равночисленных оценочных зонах в главе XIII, в то же самое время относятся и к более дифференцированным способам количественной характеристики индивидуальных вариантов посредством их децисигмальной оценки и перцентильяжа.

Децисигмальные отклонения отдельных вариантов V от среднего арифметического M вычисляются по формуле:

$$d = 10 \left(\frac{V - M}{\sigma} \right) *$$

*) Если по самому смыслу вопроса качество вариант повышается вместе с понижением выражающих это качество чисел, то децисигмальная оценка вычисляется по формуле:

$$d = 10 \left(\frac{M - V}{\sigma} \right).$$

причем эту величину d достаточно определять лишь в целых числах, отбрасывая все десятичные знаки после запятой по существующим для этого правилам. Пусть, например, при $M=10$ и $\sigma=\pm 3$ мы захотели бы дать децисигмальную оценку отдельной варианты $V=5$. Подставляя эти данные в указанную выше формулу, получим:

$$d = 10 \left(\frac{5-10}{3} \right) = -16,67$$

или окончательно (после округления):

$$d = -17.$$

Полученный результат оценки показывает, что заданная варианта (5) меньше среднего арифметического M (знак —) на 17 десятых долей σ (т. е. на 1,7 σ).

Такое отклонение (почти на 2 σ) очевидно при всех критериях оценки можно считать большим, и заданную варианту ($V=5$) в этом примере мы должны отнести к категории очень малых (или очень плохих, в зависимости от смысла оценки).

Вообще конечная интерпретация децисигмальных оценок (d) зависит от соображений, изложенных выше в применении к равномерным оценочным зонам (см. гл. XII). Так, в частности, при разбивке материала всего лишь на 3 зоны, соответствующие понятиям: малый, средний и большой (или плохой, средний и хороший), условная „норма“ будет лежать в пределах от -10 до $+10$ децисигм (т. е. от $M-3$ до $M+3$); при разбивке этого же материала на 5 равномерных оценочных зон такая „норма“ будет заключаться в несколько более суженных пределах: от -6 до $+6$ децисигм (от $M-2/3\sigma$ до $M+2/3\sigma$) и т. д. (см. гл. XII). Максимальные же пределы варьирования децисигмальных оценок при нормальном распределении будут ограничены значениями d от -30 до $+30$ или в расчете на асимметрию и эксцесс — в пределах от -40 до $+40$ децисигм.

Если в заданном ряду повторяется ограниченное (небольшое) количество отдельных значений V , то для каждого из них удобно заранее вычислить соответствующую децисигмальную оценку d и по этим данным составить оценочную таблицу.

Пусть, например, в результате какого-либо психотехнического испытания, состоящего в выполнении, скажем, 20-ти заданных операций, оказалось, что:

15 операций (75%)	выполнил	1 человек
14 " (70%)	"	9 "
13 " (65%)	"	45 "
12 " (60%)	"	70 "
11 " (55%)	"	45 "
10 " (50%)	"	9 "
9 " (45%)	"	1 "

Приняв за варианты V эти процентные количества правильно выполненных заданий, получим следующий вариационный ряд:

варианта $V=75$	имеет частоту $p=1$
" $V=70$	" " $p=9$
" $V=65$	" " $p=45$
" $V=60$	" " $p=70$
" $V=55$	" " $p=45$
" $V=50$	" " $p=9$
" $V=45$	" " $p=1$

Для этого ряда $M=60$ и $\sigma=\pm 5$, в чем нетрудно убедиться, выполнив соответствующие вычисления по способу сумм (при $k=5$).

В заданном ряду встречается лишь 7 различных значений вариант V (от 75 до 45 с промежутками в 5%). Вычислим для каждого из них соответствующую децисигмальную оценку d .

Для самого большого значения $V=75$:

$$d = 10 \left(\frac{75-60}{5} \right) = +30.$$

Для следующего значения $V=70$:

$$d = 10 \left(\frac{70-60}{5} \right) = +20 \text{ и т. д.}$$

Нетрудно убедиться в том, что:

для $V=65$	оценка $d=+10$
" $V=60$	" $d=0$
" $V=55$	" $d=-10$
" $V=50$	" $d=-20$
" $V=45$	" $d=-30$

Так как в заданном ряду встречаются только эти 7 значений V , то для производства оценки всех 180 вариантов удобно заранее составить специальную оценочную таблицу, которой в данном случае мы можем придать следующий вид:

V	d
75	+30
70	+20
65	+10
60	0
55	-10
50	-20
45	-30

В этой таблице слева выписаны все 7 могущих встретиться значений наших вариантов V , а справа дана их децисигмальная оценка d .

Проглядывая числа столбца d , мы замечаем, что они отличаются друг от друга на постоянную величину (в данном частном случае на 10 децисигм). Обозначив эту постоянную разницу между числами d через Δ (дельта), а постоянную же разницу между соседними значениями самих вариантов V по аналогии с классовым промежутком через k , будем иметь следующее соотношение между Δ и k :

$$\Delta = 10 \frac{k}{\sigma}.$$

В нашем примере $k=5$ и $\sigma=5$, следовательно:

$$\Delta = 10 \frac{5}{5} = 10,$$

что и оправдывается на числах нашего примера.

Существование постоянной разницы Δ между отдельными децисигмальными отклонениями d в некоторых случаях может в значительной степени облегчить работу по составлению такой оценочной таблицы. Вместо того, чтобы для всех последовательных значений V вычислять соответствующие им оценки d по формуле:

$$d = 10 \left(\frac{V - M}{\sigma} \right)$$

мы можем определить d только для двух крайних значений V (самого большого и самого малого), а все промежуточные числа столбца d получить просто путем последовательного прибавления постоянной разности:

$$\Delta = 10 \frac{k}{\sigma}$$

к наименьшему значению V или путем вычитания Δ из наибольшего значения V .

Пусть, например, требуется составить децисигмальную оценочную таблицу для поперечного диаметра головы мальчиков в возрасте от 12 до 13 лет по данным:

$$M = 15,09 \text{ см} \quad \sigma = \pm 0,56 \text{ см}.$$

Допустим, что измерения этого диаметра производятся с точностью до полусантиметра, и что в этом возрасте диаметр головы колеблется приблизительно в пределах от 17 до 13 см *).

Вычислим, прежде всего, децисигмальные оценки d для двух крайних значений V .

Для наибольшего значения $V=17$ будем иметь:

$$d = 10 \left(\frac{17 - 15,09}{0,56} \right) = +34,11,$$

а для наименьшего значения $V=13$:

$$d = 10 \left(\frac{13 - 15,09}{0,56} \right) = -37,32.$$

Оба эти значения d мы пока не будем округлять (посредством отбрасывания 2-х лишних знаков после запятой). Точно также вычислим с достаточным (тем же самым) числом знаков и постоянную разницу:

$$\Delta = 10 \frac{0,50}{0,56} = 8,93.$$

При этих условиях все остальные значения d можно определить путем последовательного вычитания $\Delta = 8,93$ из наибольшего значения $d = +34,11$.

*) Если эти эмпирические пределы варьирования заранее неизвестны, то в предположении нормального распределения, их легко определить приблизительно:

$$\text{от } M + 3\sigma = 16,77 \text{ до } M - 3\sigma = 13,41.$$

Вычитая из 34,11 в первый раз эту величину 8,93, получим:

$$34,11 - 8,93 = +25,18.$$

Вычитая во второй раз эту же величину 8,93, но теперь уже не из +34,11, а из только что полученного числа +25,18, будем иметь:

$$25,18 - 8,93 = +16,25.$$

После следующего вычитания получим:

$$16,25 - 8,93 = +7,32.$$

Поступая также и дальше, найдем следующий ряд постепенно уменьшающихся чисел d (см. таблицу):

V	d
17	+34,11
16,5	+25,18
16	+16,25
15,5	+7,32
15	-1,61
14,5	-10,54
14	-19,47
13,5	-28,40
13	-37,33

Эти числа, очевидно, и представляют собой искомые децисигмальные оценки, соответствующие 9-ти отдельным значениям заданных вариантов V (от 17 до 13 через каждые $1/2$ см), которые здесь выписаны слева от своих оценок.

Последнюю (самую нижнюю) оценку $d = -37,33$ легко можно проверить по следующей контрольной формуле:

$$d_b = d_a - N\Delta,$$

где через d_b обозначена самая нижняя проверяемая оценка наименьшей варианты (у нас $-37,33$), через d_a — самая верхняя оценка наибольшей

варианты (у нас +34,11), через N — общее количество выполненных нами последовательных вычитаний (в данном случае, следовательно, 8), а через Δ — вычисленная ранее постоянная разность между двумя соседними оценками (8,93). Итак в нашем примере:

$$d_b = 34,11 - 8 \cdot 8,93 = -37,33,$$

что мы и получили в действительности.

Эта вычисленная нами оценка наименьшей варианты, оказавшаяся равной $-37,33$, несколько не совпала с ранее по-

лученным ее значением, вычисленным непосредственно по формуле:

$$d = 10 \left(\frac{V - M}{\sigma} \right) = -37,32.$$

Эта разница (в сотых долях) легко объясняется здесь той неизбежной неточностью, с которой мы вычислили самую разницу Δ . Отбросив у $\Delta = 8,93$ третий и все последующие знаки после запятой, мы вычитали далее эту не вполне точную величину 8 раз подряд и естественно накопили в конечном итоге некоторую ошибку, которая и сказалась здесь на 2-м знаке. Если бы это вычитание пришлось выполнять не 8 раз, а много больше (напр. 40 раз), то накапливающаяся в одном и том же направлении ошибка могла бы исказить наш конечный результат даже и в первом знаке после запятой, но целое число децисигм все же от этого, очевидно, не пострадало бы вовсе.

В расчете на такую неизбежную неточность мы и взяли ранее два запасные знака, которые теперь (в конце вычислений) можно отбросить, округлив все оценки d до целых чисел. В результате получим следующую окончательную оценочную таблицу:

Представим себе теперь, что подлежащие нашей оценке варианты V могут принимать не только целые (или целые с половинами, как в последнем примере), но какие угодно промежуточные значения. В этом случае оценочная таблица оказалась бы слишком громоздкой, и пользоваться ею было бы неудобно.

Пусть, например, при $M=20$ и $\sigma = \pm 1,5$ отдельные варианты V были измерены с точностью до десятых долей соответствующей единицы (сантиметра, килограмма и т. д.), причем фактически они варьировали приблизительно в пределах от 25 до 15.

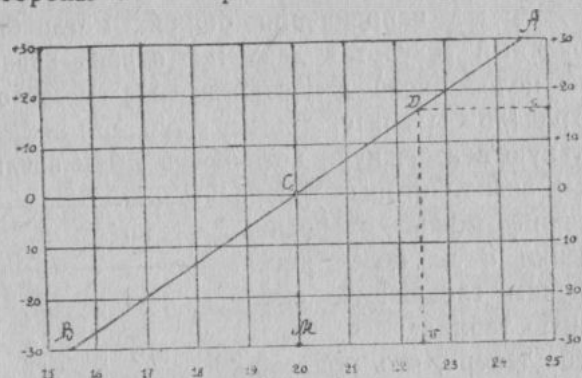
Так как в числе вариантов, подлежащих нашей оценке, могут встретиться любые промежуточные их значения (напр. 16,8 или 19,3), то всего таких различных значений V в указанном интервале окажется около 100. Само собой разумеется

V	d
17	+34
16,5	+25
16	+16
15,5	+8
15	-2
14,5	-11
14	-19
13,5	-28
13	-37

что оценочная таблица, состоящая из 100 строчек, являлась бы слишком громоздкой.

Составим по этим данным „оценочный график“, представляющий собой вычерченный на листе миллиметровой бумаги прямоугольник высотой, например, в 6 см, а шириной около 10 см (см. черт. 88, уменьш. в $1\frac{1}{2}$ раза).

Ширина этого графика определяется здесь количеством подлежащих оценке отдельных значений вариант V из расчета 1 мм на каждое такое значение ($100 \text{ мм} = 10 \text{ см}$), по высоте же такой график удобно делать в 6 см для того, чтобы при размахе колебаний отдельных вариантов V в обе стороны от M приблизительно на 3σ мы имели возможность каждую децисигму приравнять 1 мм.



Черт. 88.

На этом графике по горизонтальному направлению мы будем откладывать варианты V , а по вертикальному — их децисигмальные оценки d .

Начертив такой прямоугольник и проставив у нижней его стороны и с боков масштабные цифры, как это показано на черт. 88, мы должны далее провести наклонную прямую линию AB , определяющую искомую зависимость между различными значениями V и их децисигмальными оценками d . Делается это так:

Вычислим два значения V , соответствующие предельным отклонениям в 3σ от M (т. е. значениям d , равным $+30$ и -30 децисигм *).

$$M + 3\sigma = 20 + 3 \cdot 1,5 = 24,5$$

$$M - 3\sigma = 20 - 3 \cdot 1,5 = 15,5$$

*) В расчете на размах колебаний этих вариантов в обе стороны от M на 4 сигмы (а всего стало быть на 8σ) высота оценочного графика должна быть сделана равной 8 см (80 мм).

Первому из этих значений варианты V будет соответствовать на нашем графике точка A (для нее $V = 24,5$, а $d = +30$), второму же значению V — точка B (для которой $V = 15,5$ а $d = -30$). Через эти две точки мы и проведем по линейке прямую AB , которая, очевидно, должна пройти также и через точку C . В этой точке пересекаются горизонтальная прямая, для которой $d = 0$, и вертикаль, проходящая через точку M , соответствующую среднему арифметическому данного признака (см. черт. 88). Обязательное прохождение прямой AB через точку C служит проверкой правильности построения оценочного графика.

При оценке каждой отдельной варианты мы должны отыскать по нижнему масштабу соответствующую ей точку V (пусть, например, $V = 22,4$), мысленно восставить из нее перпендикуляр (что легко сделать по направлению клеток миллиметровой бумаги) и через точку D пересечения этого перпендикуляра с наклонной прямой AB мысленно же провести горизонтальную прямую вправо или влево (куда удобнее) до соответствующего края графика (обе эти линии на черт. 88 проведены пунктиром). Положение конечной точки E этой горизонтальной прямой и определит по вертикальному масштабу края графика искомую децисигмальную оценку нашей варианты V . В данном случае для $V = 22,4$ оценка $d = +16$.

Тот же самый результат мы получили бы, очевидно, и по формуле:

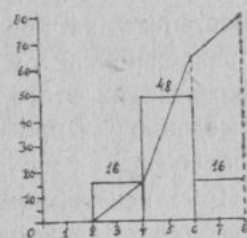
$$d = 10 \left(\frac{V - M}{\sigma} \right) = 10 \left(\frac{22,4 - 20}{1,5} \right) = +16,$$

но вычислять эти оценки конечно труднее, чем определять их с помощью графика.

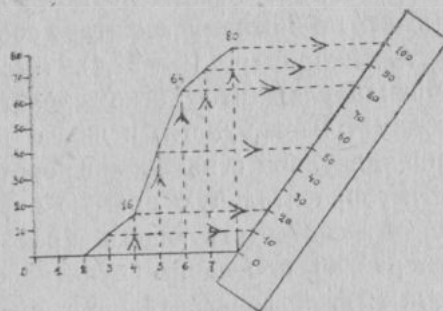
При оценке отдельных вариантов по принципу перцентилежа следовало бы разбить предварительно заданный ряд (гистограмму) на 100 равновеликих участков (перцентилей), подобно тому как это мы делали при разбивке ряда на 4 части для определения квартиля Q (см. гл. XI). Отсчитав затем слева направо порядковый номер того перцентилея, в котором оказалась оцениваемая варианта V , мы и получим ее „перцентильную“ оценку. Таким образом при данном способе оценки единичных вариантов V мы

в сущности определяем ничто иное, как процентное количество тех вариант (из общего их числа n), которые по своей величине оказываются меньше данной варианты V , т. е. на гистограмме отмечаются левее ее. Отсюда ясно, что перцентильная оценка, оказавшаяся равной 50, соответствует медиане ряда (Med), так как медиана как раз и представляет собой границу, отсекающую слева от себя 50% всей площади данной гистограммы.

Однако действительная разбивка гистограмм на 100 таких тонких полосок конечно не является необходимой. Перцентильную оценку всегда можно выполнить графическим путем, подобно тому как это мы делали при определении



Черт. 89.



Черт. 90.

процентного количества вариант, попавших в границы отдельных оценочных зон (см. гл. XV).

Пусть, например, имеем вариационный ряд, гистограмма которого вместе с соответствующей интегральной кривой представлена на черт. 89.

Здесь подлежащие нашей оценке отдельные значения вариант V колеблются в пределах от 2 до 8. Зададимся целью определить перцентильные оценки, например, для всех целых значений этих вариант:

2, 3, 4, 5, 6, 7 и 8.

Для этого следует из всех точек, соответствующих по горизонтальному масштабу заданным вариантам (см. черт. 90), восстановить ряд перпендикуляров до пересечения их с интегральной кривой и через эти точки пересечения провести ряд прямых, параллельных горизонтальной оси, как это показано на черт. 90.

Далее, поставив наклонно линейку, разделенную на 100 равных частей (например миллиметров), следует заметить на ней те точки, в которых проведенные на черт. 90 горизонтальные линии пересекаются с краем линейки. Эти точки в масштабе самой линейки и определяют перцентильные оценки соответствующих вариант.

Так, например, перпендикуляр, восстановленный из точки, соответствующей variante $V=7$ (см. черт. 90), после пересечения с интегральной кривой (в ее участке между цифрами 64 и 80) направляется вправо и пересекает наклонно поставленную линейку в точке, соответствующей 90-му делению ее масштаба. Следовательно, перцентильная оценка варианты $V=7$ в условиях нашего примера = 90.

Из того же чертежа 90 нетрудно видеть, например, что для варианты $V=6$ ее перцентильная оценка = 80, для варианты $V=3$ перц. оц. = 10 и т. д.

Наклонную линейку, по которой производится окончательный отсчет перцентильных оценок, следует ставить на чертеже так, чтобы нулевое деление этой линейки пришлось где-либо на горизонтальной оси чертежа, а верхнее (сотое) ее деление — на продолжении самой верхней горизонтальной прямой, проведенной через крайнюю (самую высокую) точку интегральной кривой.

При интерпретации перцентильных оценок вопрос о „норме“ решается также в соответствии с соображениями, изложенными выше в применении к равночисленным оценочным зонам. Чаще всего за такую норму принимают все значения перцентильных оценок в пределах от 25 до 75 (что дает 50% вариант в „норме“). При желании выделить нормой лишь 40% общего числа всех вариант, за границы этой нормы следует принять 30-й и 70-й перцентили и т. д. (см. гл. XII).

Вообще при интерпретации перцентильных оценок полезно иметь в виду, что оценки эти представляют собой ничто иное, как порядковые номера соответствующих вариант при условии, что все эти варианты были расположены в возрастающий ряд, а всего их было взято 100 (если бы мы имели не 100, а 200 вариант, то, например, порядковый № 30 соответствовал бы перцентильной оценке = 15 и т. д.).

Наряду с децисигмальной оценкой и перцентиляжем иногда применяются также и другие способы числовой характеристики отдельных вариантов V , в которых, например, отклонения этих вариантов выражаются также в децисигмах, но отсчет производится не от среднего арифметического M , а от моды Mo , или, в другом случае, оцениваемая варианта определяется в процентах от среднего арифметического M . Первый из этих способов по своей технике вполне аналогичен описанному выше методу обычной децисигмальной оценки, а второй сводится к простому вычислению процентов.

Полезно запомнить, что в этом последнем случае середина „нормы“ соответствует 100%, а числовые значения отдельных оценок колеблются в обе стороны от 100 в некоторых вообще говоря неопределенных пределах *).

XVII. Индивидуальный профиль и оценочный трафарет.

Представим себе, что достаточно большая и вполне однородная группа индивидуумов была подвергнута какому-нибудь обследованию или испытанию по нескольким (напр. по пяти) отдельным признакам, которые мы условно обозначим здесь буквами: A, B, C, D и E . Этими признаками могут служить, например, те или иные размеры тела, числовые результаты различных психотехнических испытаний и проч. Допустим, далее, что для всей массы этих объектов исследования по каждому из этих пяти варьирующих признаков было вычислено среднее арифметическое M и среднее квадратическое отклонение σ , причем оказалось, что:

для признака	A :	$M = 15$	$\sigma = \pm 4$	
"	"	B :	$M = 6$	$\sigma = \pm 2$
"	"	C :	$M = 8$	$\sigma = \pm 3$
"	"	D :	$M = 3$	$\sigma = \pm 2$
"	"	E :	$M = 22$	$\sigma = \pm 5$

*) Средние пределы колебаний таких процентных оценок (их σ) определяются величиной вариационного коэффициента (v) данного признака.

Для признака A : Для признака B : Для признака C : Для признака D : Для признака E :

V	d	V	d	V	d	V	d	V	d
27	+30	12	+30	17	+30	9	+30	30	+16
26	+28	11,5	+28	16	+27	8,5	+28	29	+14
25	+25	11	+25	15	+23	8	+25	28	+12
24	+23	10,5	+23	14	+20	7,5	+23	27	+10
23	+20	10	+20	13	+17	7	+20	26	+8
22	+18	9,5	+18	12	+13	6,5	+18	25	+6
21	+15	9	+15	11	+10	6	+15	24	+4
20	+13	8,5	+13	10	+7	5,5	+13	23	+2
19	+10	8	+10	9	+3	5	+10	22	0
18	+8	7,5	+8	8	0	4,5	+8	21	-2
17	+5	7	+5	7	-3	4	+5	20	-4
16	+3	6,5	+3	6	-7	3,5	+3	19	-6
15	0	6	0	5	-10	3	0	18	-8
14	-3	5,5	-3	4	-13	2,5	-3	17	-10
13	-5	5	-5	3	-17	2	-5	16	-12
12	-8	4,5	-8	2	-20	1,5	-8	15	-14
11	-10	4	-10	1	-23	1	-10	14	-16
10	-13	3,5	-13	0	-27	0,5	-13	13	-18
9	-15	3	-15			0	-15	12	-20
8	-18	2,5	-18					11	-22
7	-20	2	-20					10	-24
6	-23	1,5	-23					9	-26
5	-25	1	-25					8	-28
4	-28	0,5	-28					7	-30
3	-30	0	-30						

По этим стандартным данным можно составить пять децисигмальных оценочных таблиц, которые приведены на предыдущей странице (173).

Эти таблицы составлены здесь в том предположении, что признаки A, B, C и D могут варьировать от 0 и выше бесконечно (фактически приблизительно от $M-3\sigma$ до $M+3\sigma$), а признак E ограничен в своих колебаниях максимальным пределом $V=30$. Кроме того здесь предполагается, что варианты рядов A, C и E принимают только целые значения, а признаки B и D измеряются с точностью до 0,5 соответствующих единиц.

Имея в своем распоряжении такие таблицы (или соответствующие им оценочные графики), мы можем охарактеризовать состояние или достижение каждого отдельного индивидуума по всем пяти этим признакам. Пусть, например, данный индивидуум обнаружил при измерении (или испытании) следующие числовые результаты V :

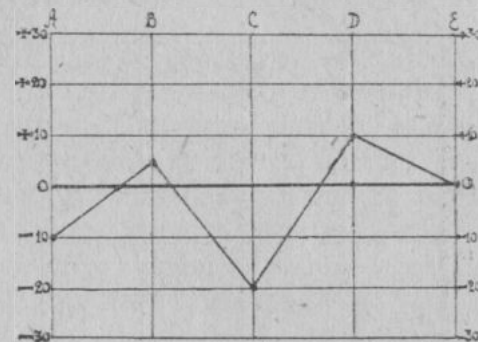
по признаку A :	$V=11$
" "	B : $V=7$
" "	C : $V=2$
" "	D : $V=5$
" "	E : $V=22$

Оценивая эти частные данные по соответствующим таблицам, мы непосредственно получаем их децисигмальные оценки d . Нетрудно убедиться в том, что:

для признака A варианта $V=11$	получает оценку $d=-10$
" " B " $V=7$	" " $d=+5$
" " C " $V=2$	" " $d=-20$
" " D " $V=5$	" " $d=+10$
" " E " $V=22$	" " $d=0$

Эти пять оценок d , относящихся к одному и тому же индивидууму, удобно, в целях наглядности, представить в графической форме. Для этого построим масштабную сетку, состоящую из ряда вертикальных и горизонтальных прямых, составляющих изображенный на черт. 91 прямоугольник. Здесь вертикальные линии соответствуют пяти подлежащим нашей оценке признакам: A, B, C, D и E , а цифрами по бокам сетки указан децисигмальный масштаб

(от $+30$ до -30). Отложив $d=-10$ по вертикали A , $d=+5$ по следующей справа вертикали B и т. д., мы наметим на сетке пять точек, соответствующих вычисленным выше значениям d . Соединив эти точки отрезками прямой, как это показано на черт. 91, мы и получим характеризующую данного индивидуума ломаную линию, называемую обычно „индивидуальным профилем“.



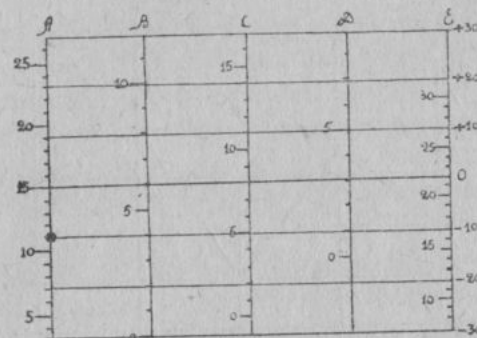
Черт. 91.

Положение и форма такого профиля с чрезвычайной наглядностью и вполне достаточной точностью позволяют судить об относительной величине всех оцениваемых признаков. В данном случае, например, весь профиль в целом лежит несколько ниже средней линии, соответствующей оценке $d=0$, что указывает на результаты испытания или измерения, в общем не достигающие относительной „нормы“ (малые размеры тела, плохие результаты психотехнического испытания и проч.). Кроме того данный профиль имеет ярко выраженную зигзагообразную форму (сильную разбросанность определяющих его точек), что свидетельствует о неоднородности в проявлениях этих пяти признаков (непропорциональное развитие отдельных частей тела или разнообразие результатов испытания по отдельным тестам).

В целях упрощения техники построения большего количества таких индивидуальных профилей можно на самой сетке или на отдельном масштабном „трафарете“ заранее нанести деления, соответствующие числовым результатам отдельных измерений, вследствие чего при построении профиля совершенно отпадает надобность в каких бы то ни было дополнительных оценочных таблицах или графиках.

На черт. 92 представлена та же самая сетка, которой мы уже пользовались выше, но на всех пяти ее вертикалях нанесены масштабные деления, соответствующие числовым результатам непосредственных измерений V .

Так как для признака A среднее арифметическое $M = 15$, то соответствующая этому значению черточка с цифрой 15 и поставлена здесь как раз по середине вертикали A , т. е. на уровне той горизонтальной прямой, для которой децисигмальная оценка



Черт. 92.

$d = 0$ (см. децисигмальный масштаб справа на полях сетки). Так как, далее, для этого признака $\sigma = \pm 4$, то в промежутке между двумя соседними горизонтальными прямыми (расстояние между которыми соответствует 10 децисигмам, т. е. одной целой сигме) должно уместиться 4 таких мелких масштабных отрезка, как это и показано на черт. 92 по вертикали A . Отложив такие же (равные друг другу) отрезки сверху и книзу от черточки с цифрой 15 и поставив у каждого 5-го деления соответствующую ему цифру, мы и получим масштаб абсолютных значений V для признака A .

Проверим правильность построения этого масштаба. Пусть, например, нам нужно пометить на этой вертикали A значение $V = 11$. Масштабное деление, соответствующее этой величине признака (см. жирную точку), оказалось здесь точно на уровне горизонтальной прямой, для которой $d = -10$. Но и по оценочной таблице для признака A значению $V = 11$ соответствует также $d = -10$. Следовательно наш масштаб верен.

При построении масштабных делений на второй вертикали вспомним, что для признака B его $\sigma = \pm 2$. Следовательно тот же самый сигмальный отрезок (т. е. постоянное расстояние между двумя соседними горизонтальными прямыми) мы должны теперь разделить на 2 равные части. Тогда в одной сигме и будет содержаться 2 единицы масштаба, как это требуется для признака B . Проставив, руководствуясь этим, все масштабные деления на вертикали B , вспомним, что в данном случае $M = 6$. Следовательно на

уровне средней горизонтальной линии сетки (для которой $d = 0$) должна оказаться черточка, соответствующая значению $V = 6$. Имея это в виду, мы и проставим у каждой пятой черточки соответствующие масштабные цифры, как это показано на черт. 92.

При построении масштаба по вертикали C мы должны разделить каждый сигмальный отрезок на 3 равные части (так как здесь $\sigma = \pm 3$), а цифры у черточек проставить с тем расчетом, чтобы к средней линии примыкала черточка, соответствующая значению 8 (так как в данном случае $M = 8$).

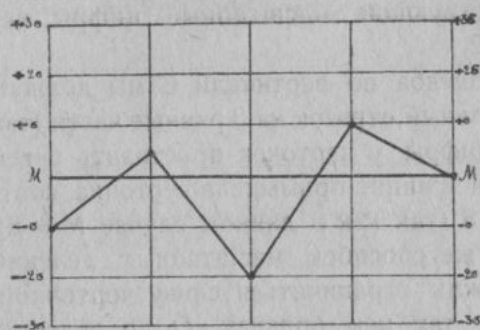
При построении тем же способом масштабных делений на вертикали D мы должны ограничиться снизу черточкой, соответствующей $V = 0$, так как признак D по условию варьирует лишь от 0 и выше (не принимая отрицательных значений). Для признака же E , ограниченного, наоборот, максимальным значением $V = 30$, масштабные деления не дойдут до верха сетки и окончатся на уровне последнего деления с цифрой 30 (см. черт. 92).

По вертикалям B и D можно было бы наметить также и половины целых значений V , но на нашей сетке эти половины делений не проставлены, чтобы не мешать наглядности чертежа.

Имея в своем распоряжении такую сетку с заранее нанесенными на ней масштабными делениями, мы без труда находим по ней отдельные черточки, соответствующие значениям оцениваемых вариантов V , и можем быть вполне уверены в том, что высота этих черточек будет совершенно точно соответствовать искомым децисигмальным оценкам d , если, конечно, все масштабные деления нанесены правильно. Все точки, которые придутся на средней горизонтальной линии сетки, будут соответствовать оценке $d = 0$, т. е. указывать на совпадение этих значений V со средними арифметическими M . Точки, располагающиеся на уровне следующей сверху горизонтальной линии, дадут оценку $d = +10$, что соответствует отклонению V от M ровно на одну σ в сторону превышения и т. д. На этом основании иногда взамен децисигмального масштаба справа на полях сетки помечаются условные обозначения:

$$+3\sigma \quad +2\sigma \quad +\sigma \quad M \quad -\sigma \quad -2\sigma \quad -3\sigma$$

указывающие на совпадение данной варианты с M или на ее отклонение от M на определенное количество сигм, как это и представлено на черт. 93, в котором средняя линия MM



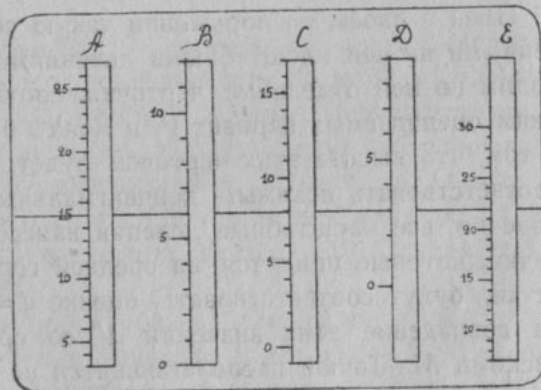
Черт. 93.

(соответствующая оценке $d=0$) сделана для ясности несколько толще всех остальных. Взамен нанесения масштабных делений на самой сетке, можно заготовить с этой целью особый оценочный трафарет, представляющий собой прямоугольный кусок картона с рядом вырезанных на нем узких вертикальных полосок (прорезей). Высота этих прорезей должна соответствовать высоте самой децисигмальной сетки, а промежутки между ними — расстояниям между отдельными ее вертикалями. По бокам каждой такой прорези и наносятся масштабные деления, отвечающие результатам непосредственного измерения соответствующих признаков V , а на среднем уровне (для

которого $d=0$), проводится через весь трафарет горизонтальная прямая (см. черт. 94).

При вычерчивании индивидуального профиля этот трафарет накладывается на масштабную сетку так, чтобы средняя горизонтальная линия трафа-

рета совпала с соответствующей средней горизонталью сетки (т. е. с линией MM , для которой $d=0$), а отдельные вертикали сетки точно пришлись в прорезях трафарета. Далее, руководствуясь масштабами, нанесенными на трафа-



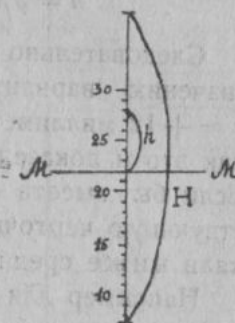
Черт. 94.

рете, следует проставить на сетке ряд точек через прорези трафарета и, сняв трафарет с сетки, соединить все эти точки отрезками прямых линий.

При вычерчивании антропометрических профилей такие трафареты удобно заготавливать для каждой возрастно-половой группы детей отдельно, а сетки делать однотипными для всех этих групп.

В приведенном выше примере нанесение на сетку или на трафарет масштабных делений не представляло особого труда, так как все стандарты (M и σ) были намеренно подобраны в виде целых чисел. В том же общем случае, когда это условие не соблюдено, отдельные черточки наших масштабов, очевидно, не придутся уже точно против горизонтальных прямых сетки, так как в одном сигмальном отрезке теперь будет содержаться некоторое дробное число таких масштабных делений.

Покажем общий способ разбивки отдельных вертикалей на мелкие масштабные деления, пояснив его сперва на разобранным уже выше примере. Пусть, в частности, требуется установить масштаб для крайней справа вертикали E по стандартам $M=22$ и $\sigma=\pm 5$ (см. черт. 95).



Черт. 95.

Обозначим полную высоту всей вертикали через H (пусть, напр., $H=60$ миллиметрам), высоту какой-либо отдельной черточки над средней линией MM через h , а величину оцениваемой варианты, которой эта черточка соответствует — через V (у нас, например, $V=28$).

При этих обозначениях искомая высота данной черточки (h) может быть вычислена по следующей формуле:

$$h = \left[\frac{H}{6\sigma} \right] V - \left[\frac{HM}{6\sigma} \right].$$

Здесь величины, поставленные в квадратные скобки [], являются постоянными для всех масштабных черточек данной вертикали, а V и h представляют собой переменные

величины (для отдельных черточек различные). В нашем примере при $M=22$, $\sigma=\pm 5$ и $H=60$:

$$\left[\frac{H}{6\sigma}\right] = \frac{60}{6 \cdot 5} = 2 \quad \left[\frac{HM}{6\sigma}\right] = \frac{60 \cdot 22}{6 \cdot 5} = 44.$$

Подставляя эти частные данные (2 и 44) в указанную выше формулу для вычисления h , получим:

$$h = 2V - 44.$$

По этой формуле и могут быть вычислены (в миллиметрах) высоты h всех черточек вертикали E , соответствующие отдельным значениям вариант V .

Так, например, для $V=28$:

$$h = 2V - 44 = 2 \cdot 28 - 44 = +12 \text{ мм.}$$

Следовательно масштабную черточку, соответствующую значению варианты $V=28$, нужно провести на высоте $h=+12$ миллиметров выше (знак $+$) средней линии MM , как это и показано в уменьшенном масштабе на черт. 95. Если бы высота h получилась со знаком $-$, то соответствующую черточку следовало бы отложить на этой вертикали ниже средней линии MM .

Например для $V=13$:

$$h = 2V - 44 = 2 \cdot 13 - 44 = -18 \text{ мм.},$$

т. е. для этой варианты (13) масштабная черточка будет лежать ниже средней линии MM на расстоянии от нее $=18$ миллиметров.

Выполнив аналогичные вычисления для всех масштабных черточек вертикали E , получим следующие высоты h , приведенные здесь рядом с соответствующими значениями вариант V (см. табл. на стр. 181). По этим данным чертежник и нанесет все масштабные деления на вертикаль сетки или на край прорези оценочного трафарета, обозначив эти деления соответствующими масштабными цифрами, например через каждую 5-ую черточку.

Чтобы не тратить лишнего времени на вычисление высоты h для всех масштабных делений данной вертикали по формуле:

$$h = \left[\frac{H}{6\sigma}\right] V - \left[\frac{HM}{6\sigma}\right]$$

можно ограничиться определением этих высот лишь для двух крайних черточек: самой верхней и самой нижней, а высоты всех остальных найти путем последовательного вычитания постоянной разницы Δ из наибольшей высоты h .

Эта постоянная разница между высотами двух соседних масштабных черточек вычисляется по формуле:

$$\Delta = \frac{kH}{6\sigma}$$

где через k по аналогии с классовым промежутком обозначена постоянная разница между двумя смежными значениями самих вариант V , для которых эти черточки помечаются. В нашем последнем примере отдельные черточки были проведены на вертикали E для всех целых значений V , т. е. с промежутками $k=1$ (если бы на масштабе отмечались также и половины V , то k равнялось бы 0,5, а при пометке черточек только, например, для четных значений V , мы имели бы $k=2$).

Итак, для нашего примера $k=1$, $H=60$, $\sigma=\pm 5$, откуда:

$$\Delta = \frac{kH}{6\sigma} = \frac{1 \cdot 60}{6 \cdot 5} = 2 \text{ мм.}$$

Следовательно высоты h отдельных черточек будут уменьшаться ровно на 2 мм, что мы и наблюдаем в действительности. Вычитая последовательно эту постоянную разницу (2 мм) из наибольшей высоты $h=+16$ мм, мы получим ряд постепенно уменьшающихся чисел, совпадающих с вычисленными ранее высотами h .

В разобранным примере оценочный трафарет был рассчитан на предельный размах колебаний отдельных вариант V в обе стороны от M на 3σ , а всего, стало быть, на 6 сигм. Обычно этот интервал несколько расширяют и принимают

V	h
30	+16
29	+14
28	+12
27	+10
26	+8
25	+6
24	+4
23	+2
22	0
21	-2
20	-4
19	-6
18	-8
17	-10
16	-12
15	-14
14	-16
13	-18
12	-20
11	-22
10	-24
9	-26
8	-28
7	-30

равным 8σ (от $M - 4\sigma$ до $M + 4\sigma$). В этом случае формулы для вычисления h и Δ будут иметь следующий вид:

$$h = \left[\frac{H}{8\sigma} \right] V - \left[\frac{HM}{8\sigma} \right] \text{ и } \Delta = \frac{kH}{8\sigma}.$$

Далее, практика показала, что полную высоту трафарета H удобно делать в 120 мм (по 15 мм на каждую сигму). Подставляя в указанную выше формулу $H = 120$, получим:

$$h = \left[\frac{15}{\sigma} \right] V - \left[\frac{15M}{\sigma} \right] \quad \Delta = \frac{15k}{\sigma}.$$

По этим формулам чаще всего и производится вычисление масштабных ординат (высот) оценочного трафарета.

Если с увеличением числовых значений признака V соответствующие оценки d наоборот уменьшаются (напр. при оценке числа неправильных ответов испытуемого), то при вычислении масштабных ординат (высот) h оценочного трафарета следует пользоваться уже другими формулами;

$$h = \left[\frac{15M}{\sigma} \right] - \left[\frac{15}{\sigma} \right] V \quad \Delta = -\frac{15k}{\sigma}.$$

Способ пользования этими формулами остается прежним.

XVIII. Пример вычисления масштабных ординат оценочного трафарета.

Пусть требуется заготовить оценочный трафарет для вычерчивания индивидуальных профилей, характеризующих, напр., следующие 5 антропометрических признаков:

1. Рост.
2. Окружность груди.
3. Емкость легких.
4. Окружность головы.
5. Вес.

Здесь рост, окружность груди и окружность головы измеряются в сантиметрах, емкость легких — в кубических сантиметрах, а вес — в килограммах.

Допустим, что для достаточно большой и вполне однородной группы детей, например для мальчиков от 13 до

14 лет, по каждому из этих пяти признаков были вычислены стандарты M и σ , причем оказалось, что:

для роста:	$M = 142,86 \text{ см}$	$\sigma = \pm 8,22 \text{ см}$
для окружности груди:	$M = 66,61 \text{ см}$	$\sigma = \pm 3,78 \text{ см}$
для емкости легких:	$M = 2,272 \text{ куб. см}$	$\sigma = \pm 460 \text{ куб. см}$
для окружн. головы:	$M = 52,68 \text{ см}$	$\sigma = \pm 1,45 \text{ см}$
для веса:	$M = 35,34 \text{ кг}$	$\sigma = \pm 5,77 \text{ кг}$

Начнем расчет масштабных ординат с той вертикали, которая будет соответствовать росту. Для этого прежде всего вычислим ожидаемые максимальные пределы варьирования роста в предположении, что эти пределы определяются размахом колебаний соответствующих вариантов V в обе стороны от M на 4σ . Так как для роста $M = 142,86$ и $\sigma = \pm 8,22$, то верхний предел $M + 4\sigma = 142,86 + 4 \cdot 8,22 = 175,74 \text{ см}$; нижний предел $M - 4\sigma = 142,86 - 4 \cdot 8,22 = 109,98 \text{ см}$.

Эти числа (175,74 и 109,98) теперь нужно округлить, но так, чтобы они не вышли за пределы $M \pm 4\sigma$, и отдельные масштабные черточки не оказались бы вследствие этого выше или ниже прорезей трафарета. Для этого верхний предел варьирования роста (175,74 см) при таком округлении следует уменьшить, а нижний предел (109,98 см) наоборот несколько увеличить. В результате получим два целых числа: 175 и 110, в границах которых и будут варьировать числовые результаты измерения роста отдельных детей этой группы.

Теперь посмотрим, сколько отдельных масштабных черточек поместится при этих условиях на краю пробези трафарета, соответствующей росту. Если проводить эти черточки для каждого отдельного сантиметра, то общее число таких масштабных делений будет равно разности:

$$175 - 110 = 65.$$

Так как общую высоту трафарета мы решили сделать в 120 мм (12 см), то на долю каждого из этих 65 масштабных делений придется приблизительно по:

$$\frac{120}{65} = 1,85 \text{ мм.}$$

Допустим, что такой промежуток между двумя соседними черточками мы признали слишком малым и, вследствие этого, пожелали бы сделать масштаб более крупным. Если провести по одной черточке для каждого двух сантиметров роста, то общее количество таких черточек уменьшится, очевидно, также вдвое, т. е. в промежутке от 175 до 110 их окажется уже 33 (вместо 65). Так как при этих условиях удобнее отмечать черточками четное число сантиметров (чтобы в этот масштаб попали круглые числа: 140, 150 и т. д.), то установленные выше пределы наибольшего колебания роста (от 175 до 110) придется еще сузить, уменьшив 175 до 174, а 110 оставив без изменения (так как это число и без того уже четное). Итак, окончательно, на масштаб первой прорези нашего трафарета попадут черточки, соответствующие значениям роста в пределах от 174 до 110 с промежутками этих значений через каждые 2 см (всего 33 такие черточки). Следовательно, в данном случае постоянная разность между двумя соседними, отмечаемыми по масштабу значениями V , т. е.

$$k = 2.$$

После выполнения этих предварительных расчетов вычислим выражения, стоящие в квадратных скобках в формуле для определения высот h :

$$h = \left[\frac{15}{\sigma} \right] V - \left[\frac{15M}{\sigma} \right],$$

а также точное значение постоянной разницы Δ между двумя смежными высотами h (т. е. расстояние в миллиметрах между двумя соседними масштабными черточками) по формуле:

$$\Delta = \frac{15k}{\sigma}.$$

В нашем примере для роста $M = 142,86$, $\sigma = \pm 8,22$ и $k = 2$. Следовательно:

$$\begin{aligned} \left[\frac{15}{\sigma} \right] &= \frac{15}{8,22} = 1,82 \\ \left[\frac{15M}{\sigma} \right] &= \frac{15 \cdot 142,86}{8,22} = 260,69 \\ \Delta &= \left[\frac{15k}{\sigma} \right] = \frac{15 \cdot 2}{8,22} = 3,65. \end{aligned}$$

Подставляя эти числа в формулу для определения высот h , получим:

$$h = 1,82 V - 260,69.$$

По этой формуле нам нужно теперь вычислить высоту двух масштабных черточек: самой верхней (для наибольшего роста $V = 174$) и самой нижней (для наименьшего $V = 110$).

Для $V = 174$ будем иметь:

$$h = 1,82 V - 260,69 = 1,82 \cdot 174 - 260,69 = +55,99,$$

а для $V = 110$ получим:

$$h = 1,82 V - 260,69 = 1,82 \cdot 110 - 260,69 = -60,49.$$

Итак, самый большой рост (174 см) должен быть помещен на нашем масштабе черточкой на высоте 55,99 мм над средней линией трафарета, а самый малый рост (110 см) — на высоте 60,49 мм под этой средней линией.

Здесь нижняя черточка (для которой высота $h = -60,49$) оказалась все же за пределами прорези, соответствующий нижний край которой при полной ее высоте в 120 мм должен отстоять от середины этой прорези ровно на 60 мм (а здесь это расстояние получилось по абсолютной своей величине больше на 0,49 мм). Эта явная (но небольшая) невязка легко объясняется здесь некоторой неточностью всех вычислений, ограниченных лишь двумя десятичными знаками после запятой.

Весь промежуток между двумя этими крайними черточками следует теперь разбить на 32 равные части, причем 33 отграничивающие их черточки будут отстоять друг от друга на постоянное расстояние:

$$\Delta = 3,65 \text{ мм.}$$

На эту же постоянную величину будут отличаться друг от друга и высоты h каждой пары двух соседних масштабных черточек. Для определения этих высот мы должны из высоты самой верхней черточки $h = +55,99$ вычесть $\Delta = 3,65$, затем из полученной разности $+52,34$ еще раз

вычесть ту же самую Δ и т. д., как это показано здесь на примере нескольких последовательных вычитаний:

$$\begin{array}{r} +55,99 \\ -3,65 \\ \hline +52,34 \\ -3,65 \\ \hline +48,69 \\ -3,65 \\ \hline +45,04 \text{ и т. д.} \end{array}$$

После последнего 32-го вычитания мы получим высоту самой нижней (33-ей по счету) черточки $h = -60,81$, которую полезно проверить по контрольной формуле:

$$h_b = h_a - N\Delta,$$

где через h_b обозначена искомая высота h самой нижней масштабной черточки, через h_a — высота самой верхней черточки, а через N — общее количество выполненных нами вычитаний. В данном случае $h_a = +55,99$, а $N = 32$. Следовательно:

$$h_b = +55,99 - 32 \cdot 3,65 = -60,81,$$

что мы и получили в действительности после 32-го вычитания *).

Найденная этим способом высота нижней черточки ($-60,81$) несколько не совпала с вычисленной ранее ее высотой $h = -60,49$. Эта разница в $0,32$ мм легко объясняется здесь той неизбежной неточностью, которая содержится в самой разности Δ (имеющей лишь 2 верных знака после запятой). Вычитая разность Δ 32 раза подряд, мы все время накапливали эту ошибку в одном и том же направлении, в результате чего и получилась такая разница в $0,32$ мм. Практически эта невязка конечно не может иметь существенного значения.

Выполнив все эти вычисления, проверив их по контрольной формуле и оценив величину накопившейся неточности

*) В целях проверки этих вычислений можно наряду с вычислением высот h для двух крайних значений V определить по обычной формуле также и несколько промежуточных их значений.

при определении высоты h последней (самой нижней) масштабной черточки, запишем все эти данные друг под другом рядом с теми значениями оцениваемых вариантов V , которым эти черточки соответствуют по сигмальному масштабу.

В результате получим представленную здесь табличку, руководствуясь которой чертежник и нанесет все 33 масштабные черточки на край соответствующей прорези оценочного трафарета.

При вычислении масштабных ординат для второй прорези определим наибольший и наименьший размер окружности груди:

$$M + 4\sigma = 66,61 + 4 \cdot 3,78 = 81,73 \text{ см}$$

$$M - 4\sigma = 66,61 - 4 \cdot 3,78 = 51,49 \text{ см.}$$

После округления этих чисел и их вычитания друг из друга получим:

$$81 - 52 = 29.$$

Если отмечать масштабной черточкой каждый отдельный сантиметр, то на всей вертикали поместится 30 таких черточек, ограничивающих 29 масштабных делений. Этот масштаб не будет очень мелким.

Вычислим выражения, стоящие в квадратных скобках (см. формулу для определения высот h):

$$\left[\frac{15}{\sigma} \right] = \frac{15}{3,78} = 3,97$$

$$\left[\frac{15M}{\sigma} \right] = \frac{15 \cdot 66,61}{3,78} = 264,33.$$

Следовательно, высоты h отдельных масштабных черточек могут быть вычислены по формуле:

$$h = 3,97 V - 264,33.$$

V	h
174	+ 55,99
172	+ 52,34
170	+ 48,69
168	+ 45,04
166	+ 41,39
164	+ 37,74
162	+ 34,09
160	+ 30,44
158	+ 26,79
156	+ 23,14
154	+ 19,49
152	+ 15,84
150	+ 12,19
148	+ 8,54
146	+ 4,89
144	+ 1,24
142	— 2,41
140	— 6,06
138	— 9,71
136	— 13,36
134	— 17,01
132	— 20,66
130	— 24,31
128	— 27,96
126	— 31,61
124	— 35,26
122	— 38,91
120	— 42,56
118	— 46,21
116	— 49,86
114	— 53,51
112	— 57,16
110	— 60,81

Для наибольшего значения $V=81$:

$$h = 3,97 \cdot 81 - 264,33 = +57,24,$$

а для наименьшего $V=52$:

$$h = 3,97 \cdot 52 - 264,32 = -57,89.$$

Постоянная разница между этими высотами при $k=1$ (так как черточки мы решили ставить для значений V , отличающихся друг от друга ровно на 1 см):

$$\Delta = \frac{15k}{\sigma} = \frac{15 \cdot 1}{3,78} = 3,97.$$

Вычитая эту Δ из наибольшей высоты $h = +57,24$, получим:

$$+57,24 - 3,97 = +53,27.$$

Вычитая Δ из только что полученной разности $+53,27$, будем иметь:

$$+53,27 - 3,97 = +49,30.$$

Продолжив эту операцию до конца, после 29-го вычитания получим высоту самой нижней масштабной черточки (для $V=52$):

$$h = -57,89 \text{ (см. таблицу в конце главы).}$$

Но и по контрольной формуле:

$$h_b = h_a - N\Delta$$

при $N=29$ будем иметь:

$$h_b = -57,89.$$

Следовательно все вычисления сделаны правильно.

В данном случае ордината h самой нижней черточки ($-57,89$), вычисленная по основной формуле, совершенно точно совпала с результатом ее определения путем ряда последовательных вычитаний.

При вычислении масштабных ординат для 3-ей прорези оценочного трафарета (для емкости легких) удобно принять $k=100$. В этом случае варианты V будут изменяться в пределах от 4.000 до 500. Если производить вычисления с двумя знаками после запятой, то, в виду необходимости умножать

числовой коэффициент при V на трех- и четырехзначные числа, конечная ошибка будет накапливаться очень сильно и может сказаться даже на целых числах (у высоты h). Поэтому для емкости легких мы будем здесь вычислять всюду 4 десятичных знака после запятой.

$$\left[\frac{15}{\sigma} \right] = \frac{15}{460} = 0,0326$$

$$\left[\frac{15M}{\sigma} \right] = \frac{15,2272}{460} = 74,0870$$

$$\left[\frac{15k}{\sigma} \right] = \frac{15 \cdot 100}{460} = 3,2609.$$

Следовательно:

$$h = 0,0326 V - 74,0870$$

$$\text{и } \Delta = 3,2609.$$

Для $V=4.000$:

$$h = 0,0326 \cdot 4.000 - 74,0870 = +56,3130.$$

для $V=500$:

$$h = 0,0326 \cdot 500 - 74,0870 = -57,7870.$$

После 35-го вычитания $\Delta = 3,2609$ для самой нижней черточки (соответствующей $V=500$) получим:

$$h = -57,8185,$$

что довольно близко совпадает с только что вычисленным по основной формуле этим же значением:

$$h = -57,7870.$$

Для четвертого признака (окружности головы) при размахе колебаний отдельных вариантов V в пределах от 58 до 47 удобно взять $k=0,5$. В этом случае будем иметь следующую конечную формулу:

$$h = 10,34 V - 544,97$$

при $\Delta = 5,17$.

Для вычисления масштабных ординат h здесь придется сделать 22 вычитания, причем конечный результат сойдется совершенно точно.

При вычислении масштабных ординат V для 5-го признака (веса) возьмем $k=2$, а пределы его варьирования от 58 до 14. Получим формулу:

$$h = 2,60 V - 91,87$$

при $\Delta = 5,20$.

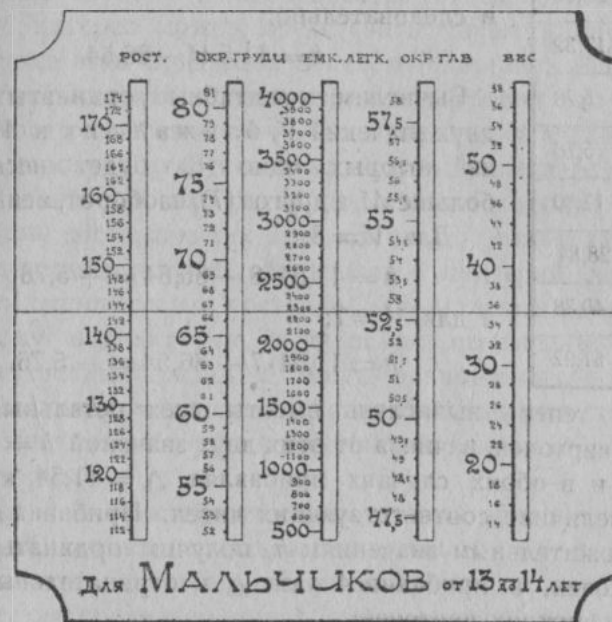
После 22-го вычитания получим результат без невязки.

Все эти конечные данные полезно записать в особую таблицу, по которой чертежник и будет наносить масштабы на края оценочного трафарета. Таблица эта в данном случае будет иметь следующий вид:

Рост		Окр. груди		Емкость легких		Окр. головы		Вес	
V	d	V	d	V	d	V	d	V	d
174	+ 55,99	81	+ 57,24	4.000	+ 56,3130	58	+ 54,75	58	+ 58,93
172	+ 52,34	80	+ 53,27	3.900	+ 53,0521	57,5	+ 49,58	56	+ 53,73
170	+ 48,69	79	+ 49,30	3.800	+ 49,7912	57	+ 44,41	54	+ 48,53
168	+ 45,04	78	+ 45,33	3.700	+ 46,5303	56,5	+ 39,24	52	+ 43,33
166	+ 41,39	77	+ 41,36	3.600	+ 43,2694	56	+ 34,07	50	+ 38,13
164	+ 37,74	76	+ 37,39	3.500	+ 40,0085	55,5	+ 28,90	48	+ 32,93
162	+ 34,09	75	+ 33,42	3.400	+ 36,7476	55	+ 23,73	46	+ 27,73
160	+ 30,44	74	+ 29,45	3.300	+ 33,4867	54,5	+ 18,56	44	+ 22,53
158	+ 26,79	73	+ 25,48	3.200	+ 30,2258	54	+ 13,39	42	+ 17,33
156	+ 23,14	72	+ 21,51	3.100	+ 26,9649	53,5	+ 8,22	40	+ 12,13
154	+ 19,49	71	+ 17,54	3.000	+ 23,7040	53	+ 3,05	38	+ 6,93
152	+ 15,84	70	+ 13,57	2.900	+ 20,4431	52,5	- 2,12	36	+ 1,73
150	+ 12,19	69	+ 9,60	2.800	+ 17,1822	52	- 7,29	34	- 3,47
148	+ 8,54	68	+ 5,63	2.700	+ 13,9213	51,5	- 12,46	32	- 8,67
146	+ 4,89	67	+ 1,66	2.600	+ 10,6604	51	- 17,63	30	- 13,87
144	+ 1,24	66	- 2,31	2.500	+ 7,3995	50,5	- 22,80	28	- 19,07
142	- 2,41	65	- 6,28	2.400	+ 4,1386	50	- 27,97	26	- 24,27
140	- 6,06	64	- 10,25	2.300	+ 0,8787	49,5	- 33,14	24	- 29,47
138	- 9,71	63	- 14,22	2.200	- 2,3832	49	- 38,31	22	- 34,67
136	- 13,36	62	- 18,19	2.100	- 5,6441	48,5	- 43,48	20	- 39,87
134	- 17,01	61	- 22,16	2.000	- 8,9050	48	- 48,65	18	- 45,07
132	- 20,66	60	- 26,13	1.900	- 12,1659	47,5	- 53,82	16	- 50,27
130	- 24,31	59	- 30,10	1.800	- 15,4268	47	- 58,99	14	- 55,47
128	- 27,96	58	- 34,07	1.700	- 18,6877				
126	- 31,61	57	- 38,04	1.600	- 21,9486				
124	- 35,26	56	- 42,01	1.500	- 25,2095				
122	- 38,91	55	- 45,98	1.400	- 28,4704				
120	- 42,56	54	- 49,95	1.300	- 31,7313				
118	- 46,21	53	- 53,92	1.200	- 34,9922				
116	- 49,86	52	- 57,89	1.100	- 38,2531				
114	- 53,51			1.000	- 41,5140				
112	- 57,16			900	- 44,7749				
110	- 60,81			800	- 48,0358				
				700	- 51,2967				
				600	- 54,5596				
				500	- 57,8185				

Здесь жирным шрифтом обозначены те цифры, которые чертежник должен проставить на масштабе крупно против несколько удлинненных черточек. В рукописных таблицах эти цифры можно или подчеркивать или писать красными чернилами.

Составленный по этой таблице оценочный трафарет в уменьшенном виде представлен на чертеже 96.



Черт. 96.

При всех описанных выше расчетах мы накапливали неизбежную ошибку, которая постепенно возрастала в направлении сверху вниз. Таким образом наиболее неточными здесь являются масштабные ординаты самых нижних черточек.

Однако во многих случаях такое скопление ошибок только в нижней части трафарета представляет некоторое неудобство (например при оценке измерений или испытаний, произведенных в группе отстающих детей).

Чтобы избежать этого неудобства, можно несколько изменить самый порядок вычислений. Пусть например при $M = 7,5$ и $\sigma = \pm 1,3$ мы имеем размах колебаний отдельных вариантов

в пределах от 12 до 3 через каждую единицу ($k=1$).

Здесь:

V	d
12	+ 51,94
11	+ 40,40
10	+ 28,86
9	+ 17,32
8	+ 5,78
7	- 5,76
6	- 17,30
5	- 28,84
4	- 40,38
3	- 51,92

$$\left[\frac{15}{\sigma}\right] = \frac{15}{1,3} = 11,54$$

$$\left[\frac{15M}{\sigma}\right] = \frac{15 \cdot 7,5}{1,3} = 86,54$$

$$\left[\frac{15k}{\sigma}\right] = \frac{15 \cdot 1}{1,3} = 11,54$$

и следовательно:

$$h = 11,5V - 86,54.$$

Вычислим масштабные ординаты h для двух значений V , ближайших к $M=7,5$, из которых одно (8) будет несколько больше M , а другое (7), наоборот, меньше M .

Для $V=8$:

$$h = 11,54 \cdot 8 - 86,54 = +5,78,$$

а для $V=7$:

$$h = 11,54 \cdot 7 - 86,54 = -5,76.$$

Будем теперь вычислять высоты всех остальных масштабных черточек, начиная от этих двух значений $h = +5,78$ и $-5,76$ и в обоих случаях прибавляя $\Delta = 11,54$ к абсолютной величине соответствующих чисел. Прибавив 4 раза Δ к положительным значениям h , получим ординаты верхних черточек, а прибавив 4 раза Δ к отрицательным h — ординаты нижних черточек:

для $V=8$	+ 5,78 11,54	для $V=7$	- 5,76 11,54
для $V=9$	+ 17,32 11,54	для $V=6$	- 17,30 11,54
для $V=10$	+ 28,86 11,54	для $V=5$	- 28,84 11,54
для $V=11$	+ 40,40 11,54	для $V=4$	- 40,38 11,54
для $V=12$	+ 51,94	для $V=3$	- 51,92

Вычислив ординаты h этим способом, мы получим симметричное распределение ошибок кверху и книзу от M , причем конечная неточность будет здесь очевидно меньше, чем при скоплении всех ошибок только вниз.

XIX. Комплексная оценка.

В предыдущих главах было дано краткое описание основных методов децисигмальной и перцентильной оценки отдельных (изолированных) признаков (например только роста или веса отдельных детей, результатов испытания по каждому тесту порознь и т. д.). Однако в процессе изучения индивидуальных свойств и особенностей отдельной личности не малый интерес может представить также и оценка всего комплекса этих признаков в целом (например общего физического развития ребенка, среднего уровня его интеллектуального развития и проч.). Количественная характеристика такой совокупности связанных в единый комплекс отдельных наблюдений или измерений при умелом подборе всех компонентов, составляющих этот комплекс, может дать в руки исследователя вполне объективный критерий для оценки сложных процессов и состояний, свойственных каждому отдельному индивидууму, и на основе выявленных при этом количественных градаций сделать возможным сопоставление этих индивидуумов как друг с другом, так и со стандартной их массой в целях установления той или иной относительной „нормы“.

Одним из простейших случаев такой объединенной (комплексной) оценки нескольких признаков может служить уже неоднократно приводимый нами пример подсчета общего количества правильно выполненных заданий при каком-либо психотехническом испытании. Каждое такое задание в сущности представляет собою особую операцию, выявляющую определенные особенности испытуемого. Если мы предполагаем эти задания вполне однородными и равноценными по трудности выполнения, то делаем это лишь для технического упрощения дальнейших вычислений, которые благодаря такому допущению и сводятся здесь к простому подсчету общего числа правильных решений. В данном случае, следовательно, весь комплекс оцениваемых признаков характеризуется суммой изолированных оценок, каждая из которых представляет собой альтернативу: решил (1) или не решил (0).

Если все испытуемые были поставлены в совершенно одинаковые условия, и им предлагалось одно и то же количество заданий, то такие суммарные оценки окажутся

вполне сравнимыми друг с другом. При различном же числе заданий эти суммы необходимо или выразить в процентах от общего количества всех предложенных заданий (т. е. привести их к 100) или же разделить на сумму всех заданий (т. е. привести к единице).

Например, если в одном случае испытуемый решил 5 задач из 10, а в другом 8 задач из 20, то при условии одинаковой трудности всех этих задач в первом случае мы имеем в среднем $\frac{5}{10} = 0,5$ правильных решений, а во втором только $\frac{8}{20} = 0,4$ решений.

Этот способ комплексной оценки в отличие от предыдущего (суммарной оценки) мы будем называть в дальнейшем средней оценкой.

Но и суммарная и средняя оценки комплекса признаков при сколько-нибудь заметном различии в относительной трудности отдельных операций могут дать в конечном итоге совершенно неверные результаты, если при таком объединении изолированных характеристик мы не учтем также и степень трудности каждого отдельного задания.

Соответствующая поправка сводится здесь к введению так называемых „коэффициентов трудности“, зависящих от числа правильных или неправильных решений данной задачи всей массой испытуемых в целом.

Пусть, например, испытуемым предлагалось всего лишь 2 задачи заведомо неодинаковой трудности. Первую из них решили, скажем, 80 человек из 100 (легкая задача), а вторую — только 40 человек из 100 (трудная задача). Предположим, что мы хотим оценить результат испытания двух людей, из которых один решил только первую задачу, а другой — только вторую.

Если учитывать лишь количество правильных решений, то оба эти испытуемые получат совершенно одинаковую оценку: оба решили по одной задаче из 2-х им предложенных. Но этот результат, конечно, явно ошибочен. Первый испытуемый решил легкую задачу, а второй — трудную, поэтому результаты их испытания не могут считаться равнозначными.

Введем сюда коэффициенты трудности, которые для простоты сделаем пропорциональными количеству непра-

вильных решений, обнаруженных у всех 100 человек, подвергавшихся этому испытанию.

Первую задачу решили 80 человек из 100. Следовательно не решили ее 20 человек. Во втором случае имеем, наоборот, 40 правильных решений и стало быть 60 неправильных.

Приняв эти числа (20 и 60) за коэффициенты трудности, мы и можем охарактеризовать ими результат всего испытания в целом. Следовательно первый испытуемый, решивший легкую задачу, получит условную оценку 20, а второй, решивший задачу трудную, будет оценен коэффициентом 60. Если бы какой-либо испытуемый решил обе задачи, то оценка его равнялась бы $20 + 60 = 80$ (максимум), при отсутствии же решений и оценка его $= 0$ (минимум).

Таким образом при введении этих коэффициентов, взамен простого счета единиц здесь складываются соответствующие коэффициенты трудности, а комплексная оценка выражается суммой таких коэффициентов.

При желании выразить конечную оценку не в виде суммы, а соответствующей средней величиной, эти суммы коэффициентов трудности, получаемые в каждом отдельном случае, необходимо еще разделить на полную сумму всех возможных коэффициентов (в нашем примере следовательно на $20 + 60 = 80$). Тогда испытуемый, решивший легкую задачу, получит среднюю оценку:

$$\frac{20}{80} = 0,25,$$

а испытуемый, давший одно решение трудной задачи:

$$\frac{60}{80} = 0,75.$$

При решении всех задач оценка будет $0,25 + 0,75 = 1$, а при отсутствии решений — нуль (0).

Чтобы не усложнять вычисления необходимостью в конечном итоге делить каждую сумму коэффициентов трудности на максимальное значение этой же суммы (при определении средней оценки), иногда такому делению подвергают уже самые коэффициенты трудности. В нашем примере, следовательно, за коэффициенты трудности мы можем принять не 20 и 60, а уменьшенные в 80 раз числа: 0,25 и 0,75,

и тогда сумму этих коэффициентов нам не надо будет делить на ее максимальное значение (так как в этом случае $0,25 + 0,75 = 1$).

В тех случаях, когда отдельные объединенные оценки вычисляются децисигмальным методом, надобность в каких бы то ни было поправочных коэффициентах очевидно совершенно отсутствует, так как при децисигмальной оценке относительная трудность испытания учитывается уже автоматически (вернее в данном случае сами децисигмальные оценки d являются уже своего рода коэффициентами трудности).

Пусть, например, группе испытуемых в 200 человек были предложены 4 задачи, причем:

1-ую задачу	решили	160 человек
2-ую	"	20 "
3-тью	"	160 "
4-ую	"	100 "

Обозначим общее количество правильно решенных задач через p , а общее количество всех испытуемых — через n . Тогда в нашем примере $n = 200$, а число верных решений:

для 1-ой задачи	$p = 160$
" 2	" $p = 20$
" 3	" $p = 160$
" 4	" $p = 100$

Вычислим для сравнения коэффициент трудности t по ранее указанному способу, т. е. пропорционально количеству неправильно решенных задач ($n - p$):

для 1-ой задачи	это число	$n - p = 200 - 160 = 40$
" 2	"	" $n - p = 200 - 20 = 180$
" 3	"	" $n - p = 200 - 160 = 40$
" 4	"	" $n - p = 200 - 100 = 100$

Сумма всех неправильных решений, т. е. $40 + 180 + 40 + 100 = 360$.

Разделив число ошибочных решений ($n - p$) каждой отдельной задачи на эту общую сумму всех неправильных

решений $\Sigma(n - p) = 360$, мы и получим коэффициенты трудности (t) для отдельных задач по формуле:

$$t = \frac{n - p}{\Sigma(n - p)}.$$

Для 1-ой задачи:

$$t_1 = \frac{40}{360} = 0,11.$$

Для 2-ой задачи:

$$t_2 = \frac{180}{360} = 0,5.$$

Для 3-ей задачи:

$$t_3 = \frac{40}{360} = 0,11.$$

Для 4-ой задачи:

$$t_4 = \frac{100}{360} = 0,28.$$

Таким образом испытуемый, решивший например 1-ую и 4-ую задачу, получит оценку:

$$0,11 + 0,28 = 0,39.$$

Испытуемый, решивший последние три задачи — оценку:

$$0,50 + 0,11 + 0,28 = 0,89$$

и т. д.

Заменим теперь эти коэффициенты трудности t соответствующими децисигмальными оценками d .

Теоретический расчет, подробностей которого мы, однако, касаться здесь не можем, дает в конечном результате следующую формулу для вычисления децисигмальной оценки d каждого правильного решения одной задачи, если из общего числа n испытуемых ее решили p человек:

$$d = \pm 10 \sqrt{\frac{n - p}{p}}.$$

Децисигмальная же оценка каждого ошибочного решения этой задачи может быть получена по формуле:

$$d = -10 \sqrt{\frac{p}{n - p}}.$$

Будем оценивать конечный результат испытания, учитывая правильные решения (можно было бы конечно поступить

и обратно, но тогда малые значения оценочных чисел указывали бы наоборот на хорошие результаты испытания).

Для 1-ой задачи при $n=200$ и $p=160$ децисигмальная оценка правильного решения:

$$d_1 = +10 \sqrt{\frac{200-160}{160}} = +5.$$

Для 2-ой задачи при $p=20$:

$$d_2 = +10 \sqrt{\frac{200-20}{20}} = +30.$$

Для 3-ей задачи при $p=160$:

$$d_3 = +10 \sqrt{\frac{200-160}{160}} = +5.$$

Для 4-ой задачи при $p=100$:

$$d_4 = +10 \sqrt{\frac{200-100}{100}} = +10.$$

Децисигмальная же оценка неправильных решений тех же самых задач, вычисленная по формуле:

$$d = -10 \sqrt{\frac{p}{n-p}},$$

для 1-ой задачи при $p=160$:

$$d_1 = -10 \sqrt{\frac{160}{200-160}} = -20.$$

для 2-ой задачи при $p=20$:

$$d_2 = -10 \sqrt{\frac{20}{200-20}} = -3,33.$$

для 3-ей задачи при $p=160$:

$$d_3 = -10 \sqrt{\frac{160}{200-160}} = -20.$$

для 4-ой задачи при $p=100$:

$$d_4 = -10 \sqrt{\frac{100}{200-100}} = -10.$$

Эти данные мы можем облечь в форму следующей оценочной таблицы:

	Децисигмальная оценка	
	правильного решения:	неправильного решения:
1-ая задача	+ 5	- 20
2-ая задача	+ 30	- 3,33
3-ья задача	+ 5	- 20
4-ая задача	+ 10	- 10

Если отдельный испытуемый правильно решил лишь 1-ую и 2-ую задачу, то для получения его суммарной оценки по всей серии этих испытаний следует сложить обе децисигмальные оценки правильных решений этих задач (т. е. +5 и +30) и присоединить к этой сумме отрицательные децисигмальные оценки всех остальных (нерешенных) задач (т. е. -20 и -10). В результате получим сумму:

$$+5 + 30 - 20 - 10 = +5.$$

Это число (+5) и определяет здесь сумму всех 4-х децисигмальных оценок, характеризующих результаты данного испытания по отдельным заданиям.

Если такая комплексная оценка производится с целью сравнения отдельных испытуемых, поставленных в различные условия (например при неодинаковом общем количестве предложенных им задач), то конечный результат каждого индивидуального испытания удобнее выразить не суммарной, а средней комплексной его оценкой, т. е. разделить суммарную оценку на число всех заданий (в данном случае на 4).

В применении к последнему примеру такая средняя оценка будет следовательно равна:

$$\frac{+5}{4} = +1,25 \text{ децисигм.}$$

Чтобы не усложнять работу этими делениями на 4 каждой суммарной оценки порознь, можно выполнить все деления

уже в самой оценочной таблице. Разделив в ней каждую оценку d на 4, получим:

	Децисигмальная оценка	
	правильного решения:	неправильного решения:
1-ая задача	+ 1,25	— 5,00
2-ая задача	+ 7,50	— 0,83
3-ья задача	+ 1,25	— 5,00
4-ая задача	+ 2,50	— 2,50

Оценивая по этой таблице того же самого испытуемого (решившего 1-ую и 2-ую задачи), получим в сумме:

$$+ 1,25 + 7,50 - 5,00 - 2,50 = + 1,25,$$

т. е. совершенно тот же результат, который мы имели и раньше.

Тот же принцип объединения изолированных характеристик применим, очевидно, и к случаю вывода суммарной или средней оценки для комплекса количественных признаков. Пусть, например, группа в n испытуемых была обследована по N количественным признакам, причем конечные результаты этого обследования выразились в следующих абсолютных числах:

Испытуемые:	Испытания:				
	I	II	III	IV	V
№ 1	26	16	34	19	40
№ 2	22	18	42	7	25
№ 3	20	15	30	10	20
№ 4	18	14	18	13	30
№ 5	14	12	26	1	10

Здесь для простоты и наглядности взято всего только 5 испытуемых и 5 испытаний, что соответствует значениям:

$$n=5 \text{ и } N=5.$$

В действительности, конечно, общее число испытуемых бывает значительно большим.

Для замены выписанных в этой таблице абсолютных результатов испытания V их децисигмальными оценками d вычислим для каждого испытания M и σ , причем при определении σ воспользуемся формулой без поправки на малое число измерений (n):

$$\sigma = \pm \sqrt{\frac{\sum x^2}{n}},$$

так как n , вообще говоря, должно быть велико, а у нас здесь n взято = 5 лишь для простоты вычислений.

В результате получим:

для I-го испытания	$M=20$	$\sigma = \pm 4$
" II "	$M=15$	$\sigma = \pm 2$
" III "	$M=30$	$\sigma = \pm 8$
" IV "	$M=10$	$\sigma = \pm 6$
" V "	$M=25$	$\sigma = \pm 10$

Вычисляя оценки d для каждого результата испытания по формуле:

$$d = 10 \frac{V - M}{\sigma}$$

или по заготовленным для этой цели специальным таблицам или графикам, получим следующие децисигмальные оценки, занесенные в оценочную ведомость рядом с заданными числами V (см. табл. на след. стр. 202).

Сложив в каждой строчке все 5 децисигмальных оценок d с учетом их знаков (+ или —), получим для каждого испытуемого его суммарную оценку (S), а разделив ее на общее число всех испытаний N — соответствующую среднюю оценку (H), как это и сделано в двух последних столбцах нашей оценочной ведомости *). Снизу здесь указаны M и σ для каждого отдельного испытания.

*) Точно также следует поступать и при выводе суммарных (S) и средних (H) комплексных оценок при перцентиле. Полезно лишь помнить, что средние перцентиле оценки (H) будут варьировать уже не в обычных пределах (от 0 до 100), а в несколько суженном интервале.

Испытуемые	Испытания:										Оценка	
	I		II		III		IV		V		суммарная S	средняя H
	V	d	V	d	V	d	V	d	V	d		
№ 1	26	+15	16	+5	34	+5	19	+15	40	+15	+55	+11
№ 2	22	+5	18	+15	42	+15	7	-5	25	0	+30	+6
№ 3	20	0	15	0	30	0	10	0	20	-5	-5	-1
№ 4	18	-5	14	-5	18	-15	13	+5	30	+5	-15	-3
№ 5	14	-15	12	-15	26	-5	1	-15	10	-15	-65	-13
$M=20$ $\sigma=\pm 4$		$M=15$ $\sigma=\pm 2$		$M=30$ $\sigma=\pm 8$		$M=10$ $\sigma=\pm 6$		$M=25$ $\sigma=\pm 10$				

Значение средних комплексных оценок вполне ясно. Каждая из них определяет то количество децисигм, на которое в среднем отклоняются результаты всех 5 испытаний каждого отдельного испытуемого от общего уровня успехов, проявленных всей группой объектов исследования вместе взятых. Суммарная же оценка имеет, очевидно, лишь условное значение. Там, где она оказалась больше, и общий результат всего комплекса испытаний был лучше и наоборот. Но, чтобы сравнить эти данные с какими-либо другими результатами испытания, необходимо или обратить суммарные оценки в средние (путем деления на соответствующее число испытаний N) или же охарактеризовать их в тех или иных оценочных зонах по способу, описанному в предыдущих главах.

Вычислив среднее арифметическое (M_0) и среднее квадратическое отклонение (σ_0) для всего ряда суммарных оценок:

$$+55 \quad +30 \quad -5 \quad -15 \quad \text{и} \quad -65$$

по формуле без поправки на малое число измерений (так как n должно быть велико), получим:

$$M_0 = 0 \quad \text{и} \quad \sigma_0 = \pm 40,99$$

Разбив этот ряд на 5 равномерных зон, получим следующую оценочную таблицу *):

1	2	3	4	5
$M_0 - 2\sigma_0$	$M_0 - \frac{2}{3}\sigma_0$	$M_0 + \frac{2}{3}\sigma_0$	$M_0 + 2\sigma_0$	
-81,98	-27,33	+27,33	+81,98	

Таким образом суммарная оценка:

+55	попадает	в зону балла	4
+30	"	"	4
-5	"	"	3
-15	"	"	3
-65	"	"	2.

Этими условными баллами и может быть оценен комплекс всех 5 испытаний каждого отдельного индивидуума в случае необходимости сопоставлять в дальнейшем эти характеристики с результатами других испытаний.

Указанный прием вычисления комплексных оценок (суммарных и средних) безоговорочно применим лишь в том частном случае, когда отдельные признаки в их отношении ко всему оцениваемому комплексу в целом являются вполне равнозначными. Если же эти объединяемые в общей оценке отдельные компоненты комплекса оказываются неравноценными друг другу, необходимо их уравнивать искусственно путем введения особых „коэффициентов значимости“ q .

Пусть, например, комплекс некоторого числа признаков был подобран из таких компонентов, которые определяют различные стороны и проявления интеллектуального развития исследуемых. Вполне понятно, что при этих условиях мыслительные функции будут иметь первостепенное значение, в то время как, скажем, темп механической работы очевидно второстепенное. Такие же функции, как сосредоточение, займут в этом комплексе признаков некоторое промежуточное место.

*) Если бы мы имели здесь не децисигмальные, а перцентильные оценки, то эту разбивку следовало бы выполнить по принципу „равночисленных“ зон (см. стр. 126).

Чтобы уравнивать все эти неодинаковые по своему смыслу и значению характеристики умственного развития, необходимо перед суммацией их децисигмальных оценок d принять во внимание каждую такую оценку либо в усиленной, либо в ослабленной степени в зависимости от того значения, которое придается данной оцениваемой функции. Такое искусственное неодинаковое увеличение отдельных слагаемых удобнее всего выполнить путем умножения их на соответствующие коэффициенты q , установленные в зависимости от относительной значимости этих изолированных оценок. В нашем случае мы можем с этой целью умножать например оценку мыслительных операций на 2, оценку сосредоточения — на 1 (т. е. оставлять без изменения), а оценку темпа механической работы — на 0 (т. е. вовсе не принимать ее во внимание при выводе суммарной оценки). Сложив для каждого испытуемого эти произведения отдельных оценок d на соответствующие коэффициенты значимости q (2, 1 или 0), мы и получим „взвешенную“ суммарную оценку S , характеризующую интеллектуальное развитие каждого отдельного испытуемого по комплексу всех учтенных и уравненных признаков.

При выводе же средней комплексной оценки мы должны в этом случае разделить суммарную оценку S уже не на число признаков N , а на сумму всех коэффициентов значимости Σq .

Пусть в разобранным выше числовом примере (см. оценочную ведомость на стр. 202) мы имели бы следующие коэффициенты значимости q :

для I-го признака	$q_1 = 3$
„ II „	$q_2 = 1$
„ III „	$q_3 = 0$
„ IV „	$q_4 = 4$
„ V „	$q_5 = 2$

Возьмем для вывода комплексной характеристики испытуемого № 2, у которого изолированные оценки оказались:

по I признаку	$d_1 = + 5$
„ II „	$d_2 = + 15$
„ III „	$d_3 = + 15$
„ IV „	$d_4 = - 5$
„ V „	$d_5 = 0$

Перемножив эти оценки d на соответствующие коэффициенты значимости q , получим:

$$\begin{aligned} d_1 q_1 &= + 5 \cdot 3 = + 15 \\ d_2 q_2 &= + 15 \cdot 1 = + 15 \\ d_3 q_3 &= + 15 \cdot 0 = 0 \\ d_4 q_4 &= - 5 \cdot 4 = - 20 \\ d_5 q_5 &= 0 \cdot 2 = 0 \end{aligned}$$

сумма этих произведений $S = + 10$.

Средняя оценка H этого же испытуемого может быть получена посредством деления суммарной его оценки S (+ 10) на общую сумму всех коэффициентов значимости Σq . У нас:

$$\Sigma q = 3 + 1 + 0 + 4 + 2 = 10.$$

Следовательно средняя оценка здесь будет равна:

$$H = \frac{+ 10}{10} = + 1,$$

Итак, при заданных условиях средняя комплексная оценка H испытуемого № 2 оказалась равной + 1, в то время как без учета значимости отдельных признаков мы получили бы $H = + 1,6$ (см. стр. 202). Таким образом эта оценка здесь сильно понизилась. Это и понятно, если принять во внимание, что высшие оценки d (+ 15) оказались у этого испытуемого по наименее значимым признакам, для которых q принято было равным лишь 1 и 0, в то время как отрицательная оценка — 5 получена им по признаку, для которого q имеет наибольшую значимость ($q = 4$).

В описанных выше примерах коэффициенты значимости q устанавливались совершенно произвольно, исходя лишь из априорного учета тех влияний, которые могут иметь отдельные признаки на весь их комплекс, подлежащий оценке в целом. Однако в некоторых случаях такие коэффициенты q могут быть получены и эмпирически.

С этой целью за коэффициенты значимости q можно, например, принять коэффициенты корреляции r между результатами соответствующих испытаний и фактической рабочей продукцией, выявленной на какой-либо пробной группе испытуемых.

Пусть имеем 3 испытания: I, II и III, с помощью которых мы намерены определять пригодность отдельных испытуемых к какой-либо профессии. Отобрав достаточно характерную и не очень малую группу лиц, уже работающих на данном поприще, мы подвергаем их всем трем испытаниям и одновременно отмечаем тем или иным способом качество их работы (например условными №№ по успешности этой работы или по ее количеству). Прокоррелировав затем числовые результаты каждого из 3-х испытаний порознь с этими характеристиками действительной рабочей продукции, мы получим 3 коэффициента корреляции r_1 , r_2 и r_3 .

Пусть в частности:

$$r_1 = +0,8 \quad r_2 = +0,3 \quad r_3 = -0,1.$$

Эти коэффициенты показывают, что наличие первого признака для успеха в данном виде работы имеет очень большое значение ($r_1 = +0,8$), наличие второго признака уже не так существенно ($r_2 = +0,3$), а наличие третьего даже несколько понижает качество работы ($r_3 = -0,1$). Исходя из этих соображений, мы и можем принять здесь полученные числа r за коэффициенты значимости q .

Пусть, например, отдельный испытуемый (уже не из числа тех, на которых производился пробный опыт) дал следующие децисигмальные оценки d :

$$+12 \quad -8 \quad +2.$$

Помножив эти оценки на коэффициенты значимости q :

$$+0,8 \quad +0,3 \quad -0,1$$

получим следующие произведения:

$$(+12)(+0,8) = +9,6$$

$$(-8)(+0,3) = -2,4$$

$$(+2)(-0,1) = -0,2$$

$$\text{сумма их } S = +7,0.$$

Итак, суммарная комплексная оценка этого испытуемого S равна здесь $+7$ децисигмам. Для вычисления соответствующей средней его оценки H необходимо полученную здесь величину $S = +7$ разделить на сумму всех коэффициентов значимости:

$$+0,8 + 0,3 - 0,1 = +1.$$

Следовательно в данном частном случае и средняя комплексная оценка $H = +7$ децисигмам.

Аналогично можно поступать и с перцентильными оценками (с соблюдением указанных ранее правил и с учетом особенностей перцентильжа, приведенных в примечаниях к этой главе).

XX. Коэффициент дисгармоничности и способ гипотенуз.

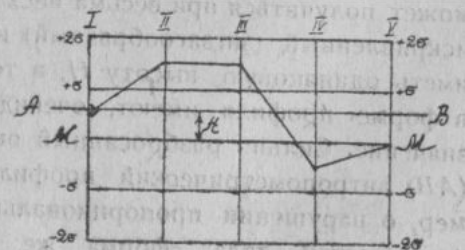
В приведенном выше примере вывода суммарной и средней комплексной оценки для 5 испытуемых по 5-ти признакам мы получили в конечном итоге следующую оценочную ведомость:

Испы- туемые	И с п ы т а н и я										Суммарная оценка S	Средняя оценка H
	I		II		III		IV		V			
	V	d	V	d	V	d	V	d	V	d		
№ 1 .	26	+ 15	16	+ 5	34	+ 5	19	+ 15	40	+ 15	+ 55	+ 11
№ 2 .	22	+ 5	18	+ 15	42	+ 15	7	- 5	25	0	+ 30	+ 6
№ 3 .	20	0	15	0	30	0	10	0	20	- 5	- 5	- 1
№ 4 .	18	- 5	14	- 5	18	- 15	13	+ 5	30	+ 5	- 15	- 3
№ 5 .	14	- 15	12	- 15	26	- 5	1	- 15	10	- 15	- 65	- 13

Для каждого из этих 5-ти испытуемых мы можем построить индивидуальный профиль, причем, например, для испытуемого № 2, получившего оценки:

$$+5 \quad +15 \quad +15 \quad -5 \quad \text{и } 0,$$

такой профиль * будет иметь форму, указанную на черт. 97. Посмотрим, какое графическое толкование мы можем дать



Черт. 97.

средней комплексной оценке H всех 5-ти испытаний этого индивидуума, которая в данном случае получилась равной $+6$ децисигмам.

Так как средняя комплексная оценка H характеризует собою среднюю величину всех 5-ти децисигмальных оценок d , которые на профиле представлены его ординатами (т. е. расстояниями отдельных точек профиля до средней горизонтальной линии сетки MM), то эта комплексная оценка $(+6)$ будет, очевидно, определять здесь среднюю высоту всего профиля в целом над той же самой центральной горизонталью MM . Если провести прямую AB на высоте $H = +6$ децисигмам, то положение этой прямой на сетке (выше или ниже) и будет характеризовать среднюю высоту всей ломаной линии профиля (см. черт. 97).

Итак, средняя оценка комплекса признаков в графической ее интерпретации соответствует расстоянию H между средней линией самого профиля (AB) и средней линией сетки (MM). Чем больше по своей абсолютной величине эта средняя комплексная оценка, тем дальше от средней горизонтали сетки расположится и средняя линия профиля (а стало быть и весь профиль в целом). При этом положительным значениям средней оценки H будет соответствовать расположение всего профиля в верхней части сетки, а отрицательным — в нижней.

Но кроме высоты профиля H , характеризующей свойственную данному индивидууму среднюю величину его оценок по отдельным признакам, во многих случаях представляется необходимым выяснить также и степень различия (количественного разнообразия) всех изолированных оценок d , из которых складывается эта комплексная характеристика H . Дело в том, что одна и та же средняя высота профиля H может получиться при весьма несходных условиях: и сильно искривленный (зигзагообразный) и ровный профиль могут иметь одинаковую высоту H , в то время как эти различия в форме профиля имеют, очевидно, весьма существенное значение. Сильно разбросанный около своей средней линии (AB) антропометрический профиль свидетельствует, например, о нарушении пропорциональности в развитии отдельных частей тела; форма же этого профиля, приближающаяся к горизонтальной линии, указывает, наобо-

рот, на идеально правильное соотношение всех частей тела.

Эту разбросанность (рассеяние) отдельных точек профиля около его средней линии (AB) мы можем количественно охарактеризовать особым „коэффициентом дисгармоничности“ D , вычисляемым по принципу, аналогичному определению среднего квадратического отклонения σ , причем в этих вычислениях средняя высота профиля H будет играть роль среднего арифметического M .

В нашем примере испытуемый № 2 имел следующие 5 изолированных оценок d :

$$+5 \quad +15 \quad +15 \quad -5 \quad 0,$$

т. е. средняя комплексная его оценка $H = +6$.

Приняв эти оценки (d) за варианты V , а комплексную оценку (H) за среднее арифметическое M , вычислим по аналогии с центральными (x) „дисгармонические“ отклонения Δ отдельных децисигмальных оценок d от общей комплексной оценки $H = +6$.

Так как оценка первого признака $d_1 = +5$, то ее дисгармоническое отклонение:

$$\Delta_1 = d_1 - H = +5 - 6 = -1.$$

Аналогично вычислим и все остальные дисгармонические отклонения Δ :

$$\begin{array}{ll} \text{Для } d_2 = +15 & \Delta_2 = d_2 - H = +15 - 6 = +9 \\ d_3 = +15 & \Delta_3 = d_3 - H = +15 - 6 = +9 \\ d_4 = -5 & \Delta_4 = d_4 - H = -5 - 6 = -11 \\ d_5 = 0 & \Delta_5 = d_5 - H = 0 - 6 = -6. \end{array}$$

Итак, комплекс оценок d испытуемого № 2 в нашем примере имеет следующие дисгармонические отклонения Δ :

$$-1 \quad +9 \quad +9 \quad -11 \quad \text{и} \quad -6,$$

которые, подобно центральным отклонениям (x) в сумме равняются 0.

Возвысив все эти отклонения (Δ) в квадрат, получим:

$$1 \quad 81 \quad 81 \quad 121 \quad \text{и} \quad 36.$$

Сумма квадратов дисгармонических отклонений $\Sigma \Delta^2$ (аналогичная сумме квадратов центральных отклонений Σx^2) равна здесь 320:

$$\Sigma \Delta^2 = 320.$$

Разделив эту сумму на общее число всех оценок $N = 5$, будем иметь:

$$\frac{\Sigma \Delta^2}{N} = \frac{320}{5} = 64.$$

После извлечения из этого числа квадратного корня мы и получим искомый коэффициент дисгармоничности:

$$D = \pm \sqrt{64} = \pm 8 \text{ децисигм.}$$

Итак, коэффициент D вычисляется по формуле:

$$D = \pm \sqrt{\frac{\Sigma \Delta^2}{N}},$$

где $\Delta = d - H$.

Полученный здесь результат ($D = \pm 8$) показывает, что индивидуальный профиль испытуемого № 2 разбросан около своей средней линии AB в обе стороны от нее ($\pm$) в среднем на 8 децисигм.

Очевидно, для профиля, имеющего форму горизонтальной прямой линии (т. е. для случая совершенно одинаковых оценок d), коэффициент $D = 0$.

Чем больше по своей абсолютной величине этот коэффициент D , тем большее разнообразие наблюдается и у отдельных оценок d .

Полную характеристику индивидуального профиля мы будем в дальнейшем изображать символическим сочетанием букв:

$$H \pm D,$$

которое, например, в применении к испытуемому № 2 будет иметь следующее числовое значение:

$$H \pm D = +6 \pm 8.$$

Для испытуемого № 3 (см. черт. 98) его децисигмальные оценки d по 5-ти изолированным признакам оказались равными:

$$0 \quad 0 \quad 0 \quad 0 \quad -5,$$

а средняя комплексная оценка $H = -1$.

Здесь при вычислении дисгармонических отклонений Δ из каждой оценки d придется вычитать отрицательную величину $H = -1$, что соответствует прибавлению к d той же величины H , но уже с обратным знаком ($+$). Итак:

$$\Delta_1 = 0 + 1 = +1$$

$$\Delta_2 = 0 + 1 = +1$$

$$\Delta_3 = 0 + 1 = +1$$

$$\Delta_4 = 0 + 1 = +1$$

$$\Delta_5 = -5 + 1 = -4.$$

Возвысив все четыре отклонения Δ в квадрат, получим:

$$1 \quad 1 \quad 1 \quad 1 \quad 16.$$

Сумма этих квадратов $\Sigma \Delta^2 = 20$.

Следовательно коэффициент дисгармоничности:

$$D = \pm \sqrt{\frac{20}{5}} = \pm \sqrt{4} = \pm 2$$

или, так как $H = -1$, то в окончательной символической форме:

$$H \pm D = -1 \pm 2.$$

Этот результат показывает, что оцениваемый профиль опущен (знак $-$ у H) относительно средней линии сетки (MM) в среднем на 1 децисигму и разбросан около своей средней линии (AB) в обе стороны от нее ($\pm$) в среднем на 2 децисигмы (см. черт. 98).

Значение H , близкое к 0, и сравнительно малая величина D указывают на принадлежность данного испытуемого к категории „нормальных“ и на почти полную однородность его оценок по всем 5-ти испытаниям.

Нетрудно убедиться в том, что в нашем примере:

$$\text{для испытуемого № 1} \quad H \pm D = +11 \pm 4,90^*)$$

$$\text{„ „ „ № 2} \quad H \pm D = +6 \pm 8$$

$$\text{„ „ „ № 3} \quad H \pm D = -1 \pm 2$$

$$\text{„ „ „ № 4} \quad H \pm D = -3 \pm 7,48^{**})$$

$$\text{„ „ „ № 5} \quad H \pm D = -13 \pm 4$$

*) Точно: $D = \pm \sqrt{24}$.

**) Точно: $D = \pm \sqrt{56}$.



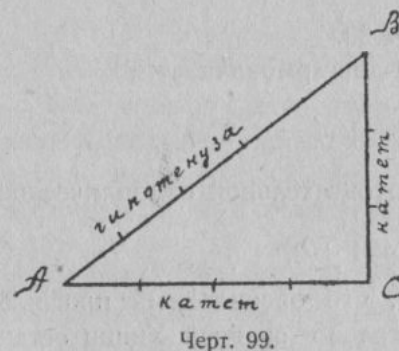
Черт. 98.

Вычисление коэффициентов дисгармоничности по формуле:

$$D = \pm \sqrt{\frac{\sum \Delta^2}{N}}$$

при сколько-нибудь значительном количестве объединяемых в одном комплексе признаков (N) и при очень большом числе испытуемых (n), для которых эти коэффициенты вычисляются, представляет собою настолько кропотливую и громоздкую операцию, что без каких-либо технических ее упрощений такой детальный способ комплексной характеристики индивидуумов на практике может оказаться просто неосуществимым.

Единственным и весьма действительным выходом из создавшегося здесь затруднения служит применение чрезвычайного экономного по затратам времени и в сущности очень несложного графического приема определения коэффициентов D , основанного на известной „теореме Пифагора“.



AB существует точная количественная зависимость, определяемая формулой:

$$AB = \sqrt{AC^2 + BC^2}$$

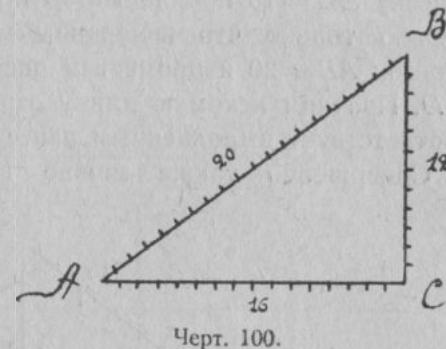
Эта формула показывает, что для вычисления длины отрезка AB (гипотенузы) следует измерить в каких-либо единицах (например в сантиметрах) оба катета AC и CB , затем каждое из полученных чисел возвысить в квадрат, сложить их друг с другом и из полученной суммы извлечь квадратный корень. Так, например, в прямоугольном треугольнике, изображенном на черт. 99, катет AC сделан длиной в 4 см ($AC = 4$), а катет BC по высоте равен 3 см ($BC = 3$). Следовательно расстояние между точками A и B , т. е. гипотенуза нашего прямоугольного треугольника:

$$AB = \sqrt{4^2 + 3^2} = \sqrt{16 + 9} = \sqrt{25} = 5 \text{ см.}$$

И действительно в отрезке AB укладывается ровно 5 см, в чем нетрудно убедиться непосредственным его измерением.

На основании этих соображений мы всегда можем заменить вычисление квадратного корня из суммы квадратов двух величин AC и BC простым измерением (по масштабу) гипотенузы AB такого прямоугольного треугольника ABC , в котором оба катета AC и BC соответствуют по своей длине и высоте двум заданным величинам.

К аналогичному (но дважды повторенному) построению можно прибегнуть и при вычислении квадратного корня из суммы квадратов трех величин, например 12, 16 и 15.



Непосредственное вычисление по формуле дает здесь следующие результаты:

$$\sqrt{12^2 + 16^2 + 15^2} = \sqrt{144 + 256 + 225} = \sqrt{625} = 25.$$

Разобьем весь процесс этого вычисления на 2 части: сперва определим числовое значение квадратного корня из суммы квадратов двух первых по порядку величин: 12 и 16:

$$\sqrt{12^2 + 16^2} = \sqrt{144 + 256} = \sqrt{400} = 20,$$

а затем уже найдем значение корня из суммы квадратов только что найденной величины (20) и оставшейся (третьей) величины 15:

$$\sqrt{20^2 + 15^2} = \sqrt{400 + 225} = \sqrt{625} = 25.$$

В конечном итоге мы и придем, очевидно, всегда к тому же самому результату (25), который был получен раньше при вычислении полного выражения:

$$\sqrt{12^2 + 16^2 + 15^2} = 25.$$

Графически этот способ вычисления соответствует двум следующим построениям (см. чертежи 100 и 101).

К вертикально поставленному отрезку $BC = 12$ делениям какого-либо произвольного масштаба (черт. 100) прикладываем горизонтальный отрезок $AC = 16$ делениям того же самого масштаба и измеряем расстояние между точками A и B , т. е. гипотенузу AB , оказавшуюся $= 20$ таким делениям. Далее (черт. 101) к вертикально поставленному отрезку $DB = 15$ прикладываем по горизонтальному направлению только что измеренный перед тем (на черт. 100) отрезок $AB = 20$ и промеряем расстояние между точками A и D . Получим искомую длину отрезка $AD = 25$, что вполне соответствует выполненным ранее вычислениям.

Совершенно также можно выполнить и построения, заменяющие собою непосредственные вычисления квадратного корня из суммы квадратов четырех и более величин.

Применим этот графический прием, могущий быть названным „способом гипотенуз“*, к определению числового значения квадратного корня из

суммы квадратов всех дисгармонических отклонений Δ , т. е. выражения:

$$\sqrt{\sum \Delta^2}$$

входящего в числитель окончательной формулы для определения коэффициента дисгармоничности D .

В самом деле:

$$D = \sqrt{\frac{\sum \Delta^2}{N}}$$

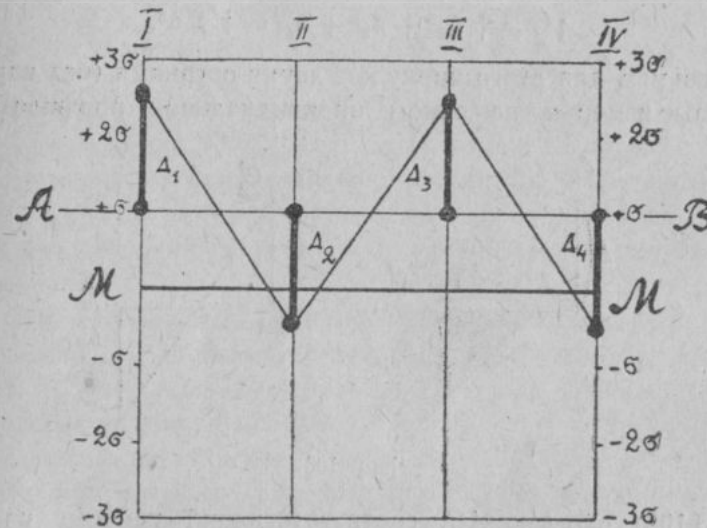
* О применении „способа гипотенуз“ в опытной агрономии см. мою статью „Графический метод обработки результатов полевого опыта“. Труды по прикл. ботанике, генетике и селекции. Т. XXI, вып. 1-й, 1928—1929. Стр. 597—632.

или, извлекая корень в числителе и в знаменателе отдельно:

$$D = \frac{\sqrt{\sum \Delta^2}}{\sqrt{N}}$$

Поэтому при определении коэффициента D мы можем найти сперва числовое значение $\sqrt{\sum \Delta^2}$ и затем уже разделить его на $\sqrt{N}$.

Пусть, например, имеем индивидуальный профиль, представленный на чертеже 102.



Черт. 102.

Отдельные децисигмальные оценки d здесь имеют следующие числовые значения:

$$+25 \quad -5 \quad +25 \quad -5.$$

Сумма их (суммарная комплексная оценка):

$$S = +40,$$

а средняя комплексная оценка (средняя высота всего профиля в целом):

$$H = \frac{+40}{4} = +10.$$

Проведя горизонтальную прямую на высоте $H = +10$ децисигмам над средней линией сетки (MM), найдем положение средней линии профиля AB .

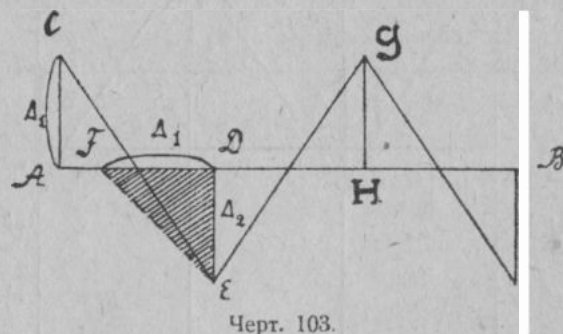
Здесь дисгармонические отклонения, представляющие собой расстояния отдельных вершин профиля до прямой AB , изображены жирными вертикальными отрезками, обозначенными буквами $\Delta_1, \Delta_2, \Delta_3$ и Δ_4 . Нетрудно убедиться в том, что:

$$\begin{aligned}\Delta_1 &= +15 & \Delta_3 &= +15 \\ \Delta_2 &= -15 & \Delta_4 &= -15.\end{aligned}$$

Для вычисления коэффициента дисгармоничности D определим по „способу гипотенуз“ величину:

$$\sqrt{\Delta_1^2 + \Delta_2^2 + \Delta_3^2 + \Delta_4^2} = \sqrt{\Sigma \Delta^2}.$$

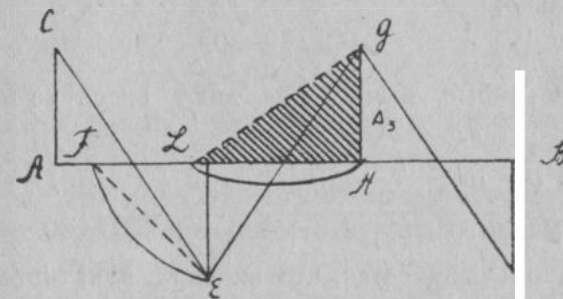
Возьмем для этого циркуль с двумя остриями (без карандаша) и измерим на самом индивидуальном профиле (см.



черт. 103) отрезок AC , соответствующий первому из 4-х заданных дисгармонических отклонений Δ_1 . Далее отложим по средней линии профиля AB отрезок DF , равный только что измеренному перед тем первому отрезку AC (оба эти равные отрезки помечены на черт. 103 дугами). Получим прямоугольный треугольник FDE , в котором катет $FD = \Delta_1$, а катет $DE = \Delta_2$. В этом треугольнике нам нужно теперь измерить циркулем его гипотенузу FE . Для этого, не снимая с чертежа ту ножку циркуля, которая при отложении отрезка FD попала в точку F , другую его ножку мы приподнимаем над чертежом (под точкой D) и, раздвигая циркуль настолько, насколько это понадобится, ставим второе его острие в точку E (являющуюся вершиной второго отрезка $DE = \Delta_2$).

После этого циркуль совершенно снимается с чертежа, а измеренный им отрезок FE переносится опять на сред-

нюю линию профиля (AB) к основанию H третьего отрезка $HG = \Delta_3$ (см. черт. 104). Здесь отрезок HL в точности равен

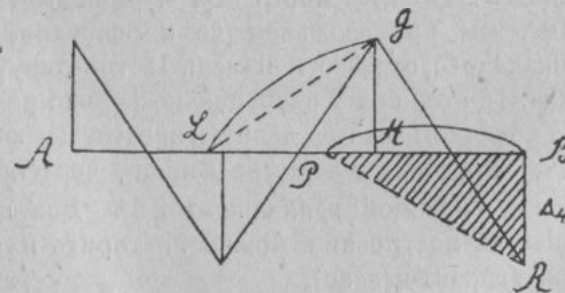


Черт. 104.

полученному ранее отрезку FE , который служил гипотенузой треугольника FDE (эти отрезки помечены дугами). Далее мы снимаем острие циркуля с точки H и, не сдвигая второго острия с точки L , переносим освободившуюся ножку циркуля из точки H в точку G , т. е. измеряем новую гипотенузу LG .

Отложив, наконец, эту гипотенузу LG на той же прямой AB влево от основания B последнего отрезка $BR = \Delta_4$ (см. черт. 105), мы получим отрезок $BP (= LG)$, который служит катетом треугольника PBR .

Сняв ножку циркуля с точки B и поставив ее в вершину последнего отрезка (т. е. в точку R), мы и измерим гипотенузу PR этого последнего треугольника.



Черт. 105.

На основании высказанных ранее соображений длина этого отрезка PR в нашем децисигмальном масштабе и будет соответствовать искомой величине:

$$\sqrt{\Sigma \Delta^2}.$$

В данном случае отрезок $PR = 30$ децисигмам, в чем нетрудно убедиться путем непосредственного его измерения. Следовательно:

$$\sqrt{\Sigma \Delta^2} = 30.$$

Но и обычные вычисления дают здесь те же самые результаты:

$$\begin{aligned} \sqrt{(+15)^2 + (-15)^2 + (+15)^2 + (-15)^2} = \\ \sqrt{225 + 225 + 225 + 225} = \sqrt{900} = 30. \end{aligned}$$

Само собой разумеется, что никаких пунктирных линий при таком построении проводить на профиле не приходится. Эти пунктиры и штриховка отдельных треугольников сделаны на чертежах 103, 104 и 105 лишь для ясности и наглядности описания самого порядка последовательных измерений всех гипотенуз.

Все эти построения производятся здесь только посредством ряда отмеров острями циркуля всех расстояний на самом индивидуальном профиле, без какого-бы то ни было прочерчивания дополнительных прямых по линейке, кроме основной средней линии профиля AB .

Итак, для построения величины $\sqrt{\Sigma \Delta^2}$ мы должны измерить циркулем отрезок, соответствующий первому дисгармоническому отклонению (Δ_1), и приложить его к основанию отрезка, соответствующего второму отклонению (Δ_2). Измерив гипотенузу полученного при этом прямоугольного треугольника, мы прикладываем ее к основанию следующего вертикального отрезка и измеряем гипотенузу нового треугольника. Повторив эту операцию несколько раз подряд, получим в конечном итоге длину гипотенузы последнего треугольника, которая в масштабе данного чертежа и будет соответствовать искомой величине $\sqrt{\Sigma \Delta^2}$. Для проверки результатов этих построений можно повторить их в обратном порядке (справа-налево).

Определив величину $\sqrt{\Sigma \Delta^2}$, которая в нашем примере оказалась $= 30$ децисигмам, следует разделить ее на $\sqrt{N}$. Так как у нас $N = 4$, то:

$$D = \pm \frac{\sqrt{\Sigma \Delta^2}}{\sqrt{N}} = \pm \frac{30}{\sqrt{4}} = \pm \frac{30}{2} = \pm 15.$$

Итак, в нашем примере средняя высота профиля $H = +10$, а коэффициент дисгармоничности $D = \pm 15$ или:

$$H \pm D = +10 \pm 15.$$

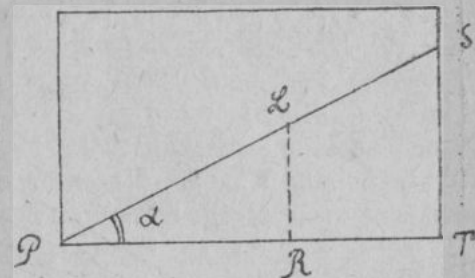
В общем случае, когда корень из числа N не извлекается нацело, такое деление последней гипотенузы $PR = \sqrt{\Sigma \Delta^2}$ на $\sqrt{N}$ удобно производить также графическим путем. Для этого следует заранее вычертить на листе миллиметровой бумаги прямоугольник произвольного размера и из левого нижнего его угла P провести наклонную прямую PS , как это показано на черт. 106.

Положение точки S , через которую проходит прямая PS определяется для каждого данного числа N по следующей формуле:

$$ST = \sqrt{\frac{(PT)^2}{N}}.$$

В нашем примере $N = 4$, а основание прямоугольника PT взято равным 5 см. Итак:

$$N = 4 \quad PT = 5.$$



Черт. 106.

Следовательно высота искомого отрезка:

$$ST = \sqrt{\frac{(PT)^2}{N}} = \sqrt{\frac{5^2}{4}} = \sqrt{\frac{25}{4}} = \sqrt{6,25} = 2,5 \text{ см.}$$

Отложив сверху от точки T отрезок $TS = 2,5$ см, мы и получим на противоположном его конце точку S . Далее точку S следует соединить прямой линией с левым нижним углом этого прямоугольника, т. е. с точкой P .

Этот представленный на черт. 106 „расчетный график“ заготовлен для $N = 4$ (при другом значении N наклонная прямая PS прошла бы, очевидно, уже под другим углом к основанию этого прямоугольника).

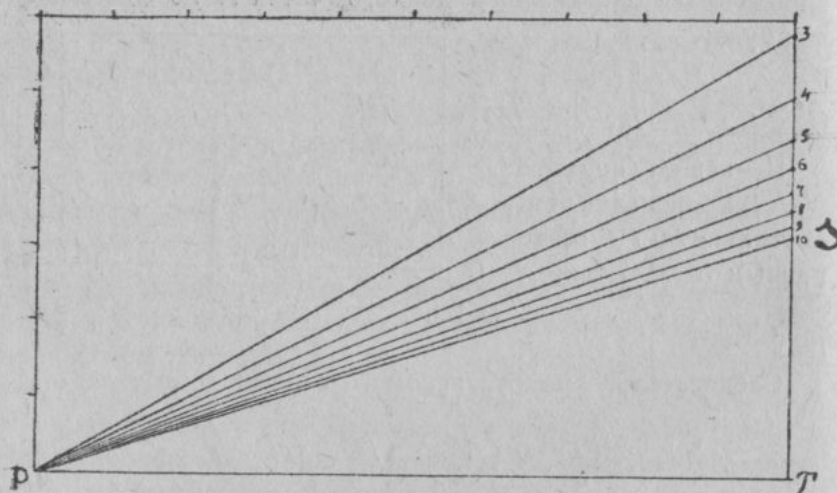
*) Можно было бы вывести величину PT из под знака радикала, но в этом виде формула представляет большие практические удобства при производстве вычислений.

Для деления отрезка $PR = \sqrt{\Sigma \Delta^2}$ на $\sqrt{N}$ мы откладываем его циркулем по нижнему основанию прямоугольника PT вправо от точки P и из противоположного конца R этого отрезка PR по направлению клеток миллиметровой бумаги измеряем расстояние RL кверху до наклонной прямой PS . Этот отрезок RL в первоначальном (децисигмальном) масштабе и будет равен искомому коэффициенту дисгармоничности D . В данном случае отрезок:

$$RL = 15 \text{ децисигмам,}$$

что вполне совпадает с вычисленным выше значением коэффициента D .

Так как отдельные индивидуальные профили, для кото-



Черт. 107.

рых приходится определять коэффициенты дисгармоничности D , чаще всего характеризуют некоторое постоянное количество признаков (N), то для окончательного деления $\sqrt{\Sigma \Delta^2}$ на $\sqrt{N}$ достаточно бывает заготовить на расчетном графике лишь одну единственную наклонную прямую. В расчете же на различные числа N в таком прямоугольнике можно провести целый пучок исходящих из точки P наклонных прямых линий, каждая из которых будет соответствовать определенному значению N .

На черт. 107 представлен такой график, рассчитанный на значения N в пределах от 3 до 10. Цифрами справа

указано значение N для каждой отдельной наклонной прямой.

Способ построения этих наклонных линий и использование их при делении $\sqrt{\Sigma \Delta^2}$ на $\sqrt{N}$ остается прежним. Например, самая нижняя из этих линий, соответствующая значению $N=10$, проводится через точки P и S , причем последняя ставится на правой стороне прямоугольника на высоте:

$$ST = \sqrt{\frac{(PT)^2}{N}}.$$

Здесь основание графика $PT = 10 \text{ см}$, а число $N = 10$, следовательно:

$$ST = \sqrt{\frac{10^2}{10}} = \sqrt{\frac{100}{10}} = \sqrt{10} = 3,16 \text{ см.}$$

Итак, наклонную прямую PS , соответствующую значению $N=10$, следует провести через точку P и точку S , поднятую над точкой T на высоту $ST = 3,16 \text{ см}$.

Аналогично проводятся на графике и все остальные наклонные прямые.

В частности, например, при основании прямоугольника $PT = 20 \text{ см}$ высота точки S над точкой T для различных значений числа N будет иметь следующие частные значения:

при $N = 2$ высота $ST = 14,14 \text{ см}$

"	$N = 3$	"	$ST = 11,55$	"
"	$N = 4$	"	$ST = 10,00$	"
"	$N = 5$	"	$ST = 8,94$	"
"	$N = 6$	"	$ST = 8,16$	"
"	$N = 7$	"	$ST = 7,56$	"
"	$N = 8$	"	$ST = 7,07$	"
"	$N = 9$	"	$ST = 6,67$	"
"	$N = 10$	"	$ST = 6,32$	"
"	$N = 11$	"	$ST = 6,03$	"
"	$N = 12$	"	$ST = 5,77$	"
"	$N = 13$	"	$ST = 5,55$	"
"	$N = 14$	"	$ST = 5,35$	"
"	$N = 15$	"	$ST = 5,16$	"
"	$N = 16$	"	$ST = 5,00$	"
"	$N = 17$	"	$ST = 4,85$	"
"	$N = 18$	"	$ST = 4,71$	"
"	$N = 19$	"	$ST = 4,59$	"
"	$N = 20$	"	$ST = 4,47$	"

XXI. Некоторые свойства изолированных и комплексных оценок.

В приведенном выше числовом примере оценки 5-ти испытуемых по 5-ти признакам мы получили следующие конечные результаты (см. стр. 207 и 211).

Испытуемые:	Оценки							
	Изолированная:					Комплексная:		
	d_1	d_2	d_3	d_4	d_5	S	H	D
№ 1	+15	+5	+5	+15	+15	+55	+11	$\pm 4,90$
№ 2	+5	+15	+15	-5	0	+30	+6	± 8
№ 3	0	0	0	0	-5	-5	-1	± 2
№ 4	-5	-5	-15	+5	+5	-15	-3	$\pm 7,48$
№ 5	-15	-15	-5	-15	-15	-65	-13	± 4

Здесь буквами: d_1, d_2, d_3, d_4 и d_5 обозначены как всегда децисигмальные оценки этих испытуемых по отдельным изолированным признакам, а буквами S, H и D — их комплексные оценки: S — суммарная, H — средняя, а D — коэффициент дисгармоничности.

При выводе всех этих комплексных оценок мы выполняли различные требуемые правилами действия в направлении горизонтальных строчек этой таблицы (напр. при выводе суммарной оценки складывали все изолированные оценки d , полученные данным испытуемым по всем 5-ти испытаниям).

Выясним некоторые соотношения, которые получаются здесь при выполнении определенных действий над этими же числами по направлению вертикальных столбцов нашей таблицы.

Прежде всего следует обратить внимание на то, что алгебраическая сумма этих чисел во всех столбцах, кроме крайнего (D), в точности равна нулю:

$$\Sigma d = 0 \quad \Sigma S = 0 \quad \text{и} \quad \Sigma H = 0,$$

в чем нетрудно убедиться, выполнив эти сложения в действительности.

Так например в первом столбце d_1 сумма всех изолированных децисигмальных оценок по I признаку:

$$\Sigma d_1 = -15 - 5 + 0 + 5 + 15 = 0.$$

То же самое получим и в столбцах d_2, d_3, d_4 и d_5 , а также в двух предпоследних: S и H .

В нашем примере все эти суммы равны нулю совершенно точно, так как при вычислении отдельных оценок d, S и H нам не пришлось отбрасывать лишних знаков после запятой (все деления были выполнены нацело). Если же эти оценки вычислялись бы приближенно, то и равенство нулю их сумм носило бы также не вполне строгий характер. При этом отличие всех сумм от нуля было бы тем большим, чем большее количество (n) таких неточных оценок мы подвергали бы этому суммированию.

Но если суммы чисел ($\Sigma d, \Sigma S$ и ΣH) в каждом из вертикальных столбцов $= 0$, то этим же свойством будут обладать очевидно и их средние арифметические (M_d, M_S, M_H), так как всякое среднее арифметическое равно сумме чисел, деленной на их количество (n).

В случае приближенного вычисления отдельных оценок d, S и H точность этих средних арифметических, вычисляемых по вертикальным столбцам, уже не будет находиться в зависимости от их количества n , так как накопление ошибок при увеличении n всегда окажется компенсированным в процессе дальнейшего деления суммы всех оценок на это большое число n .

Равенство нулю средних арифметических по отдельным столбцам может служить средством проверки суммарных и средних комплексных оценок (S и H) *).

Второе весьма существенное свойство децисигмальных оценок d заключается в постоянстве суммы их квадратов Σd^2 , которая в каждом из столбцов d_1, d_2, d_3, d_4 и d_5 в точности равна $100n$. В самом деле для крайнего, например, столбца (d_1) отдельные оценки равны:

$$+15 \quad +5 \quad 0 \quad -5 \quad -15.$$

*) Средние арифметические перцентильных оценок $= 50$.

а квадраты их в сумме составляют:

$$225 + 25 + 0 + 25 + 225 = 500.$$

Так как всего этих оценок здесь 5 ($n=5$), то $100n$ также $=500$. Нетрудно убедиться в том, что и для всех остальных столбцов суммы квадратов их оценок d будут $=100n$. Итак:

$$\sum d^2 = 100n.$$

Точность этого равенства также находится в зависимости от точности самих оценок d и общего их количества n .

Из постоянства суммы квадратов всех d вытекает и постоянство средних квадратических отклонений (σ), вычисленных для чисел каждого из первых пяти вертикальных столбцов нашей таблицы. Эти сигмы оказываются здесь всегда равными 10. В самом деле, для чисел крайнего левого столбца (d_1):

$$+15 \quad +5 \quad 0 \quad -5 \quad -15$$

среднее квадратическое отклонение $\sigma = \pm 10$ *).

То же самое будем иметь и во всех остальных столбцах, в которых у нас были записаны изолированные оценки d . Здесь точность этого результата зависит только от точности вычисления отдельных оценок d , но уже не от их общего количества n , так как накопление ошибок при увеличении n компенсируется последующим делением $\sum d^2$ на это число n (под корнем).

Итак, среднее квадратическое отклонение ряда чисел, представляющих собой децисигмальные оценки по одному изолированному признаку для всей группы объектов исследования (на которой был выработан критерий этой оценки) всегда $=10$ (децисигмам). Отсюда ясно, что децисигмальные оценки d по отдельным изолированным признакам варьируют всегда в одних и тех же постоянных пределах **).

*) Здесь σ вычислена по формуле без поправки на малое число измерений n , так как число это предполагается достаточно большим.

**) Перцентильные оценки при достаточно большом их количестве будут варьировать в пределах каждого отдельного признака в среднем на $\sigma = \pm \frac{50}{\sqrt{3}} = \pm 28,87$ перцентилей (если эту величину определить по обычным правилам вычисления σ).

Теперь посмотрим, какие соотношения и количественные закономерности наблюдаются у комплексных оценок S , H и D , если их учитывать для всей совокупности n объектов исследования в целом (т. е. также по вертикальным столбцам).

Здесь прежде всего следует отметить весьма важную зависимость, которая существует между суммой квадратов средних оценок H и суммой квадратов коэффициентов дисгармоничности D , т. е. между $\sum H^2$ и $\sum D^2$. На нашем числовом примере легко убедиться в том, что:

$$\sum H^2 + \sum D^2 = 100n.$$

В самом деле, средние оценки H всех пяти испытуемых у нас оказались равными (снизу вверх):

$$-13 \quad -3 \quad -1 \quad +6 \quad +11,$$

поэтому квадраты их выразятся следующими числами:

$$169 \quad 9 \quad 1 \quad 36 \quad 121,$$

а сумма квадратов всех H , т. е.

$$\sum H^2 = 336.$$

Коэффициенты дисгармоничности D для этих же 5-ти испытуемых у нас получились равными:

$$\pm 4 \quad \pm 7,48 \quad \pm 2 \quad \pm 8 \quad \text{и} \quad \pm 4,90,$$

а квадраты этих коэффициентов (см. прим. на стр. 211):

$$16 \quad 56 \quad 4 \quad 64 \quad 24,$$

откуда сумма квадратов всех D , т. е.

$$\sum D^2 = 164.$$

Следовательно:

$$\sum H^2 + \sum D^2 = 336 + 164 = 500.$$

Но и $100n = 100 \cdot 5 = 500$, что вполне подтверждает высказанное выше правило.

Какие отсюда могут быть сделаны выводы?

Если величины $\sum H^2$ и $\sum D^2$ в сумме равны постоянному числу $100n$, то увеличение каждой из них может происходить только лишь за счет соответствующего уменьшения

другой. Следовательно там, где средние оценки H будут варьировать сильно (ΣH^2 будет иметь большое значение), там большинство коэффициентов дисгармоничности D окажутся близкими к 0 (ΣD^2 будет иметь малое значение) и наоборот: слабое варьирование средних оценок H всегда будет указывать на значительную дисгармоничность большинства оцениваемых индивидуумов.

Если, в частности, все n испытуемых имели бы совершенно одинаковые оценки d по каждому из N изолированных признаков, то все коэффициенты D (а стало быть и сумма их квадратов ΣD^2) = 0. В этом случае сумма квадратов средних оценок H т. е. ΣH^2 имела бы наибольшее из возможных своих значений:

$$\Sigma H^2 = 100 n,$$

т. е. средние оценки H варьировали бы в тех же самых пределах, в каких варьируют и децисигмальные оценки d по каждому изолированному признаку (см. стр. 224).

Так как в общем случае отдельные коэффициенты D не = 0, то и ΣH^2 обычно оказывается меньше 100 n . Отсюда ясно, что варьирование средних оценок H всегда меньше варьирования децисигмальных оценок d .

Таким образом, если за „норму“ отдельных признаков мы условно принимаем например интервал варьирования децисигмальных оценок d от +10 до -10 (что соответствует колебаниям отдельных вариантов V в пределах от $M + \sigma$ до $M - \sigma$), то границы „нормы“ для средних комплексных оценок H должны быть всегда несколько сужены. Чтобы определить эти границы, в каждом отдельном случае необходимо вычислить среднее квадратическое отклонение (σ) для всего ряда фактически полученных средних оценок H^*). В нашем примере для ряда чисел:

$$-13 \quad -3 \quad -1 \quad +6 \quad +11$$

*) Для того чтобы судить о сужении размаха колебаний средних оценок (S) при перцентилеже, следует или разбить их на „равночисленные“ зоны (см. стр. 126) или вторично перцентилировать их с помощью интегральной кривой (см. стр. 170).

среднее квадратическое отклонение приблизительно = 8,2 (а не 10, как это имело место в применении к любому ряду изолированных децисигмальных оценок d).

Следовательно, границами „нормы“ для всего комплекса 5-ти признаков в целом будут служить числа:

$$-8,2 \quad \text{и} \quad +8,2,$$

если за „норму“ оценок d по отдельным изолированным признакам был взят интервал:

$$\text{от } -10 \text{ до } +10.$$

Для суммарных оценок S эта зона „нормы“ должна быть наоборот расширена путем умножения обеих границ +8,2 и -8,2 на число объединяемых признаков N (т. е. в данном случае на 5). Следовательно в нашем примере при $N=5$ область нормальных значений суммарных оценок S будет л жать в пределах:

$$\text{от } -41 \text{ до } +41.$$

Эти же границы мы могли бы получить, очевидно, и непосредственно, путем вычисления среднего квадратического отклонения (σ) для ряда суммарных оценок S , но приближаясь к рассмотрению ряда средних оценок H . В самом деле, в нашем примере для чисел S :

$$-65 \quad -15 \quad -5 \quad +30 \quad +55$$

среднее квадратическое отклонение σ приблизительно = 41, что и определяет здесь границы „нормы“ в пределах:

$$\text{от } -41 \text{ до } +41.$$

Это совпадение конечных результатов вычислений объясняется здесь зависимостью, которая существует между средним квадратическим отклонением (σ) ряда суммарных и средних оценок. Сигма ряда S всегда в N раз больше сигмы ряда H (у нас сигма ряда S равна 41, а сигма ряда H равна 8,2, т. е. как раз в 5 раз больше).

Одним из весьма важных свойств децисигмальных оценок d , кроме указанного выше постоянства пределов их варьирования, служит возможность довольно существенных упрощений вычислительной техники при опреде-

лении коэффициента корреляции r между двумя рядами таких оценок. Вспомним (см. I часть этого руководства), что коэффициент корреляции r вычисляется по формуле:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}} \quad \text{или} \quad r = \frac{\sum xy}{n \sigma_x \sigma_y}$$

Если же коррелирующими вариантами V служат два ряда децисигмальных оценок: d_x и d_y , то коэффициент r может быть вычислен по следующей более простой формуле:

$$r = \frac{\sum d_x d_y}{100 n}$$

Пусть, например, нам требуется вычислить коэффициент корреляции r между результатами V_x и V_y двух испытаний:

Испы- туемый	I испытание		II испытание	
	Результаты V_x	Оценка d_x	Результаты V_y	Оценка d_y
№ 1 . . .	26	+15	19	+15
№ 2 . . .	22	+5	7	-5
№ 3 . . .	20	0	10	0
№ 4 . . .	18	-5	13	+5
№ 5 . . .	14	-15	1	-15

Коррелируя числовые результаты непосредственного измерения V_x и V_y , получим следующий расчет:

V_x	V_y	x	x^2	y	y^2	xy
26	19	+6	36	+9	81	+54
22	7	+2	4	-3	9	-6
20	10	0	0	0	0	0
18	13	-2	4	+3	9	-6
14	1	-6	36	-9	81	+54
100	50	0	80	0	180	+96

Здесь $\sum V_x = 100$, а $\sum V_y = 50$, следовательно:

$$M_x = \frac{\sum V_x}{n} = \frac{100}{5} = 20.$$

$$M_y = \frac{\sum V_y}{n} = \frac{50}{5} = 10.$$

Вычислив все центральные отклонения x и y , возвысив их в квадрат и сложив, получим в сумме:

$$\sum x^2 = 80 \quad \sum y^2 = 180.$$

Перемножив попарно x на y и подведя итог столбца xy , получим в сумме:

$$\sum xy = +96.$$

Следовательно искомый коэффициент корреляции:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}} = \frac{+96}{\sqrt{80 \cdot 180}} = \frac{+96}{\sqrt{14400}} = \frac{+96}{120} = +0,8.$$

Если бы у нас были предварительно вычислены оба средние квадратичные отклонения (σ) этих рядов: $\sigma_x = \sqrt{\frac{\sum x^2}{n}} = \sqrt{\frac{80}{5}} = \sqrt{16} = \pm 4$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n}} = \sqrt{\frac{180}{5}} = \sqrt{36} = \pm 6,$$

то коэффициент корреляции r мы могли бы определить и по другой формуле:

$$r = \frac{\sum xy}{n \sigma_x \sigma_y} = \frac{+96}{5 \cdot 4 \cdot 6} = \frac{+96}{120} = +0,8.$$

Покажем теперь, какие упрощения можно внести в процесс вычисления r , если наряду с вариантами V_x и V_y мы имеем также и их децисигмальные оценки d_x и d_y . Перемножив попарно d_x на d_y и сложив все эти произведения, получим:

$$\sum d_x d_y = +400.$$

d_x	d_y	$d_x d_y$
+15	+15	+225
+5	-5	-25
0	0	0
-5	+5	-25
-15	-15	+225
0	0	+400

откуда искомый коэффициент корреляции:

$$r = \frac{\sum d_x d_y}{100 n} = \frac{+400}{100 \cdot 5} = +0,8.$$

Этот способ, очевидно, много проще предыдущего.

Применяя этот упрощенный прием к вычислению всех коэффициентов корреляции r между отдельными испытаниями в нашем последнем примере (см. стр. 222), получим:

$$\begin{array}{lll} r_{12} = +0,80 & r_{21} = +0,80 & r_{31} = +0,60 \\ r_{13} = +0,60 & r_{23} = +0,80 & r_{32} = +0,80 \\ r_{14} = +0,80 & r_{24} = +0,40 & r_{34} = 0 \\ r_{15} = +0,85 & r_{25} = +0,55 & r_{35} = +0,15 \\ r_{41} = +0,80 & r_{51} = +0,85 & \\ r_{42} = +0,40 & r_{52} = +0,55 & \\ r_{43} = 0 & r_{53} = +0,15 & \\ r_{45} = +0,95 & r_{54} = +0,95, & \end{array}$$

Здесь цифровые пометки у буквы r обозначают те признаки, которые в данном случае коррелируются. Так например r_{12} обозначает коэффициент корреляции между признаками I и II, а r_{35} — между III и V и т. д. Ясно, что прямой и обратный порядок этих цифровых значков (например 1 и 2 или, наоборот, 2 и 1) указывает на один и тот же коэффициент корреляции (между I и II признаком). В самом деле, у нас здесь и $r_{35} = +0,15$ и $r_{53} = +0,15$. Точно также равны друг другу и все остальные коэффициенты корреляции r , отличающиеся лишь порядком своих значков. Например r_{24} и r_{42} или r_{13} и r_{31} и т. д.

Все эти коэффициенты корреляции r можно представить также и в виде следующей таблицы:

	I	II	III	IV	V
I	—	+0,80	+0,60	+0,80	+0,85
II	+0,80	—	+0,80	+0,40	+0,55
III	+0,60	+0,80	—	0	+0,15
IV	+0,80	+0,40	0	—	+0,95
V	+0,85	+0,55	+0,15	+0,95	—

Числа, стоящие в клетках этой таблицы, представляют собой коэффициенты корреляции r между двумя признаками, обозначенными в заголовках соответствующей строки и столбца. Так например число $+0,85$, записанное в левом нижнем углу этой таблицы, относится к столбцу с заголовком I и к строке с заголовком V. Следовательно коэффициент корреляции между I и V признаком $r_{15} = +0,85$.

Все эти коэффициенты понадобятся еще нам в дальнейшем (в главе об интеркомпонентной корреляции).

Рассмотрим теперь пару признаков, например I и V, для которых коэффициент корреляции $r_{15} = +0,85$ (в дальнейшем мы будем обозначать его просто буквой r без всяких цифровых значков). Выведем средние оценки H для всех пяти испытуемых по этим двум признакам: I и V.

Выше было указано, что ряд чисел H всегда варьирует в более узких пределах, чем оба ряда децисигмальных оценок d_1 и d_5 (см. стр. 226).

Покажем, что между средним квадратическим отклонением σ_H , вычисленным для ряда чисел H , и коэффициентом корреляции r между двумя рядами изолированных оценок d_1 и d_5 существует следующая постоянная зависимость:

$$\sigma_H = \sqrt{50(1+r)}.$$

В нашем примере для 5-ти чисел H :

$$-15 \quad 0 \quad -2,5 \quad +2,5 \quad +15$$

среднее квадратическое отклонение:

$$\sigma_H = \pm 9,62$$

(в то время как σ для ряда изолированных децисигмальных оценок d , как это уже было указано выше, всегда $= \pm 10$).

Испы- туемый	Оценки		Средняя оценка H
	d_1	d_5	
№ 1.	+15	+15	+15
№ 2.	+5	0	+2,5
№ 3.	0	-5	-2,5
№ 4.	-5	+5	0
№ 5.	-15	-15	-15

Но то же самое значение σ_H мы можем получить и по приведенной выше формуле (при $r = +0,85$):

$$\sigma_H = \sqrt{50(1+r)} = \sqrt{50(1+0,85)} = \pm 9,62,$$

что и подтверждает ее справедливость.

Итак, при объединении двух признаков средние пределы варьирования комплексных оценок H находятся в определенной зависимости от корреляции между этими признаками. Чем больше r , тем больше и σ_H и обратно.

В частности, при полной прямой корреляции между обоими признаками, т. е. при $r = +1$:

$$\sigma_H = \sqrt{50(1+r)} = \sqrt{50(1+1)} = \sqrt{100} = \pm 10.$$

В этом случае, следовательно, средние оценки будут варьировать совершенно так же, как варьируют изолированные оценки d .

При отсутствии корреляции, т. е. при $r = 0$:

$$\sigma_H = \sqrt{50(1+r)} = \sqrt{50(1+0)} = \sqrt{50} = \pm 7,07.$$

При полной обратной корреляции, т. е. при $r = -1$:

$$\sigma_H = \sqrt{50(1+r)} = \sqrt{50(1-1)} = \sqrt{0} = 0,$$

т. е. в этом случае средние оценки H не варьировали бы вовсе (сохраняли бы для всех испытуемых одну и ту же постоянную величину, всегда = 0).

В заключение на том же самом числовом примере рассмотрим зависимость между средней величиной всех дисгармонических отклонений Δ^*) и тем же коэффициентом корреляции r .

Испытуемый № 1 по обоим признакам (I и V) получил совершенно одинаковую оценку d_1 и d_5 , оказавшуюся равной +15 (см. стр. 231).

Поэтому и его средняя оценка H тоже = +15. В данном случае обе изолированные оценки d_1 и d_5 точно совпадают со средней оценкой H и, стало быть, оба дисгармонические отклонения Δ_1 и Δ_5 этих оценок равны здесь 0.

*) См. стр. 204.

У второго испытуемого $d_1 = +5$, $d_5 = 0$, а $H = +2,5$. Здесь каждая изолированная оценка отличается от средней оценки H на 2,5 децисигмы (одна из них больше H , а другая, наоборот, меньше). Поэтому для испытуемого № 2 его дисгармонические отклонения: $\Delta_1 = +2,5$ и $\Delta_5 = -2,5$.

Для испытуемого № 3 оценки $d_1 = 0$, $d_5 = -5$ и $H = -2,5$. Следовательно дисгармоническое отклонение этих оценок $\Delta_1 = +2,5$ и $\Delta_5 = -2,5$.

Нетрудно убедиться в том, что:

для испытуемого № 4 (при $H = 0$) $\Delta_1 = -5$ $\Delta_5 = +5$

" " № 5 (" $H = -15$) $\Delta_1 = 0$ и $\Delta_5 = 0$.

Итак, у отдельных испытуемых изолированные оценки d отличаются здесь от соответствующих средних оценок H на величину следующих дисгармонических отклонений:

$$0 \quad \pm 2,5 \quad \pm 2,5 \quad \pm 5 \quad 0.$$

Среднюю величину этих отклонений, в виду различия в их знаках (которые сами по себе нас не интересуют), удобно определить по способу, аналогичному вычислению σ , т. е. возвысить все Δ в квадрат, сложить эти квадраты, разделить их сумму на n и извлечь квадратный корень.

В данном случае квадраты Δ равны:

$$0 \quad 6,25 \quad 6,25 \quad 25 \quad 0,$$

а сумма квадратов:

$$\Sigma \Delta^2 = 37,5.$$

Разделив эту сумму на $n = 5$, получим:

$$\frac{\Sigma \Delta^2}{n} = \frac{37,5}{5} = 7,5.$$

Из этого числа следует извлечь теперь квадратный корень:

$$\sqrt{7,5} = \pm 2,74.$$

Эту среднюю величину различий между отдельными изолированными оценками и соответствующими им комплексными оценками H для всех n испытуемых мы обозначим здесь буквами δ (дельта). Итак в нашем примере:

$$\delta = \pm 2,74.$$

Между δ и r существует постоянная зависимость, определяемая формулой:

$$\delta = \pm \sqrt{50(1-r)}.$$

В нашем примере $r = +0,85$. Следовательно:

$$\delta = \pm \sqrt{50(1-0,85)} = \pm 2,74,$$

что мы и получили уже выше при непосредственном вычислении δ из отдельных дисгармонических отклонений Δ .

Итак средняя величина различий между изолированными и средними оценками всех испытуемых (т. е. δ) находится в определенной обратной зависимости от корреляции между объединяемыми признаками. Чем теснее эти признаки коррелируют друг с другом, тем точнее средняя оценка H характеризует у отдельных испытуемых обе их изолированные оценки вместе взятые.

В частности, при полной прямой корреляции этих признаков, т. е. при $r = +1$:

$$\delta = \sqrt{50(1-r)} = \sqrt{50(1-1)} = 0.$$

Следовательно при наличии строгой функциональной зависимости между двумя признаками все средние оценки H совершенно точно совпадают с обеими изолированными оценками d_1 и d_2 (различие между ними равно 0).

При $r = 0$:

$$\delta = \sqrt{50(1-r)} = \sqrt{50(1-0)} = \pm 7,07.$$

При $r = -1$:

$$\delta = \sqrt{50(1-r)} = \sqrt{50(1+1)} = \pm 10,$$

т. е. при полной обратной корреляции объединяемых признаков средняя величина различий между изолированными и средними оценками отдельных испытуемых достигает своего максимального значения = 10 децисигмам или одной сигме.

XXII. Интеркомпонентная корреляция.

Существование определенной зависимости между δ и r позволяет судить о том, в какой мере комплексная оценка двух признаков H может заменить собой обе изолированные их оценки, например d_1 и d_2 (см. предыдущую главу). Если средняя оценка комплекса в большинстве случаев будет мало отличаться от изолированных оценок отдельных его компонентов, такая замена их единой оценкой H может считаться законной и целесообразной. Если же различия между изолированными оценками d и общей оценкой H в среднем будут значительными — самое объединение отдельных компонентов такого разрозненного комплекса потеряет, очевидно, всякий практический смысл.

Поясним эту мысль следующим примером. Пусть какой-либо вид работы, например чтение или счет, оценивается нами с двух точек зрения: в отношении скорости и в отношении правильности. Оба эти признака, характеризуя различные стороны одного и того же процесса, могут, очевидно, получить здесь также и единую комплексную оценку H , которая в данном случае будет определять конечную продукцию этой работы в целом, независимо от тех или иных отдельных ее проявлений и особенностей в частности.

Предположим, что скорость и правильность этой работы коррелируют между собой очень сильно, т. е. быстрое ее исполнение чаще всего сопровождается также и малым количеством ошибок и обратно: медленный темп работы в большинстве случаев оказывается связанным с частыми и обильными ошибками. Сделаем допущение (конечно совершенно произвольное), что коэффициент корреляции между скоростью и правильностью:

$$r = +0,92.$$

Тогда на основании указанной выше зависимости между δ и r *)

$$\delta = \sqrt{50(1-r)} = \sqrt{50(1-0,92)} = \sqrt{50 \cdot 0,08} = \sqrt{4} = \pm 2.$$

Итак, при этих условиях (т. е. при $r = +0,92$) средняя величина различий между изолированными оценками пра-

*) См. стр. 234.

вильности и скорости с одной стороны и комплексными оценками H с другой равна здесь 2 децисигмам.

Пусть, в частности, какой-либо отдельный испытуемый получил среднюю оценку $H = +5$. Спрашивается, за счет каких двух изолированных оценок скорости и правильности или, вернее, за счет какой их комбинации могла получиться эта средняя оценка H ? Так как $\delta = \pm 2$, то изолированные оценки скорости и правильности работы отдельных испытуемых в среднем будут здесь колебаться в сравнительно узких пределах от $+7$ до $+3$ (эти две оценки и отличаются от $H = +5$ на 2 децисигмы). Поэтому у данного испытуемого его средняя оценка $H = +5$ могла получиться при объединении довольно близких друг к другу изолированных оценок, каждая из которых свидетельствует в сущности приблизительно об одном и том же уровне проявленного им успеха.

Но совершенно иную картину мы получили бы в предположении малой корреляции между скоростью и правильностью этой работы. Например при $r = +0,02$:

$$\delta = \sqrt{50(1-r)} = \sqrt{50(1-0,02)} = \sqrt{50 \cdot 0,98} = \sqrt{49} = \pm 7.$$

Стало быть испытуемый, получивший ту же самую среднюю оценку $H = +5$, мог иметь здесь две существенно различные изолированные оценки по отдельным компонентам этого комплекса: $+12$ и -2 (эти оценки и отличаются от $H = +5$ на величину ± 7 децисигм). При этом мы совершенно не знаем, которая из таких оценок относится к скорости работы, а которая к ее правильности. Таким образом средняя оценка H при слишком больших значениях δ теряет свое основное значение общей характеристики всего комплекса признаков в целом, так как не может уже заменить собою столь несходные друг с другом изолированные характеристики отдельных его компонентов.

Итак, объединение двух варьирующих признаков в одной общей их комплексной оценке H имеет реальный смысл и практическое значение только лишь при малых значениях δ , т. е. при достаточно большой корреляции между этими признаками.

Если коэффициент корреляции r близок к 0 или оказывается величиной отрицательной, объединение двух при-

знаков становится невозможным, и оценивать их приходится только порознь.

По аналогии с комплексом, состоящим всего лишь из двух компонентов, такой критерий объединения можно применить и к большему числу признаков, для которых предполагается вычислять средние или суммарные комплексные оценки H и S . Общий принцип здесь тот же самый: если все компоненты данного комплекса тесно связаны друг с другом, вычисление общих оценок H и S является весьма целесообразным; при слабой же корреляции между этими изолированными признаками объединение их становится уже невозможным.

Для суждения о большей или меньшей связи между многими варьирующими признаками установим понятие о „внутрикомплексном“ коэффициенте корреляции, определяющем зависимость между всеми компонентами данного комплекса в целом. В нашем числовом примере, неоднократно приводимом в нескольких последних главах (см. стр. 207, 222, и 230), мы получили следующие коэффициенты корреляции r между отдельными парами этих признаков (назовем эти коэффициенты „парными“):

	I	II	III	IV	V
I	—	+0,80	+0,60	+0,80	+0,85
II	+0,80	—	+0,80	+0,40	+0,55
III	+0,60	+0,80	—	0	+0,15
IV	+0,80	+0,40	0	—	+0,95
V	+0,85	+0,55	+0,15	+0,95	—

Здесь имеется всего 20 таких „парных“ коэффициентов корреляции r , причем все они повторяются по 2 раза. Следовательно неповторяющихся (различных) значений r здесь имеется только 10.

Эти 10 коэффициентов расположились в нашей таблице в правой верхней ее части. Рассматривая записи в этих клетках, мы замечаем, что коэффициенты r колеблются

в пределах от 0 до 0,95, оставаясь все время числами положительными. Отсюда следует, что отдельные компоненты данного комплекса различным образом коррелируют друг с другом (одни больше, другие меньше). Чтобы составить себе представление о том, как связаны эти компоненты друг с другом в среднем, можно вычислить среднее арифметическое из всех этих 10 „парных“ коэффициентов корреляции r . Это среднее арифметическое мы обозначим здесь буквою ρ (ро) со значком 0 (ноль), т. е. символом ρ_0 , которым и будем в дальнейшем обозначать „внутрикомплексный“ коэффициент корреляции для совокупности N варьирующих признаков (здесь $N=5$). В нашем примере сумма всех 10 „парных“ коэффициентов корреляции r , т. е.:

$$\Sigma r = +0,80 + 0,60 + 0,80 + 0,80 + 0,40 + 0 + 0,85 + 0,55 + 0,15 + 0,95 = +5,9,$$

следовательно:

$$\rho_0 = \frac{+5,9}{10} = +0,59.$$

Этот коэффициент показывает, что все 5 компонентов нашего комплекса в общем связаны друг с другом довольно тесно, т. е. что высокие оценки d по отдельным признакам чаще всего сопровождаются высокими же оценками и по большинству остальных признаков, точно также как и низкие оценки d у большинства испытуемых, как общее правило, получаются почти по всем признакам одновременно.

Наличие этого обстоятельства и дает нам право объединить все 5 изолированных оценок в одной общей комплексной оценке S или H , так как сравнительно высокий коэффициент ρ_0 гарантирует здесь достаточную однородность этих отдельных изолированных оценок d .

Однако столь сложный способ вычисления „внутрикомплексного“ коэффициента ρ_0 , получаемого в данном случае из 10 „парных“ коэффициентов r , практически является чрезвычайно обременительным.

Укажем другой, значительно более легкий способ определения ρ_0 с помощью среднего квадратического отклонения σ_0 , вычисленного для ряда суммарных оценок S . В на-

шем примере (см. табл.) эти суммарные оценки оказались следующими:

$$-65 \quad -15 \quad -5 \quad +30 \quad +55.$$

Испытуемые	Оценки						
	Изолированная					Комплексная	
	d_1	d_2	d_3	d_4	d_5	S	H
№ 1	+15	+5	+5	+15	+15	+55	+11
№ 2	+5	+15	+15	-5	0	+30	+6
№ 3	0	0	0	0	-5	-5	-1
№ 4	-5	-5	-15	+5	+5	-15	-3
№ 5	-15	-15	-5	-15	-15	-65	-13

Для данного ряда 5-ти чисел среднее квадратическое отклонение:

$$\sigma_0 = \sqrt{1680} = 40,99,$$

а квадрат этой величины:

$$\sigma_0^2 = 1680.$$

Имея в своем распоряжении этот квадрат сигмы (σ_0^2), мы можем вычислить ρ_0 по следующей упрощенной формуле:

$$\rho_0 = \frac{\sigma_0^2 - 100N}{100N(N-1)}$$

где через N обозначено общее количество признаков (у нас $N=5$).

Подставляя сюда наши числовые данные, получим:

$$\rho_0 = \frac{1680 - 100 \cdot 5}{100 \cdot 5 \cdot (5-1)} = \frac{1680 - 500}{500 \cdot 4} = \frac{+1180}{2000} = +0,59,$$

т. е. совершенно тот же результат, который мы имели и раньше.

Итак, при решении вопроса о целесообразности комплексной оценки n испытуемых по N изолированным признакам следует вычислить для каждого испытуемого его суммарную оценку S , затем определить среднее квадрати-

Подставляя сюда $\omega = +0,99$, $\sigma_0 = 40,99$ и $N = 5$, получим

$$\rho = \frac{40,99 \cdot (+0,99) - 10}{10(5-1)} = \frac{40,58 - 10}{40} = +0,76,$$

что вполне соответствует найденному ранее значению ρ при вычислении этого коэффициента, как среднего арифметического из 4-х „парных“ коэффициентов r .

Выполнив аналогичные вычисления для всех остальных признаков, получим следующие 5 групповых коэффициентов корреляции:

для I-го признака	$\rho = +0,76$
„ II „	$\rho = +0,64$
„ III „	$\rho = +0,39$
„ IV „	$\rho = +0,54$
„ V „	$\rho = +0,62$

Отсюда видно, что III признак (для которого коэффициент $\rho = +0,39$) менее всего гармонирует с группой всех остальных компонентов данного комплекса.

Полезно запомнить, что среднее арифметическое из всех „групповых“ коэффициентов корреляции ρ равняется „внутрикомплексному“ коэффициенту ρ_0 . В самом деле, сумма этих 5-ти „групповых“ коэффициентов:

$$\Sigma \rho = +0,76 + 0,64 + 0,39 + 0,54 + 0,62 = +2,96.$$

откуда внутрикомплексный коэффициент:

$$\rho_0 = \frac{\Sigma \rho}{N} = \frac{+2,96}{5} = +0,59,$$

что мы имели и раньше.

Эта формула может служить средством проверки коэффициентов ρ .

Если бы один из групповых коэффициентов по сравнению с остальными оказался слишком малым, то соответствующий признак, как негармонирующий с данным комплексом, следовало бы исключить и среднюю оценку H выводить только из децисигмальных оценок (d) оставшихся компонентов этого комплекса, тесно связанных со всеми другими.

В целях выяснения причин этих несоответствий можно обычным путем вычислить все „парные“ коэффициенты кор-

реляции между данным дисгармонирующим признаком и всеми остальными. В нашем примере в применении к III признаку, для которого групповой коэффициент $\rho = +0,39$, связь его с остальными 4-мя признаками этого комплекса определяется следующими „парными“ коэффициентами корреляции:

между III и I	$r_{31} = +0,60$
„ III „ II	$r_{32} = +0,80$
„ III „ IV	$r_{34} = 0$
„ III „ V	$r_{35} = +0,15$

Среднее арифметическое из этих 4-х чисел и дает здесь вычисленный ранее коэффициент ρ :

$$\rho = \frac{+0,60 + 0,80 + 0 + 0,15}{4} = \frac{+1,55}{4} = +0,39.$$

Аналогично мы могли бы поступить и с другими групповыми коэффициентами ρ , расчленив их на соответствующие „парные“ коэффициенты r , из которых они составлены.

Итак, данный метод „интеркомпонентной“ корреляции (т. е. корреляции между отдельными компонентами комплекса варьирующих признаков) позволяет устанавливать искомые связи сразу и непосредственно для всего комплекса в целом, переходя к деталям корреляционного анализа лишь постепенно и притом только по мере возникновения в том действительной надобности (в зависимости от результатов предыдущей фазы этого анализа).

Последовательность этих операций здесь, стало быть, такова. Сперва вычисляется „внутрикомплексный коэффициент корреляции“ ρ_0 . Если он оказывается достаточно большим, то на этом анализ материала и заканчивается. При малой величине ρ_0 вычисляются все групповые коэффициенты ρ . Те признаки, для которых коэффициенты ρ получаются слишком низкими, из данного комплекса исключаются, а для оставшихся признаков выводится их средняя оценка H (или весь комплекс разбивается на те или иные составные части, каждая из которых будет иметь достаточно высокий коэффициент ρ_0). Если при этом возникает надобность в подробном выяснении связи между отдельными ком-

понентами целого комплекса, то вычисляются соответствующие „парные“ коэффициенты корреляции r .

О том, какую экономию времени и труда вносит этот порядок вычислений по сравнению с обычным, можно видеть из следующего примера. Пусть мы имеем комплекс признаков, состоящий из 11 компонентов (например совокупность 11 антропометрических признаков или ряд психотехнических испытаний по 11 направлениям). Для общей характеристики связи между этими компонентами мы должны были бы каждый признак прокоррелировать со всеми остальными, т. е. вычислить 55 „парных“ коэффициентов корреляции r и из этих 55 чисел вычислить среднее арифметическое. Этим сложным путем мы и получили бы „внутрикомплексный“ коэффициент корреляции r_0 . При пользовании же описанным выше упрощенным приемом представляется возможным вовсе избежать определения таких „парных“ коэффициентов r , сведя эту работу к вычислению лишь одного среднего квадратического отклонения (σ_0) для ряда суммарных оценок S . Точно также и при вычислении групповых коэффициентов r взамен предварительного определения 10 „парных“ коэффициентов r мы можем найти лишь один „вспомогательный“ коэффициент корреляции ω между оценками d и S . Отдельные же „парные“ коэффициенты r нам приходится вычислять здесь только тогда, когда в этом возникает действительная надобность.

В приведенных выше примерах общее число объектов индивидуальной оценки было крайне незначительным ($n=5$). В нормальных условиях при больших значениях n все эти вычисления следует производить по способу сумм, разбивая ряды оценок d и S на классы (от — до) и выполняя все обычные операции, связанные с определением среднего квадратического отклонения (σ_0) и коэффициента корреляции (ω). При этом следует иметь в виду, что для точности конечных значений коэффициентов r_0 и r необходимо выбирать классовый промежуток (k) как можно мельче.

Если в результате интеркомпонентного анализа общий (генеральный) комплекс признаков был разбит нами на несколько составных частей (частных комплексов), представляется иногда интересным определить зависимость также и между каждыми двумя такими частными комплексами

путем вычисления „междукомплексных“ коэффициентов корреляции r_{12} (или r_{13} и т. д.).

Пусть в нашем числовом примере весь генеральный комплекс, состоящий из 5 признаков, был разбит на 2 частных комплекса, причем в первый из них вошли два признака: I и II, а в другой — три остальные, т. е. III, IV и V.

Чтобы охарактеризовать зависимость между этими частными комплексами, выпишем все „парные“ коэффициенты корреляции r между признаками, входящими в I-й комплекс, и признаками, входящими во 2-й комплекс. В нашем примере (см. стр. 237) коэффициенты корреляции между:

I и III признаками	$r_{13} = +0,60$
I „ IV „	$r_{14} = +0,80$
I „ V „	$r_{15} = +0,85$
II „ III „	$r_{23} = +0,80$
II „ IV „	$r_{24} = +0,40$
II „ V „	$r_{25} = +0,55$

Среднее арифметическое из этих 6 чисел мы и примем здесь за „междукомплексный“ коэффициент корреляции для 1-го и 2-го частного комплекса (r_{12}). Так как сумма этих 6-ти „парных“ коэффициентов r :

$$\Sigma r = +0,60 + 0,80 + 0,85 + 0,80 + 0,40 + 0,55 = +4,$$

то среднее арифметическое:

$$r_{12} = \frac{+4}{6} = +0,67.$$

Итак, в данном случае „междукомплексный“ коэффициент корреляции r_{12} оказался выше „внутрикомплексного“ коэффициента $r_0 = +0,59$, вычисленного для всего генерального комплекса, состоящего из 5-ти признаков.

Взамен такого сложного способа определения r_{12} , требующего предварительного вычисления 6-ти „парных“ коэффициентов r , можно применить здесь следующий упрощенный прием:

Вычислим суммарные оценки S_1 и S_2 для обоих заданных выше частных комплексов (см. стр. 239):

I-ый частн. комплекс.

Изолир. оценка		Суммарн. оценка S_1
d_1	d_2	
+15	+5	+20
+5	+15	+20
0	0	0
-5	-5	-10
-15	-15	-30

II-ой частн. комплекс.

Изолир. оценка			Суммарн. оценка S_2
d_3	d_4	d_5	
+5	+15	+15	+35
+15	-5	0	+10
0	0	-5	-5
-15	+5	+5	-5
-5	-15	-15	-35

Если среднее квадратическое отклонение суммарных оценок для первого частного комплекса обозначить через σ_1 , для второго частного комплекса — через σ_2 , число компонентов, входящих в состав первого комплекса — через N_1 , число компонентов, входящих в состав второго комплекса — через N_2 , а „вспомогательный“ коэффициент корреляции между двумя рядами суммарных оценок S_1 и S_2 — через ω_{12} , то искомый „междукомплесный“ коэффициент корреляции ρ_{12} может быть вычислен по формуле:

$$\rho_{12} = \frac{\sigma_1 \sigma_2 \omega_{12}}{100 N_1 N_2}$$

В нашем примере мы имеем следующие два ряда суммарных оценок:

оценки S_1 : -30 -10 0 +20 +20
оценки S_2 : -35 -5 -5 +10 +35.

Среднее квадратическое отклонение для 5-ти чисел первого ряда:

$$\sigma_1 = \sqrt{360} = \pm 18,97,$$

а для чисел второго ряда:

$$\sigma_2 = \sqrt{520} = \pm 22,80.$$

Коэффициент корреляции между двумя рядами этих чисел:

$$\omega_{12} = +0,92.$$

Так как в первый комплекс входят два компонента, а во второй — три, то:

$$N_1 = 2 \quad N_2 = 3.$$

Подставляя эти данные в указанную выше формулу, получим:

$$\rho_{12} = \frac{18,97 \cdot 22,80 (+0,92)}{100 \cdot 2 \cdot 3} = +0,66.$$

Этот результат очень близок к полученному ранее. Разница в последнем знаке объясняется здесь той неточностью, которая была допущена нами при вычислении σ_1 , σ_2 и ω_{12} лишь с 2-мя десятичными знаками.

Если по ходу работы наряду с отдельными „междукомплесными“ коэффициентами корреляции ρ_{12} , ρ_{23} , ρ_{13} , ρ_{24} и т. д. желательно охарактеризовать также и связь каждого отдельного частного комплекса со всем генеральным комплексом в целом, можно поступить следующим образом:

Пусть, например, частный комплекс состоит из первых трех признаков I, II и III. „Парные“ коэффициенты корреляции r , определяющие связь этих трех признаков с остальными (т. е. с IV и V), имеют у нас следующее значение (см. стр. 237):

$$\begin{array}{lll} r_{14} = +0,80 & r_{24} = +0,40 & r_{34} = +0 \\ r_{15} = +0,85 & r_{25} = +0,55 & r_{35} = +0,15. \end{array}$$

Среднее арифметическое из этих 6-ти чисел мы и примем за „комплесный“ коэффициент корреляции ρ_{01} , характеризующий связь 3-х компонентов данного частного комплекса со всеми остальными (в данном случае, следовательно, только с 2-мя) компонентами генерального комплекса.

Сумма этих парных коэффициентов r , т. е.

$$\Sigma r = +0,80 + 0,85 + 0,40 + 0,55 + 0 + 0,15 = +2,75.$$

Следовательно искомый „комплесный“ коэффициент:

$$\rho_{01} = \frac{+2,75}{6} = +0,46.$$

Этот же коэффициент можно получить и упрощенным способом, прокоррелировав суммарные оценки S , вычисленные для всех N компонентов генерального комплекса, с суммарными оценками S_1 , вычисленными для тех же испытуемых по N_1 компонентам, входящим в состав только данного частного комплекса (у нас $N=5$, а $N_1=3$).

d_1	d_2	d_3	S_1
+ 15	+ 5	+ 5	+ 25
+ 5	+ 15	+ 15	+ 35
0	0	0	0
- 5	- 5	- 15	- 25
- 15	- 15	- 5	- 35

В нашем примере:

оценки S : - 65 - 15 - 5 + 30 + 55
оценки S_1 : - 35 - 25 0 + 35 + 25.

Среднее квадратическое отклонение для чисел первого ряда:

$$\sigma_0 = 40,99,$$

а для чисел второго ряда:

$$\sigma_1 = 27,20.$$

Вспомогательный коэффициент корреляции между числами этих рядов:

$$\omega_{01} = +0,91.$$

С помощью этих данных искомый „комплексный“ коэффициент ρ_{01} может быть найден по следующей формуле:

$$\rho_{01} = \frac{\sigma_1 (\sigma_0 \omega_{01} - \sigma_1)}{100 N_1 (N - N_1)}.$$

В нашем примере, следовательно:

$$\rho_{01} = \frac{27,20 (40,99 \cdot 0,91 - 27,20)}{100 \cdot 3 (5 - 3)} = +0,46,$$

что мы имели и раньше.

„Внутрикомплексный“ коэффициент корреляции ρ_1 (или ρ_2 или ρ_3 и т. д.), характеризующий взаимную связь между всеми N_1 компонентами данного частного комплекса, может быть вычислен по аналогии с разобранным в начале этой главы „внутрикомплексным“ коэффициентом ρ_0 , относящимся

ко всему генеральному комплексу в целом (см. стр. 239), т. е. по формуле:

$$\rho_1 = \frac{\sigma_1^2 - 100 N_1}{100 N_1 (N_1 - 1)}.$$

В дальнейшем этот частный случай мы не будем упоминать вовсе.

Итак, „внутрикомплексный“ коэффициент корреляции ρ_0 , „групповой“ коэффициент ρ , „междуконкомплексный“ ρ_{12} (или ρ_{13} , ρ_{23} ... и т. д.) и „комплексный“ ρ_{01} , (или ρ_{02} , ρ_{03} и т. д.) вычисляются по следующим основным формулам:

$$\rho_0 = \frac{\sigma_0^2 - 100 N}{100 N (N - 1)}$$

$$\rho = \frac{\sigma_0 \omega - 10}{10 (N - 1)}$$

$$\rho_{12} = \frac{\sigma_1 \sigma_2 \omega_{12}}{100 N_1 N_2}$$

$$\rho_{01} = \frac{\sigma_1 (\sigma_0 \omega_{01} - \sigma_1)}{100 N_1 (N - N_1)}.$$

При этом все вычисления следует производить по способу сумм с разбивкой соответствующих рядов изолированных и комплексных оценок (d и S) на возможно большее число мелких классов (с малым классовым промежутком k).

Само собой разумеется, что вычисление коэффициентов ρ_0 , ρ , ρ_{12} и ρ_{01} удобно производить лишь в том случае, когда все изолированные оценки (d) независимо от этого уже ранее были получены в самом процессе работы. Специальное вычисление этих оценок только для интеркомпонентной корреляции является конечно нецелесообразным.

Средние ошибки (m) всех этих коэффициентов, которые по аналогии с самими коэффициентами ρ мы обозначим символами m_0 , m , m_{12} и m_{01} , могут быть вычислены по следующим приближенным формулам (дающим результат с избытком):

$$m_0 = \pm 2 \sqrt{\frac{1 - \rho_0^2}{n N (N - 1)}}$$

$$m_{12} = \pm \sqrt{\frac{2(1 - \rho_{12}^2)}{n N_1 N_2}}$$

$$m = \pm \sqrt{\frac{2(1 - \rho^2)}{n (N - 1)}}$$

$$m_{01} = \pm \sqrt{\frac{2(1 - \rho_{01}^2)}{n N_1 (N - N_1)}}.$$

Метод интеркомпонентной корреляции применим и к перцентильным оценкам (варьирующим в пределах от 0 до 100).

Общая схема вычислений остается здесь без изменений; также определяются суммарные оценки S , средние оценки H , среднее квадратическое отклонение суммарных оценок σ_0 (или σ_1 , σ_2 и т. д.) и вспомогательные коэффициенты корреляции ω между соответствующими рядами изолированных и комплексных оценок (d и S). Разница заключается лишь в конечных формулах, которые в данном случае будут иметь следующий вид:

Внутрикомплексный коэффициент:

$$\rho_0 = \frac{12 \sigma_0^2 - 10\,000 N}{10\,000 N(N-1)}.$$

Групповой коэффициент:

$$\rho = \frac{\sqrt{12} \sigma_0 \omega - 100}{100 (N-1)}.$$

Междукомплексный коэффициент:

$$\rho_{12} = \frac{12 \sigma_1 \sigma_2 \omega_{12}}{10\,000 N_1 N_2}.$$

Комплексный коэффициент:

$$\rho_{01} = \frac{12 \sigma_1 (\sigma_0 \omega_{01} - \sigma_1)}{10\,000 N_1 (N - N_1)}.$$

Но в данном случае имеется одна существенная принципиальная разница. Дело в том, что коррелируя децисигмальные оценки, мы вместе с тем выясняем и корреляцию первичных данных (вследствие линейной зависимости, существующей между результатами непосредственных измерений и их децисигмальными оценками), в то время как оперируя с оценками перцентиллярными, мы можем обнаружить лишь ту или иную зависимость между этими оценками, как таковыми, связь же между первичными вариантами таким путем уловлена быть не может.

XXIII. Особые случаи интеркомпонентной корреляции.

В предыдущей главе мы рассмотрели способы вычисления коэффициентов ρ_0 , ρ , ρ_{12} и ρ_{01} с помощью средних квадратических отклонений σ_0 , σ_1 и σ_2 и вспомогательных коэффициентов корреляции ω , ω_{12} и ω_{01} , выведенных из суммарных комплексных оценок S , S_1 и S_2 . Такой путь вычисления имеет свои выгодные стороны, так как в результате интеркомпонентного анализа может выясниться нецелесообразность объединения всех компонентов данного комплекса, и надобность в выводе средних оценок H в этом случае совершенно отпадет. Чтобы избежать лишней работы, мы и оперировали все время только с суммарными оценками S , из которых средние оценки H легко могут быть получены в дальнейшем посредством простого деления S на число объединяемых признаков N .

Однако в тех случаях, когда средние оценки H были уже выведены для всех n испытуемых, можно определить интеркомпонентные коэффициенты ρ_0 , ρ , ρ_{12} и ρ_{01} , также и при помощи средних оценок H . Для данного случая (сохраняя прежние обозначения) будем иметь следующие формулы:

$$\begin{aligned} \rho_0 &= \frac{N \sigma_0^2 - 100}{100 (N-1)} \\ \rho &= \frac{N \sigma_0 \omega - 10}{10 (N-1)} \\ \rho_{12} &= \frac{\sigma_1 \sigma_2 \omega_{12}}{100} \\ \rho_{01} &= \frac{\sigma_1 (N \sigma_0 \omega_{01} - N_1 \sigma_1)}{100 (N - N_1)}. \end{aligned}$$

В применении к перцентиллярным оценкам эти формулы будут иметь следующий вид:

$$\begin{aligned} \rho_0 &= \frac{12 N \sigma_0^2 - 10\,000}{10\,000 (N-1)} \\ \rho &= \frac{\sqrt{12} N \sigma_0 \omega - 100}{100 (N-1)} \\ \rho_{12} &= \frac{12 \sigma_1 \sigma_2 \omega_{12}}{10\,000} \\ \rho_{01} &= \frac{12 \sigma_1 (N \sigma_0 \omega_{01} - N_1 \sigma_1)}{10\,000 (N - N_1)}. \end{aligned}$$

Во всех этих случаях, следовательно, средние квадратические отклонения σ_0 , σ_1 , σ_2 и т. д. выводятся из соответствующих рядов средних оценок H (а не суммарных S , как раньше). Вспомогательные же коэффициенты корреляции ω , ω_{12} , ω_{01} и т. д. в обоих случаях окажутся одинаковыми.

Все эти формулы, как вполне аналогичные по смыслу и применению предыдущим (указанным в прошлой главе для суммарных оценок S), не требуют, очевидно, никаких дополнительных разъяснений и числовых иллюстраций.

Рассмотрим теперь самый общий случай интеркомпонентной корреляции, когда при выводе суммарных и средних оценок были приняты во внимание также и различные (неодинаковые) коэффициенты значимости q (см. стр. 203) отдельных компонентов этого комплекса (в разобранных выше примерах эти коэффициенты предполагались совершенно одинаковыми, всюду равными 1).

Пусть, например, для комплекса 5-ти признаков, много раз приводимого в предыдущих главах (см., напр., стр. 239), мы имеем следующие коэффициенты значимости q :

для I признака	$q_1 = 1$
" II "	$q_2 = 2$
" III "	$q_3 = 2$
" IV "	$q_4 = 0$
" V "	$q_5 = 3$

При этих условиях за „внутрикомплексный“ коэффициент корреляции r_0 мы можем принять „взвешенное“ среднее арифметическое из всех „парных“ коэффициентов корреляции r с „весами“, пропорциональными произведению соответствующих коэффициентов значимости q .

В нашем примере мы имели следующие 10 „парных“ коэффициентов корреляции r между отдельными признаками или 20 коэффициентов, принимая во внимание двукратную их повторяемость. (См. табл. на стр. 253).

При вычислении „взвешенного“ среднего из всех 10 чисел, записанных в клетках правой верхней части этой таблицы, следует еще до суммирования помножить каждое из них на соответствующий „вес“ p , равный произведению двух коэффициентов значимости тех признаков, к которым относится данный „парный“ коэффициент r .

	I	II	III	IV	V
I . . .	—	+ 0,80	+ 0,60	+ 0,80	+ 0,85
II . . .	+ 0,80	—	+ 0,80	+ 0,40	+ 0,55
III . . .	+ 0,60	+ 0,80	—	0	+ 0,15
IV . . .	+ 0,80	+ 0,40	0	—	+ 0,95
V . . .	+ 0,85	+ 0,55	+ 0,15	+ 0,95	—

Так, например, коэффициент корреляции между I и II признаком:

$$r_{12} = + 0,80$$

следует помножить на произведение коэффициентов значимости I-го и II-го признака q_1 и q_2 . У нас $q_1 = 1$, $q_2 = 2$; следовательно произведение их:

$$q_1 q_2 = 1 \cdot 2 = 2.$$

Это число (2), которое мы обозначим здесь символом p_{12} , и является „весом“ коэффициента корреляции r_{12} .

Помножив коэффициент r_{12} на его „вес“ p_{12} , получим:

$$p_{12} r_{12} = 2 \cdot (+ 0,80) = + 1,60.$$

Для следующего коэффициента корреляции $r_{13} = + 0,60$ (между I и III признаком) коэффициенты значимости $q_1 = 1$ и $q_3 = 2$. Следовательно „вес“ этого коэффициента:

$$p_{13} = q_1 q_3 = 1 \cdot 2 = 2.$$

Помножив $r_{13} = + 0,60$ на его „вес“ $p_{13} = 2$, получим:

$$p_{13} r_{13} = + 1,20.$$

Поступая аналогично со всеми 10-ю коэффициентами корреляции r , получим следующий расчет:

$r_{12} = + 0,80$	$p_{12} = q_1 q_2 = 1 \cdot 2 = 2$	$p_{12} r_{12} = 2 (+ 0,80) = + 1,60$
$r_{13} = + 0,60$	$p_{13} = q_1 q_3 = 1 \cdot 2 = 2$	$p_{13} r_{13} = 2 (+ 0,60) = + 1,20$
$r_{14} = + 0,80$	$p_{14} = q_1 q_4 = 1 \cdot 0 = 0$	$p_{14} r_{14} = 0 (+ 0,80) = 0$
$r_{15} = + 0,85$	$p_{15} = q_1 q_5 = 1 \cdot 3 = 3$	$p_{15} r_{15} = 3 (+ 0,85) = + 2,55$
$r_{23} = + 0,80$	$p_{23} = q_2 q_3 = 2 \cdot 2 = 4$	$p_{23} r_{23} = 4 (+ 0,80) = + 3,20$
$r_{24} = + 0,40$	$p_{24} = q_2 q_4 = 2 \cdot 0 = 0$	$p_{24} r_{24} = 0 (+ 0,40) = 0$

$$\begin{array}{lll}
 r_{25} = +0,55 & p_{25} = q_2 q_5 = 2 \cdot 3 = 6 & p_{25} r_{25} = 6(+0,55) = +3,30 \\
 r_{34} = 0 & p_{34} = q_3 q_4 = 2 \cdot 0 = 0 & p_{34} r_{34} = 0(0) = 0 \\
 r_{35} = +0,15 & p_{35} = q_3 q_5 = 2 \cdot 3 = 6 & p_{35} r_{35} = 6(+0,15) = +0,90 \\
 r_{45} = +0,95 & p_{45} = q_4 q_5 = 0 \cdot 3 = 0 & p_{45} r_{45} = 0(+0,95) = 0.
 \end{array}$$

Сложив все произведения „весов“ p на коэффициенты корреляции r , получим в сумме:

$$\Sigma pr = +1,60 + 1,20 + 0 + 2,55 + 3,20 + 0 + 3,30 + 0 + 0,90 + 0 = +12,75.$$

Эту сумму произведений p на r , т. е. Σpr , следует теперь разделить на сумму всех „весов“ p , т. е. на Σp . У нас:

$$\Sigma p = 2 + 2 + 0 + 3 + 4 + 0 + 6 + 0 + 6 + 0 = 23.$$

Разделив $\Sigma pr = +12,75$ на $\Sigma p = 23$, мы и получим „взвешенное“ среднее из всех 10 „парных“ коэффициентов корреляции r , т. е. искомый „внутрикомплексный“ коэффициент корреляции ρ_0 . В нашем примере, следовательно:

$$\rho_0 = \frac{\Sigma pr}{\Sigma p} = \frac{12,75}{23} = +0,55.$$

Тот же коэффициент ρ_0 можно получить и не прибегая к столь сложным вычислениям, требующим предварительного определения всех 10 „парных“ коэффициентов корреляции r . Для этого нам нужно иметь суммарные оценки S , вычисленные с учетом коэффициентов значимости q („взвешенные“ суммарные оценки S).

Соответствующие вычисления представлены в нижеследующей оценочной ведомости:

I признак $q_1 = 1$		II признак $q_2 = 2$		III признак $q_3 = 2$		IV признак $q_4 = 0$		V признак $q_5 = 3$		$S = \Sigma qd$
d_1	$q_1 d_1$	d_2	$q_2 d_2$	d_3	$q_3 d_3$	d_4	$q_4 d_4$	d_5	$q_5 d_5$	
+15	+15	+5	+10	+5	+10	+15	0	+15	+45	+80
+5	+5	+15	+30	+15	+30	-5	0	0	0	+65
0	0	0	0	0	0	0	0	-5	-15	-15
-5	-5	-5	-10	-15	-30	+5	0	+5	+15	-30
-15	-15	-15	-30	-5	-10	-15	0	-15	-45	-100

Здесь под заголовками признаков проставлены коэффициенты значимости q , в рубрике d выписаны децисимальные оценки, а рядом в рубрике qd — произведения этих оценок d на соответствующие коэффициенты значимости q . Сложив в каждой горизонтальной строчке все произведения d на q (т. е. числа столбцов qd), мы и получим „взвешенные“ суммарные оценки S , записанные в крайних справа клетках этих строчек (т. е. в крайнем правом столбце, озаглавленном буквой S).

Так, например, для самой верхней строчки (соответствующей оценкам испытуемого № 1) эта сумма произведений, т. е. „взвешенная“ суммарная оценка:

$$S = +15 + 10 + 10 + 0 + 45 = +80.$$

Для следующих строчек эти оценки S равны:

$$+65 \quad -15 \quad -30 \quad \text{и} \quad -100.$$

Для данного ряда взвешенных суммарных оценок S мы можем вычислить среднее квадратическое отклонение σ_0 . В нашем примере для 5-ти чисел:

$$+80 \quad +65 \quad -15 \quad -30 \quad -100^*)$$

среднее квадратическое отклонение:

$$\sigma_0 = \sqrt{4350} = \pm 65,95, \\ \text{а } \sigma_0^2 = 4350.$$

При помощи этой величины (σ_0^2) мы и можем теперь вычислить „взвешенный внутрикомплексный“ коэффициент корреляции ρ_0 по формуле:

$$\rho_0 = \frac{\sigma_0^2 - 100 \Sigma q^2}{100 (\Sigma^2 q - \Sigma q^2)}.$$

В этой формуле символом Σq^2 обозначена сумма квадратов всех коэффициентов значимости q , а символом $\Sigma^2 q$ — квадрат суммы этих коэффициентов. В нашем примере коэффициенты значимости q имели следующие числовые значения:

$$1 \quad 2 \quad 2 \quad 0 \quad 3,$$

*) Сумма их всегда равна нулю:

$$\Sigma S = 0$$

а квадраты их (q^2), следовательно:

$$1 \quad 4 \quad 4 \quad 0 \quad 9,$$

откуда сумма этих квадратов, т. е.:

$$\Sigma q^2 = 1 + 4 + 4 + 0 + 9 = 18.$$

Так как сумма всех коэффициентов значимости (q):

$$\Sigma q = 1 + 2 + 2 + 0 + 3 = 8,$$

то квадрат этой суммы, т. е.:

$$\Sigma^2 q = 8^2 = 64.$$

Итак, в нашем примере:

$$\Sigma q^2 = 18, \text{ а } \Sigma^2 q = 64.$$

Подставляя эти данные в указанную выше формулу, получим:

$$\rho_0 = \frac{4350 - 100 \cdot 18}{100(64 - 18)} = +0,55,$$

что мы имели и раньше.

Покажем теперь способ вычисления взвешенного группового коэффициента ρ . Пусть, например, этот коэффициент требуется определить для I-го признака. Корреляция этого признака с группой всех остальных определяется следующими 4-мя „парными“ коэффициентами r (см. табл. на стр. 253):

$$\begin{array}{ll} r_{12} = +0,80 & \text{с „весом“ } p_{12} = 2 \text{ (так как } q_1 = 1, \text{ а } q_2 = 2) \\ r_{13} = +0,60 & \text{„ „ } p_{13} = 2 \text{ („ „ } q_1 = 1, \text{ „ } q_3 = 2) \\ r_{14} = +0,80 & \text{„ „ } p_{14} = 0 \text{ („ „ } q_1 = 1, \text{ „ } q_4 = 0) \\ r_{15} = +0,85 & \text{„ „ } p_{15} = 3 \text{ („ „ } q_1 = 1, \text{ „ } q_5 = 3). \end{array}$$

Сумма произведений всех коэффициентов r на их веса p , т. е.:

$$\Sigma pr = 2(+0,80) + 2(+0,60) + 0(+0,80) + 3(+0,85) = +1,60 + 1,20 + 0 + 2,55 = +5,35.$$

Сумма всех весов p , т. е.:

$$\Sigma p = 2 + 2 + 0 + 3 = 7.$$

Разделив $\Sigma pr = +5,35$ на $\Sigma p = 7$, мы и получим „взвешенное“ среднее арифметическое для всех 4-х чисел r , принимаемое здесь за групповой коэффициент ρ . Итак:

$$\rho = \frac{\Sigma pr}{\Sigma p} = \frac{+5,35}{7} = +0,76. *)$$

Более простой способ определения ρ заключается в вычислении вспомогательного коэффициента корреляции ω между рядом децисигмальных оценок d данного изолированного признака (в нашем примере следовательно I-го признака) и взвешенными суммарными оценками S .

Прокоррелировав эти два ряда чисел (см. стр. 254):

$$\begin{array}{rrrrr} S: & -100 & -30 & -15 & +65 & +80 \\ d: & -15 & -5 & 0 & +5 & +15, \end{array}$$

получим „вспомогательный“ коэффициент:

$$\omega = +0,96.$$

С помощью этого коэффициента корреляции (ω) мы и можем теперь вычислить взвешенный групповой коэффициент ρ по формуле:

$$\rho = \frac{\sigma_0 \omega - 10 q}{10(\Sigma q - q)}.$$

Здесь сумма всех коэффициентов значимости $\Sigma q = 8$, а данный коэффициент значимости (I признака) $q = 1$. Кроме того, ранее (см. стран. 255) мы вычислили $\sigma_0 = \pm 65,95$.

Подставляя эти данные в указанную выше формулу, получим:

$$\rho = \frac{65,95(+0,96) - 10 \cdot 1}{10(8 - 1)} = +0,76,$$

что мы имели и раньше.

Взвешенный „междукомплексный“ коэффициент корреляции ρ_{12} , определяющий зависимость между отдельными компонентами двух частных комплексов, мы можем вычислить... как взвешенное среднее из соответствующих парных коэффициентов r .

*) Взвешенный групповой коэффициент ρ можно было бы получить и по формуле:

$$= \frac{\Sigma q r}{\Sigma q}$$

Пусть, например, первый комплекс состоит из I и II признака, а второй — из III, IV и V.

„Парные“ коэффициенты r и их веса p будут иметь здесь следующие значения (см. табл. на стр. 253):

$r_{13} = +0,60$	$p_{13} = 2$	(так как $q_1 = 1$, а $q_3 = 2$)
$r_{14} = +0,80$	$p_{14} = 0$	(„ „ $q_1 = 1$, „ $q_4 = 0$)
$r_{15} = +0,85$	$p_{15} = 3$	(„ „ $q_1 = 1$, „ $q_5 = 3$)
$r_{23} = +0,80$	$p_{23} = 4$	(„ „ $q_2 = 2$, „ $q_3 = 2$)
$r_{24} = +0,40$	$p_{24} = 0$	(„ „ $q_2 = 2$, „ $q_4 = 0$)
$r_{25} = +0,55$	$p_{25} = 6$	(„ „ $q_2 = 2$, „ $q_5 = 3$)

Перемножив эти коэффициенты r на соответствующие им веса p и сложив все эти произведения, получим в сумме:

$$\Sigma pr = +10,25.$$

Сложив все 6 весов p , получим в сумме:

$$\Sigma p = 15.$$

Откуда взвешенный междуконфлексный коэффициент корреляции:

$$r_{12} = \frac{\Sigma pr}{\Sigma p} = \frac{+10,25}{15} = +0,68.$$

Для получения этого же коэффициента r_{12} сокращенным способом необходимо иметь два ряда взвешенных суммарных оценок S_1 и S_2 по каждому из частных комплексов отдельно.

Необходимые для этого вычисления выполнены в ниже следующей оценочной ведомости:

Для 1-го частного комплекса:

I призн. $q_1 = 1$		II призн. $q_2 = 2$		S_1
d_1	$q_1 d_1$	d_2	$q_2 d_2$	
+15	+15	+5	+10	+25
+5	+5	+15	+30	+35
0	0	0	0	0
-5	-5	-5	-10	-15
-15	-15	-15	-30	-45

Для 2-го частн. комплекса:

III призн. $q_3 = 2$		IV призн. $q_4 = 0$		V призн. $q_5 = 3$		S_2
d_3	$q_3 d_3$	d_4	$q_4 d_4$	d_5	$q_5 d_5$	
+5	+10	+15	0	+15	+45	+55
+15	+30	-5	0	0	0	+30
0	0	0	0	-5	-15	-15
-15	-30	+5	0	+5	+15	-15
-5	-10	-15	0	-15	-45	-55

Прокоррелировав два ряда суммарных оценок:

$$\begin{aligned} S_1: & -45 \quad -15 \quad 0 \quad +35 \quad +25 \\ S_2: & -55 \quad -15 \quad -15 \quad +30 \quad +55, \end{aligned}$$

получим вспомогательный коэффициент корреляции:

$$\omega_{12} = +0,93$$

и оба средние квадратические отклонения этих рядов:

$$\sigma_1 = \pm 28,64$$

$$\sigma_2 = \pm 38,47.$$

С помощью этих данных можно вычислить взвешенный междуконфлексный коэффициент r_{12} по формуле:

$$r_{12} = \frac{\sigma_1 \sigma_2 \omega_{12}}{100 \Sigma q_1 \Sigma q_2}.$$

В нашем примере сумма коэффициентов значимости двух признаков, входящих в 1-ый комплекс:

$$\Sigma q_1 = 1 + 2 = 3,$$

а сумма этих коэффициентов для трех признаков 2-го комплекса:

$$\Sigma q_2 = 2 + 0 + 3 = 5.$$

Подставляя эти данные в указанную выше формулу, получим:

$$r_{12} = \frac{28,64 \cdot 38,47 (+0,93)}{100 \cdot 3 \cdot 5} = +0,68,$$

что мы имели и раньше.

Взвешенный комплексный коэффициент r_{01} , определяющий зависимость между компонентами данного частного комплекса и всеми остальными компонентами входящими в состав генерального комплекса, представляет собой взвешенное среднее из соответствующих парных коэффициентов корреляции r . Пусть, например, частный комплекс состоит из первых трех признаков: I, II и III. Связь этих признаков с двумя остальными (IV и V) определяется здесь следующими 6-тью коэффициентами r с их весами p (см. табл. на стр. 253):

$r_{14} = +0,80$	$p_{14} = 0$	(так как $q_1 = 1$, а $q_4 = 0$)
$r_{15} = +0,85$	$p_{15} = 3$	(„ „ $q_1 = 1$, „ $q_5 = 3$)
$r_{24} = +0,40$	$p_{24} = 0$	(„ „ $q_2 = 2$, „ $q_4 = 0$)
$r_{25} = +0,55$	$p_{25} = 6$	(„ „ $q_2 = 2$, „ $q_5 = 3$)
$r_{34} = +0$	$p_{34} = 0$	(„ „ $q_3 = 2$, „ $q_4 = 0$)
$r_{35} = +0,15$	$p_{35} = 6$	(„ „ $q_3 = 2$, „ $q_5 = 3$)

Перемножив r на p и сложив все эти произведения, получим в сумме:

$$\Sigma pr = +6,75.$$

Сумма же всех 6-ти весов p , т. е.:

$$\Sigma p = 15.$$

Разделив $\Sigma pr = +6,75$ на $\Sigma p = 15$, мы и получим взвешенный комплексный коэффициент корреляции:

$$\rho_{01} = \frac{\Sigma pr}{\Sigma p} = \frac{+6,75}{15} = +0,45.$$

Для вычисления коэффициента ρ_{01} упрощенным способом необходимо иметь ряд взвешенных суммарных оценок S_1 по трем признакам данного частного комплекса. Вычисление их представлено в следующей оценочной ведомости:

I призн. $q_1 = 1$		II призн. $q_2 = 2$		III призн. $q_3 = 2$		S_1
d_1	$q_1 d_1$	d_2	$q_2 d_2$	d_3	$q_3 d_3$	
+15	+15	+5	+10	+5	+10	+35
+5	+5	+15	+30	+15	+30	+65
0	0	0	0	0	0	0
-5	-5	-5	-10	-15	-30	-45
-15	-15	-15	-30	-5	-10	-55

Прокоррелировав эти оценки S_1 с суммарными оценками генерального комплекса S (см. стр. 254):

$$\begin{array}{l} S_1: \quad -55 \quad -45 \quad 0 \quad +65 \quad +35 \\ S: \quad -100 \quad -30 \quad -15 \quad +65 \quad +80, \end{array}$$

получим следующий вспомогательный коэффициент корреляции:

$$\omega_{01} = +0,92$$

и оба средние квадратические отклонения:

$$\begin{array}{l} \sigma_1 = \pm 45,83 \\ \sigma_0 = \pm 65,95. \end{array}$$

С помощью этих данных можно вычислить коэффициент ρ_{01} по формуле:

$$\rho_{01} = \frac{\sigma_1 (\sigma_0 \omega_{01} - \sigma_1)}{100 \Sigma q_1 (\Sigma q - \Sigma q_1)}.$$

Здесь сумма коэффициентов значимости трех признаков данного частного комплекса:

$$\Sigma q_1 = 1 + 2 + 2 = 5,$$

а сумма коэффициентов значимости всех 5-ти признаков, входящих в состав генерального комплекса:

$$\Sigma q = 1 + 2 + 2 + 0 + 3 = 8.$$

Следовательно:

$$\rho_{01} = \frac{45,83 (65,95 \cdot 0,92 - 45,83)}{100 \cdot 5 (8 - 5)} = +0,45,$$

что мы имели и раньше.

Итак, для вычисления „взвешенных“ коэффициентов ρ_0 , ρ , ρ_{12} и ρ_{01} служат следующие основные формулы:

$$\rho_0 = \frac{\sigma_0^2 - 100 \Sigma q^2}{100 (\Sigma^2 q - \Sigma q^2)}.$$

$$\rho = \frac{\sigma_0 \omega - 10 q}{10 (\Sigma q - q)}.$$

$$\rho_{12} = \frac{\sigma_1 \sigma_2 \omega_{12}}{100 \Sigma q_1 \Sigma q_2}.$$

$$\rho_{01} = \frac{\sigma_1 (\sigma_0 \omega_{01} - \sigma_1)}{100 \Sigma q_1 (\Sigma q - \Sigma q_1)}.$$

В применении к перцентильным оценкам эти формулы будут иметь следующий вид:

$$\rho_0 = \frac{12 \sigma_0^2 - 10\,000 \Sigma q^2}{10\,000 (\Sigma^2 q - \Sigma q^2)}.$$

$$\rho = \frac{\sqrt{12} \sigma_0 \omega - 100 q}{100 (\Sigma q - q)}.$$

$$\rho_{12} = \frac{12 \sigma_1 \sigma_2 \omega_{12}}{10\,000 \Sigma q_1 \Sigma q_2}.$$

$$\rho_{01} = \frac{12 \sigma_1 (\sigma_0 \omega_{01} - \sigma_1)}{10\,000 \Sigma q_1 (\Sigma q - \Sigma q_1)}.$$

В тех случаях, когда для генерального комплекса или для отдельных частных комплексов наряду со взвешенными суммарными оценками S были вычислены также и все

средние оценки H путем деления S на сумму всех коэффициентов значимости Σq , т. е. по формуле:

$$H = \frac{S}{\Sigma q}$$

эти же интеркомпонентные коэффициенты ρ_0 , ρ , ρ_{12} и ρ_{01} для случая децисигмальной оценки могут быть получены по следующим формулам:

$$\rho_0 = \frac{\Sigma^2 q \sigma_0^2 - 100 \Sigma q^2}{100 (\Sigma^2 q - \Sigma q^2)}$$

$$\rho = \frac{\Sigma q \sigma_0 \omega - 10 q}{10 (\Sigma q - q)}$$

$$\rho_{12} = \frac{\sigma_1 \sigma_2 \omega_{12}}{100}$$

$$\rho_{01} = \frac{\sigma_1 (\Sigma q \sigma_0 \omega_{01} - \Sigma q_1 \sigma_1)}{100 (\Sigma q - \Sigma q_1)}$$

где средние квадратические отклонения σ_0 , σ_1 , σ_2 и т. д. вычисляются из соответствующих рядов средних оценок H (а не суммарных S).

В применении же к перцентилярным оценкам эти формулы будут иметь следующий вид:

$$\rho_0 = \frac{12 \Sigma^2 q \sigma_0^2 - 10\,000 \Sigma q^2}{10\,000 (\Sigma^2 q - \Sigma q^2)}$$

$$\rho = \frac{\sqrt{12} \Sigma q \sigma_0 \omega - 100 q}{100 (\Sigma q - q)}$$

$$\rho_{12} = \frac{12 \sigma_1 \sigma_2 \omega_{12}}{10\,000}$$

$$\rho_{01} = \frac{12 \sigma_1 (\Sigma q \sigma_0 \omega_{01} - \Sigma q_1 \sigma_1)}{10\,000 (\Sigma q - \Sigma q_1)}$$

где σ_0 , σ_1 , σ_2 и т. д. вычисляются также из средних оценок H (а не из суммарных S).

XXIV. Корреляционный анализ неоднородного материала.

В прошлых главах мы рассмотрели методы комплексной оценки и интеркомпонентного анализа такого материала, который варьировал только лишь по признакам, подлежащим нашей оценке, а во всех остальных отношениях предполагался вполне однородным. Однако в действительности это условие, очевидно, соблюдается далеко не всегда и не

в полной мере. Любая группа объектов антропометрических измерений или психотехнических испытаний чаще всего различается также и по ряду других признаков, например по полу, возрасту и проч. Если эти различия окажутся значительными, то игнорирование их в процессе интеркомпонентного анализа может привести к весьма грубым ошибкам и заведомо неправильным выводам.

Для иллюстрации этих соображений приведем два числовых примера.

Пусть при каком-либо психотехническом обследовании, состоящем, скажем, из трех отдельных испытаний ($N=3$), мы имели 15 испытуемых ($n=15$), распа-
дающихся по своему возрасту на три группы: старшую (напр. от 16 до 14 лет), среднюю

(от 13 до 11 лет) и младшую (от 10 до 8 лет) по 5 человек в каждой такой группе.

Предположим, что в результате этого обследования мы получили следующие данные, выраженные в абсолютных числах (в количестве секунд, потраченных на одну операцию; в числе операций, выполненных в единицу времени; в количестве правильно решенных заданий и проч.).

Проанализируем этот материал, не принимая пока во внимание деления его на возрастные группы.

Возраст	Испытуемый	Испытания		
		I	II	III
Старший	№ 1	60	11	40
	№ 2	56	7	30
	№ 3	54	8	25
	№ 4	52	9	20
	№ 5	48	5	10
Средний	№ 6	36	23	30
	№ 7	32	19	40
	№ 8	30	20	20
	№ 9	28	21	25
	№ 10	24	17	10
Младший	№ 11	12	35	40
	№ 12	8	31	25
	№ 13	6	32	30
	№ 14	4	33	20
	№ 15	0	29	10

Для этого прежде всего вычислим среднее арифметическое M и среднее квадратическое отклонение σ по каждому тесту отдельно.

Проделав обычные выкладки с числами трех заданных вертикальных столбцов, в конечном итоге получим:

$$\begin{aligned} \text{для I-го испытания: } M_1 &= 30 & \sigma_1 &= \pm 20 \\ \text{„ II-го „ } M_2 &= 20 & \sigma_2 &= \pm 10 \\ \text{„ III-го „ } M_3 &= 25 & \sigma_3 &= \pm 10. \end{aligned}$$

Произведя с помощью этих стандартов децисигмальную оценку каждого из 15-ти испытуемых по всем трем тестам, получим результаты, занесенные в оценочную ведомость на стр. 265.

В целях выяснения вопроса о целесообразности комплексной оценки этих испытуемых по совокупности всех трех тестов одновременно, вычислим внутрикомплексный коэффициент корреляции ρ_0 .

В нашем примере сумма квадратов суммарных оценок S , т. е.:

$$\Sigma S^2 = 2676,$$

число признаков $N=3$, а число испытуемых $n=15$. Следовательно среднее квадратическое отклонение для ряда суммарных оценок (S):

$$\sigma_0 = \sqrt{\frac{2676}{15}} = \sqrt{178,4},$$

а квадрат этой величины:

$$\sigma_0^2 = 178,4.$$

Подставляя эти данные в формулу для вычисления внутрикомплексного коэффициента корреляции (см. стр. 239):

$$\rho_0 = \frac{\sigma_0^2 - 100 N}{100 N (N - 1)},$$

получим в числах:

$$\rho_0 = \frac{178,4 - 100 \cdot 3}{100 \cdot 3 (3 - 1)} = -0,20.$$

Отрицательный коэффициент ρ_0 указывает на то, что при данном способе оценки (без учета возрастных различий этих 15-ти испытуемых) связь между тремя заданными призна-

Испы- туемые	И с п ы т а н и я						Суммарн. оценка	
	I		II		III		S	S ²
	V ₁	d ₁	V ₂	d ₂	V ₃	d ₃		
№ 1	60	+ 15	11	- 9	40	+ 15	+ 21	441
№ 2	56	+ 13	7	- 13	30	+ 5	+ 5	25
№ 3	54	+ 12	8	- 12	25	0	0	0
№ 4	52	+ 11	9	- 11	20	- 5	- 5	25
№ 5	48	+ 9	5	- 15	10	- 15	- 21	441
№ 6	36	+ 3	23	+ 3	30	+ 5	+ 11	121
№ 7	32	+ 1	19	- 1	40	+ 15	+ 15	225
№ 8	30	0	20	0	20	- 5	- 5	25
№ 9	28	- 1	21	+ 1	25	0	0	0
№ 10	24	- 3	17	- 3	10	- 15	- 21	441
№ 11	12	- 9	35	+ 15	40	+ 15	+ 21	441
№ 12	8	- 11	31	+ 11	25	0	0	0
№ 13	6	- 12	32	+ 12	30	+ 5	+ 5	25
№ 14	4	- 13	33	+ 13	20	- 5	- 5	25
№ 15	0	- 15	29	+ 9	10	- 15	- 21	441
		M ₁ = 30	M ₂ = 20	M ₃ = 25	Σ S ² = 2676			
		σ ₁ = ± 20	σ ₂ = ± 10	σ ₃ = ± 10				

ками в среднем имеет очень слабый (даже обратный) характер, что несомненно лишает нас права оценивать этих индивидуумов по всем трем тестам одновременно.

Однако совершенно иная картина получится здесь при оценке отдельных испытуемых по их возрастным стандартам. Вычислив M и σ для отдельных тестов по каждому возрасту отдельно, получим следующие результаты:

	I испытание:	II испытание:	III испытание:
Старш. группа	$M = 54 \quad \sigma = \pm 4$	$M = 8 \quad \sigma = \pm 2$	$M = 25 \quad \sigma = \pm 10$
Средн. группа	$M = 30 \quad \sigma = \pm 4$	$M = 20 \quad \sigma = \pm 2$	$M = 25 \quad \sigma = \pm 10$
Младш. группа	$M = 6 \quad \sigma = \pm 4$	$M = 32 \quad \sigma = \pm 2$	$M = 25 \quad \sigma = \pm 10$

Просматривая эту таблицу, мы замечаем, что средние арифметические (M) по I-му тесту вместе с возрастом увеличиваются, по II-му тесту, наоборот, уменьшаются, а по III-му — остаются без изменения, в то время как средние квадратические отклонения (σ) в различных возрастных группах сохраняют одно и то же неизменное значение. Эти зависящие от возраста разнохарактерные изменения средних арифметических M очевидно и затемнили собою общую картину имеющейся здесь корреляции, искусственно придав ей обратный характер.

Произведя переоценку всех испытуемых по этим новым возрастным критериям, получим следующие три оценочные ведомости:

Для старшей группы:

Испы- туемые	И с п ы т а н и я						Суммарн. оценка	
	I		II		III		S	S ²
	V ₁	d ₁	V ₂	d ₂	V ₃	d ₃		
№ 1	60	+15	11	+15	40	+15	+45	2 025
№ 2	56	+5	7	-5	30	+5	+5	25
№ 3	54	0	8	0	25	0	0	0
№ 4	52	-5	9	+5	20	-5	-5	25
№ 5	48	-15	5	-15	10	-15	-45	2 025
	$M_1 = 54$ $\sigma_1 = \pm 4$		$M_2 = 8$ $\sigma_2 = \pm 2$		$M_3 = 25$ $\sigma_3 = \pm 10$		$\Sigma S^2 = 4 100$	

Здесь при $N=3$, $n=5$ и $\Sigma S^2 = 4 100$:

$$\sigma_0 = \sqrt{\frac{4 100}{5}} = \sqrt{820}$$

и следовательно: $\sigma_0^2 = 820$,

откуда искомый коэффициент:

$$\rho_0 = \frac{\sigma_0^2 - 100N}{100N(N-1)} = \frac{820 - 100 \cdot 3}{100 \cdot 3(3-1)} = +0,87.$$

Для средней группы:

Испы- туемые	И с п ы т а н и я						Суммарн. оценка	
	I		II		III		S	S ²
	V ₁	d ₁	V ₂	d ₂	V ₃	d ₃		
№ 6	36	+15	23	+15	30	+5	+35	1 225
№ 7	32	+5	19	-5	40	+15	+15	225
№ 8	30	0	20	0	20	-5	-5	25
№ 9	28	-5	21	+5	25	0	0	0
№ 10	24	-15	17	-15	10	-15	-45	2 025
	$M_1 = 30$ $\sigma_1 = \pm 4$		$M_2 = 20$ $\sigma_2 = \pm 2$		$M_3 = 25$ $\sigma_3 = \pm 10$		$\Sigma S^2 = 3 500$	

Здесь при $N=3$, $n=5$ и $\Sigma S^2 = 3 500$:

$$\sigma_0 = \sqrt{\frac{3 500}{5}} = \sqrt{700}$$

и следовательно: $\sigma_0^2 = 700$,

откуда: $\rho_0 = \frac{\sigma_0^2 - 100N}{100N(N-1)} = \frac{700 - 100 \cdot 3}{100 \cdot 3(3-1)} = +0,67.$

Для младшей группы:

Испы- туемые	И с п ы т а н и я						Суммарн. оценка	
	I		II		III		S	S ²
	V ₁	d ₁	V ₂	d ₂	V ₃	d ₃		
№ 11	12	+ 15	35	+ 15	40	+ 15	+ 45	2 025
№ 12	8	+ 5	31	— 5	25	0	0	0
№ 13	6	0	32	0	30	+ 5	+ 5	25
№ 14	4	— 5	33	+ 5	20	— 5	— 5	25
№ 15	0	— 15	29	— 15	10	— 15	— 45	2 025
	M ₁ = 6		M ₂ = 32		M ₃ = 25		Σ S ² = 4 100	
	σ ₁ = ± 4		σ ₂ = ± 2		σ ₃ = ± 10			

Здесь при $N = 3$, $n = 5$ и $\Sigma S^2 = 4 100$:

$$\sigma_0 = \sqrt{\frac{4 100}{5}} = \sqrt{820}$$

и следовательно: $\sigma_0^2 = 820$,

$$\text{откуда: } \rho_0 = \frac{\sigma_0^2 - 100 N}{100 N(N-1)} = \frac{820 - 100 \cdot 3}{100 \cdot 3(3-1)} = +0,87.$$

Итак, при группировке нашего материала по возрастам мы получили следующие значения внутрикомплексных коэффициентов корреляции:

$$\begin{aligned} \text{для старшей группы} & \quad \rho_0 = +0,87 \\ \text{„ средней „} & \quad \rho_0 = +0,67 \\ \text{„ младшей „} & \quad \rho_0 = +0,87. \end{aligned}$$

Столь высокие коэффициенты ρ_0 указывают на достаточно тесную связь между результатами всех трех испытаний (при условии оценки их по возрастным критериям), что и дает нам право в данном случае выводить для отдельных испытуемых их комплексные оценки S или H (но уже по данным возрастной оценки).

Если бы мы произвели все эти вычисления в одной общей для всех 15 испытуемых ведомости (но все же по

возрастным, а не по общим критериям оценки), то в конечном результате получили бы внутрикомплексный коэффициент корреляции ρ_0 , равный среднему арифметическому из 3-х коэффициентов ρ_0 , только что выведенных для отдельных возрастов (+0,87, +0,67 и +0,87):

$$\rho_0 = \frac{0,87 + 0,67 + 0,87}{3} = \frac{+2,41}{3} = +0,80.$$

Этот расчет представлен в нижеследующей ведомости:

Возраст	Испы- туемые	И с п ы т а н и я						Сумм. оценка	
		I		II		III		S	S ²
		V ₁	d ₁	V ₂	d ₂	V ₃	d ₃		
Старший	№ 1	60	+ 15	11	+ 15	40	+ 15	+ 45	2 025
	№ 2	56	+ 5	7	- 5	30	+ 5	+ 5	25
	№ 3	54	0	8	0	25	0	0	0
	№ 4	52	- 5	9	+ 5	20	- 5	- 5	25
	№ 5	48	- 15	5	- 15	10	- 15	- 45	2 025
Средний	№ 6	36	+ 15	23	+ 15	30	+ 5	+ 35	1 225
	№ 7	32	+ 5	19	- 5	40	+ 15	+ 15	225
	№ 8	30	0	20	0	20	- 5	- 5	25
	№ 9	28	- 5	21	+ 5	25	0	0	0
	№ 10	24	- 15	17	- 15	10	- 15	- 45	2 025
Младший	№ 11	12	+ 15	35	+ 15	40	+ 15	+ 45	2 025
	№ 12	8	+ 5	31	- 5	25	0	0	0
	№ 13	6	0	32	0	30	+ 5	+ 5	25
	№ 14	4	- 5	33	+ 5	20	- 5	- 5	25
	№ 15	0	- 15	29	- 15	10	- 15	- 45	2 025

Здесь при $N=3$, $n=15$ и $\Sigma S^2=11700$:

$$\sigma_0 = \sqrt{\frac{11700}{15}} = \sqrt{780}$$

и следовательно: $\sigma_0^2 = 780$,

$$\text{откуда: } \rho_0 = \frac{\sigma_0^2 - 100N}{100N(N-1)} = \frac{780 - 100 \cdot 3}{100 \cdot 3(3-1)} = +0,80,$$

что вполне оправдывает высказанное выше положение.

Если бы отдельные испытуемые распределились по возрастным группам неравномерно, то этот общий для всего анализируемого материала коэффициент ρ_0 оказался бы равен „взвешенному“ среднему из соответствующих коэффициентов, вычисленных по возрастам. Так, если при тех же числовых значениях этих коэффициентов (+0,87, +0,67 и +0,87) в старшую группу попали бы, скажем, 4 испытуемых, в среднюю — 8, а в младшую — 3, то коэффициент ρ_0 оказался бы равным сумме произведений отдельных возрастных коэффициентов на численность соответствующих групп, разделенной на общее число всех испытуемых:

$$\begin{aligned} \rho_0 &= \frac{(+0,87) \cdot 4 + (+0,67) \cdot 8 + (+0,87) \cdot 3}{4 + 8 + 3} = \\ &= \frac{+3,48 + 5,36 + 2,61}{15} = \frac{+11,45}{15} = +0,76, \end{aligned}$$

а не +0,80, как это мы имели ранее при совершенно одинаковой численности всех возрастных групп.

В приведенном выше примере внутрикомплексный коэффициент корреляции ρ_0 вследствие неучета возрастных группировок оказался сильно преуменьшенным и при внесении соответствующей поправки заметно увеличился. Мыслим, очевидно, и обратный случай, когда оценка по возрастным или иным критериям в конечном результате исправит этот коэффициент ρ_0 в сторону его уменьшения.

Пусть, например, имеются следующие первичные данные, обработанные по обычным правилам в приведенной ниже оценочной ведомости без учета возрастных группировок:

Испы- туемые	И с п ы т а н и я						Суммарн. оценка	
	I		II		III		S	S²
	V ₁	d ₁	V ₂	d ₂	V ₃	d ₃		
№ 1	60	+ 15	31	+ 11	40	+ 15	+ 41	1 681
№ 2	56	+ 13	35	+ 15	36	+ 11	+ 39	1 521
№ 3	54	+ 12	32	+ 12	37	+ 12	+ 36	1 296
№ 4	52	+ 11	29	+ 9	38	+ 13	+ 33	1 089
№ 5	48	+ 9	33	+ 13	34	+ 9	+ 31	961
№ 6	36	+ 3	19	— 1	28	+ 3	+ 5	25
№ 7	32	+ 1	23	+ 3	24	— 1	+ 3	9
№ 8	30	0	20	0	25	0	0	0
№ 9	28	— 1	17	— 3	26	+ 1	— 3	9
№ 10	24	— 3	21	+ 1	22	— 3	— 5	25
№ 11	12	— 9	7	— 13	16	— 9	— 31	961
№ 12	8	— 11	11	— 9	12	— 13	— 33	1 089
№ 13	6	— 12	8	— 12	13	— 12	— 36	1 296
№ 14	4	— 13	5	— 15	14	— 11	— 39	1 521
№ 15	0	— 15	9	— 11	10	— 15	— 41	1 681
	M ₁ = 30		M ₂ = 20		M ₃ = 25		Σ S² = 13 164	
	σ ₁ = ± 20		σ ₂ = ± 10		σ ₃ = ± 10			

Здесь при $N=3$, $n=15$ и $\Sigma S^2=13164$:

$$\sigma_0 = \sqrt{\frac{13164}{15}} = \sqrt{877,6} \text{ и следоват. } \sigma_0^2 = 877,6, \text{ откуда:}$$

$$\rho_0 = \frac{\sigma_0^2 - 100N}{100N(N-1)} = \frac{877,6 - 100 \cdot 3}{100 \cdot 3(3-1)} = +0,96.$$

При обработке этих же данных по трем возрастным группам отдельно получим следующие значения M и σ :

	I испытание	II испытание	III испытание
Старшая группа . .	$M = 54 \sigma = \pm 4$	$M = 32 \sigma = \pm 2$	$M = 37 \sigma = \pm 2$
Средняя группа . .	$M = 30 \sigma = \pm 4$	$M = 20 \sigma = \pm 2$	$M = 25 \sigma = \pm 2$
Младшая группа . .	$M = 6 \sigma = \pm 4$	$M = 8 \sigma = \pm 2$	$M = 13 \sigma = \pm 2$

Здесь во всех трех испытаниях средние арифметические M изменяются вместе с возрастом, что указывает на наличие прямой корреляции между возрастом и результатами отдельных испытаний. Таким образом полученный нами высокий коэффициент $\rho_0 = +0,96$ свидетельствует здесь не столько о внутренней связи между тремя компонентами заданного комплекса признаков, сколько о связи каждого из этих компонентов порознь с четвертым, неучтенным нами признаком (возрастом), влияние которого нам и предстоит теперь устранить.

Произведя переоценку всех вариант V по новым возрастным критериям M и σ и выполнив обычные вычисления, получим следующую оценочную ведомость (см. стр. 273).

Здесь при $N=3$, $n=15$ и $\Sigma S^2=5100$:

$$\sigma_0 = \sqrt{\frac{5100}{15}} = \sqrt{340} \text{ и следоват. } \sigma_0^2 = 340.$$

$$\rho_{01} = \frac{\sigma_0^2 - 100N}{100N(N-1)} = \frac{340 - 100 \cdot 3}{100 \cdot 3(3-1)} = +0,07.$$

Таким образом, устранив влияние возраста, мы получили очень слабую зависимость между отдельными компонентами заданного комплекса признаков.

Подобно внутрикомплексному коэффициенту ρ_0 такие же изменения, зависящие от группировки материала по различным дополнительным признакам, могут иметь место и по отношению к другим интеркомпонентным коэффициентам: ρ , ρ_{12} и ρ_{01} , а также и по отношению к отдельным „парным“ коэффициентам корреляции r .

Выясним, например, характер изменения „парного“ коэффициента корреляции r_{23} между II и III признаком в связи с учетом возрастных группировок. В первом из двух приведенных выше числовых примеров (см. стран. 265) мы имели следующие децисигмальные оценки d_2 и d_3 по II и III при-

Возраст	Испытуемые	Испытания						Суммарн. оценка	
		I		II		III		S	S ²
		V ₁	d ₁	V ₂	d ₂	V ₃	d ₃		
Старший	№ 1	60	+15	31	-5	40	+15	+25	625
	№ 2	56	+5	35	+15	36	-5	+15	225
	№ 3	54	0	32	0	37	0	0	0
	№ 4	52	-5	29	-15	38	+5	-15	225
	№ 5	48	-15	33	+5	34	-15	-25	625
Средний	№ 6	35	+15	19	-5	28	+15	+25	625
	№ 7	32	+5	23	+15	24	-5	+15	225
	№ 8	30	0	20	0	25	0	0	0
	№ 9	28	-5	17	-15	26	+5	-15	225
	№ 10	24	-15	21	+5	22	-15	-25	625
Младший	№ 11	12	+15	7	-5	16	+15	+25	625
	№ 12	8	+5	11	+15	12	-5	+15	225
	№ 13	6	0	8	0	13	0	0	0
	№ 14	4	-5	5	-15	14	+5	-15	225
	№ 15	0	-15	9	+5	10	-15	-25	625

знаку для всех 15 испытуемых при оценке результатов их испытания без учета возрастных группировок (см. стр. 274).

Перемножив эти оценки d_2 и d_3 попарно и сложив все полученные произведения, получим в сумме:

$$\Sigma d_2 d_3 = +210,$$

откуда, пользуясь упрощенной формулой (см. стран. 228),

$$r_{23} = \frac{\Sigma d_2 d_3}{100n} = \frac{+210}{100 \cdot 15} = +0,14.$$

Испы- туемые	d_2	d_3	$d_2 d_3$
№ 1	- 9	+ 15	- 135
№ 2	- 13	+ 5	- 65
№ 3	- 12	0	0
№ 4	- 11	- 5	+ 55
№ 5	- 15	- 15	+ 225
№ 6	+ 3	+ 5	+ 15
№ 7	- 1	+ 15	- 15
№ 8	0	- 5	0
№ 9	+ 1	0	0
№ 10	- 3	- 15	+ 45
№ 11	+ 15	+ 15	+ 225
№ 12	+ 11	0	0
№ 13	+ 12	+ 5	+ 60
№ 14	+ 13	- 5	- 65
№ 15	+ 9	- 15	- 135
$\Sigma d_2 d_3 =$			+ 210

Вычислив же эти коэффициенты для каждой из трех возрастных групп отдельно, получим следующий расчет:

Младшая группа:

d_2	d_3	$d_2 d_3$
+ 15	+ 15	+ 225
- 5	0	0
0	+ 5	0
+ 5	- 5	- 25
- 15	- 15	+ 225

$$\Sigma d_2 d_3 = + 425$$

Средняя группа:

d_2	d_3	$d_2 d_3$
+ 15	+ 5	+ 75
- 5	+ 15	- 75
0	- 5	0
+ 5	0	0
- 15	- 15	+ 225

$$\Sigma d_2 d_3 = + 225$$

Старшая группа:

d_2	d_3	$d_2 d_3$
+ 15	+ 15	+ 225
- 5	+ 5	- 25
0	0	0
+ 5	- 5	- 25
- 15	- 15	+ 225

$$\Sigma d_2 d_3 = + 400$$

Здесь при $n=5$ в отдельных группах получаются следующие так называемые „парциальные“ (частные) коэффициенты корреляции r :

Для старшей группы (от 16 до 14 лет):

$$r_{23} = \frac{\Sigma d_2 d_3}{100 n} = \frac{+ 400}{100 \cdot 5} = + 0,80.$$

Для средней группы (от 13 до 11 лет):

$$r_{23} = \frac{\Sigma d_2 d_3}{100 n} = \frac{+ 225}{100 \cdot 5} = + 0,45.$$

Для младшей группы (от 10 до 8 лет):

$$r_{23} = \frac{\Sigma d_2 d_3}{100 n} = \frac{+ 425}{100 \cdot 5} = + 0,85.$$

Если бы весь этот расчет (по новым возрастным критериям оценки) был выполнен в одной общей для всех 15-ти испытуемых оценочной ведомости, то сумма всех произведений d_2 на d_3 оказалась бы равной + 1050, откуда при $n=15$ и $\Sigma d_2 d_3 = + 1050$:

$$r_{23} = \frac{\Sigma d_2 d_3}{100 n} = \frac{+ 1050}{100 \cdot 15} = + 0,70.$$

Это же значение r_{23} можно получить и как среднее арифметическое из всех трех вычисленных ранее „парциальных“ коэффициентов корреляции:

$$r_{23} = \frac{+ 0,80 + 0,45 + 0,85}{3} = \frac{+ 2,10}{3} = + 0,70.$$

Этот средний коэффициент обычно для краткости также называется „парциальным“.

Итак, при определении корреляционной зависимости между II и III признаком в группе 15 объектов исследования, различающихся между собою также и по некоторому дополнительному признаку (возрасту), мы получили несколько искаженную (затемненную) картину искомой зависимости, характеризующейся величиной коэффициента корреляции:

$$r_{23} = + 0,14.$$

После же устранения влияния со стороны этого дополнительного фактора (т. е. как бы при полном уравнивании

возраста всех 15 испытуемых) искомая зависимость выявилась в своем чистом виде, и исправленный „парциальный“ коэффициент корреляции оказался значительно выше предыдущего:

$$r_{23} = +0,70.$$

В тех случаях, когда известными являются также и оба коэффициента корреляции между двумя данными признаками и третьим дополнительным, влияние со стороны которого требуется устранить, этот „парциальный“ коэффициент может быть вычислен и проще, не прибегая к переоценке всех вариантов с помощью новых (групповых) стандартов M и σ .

Вычислим коэффициенты корреляции между возрастом испытуемых с одной стороны и результатами II и III-го их испытания с другой.

Обозначив результаты II-го испытания через V_x , результаты III-го испытания — через V_y , а возраст испытуемых — через V_z , получим следующий обычный расчет (см. расчетную решетку). Здесь возраст от 16 до 14 лет условно принят равным в среднем 15 годам, возраст от 13 до 11 — 12 годам, а возраст от 10 до 8 лет — 9 годам.

В нашем примере (см. табл. на стр. 277) средние арифметические:

$$M_x = 20$$

$$M_z = 12.$$

Суммы квадратов центральных отклонений:

$$\Sigma \bar{x}^2 = 1500$$

$$\Sigma z^2 = 90.$$

Сумма произведений центральных отклонений:

$$\Sigma xz = -360.$$

Следовательно коэффициент корреляции между результатами II-го испытания (V_x) и возрастом испытуемых (V_z):

$$r_{xz} = \frac{\Sigma xz}{\sqrt{\Sigma \bar{x}^2 \cdot \Sigma z^2}} = \frac{-360}{\sqrt{1500 \cdot 90}} = +0,98.$$

V_x	V_z	x	x^2	z	z^2	xz
11	15	- 9	81	+ 3	9	- 27
7	15	- 13	169	+ 3	9	- 39
8	15	- 12	144	+ 3	9	- 36
9	15	- 11	121	+ 3	9	- 33
5	15	- 15	225	+ 3	9	- 45
23	12	+ 3	9	0	0	0
19	12	- 1	1	0	0	0
20	12	0	0	0	0	0
21	12	+ 1	1	0	0	0
17	12	- 3	9	0	0	0
35	9	+ 15	225	- 3	9	- 45
31	9	+ 11	121	- 3	9	- 33
32	9	+ 12	144	- 3	9	- 36
33	9	+ 13	169	- 3	9	- 39
29	9	+ 9	81	- 3	9	- 27
—	—	0	1500	0	90	- 360

Аналогичным расчетом легко убедиться в том, что коэффициент корреляции между результатами III испытания (V_y) и возрастом испытуемых (V_z), т. е.:

$$r_{yz} = 0.$$

Ранее же (см. стран. 273) мы установили, что коэффициент корреляции между результатами II-го испытания (V_x) и результатами III-го испытания (V_y), т. е.:

$$r_{xy} = +0,14$$

(выше мы обозначали этот коэффициент символом r_{23}).

При этих условиях искомый „парциальный“ коэффициент корреляции между результатами II-го и III-го испытания (при постоянном возрасте), обозначаемый обычно символом $r_{xy.z}$, может быть вычислен по следующей формуле:

$$r_{xy.z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}.$$

Подставляя в эту формулу найденные выше значения всех трех коэффициентов корреляции r_{xy} , r_{xz} и r_{yz} , будем иметь:

$$\begin{aligned} r_{xy.z} &= \frac{+0,14 - (-0,98) \cdot 0}{\sqrt{[1 - (-0,98)^2] \cdot [1 - 0]}} = \frac{+0,14 - 0}{\sqrt{(1 - 0,96) \cdot 1}} = \\ &= \frac{+0,14}{\sqrt{0,04}} = \frac{0,14}{0,2} = +0,70, \end{aligned}$$

что мы имели и раньше.

Заметим, кстати, что эта общая формула для вычисления „парциального“ коэффициента корреляции $r_{xy.z}$ с помощью трех коэффициентов: r_{xy} , r_{xz} и r_{yz} дает абсолютно точные результаты только лишь при соблюдении двух определенных условий, искусственно выполненных при составлении приведенного выше числового примера.

Первое из этих условий заключается в изменении средних арифметических всех трех коррелирующих признаков (V_x , V_y и V_z) на одну и ту же постоянную величину от группы к группе *). Вспомним, что средние арифметические (M) числовых результатов II-го испытания в нашем примере имели следующее значение:

$$\begin{aligned} \text{для старшей группы: } M &= 32 \\ \text{„ средней „ } M &= 20 \\ \text{„ младшей „ } M &= 8. \end{aligned}$$

Следовательно здесь эти средние изменяются от группы к группе на постоянную величину = 12.

*) Или на величины, пропорциональные друг другу во всех трех признаках.

Точно также и средние арифметические результатов III-го испытания:

$$\begin{aligned} \text{для старшей группы: } M &= 25 \\ \text{„ средней „ } M &= 25 \\ \text{„ младшей „ } M &= 25 \end{aligned}$$

отличаются друг от друга на постоянную величину, в данном случае = 0.

Далее, средние значения третьего признака (возраста):

$$\begin{aligned} \text{для старшей группы (от 16 до 14 лет) в среднем } &15 \text{ лет} \\ \text{„ средней „ („ 13 „ 11 „) „ } &12 \text{ „} \\ \text{„ младшей „ („ 10 „ 8 „) „ } &9 \text{ „} \end{aligned}$$

в смежных группах отличаются друг от друга также на постоянную величину = 3.

Второе условие заключается в равенстве (в пределах каждого признака) всех средних квадратических отклонений (σ), вычисленных для отдельных возрастных групп.

Вспомним, что в нашем примере для II-го испытания во всех трех группах:

$$\sigma = \pm 10$$

и для III-го испытания:

$$\sigma = \pm 10$$

(для разных испытаний эти сигмы могли бы оказаться и различными, лишь бы они были одинаковыми во всех возрастных группах для одного и того же признака).

Так как, далее, третий признак (возраст) внутри каждой отдельной группы вовсе не варьировал, то и здесь все три сигмы являются величинами постоянными, в данном случае равными 0.

Формула для вычисления „парциального“ коэффициента корреляции дает возможность исключать любой из трех признаков, связанных взаимной корреляцией.

Пусть, например, имеем три антропометрических признака: V_x , V_y и V_z , которые для краткости мы будем в дальнейшем называть просто x , y и z . Предположим, что коэффициенты корреляции:

$$\begin{aligned} \text{между признаками } x \text{ и } y \dots r_{xy} &= +0,84 \\ \text{„ „ } x \text{ „ } z \dots r_{xz} &= +0,60 \\ \text{„ „ } y \text{ „ } z \dots r_{yz} &= +0,80. \end{aligned}$$

Выключим третий признак (z), т. е. определим „парциальный“ коэффициент корреляции между признаками x и y как бы при постоянных значениях признака z :

$$\begin{aligned} r_{xy.z} &= \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} = \frac{+0,84 - 0,60 \cdot 0,80}{\sqrt{(1 - 0,6^2)(1 - 0,8^2)}} = \\ &= \frac{+0,84 - 0,48}{\sqrt{(1 - 0,36)(1 - 0,64)}} = \frac{+0,36}{\sqrt{0,64 \cdot 0,36}} = \frac{+0,36}{\sqrt{0,2304}} = \\ &= \frac{+0,36}{0,48} = +0,75. \end{aligned}$$

Следовательно при различной (варьирующей) величине признака z корреляция между x и y выражается величиной коэффициента:

$$r_{xy} = +0,84,$$

а при подборе всех объектов исследования, обладающих одной и той же постоянной величиной признака z , исправленный коэффициент корреляции (парциальный):

$$r_{xy.z} = +0,75.$$

По этим же данным можно решить и другую задачу, напр.: какова будет корреляция между признаками x и z при постоянной (искусственно уравненной) величине признака y ?

Для этого служит формула, вполне аналогичная предыдущей, в которой лишь переставлены буквенные значки x , y и z :

$$r_{xz.y} = \frac{r_{xz} - r_{xy} \cdot r_{zy}}{\sqrt{(1 - r_{xy}^2)(1 - r_{zy}^2)}}.$$

Подставляя сюда наши числовые данные, получим:

$$r_{xz.y} = \frac{+0,60 - 0,84 \cdot 0,80}{\sqrt{(1 - 0,84^2)(1 - 0,80^2)}} = -0,22.$$

Совершенно также можно выключить и признак x , вычислив „парциальный“ коэффициент корреляции между признаками y и z при постоянном x по формуле:

$$r_{yz.x} = \frac{r_{yz} - r_{yx} \cdot r_{zx}}{\sqrt{(1 - r_{yx}^2)(1 - r_{zx}^2)}}.$$

или в числах:

$$r_{yz.x} = \frac{+0,80 - 0,84 \cdot 0,60}{\sqrt{(1 - 0,84^2)(1 - 0,60^2)}} = +0,70.$$

Итак в нашем примере корреляции трех признаков: x , y и z мы получили следующие три „парциальные“ коэффициента $r_{xy.z}$, $r_{xz.y}$ и $r_{yz.x}$, которые для ясности мы сопоставим здесь с заданными коэффициентами: r_{xy} , r_{xz} и r_{yz} :

$$\begin{array}{ll} r_{xy} = +0,84 & r_{xy.z} = +0,75 \\ r_{xz} = +0,60 & r_{xz.y} = -0,22 \\ r_{yz} = +0,80 & r_{yz.x} = +0,70. \end{array}$$

При вычислении этих парциальных коэффициентов корреляции мы пользовались следующими формулами:

$$\begin{aligned} r_{xy.z} &= \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \\ r_{xz.y} &= \frac{r_{xz} - r_{xy} \cdot r_{zy}}{\sqrt{(1 - r_{xy}^2)(1 - r_{zy}^2)}} \\ r_{yz.x} &= \frac{r_{yz} - r_{yx} \cdot r_{zx}}{\sqrt{(1 - r_{yx}^2)(1 - r_{zx}^2)}}. \end{aligned}$$

XXV. Эмпирические линии регрессии.

При разбивке материала, варьирующего в отношении некоторого признака V_y на те или иные группы, отличающиеся друг от друга по величине другого признака V_x (или наоборот), во многих случаях полезно бывает выяснить характер изменения всех групповых средних (M) по обоим признакам V_y и V_x , придав конечным результатам этих вычислений наглядную графическую форму.

Пусть, например, имеем следующие два ряда коррелирующих друг с другом чисел (вариант):

V_y :	5	15	12	16	32	20	28	30	17	25
V_x :	2	2	4	4	4	6	6	6	8	8

Здесь варианты нижнего ряда (V_x) естественно распадутся на 4 группы соответственно их значениям: 2, 4, 6 и 8, причем значение:

$V_x = 2$	повторяется 2 раза	($p = 2$)
$V_x = 4$	" 3 "	($p = 3$)
$V_x = 6$	" 3 "	($p = 3$)
$V_x = 8$	" 2 "	($p = 2$).

Вычислим средние арифметические верхних чисел V_y , соответствующих каждой отдельной группе одинаковых нижних чисел V_x .

Две крайние слева пары этих вариантов, для которых $V_x = 2$, имеют следующие значения V_y :

5 и 15.

Следовательно среднее арифметическое этих чисел:

$$M_1 = \frac{5+15}{2} = \frac{20}{2} = 10.$$

Следующая (в направлении слева направо) группа трех пар этих вариантов, для которых $V_x = 4$, содержит следующие значения V_y :

12, 16 и 32.

Следовательно среднее арифметическое чисел V_y в этой группе:

$$M_2 = \frac{12+16+32}{3} = \frac{60}{3} = 20.$$

Далее для следующей группы трех вариантов V_y , для которой $V_x = 6$:

$$M_3 = \frac{20+28+30}{3} = \frac{78}{3} = 26.$$

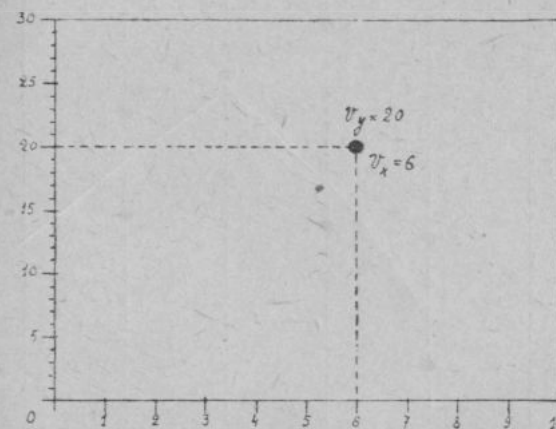
Наконец для крайней справа группы двух вариантов V_y , для которых $V_x = 8$:

$$M_4 = \frac{17+25}{2} = \frac{42}{2} = 21.$$

Итак:

для $V_x = 2$	среднее арифметическое	$M_1 = 10$
" $V_x = 4$	" "	$M_2 = 20$
" $V_x = 6$	" "	$M_3 = 26$
" $V_x = 8$	" "	$M_4 = 21$.

Все эти данные и конечные результаты вычислений удобно изобразить в графической форме, воспользовавшись для этого системой прямоугольных осей координат с произвольно выбранными масштабами (см. черт. 108).



Черт. 108.

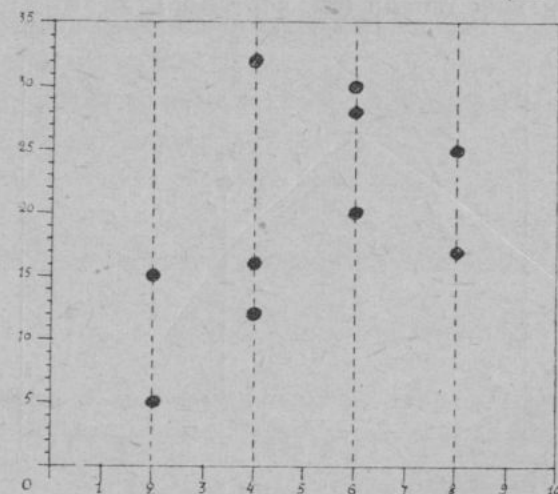
Условимся по горизонтальному масштабу (по оси абсцисс) откладывать варианты V_x , а по масштабу вертикальной оси (ординат) — варианты V_y . Так, например, пару вариант: $V_x = 6$ и $V_y = 20$ мы пометим в плоскости чертежа точкой, отстоящей на 6 делений горизонтального масштаба вправо от вертикальной оси (т. к. $V_x = 6$) и поднятой над горизонтальной осью на 20 делений вертикального масштаба (т. к. $V_y = 20$), как это и показано на чертеже 108.

Пометив этим же способом все 10 точек, соответствующих 10 парам заданных вариантов V_y и V_x , получим так называемый „точечный график“, представленный на чертеже 109. Здесь все точки, для которых V_x имеют одинаковую величину (2, 4, 6 или 8), соединены пунктирными

вертикальными прямыми, пересекающими горизонтальную ось в соответствующих масштабных делениях:

2, 4, 6 и 8.

Вспомним, что для двух вариантов V_y (5 и 15), для которых $V_x = 2$ (соответствующие этим вариантам две точки расположены на крайней слева вертикальной прямой), среднее арифметическое $M_1 = 10$. Это среднее значение (10) мы пометим на нашем чертеже в виде звездочки(*), расположив ее на высоте 10-го деления вертикального масштаба (см. черт. 110).



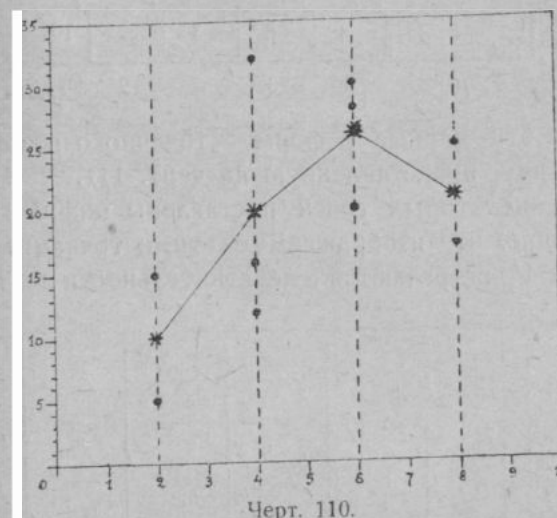
Черт. 109.

На следующей вертикали (соответствующей $V_x = 4$) мы пометим звездочкой среднее арифметическое $M_2 = 20$, вычисленное ранее для группы трех вариантов V_y : 12, 16 и 32 (см. черт. 110).

Далее на вертикали, соответствующей $V_x = 6$, пометим среднее $M_3 = 26$. Наконец, на крайней справа вертикали (для $V_x = 8$) поставим звездочку, соответствующую $M_4 = 21$. Соединив все эти звездочки отрезками прямых, получим ломаную линию, изображенную на черт. 110. Эта ломаная носит название „эмпирической линии регрессии“ *).

*) Читателю рекомендуется самостоятельно выполнить все указанные здесь (и в дальнейшем) построения в крупном масштабе на листе миллиметровой бумаги.

Таким образом эмпирическая линия регрессии высотой отдельных своих ординат наглядно показывает характер изменения групповых средних арифметических признака V_y в зависимости от изменения числовых значений другого признака V_x . Так, в частности, эмпирическая линия регрессии, например, веса по росту указывала бы на средние изменения веса людей в связи с последовательным увеличением их роста. При наличии достаточной корреляции между ростом и весом эта линия имела бы тенденцию к поднятию



Черт. 110.

в направлении слева-направо. Это означало бы, что люди низкого роста (отмеченные на горизонтальной оси слева) имеют в среднем также и незначительный вес (низкое положение линии регрессии), а у высоких людей (отмеченных справа) и средний вес достигает сравнительно больших значений (линия регрессии расположена высоко).

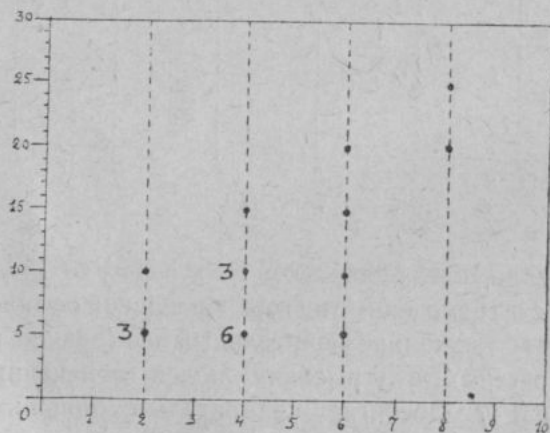
Приведем еще один числовой пример построения эмпирической линии регрессии, выразив на этот раз первичные данные в форме обычной корреляционной решетки с занесенными в ее клетках частотами p . Здесь сверху и слева в заголовках указаны числовые значения вариантов V_x и V_y , снизу подытожены частоты p_x ряда V_x , а справа—частоты p_y ряда V_y . В правом нижнем углу этой решетки записана

общая сумма всех частот p , т. е. общее число всех коррелирующих пар $n = 20$.

$V_y \backslash V_x$	2	4	6	8	p_y
25				1	1
20			1	1	2
15		1	1		2
10	1	3	1		5
5	3	6	1		10
p_x	4	10	4	2	20 = n

Облекая эти данные в форму „точечного графика“, получим картину, представленную на черт. 111.

Здесь у некоторых точек поставлены цифры: 3, 3 и 6, показывающие, что изображаемые этими точками пары вариант V_y и V_x встречаются в действительности не по одному



Черт. 111.

разу, а по несколько. Так, например, пара вариант $V_x = 2$ $V_y = 5$ по данным нашего задания встречается 3 раза. На этом основании у соответствующей точки (в левом нижнем углу чертежа) мы и поставим пояснительную цифру 3, которая, таким образом, изображает здесь частоту $p = 3$ данного сочетания: $V_x = 2$ и $V_y = 5$. Там, где эти сочетания встре-

чаются лишь по одному разу ($p = 1$), у соответствующих точек частоты p не проставлены вовсе (чтобы не загромождать чертежа лишними цифрами).

Для построения линии регрессии вычислим средние арифметические M в каждой группе вариант V_y с одинаковым значением V_x . Начнем с крайней слева группы точек, расположенных на вертикали $V_x = 2$. Там проставлены всего лишь 2 точки, соответствующие значениям: $V_y = 10$ и $V_y = 5$, но нижняя точка $V_y = 5$ изображает здесь не одно, а три таких значения V_y . Следовательно всего мы имеем в этой группе 4 варианты V_y : 10, 5, 5 и 5. Вычислим среднее арифметическое этой первой группы вариант V_y по способу сумм, приняв за начало отсчетов C например число 15 (центральное по отношению ко всем имеющимся в задании значениям вариант V_y).

Здесь при $C = 15$ итоги:

$$\begin{aligned} a_1 &= 0 & a_2 &= 0 \\ b_1 &= 7 & b_2 &= 3, \end{aligned}$$

откуда:

$$S_1 = a_1 - b_1 = 0 - 7 = -7.$$

Так как классовый промежуток $k = 5$, а общее число измерений в данной группе $n = 4$, то искомое среднее арифметическое:

$$M_1 = C + \frac{kS_1}{n} = 15 + \frac{5 \cdot (-7)}{4} = 6,25.$$

Так как при определении M_1 нам приходится вычислять лишь сумму S_1 , для чего нужны только итоги a_1 и b_1 , то крайний справа столбец этой решетки оказывается в данном случае совершенно излишним. Поэтому в дальнейшем мы вовсе не будем его вычерчивать, ограничив расчетную решетку всего только двумя вертикальными столбцами.

Ниже представлена эта же расчетная решетка в сокращенном виде, а рядом с нею три таких же решетки для вычисления средних арифметических в следующих по порядку группах вариант V_y , соответствующих значениям

V_y	p	0	0
25			
20			—
15	0	—	—
10	1	4	—
5	3	3	3
	4	7	3

вариант: $V_x = 4$, $V_x = 6$ и $V_x = 8$ (во всех случаях $C = 15$ и $k = 5$).

Для $V_x = 2$			Для $V_x = 4$			Для $V_x = 6$			Для $V_x = 8$		
V_y	p	0	V_y	p	0	V_y	p	1	V_y	p	3
25			25			25			25	1	1
20			20			20	1	1	20	1	2
15	0	—	15	1	—	15	1	—	15	0	—
10	1	4	10	3	9	10	1	2	10		
5	3	3	5	6	6	5	1	1	5		
	4	7		10	15		4	3		2	0
$a_1 = 0$			$a_1 = 0$			$a_1 = 1$			$a_1 = 3$		
$b_1 = 7$			$b_1 = 15$			$b_1 = 3$			$b_1 = 0$		
$S_1 = -7$			$S_1 = -15$			$S_1 = -2$			$S_1 = +3$		
$n = 4$			$n = 10$			$n = 4$			$n = 2$		
$M_1 = 6,25$			$M_2 = 7,50$			$M_3 = 12,50$			$M_4 = 22,50$		

Эти же вычисления можно было бы выполнить и в одной общей для всех 4-х групп сложной расчетной решетке с контрольной сводкой (последнюю в целях удобства дальнейшей обработки этого материала удобно составить из трех вертикальных столбцов):

$V_x=2$		$V_x=4$		$V_x=6$		$V_x=8$		Сводка			
V_y	p	0	p	0	p	1	p	3	p	4	1
25							1	1	1	1	1
20					1	1	1	2	2	3	—
15	0	—	1	—	1	—		—	2	—	—
10	1	4	3	9	1	2			5	15	—
5	3	3	6	6	1	1			10	10	10
	4	7	10	15	4	3	2	0	20	25	10
$S_1=-7$ $M_1=6,25$		$S_1=-15$ $M_2=7,50$		$S_1=-2$ $M_3=12,50$		$S_1=+3$ $M_4=22,50$		$S_1=-21$ $M_0=9,75$			

Способ проверки всех этих вычислений мы уже знаем. Верхние итоги a_1 в сумме должны составить итог a_1 , получившийся в сводке:

$$0 + 0 + 1 + 3 = 4;$$

сумма нижних итогов b_1 должна совпасть с итогом b_1 в сводке:

$$7 + 15 + 3 + 0 = 25,$$

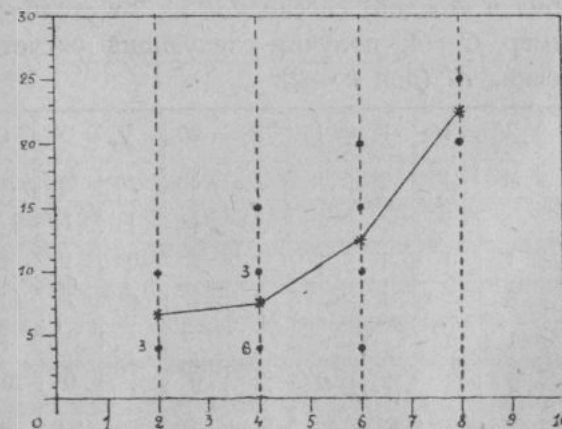
а сумма вспомогательных величин S_1 должна сойтись с той же величиной S_1 в сводке:

$$-7 - 15 - 2 + 3 = -21.$$

Среднее же арифметическое M_0 , полученное в сводке (для всех 20 вариантов V_y), представляет собой „взвешенное“ среднее из всех 4-х величин M , вычисленных в отдельных частях расчетной решетки. Помножив каждое из этих частных средних (M_1 , M_2 и т. д.) на численность соответствующей группы вариантов (для которой оно было вычислено), сложив все эти произведения и разделив эту сумму на общее число всех вариантов n , мы получим среднее арифметическое M_0 , найденное нами выше по данным контрольной сводки (т. е. в данном случае число 9,75):

$$M_0 = \frac{6,25 \cdot 4 + 7,5 \cdot 10 + 12,5 \cdot 4 + 22,5 \cdot 2}{20} = \frac{25 + 75 + 50 + 45}{20} = \frac{195}{20} = 9,75.$$

Нанеся (звездочками) все вычисленные нами средние арифметические: 6,25, 7,50, 12,50 и 22,50 на вычерченный уже ранее точечный график (см. черт. 111) и соединив эти звездочки отрезками прямых, мы и получим эмпирическую



Черт. 112.

линию регрессии, соответствующую данным нашей корреляционной решетки (см. черт. 112).

Наряду со средними арифметическими вариант V_y , соответствующих определенным (постоянным) значениям V_x , можно вычислить наоборот и ряд средних арифметических для отдельных групп вариант V_x , имеющих определенные (постоянные) значения V_y . Для этого в нашем примере придется обработать по способу сумм ряды частот p по горизонтальным строчкам совершенно так же, как это мы сделали только что в применении к вертикальным столбцам. Чтобы облегчить запись частот p из корреляционной решетки в расчетную, переписем заново наше задание, поменяв в нем местами заголовки V_x и V_y и соответствующим образом переставив во всех клетках частоты p .

Прежний порядок:						Новый порядок:						
$V_y \backslash V_x$	2	4	6	8	p_y	$V_y \backslash V_x$	5	10	15	20	25	p_x
25				1	1	8				1	1	2
20			1	1	2	6	1	1	1	1		4
15		1	1		2	4	6	3	1			10
10	1	3	1		5	2	3	1				4
5	3	6	1		10							
p_x	4	10	4	2	20	p_y	10	5	2	2	1	20

Составив по этим данным расчетную решетку из 5 частей (по 2 столбца в каждой) со сводкой из 3-х столбцов и приняв, например, $C=4$, получим следующий расчет средних арифметических M (при $k=2$):

V_x	$V_y = 5$		$V_y = 10$		$V_y = 15$		$V_y = 20$		$V_y = 25$		Сводка:		
	p	1	p	1	p	1	p	3	p	2	p	8	2
8							1	1	1	1	2	2	2
6	1	1	1	1	1	1	1	2	0	1	4	6	—
4	6	—	3	—	1	—	0	—	0	—	10	—	—
2	3	3	1	1							4	4	—
	10	3	5	1	2	0	2	0	1	0	20	4	0
	$S_1 = -2$ $M_1 = 3,60$		$S_1 = 0$ $M_2 = 4$		$S_1 = +1$ $M_3 = 5$		$S_1 = +3$ $M_4 = 7$		$S_1 = +2$ $M_5 = 8$		$S_1 = +4$ $M_0 = 4,40$		

Проверка дает здесь следующие результаты:

$$a_1: 1+1+1+3+2=8$$

$$b_1: 3+1+0+0+0=4$$

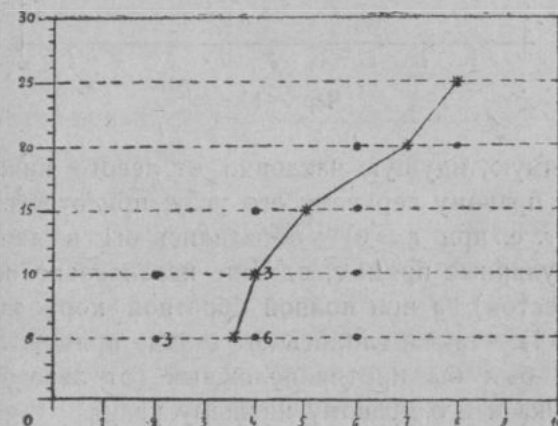
$$S_1: -2+0+1+3+2=+4$$

$$M_0 = \frac{3,6 \cdot 10 + 4 \cdot 5 + 5 \cdot 2 + 7 \cdot 2 + 8 \cdot 1}{20} =$$

$$= \frac{36 + 20 + 10 + 14 + 8}{20} = 4,40.$$

Эмпирическую линию регрессии, соответствующую этим значениям средних арифметических M , мы пометим на точечном графике, не меняя в нем положения осей координат, т. е. по-прежнему откладывая варианты V_y на вертикальной оси, а величину соответствующих средних арифметических M отсчитывая по масштабу горизонтальной оси (см. черт. 113).

Вычертив обе линии регрессии на одном общем чертеже



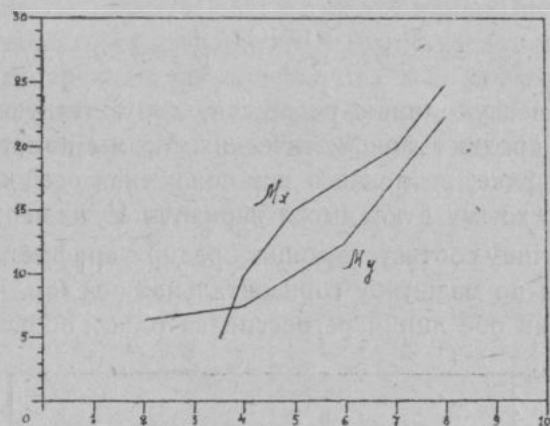
Черт. 113.

(без нанесения отдельных точек), получим следующую картину наглядного их сопоставления (см. черт. 114).

Первая из них, условно обозначенная символом M_y , дает наглядное представление о характере изменения средних арифметических признака V_y в зависимости от изменения числовых значений признака V_x , а вторая (M_x) — о характере изменения средних арифметических признака V_x в связи с изменением признака V_y . Кривая M_y обычно называется

эмпирической линией регрессии y на x (или y по x), а кривая M_x — эмпирической линией регрессии x на y (или x по y).

В данном случае при неполной корреляции между этими признаками (здесь $r = +0,72$) обе линии регрессии не совпадают друг с другом. При $r = +1$ они слились бы



Черт. 114.

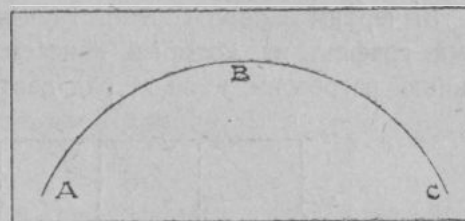
в одну прямую, идущую наклонно от левого нижнего угла графика к правому верхнему его углу; при отсутствии корреляции (т. е. при $r = 0$) — обратились бы в две взаимно перпендикулярные прямые, идущие параллельно осям координат (крестом), а при полной обратной корреляции (т. е. при $r = -1$) — также слились бы в одну прямую линию, но наклон ее был бы противоположный (от левого верхнего угла графика к его правому нижнему углу).

XXVI. Криволинейная корреляция.

Коэффициент корреляции r дает вполне точную характеристику связи двух признаков только лишь в том частном случае, когда обе эмпирические линии регрессии (y по x и x по y) имеют форму прямых линий или во всяком случае не очень сильно искривлены. При большом же изгибе линий

регрессии коэффициент r уже теряет свое значение числового показателя корреляции и в самых крайних случаях кривизны этих линий может дать явно ошибочные (преуменьшенные) указания.

Представим себе такой случай зависимости между признаками V_x и V_y , при котором эмпирическая линия регрессии y на x имеет дугообразную форму (ABC) с подъемом в середине (B) этой кривой (см. черт. 115). Левый конец этой кривой (A) по своему положению на чертеже 115 указывает на то, что малым значениям признака V_x (по горизонтальной оси) в среднем соответствуют также малые значения и признака V_y (по вертикальной оси). На этом основании корреляция здесь должна быть признана прямой. Но правый конец линии регрессии (C) свидетельствует уже о противоположном характере этой зависимости, так как большим значения признака V_x здесь соответствуют не большие, а малые значения V_y . Это обстоятельство указывает уже на корреляцию обратного типа. Таким образом в данном случае мы имеем пример прямой и обратной корреляции одновременно, в результате чего коэффициент r может оказаться равным 0, что характеризует полное отсутствие какой бы то ни было зависимости между этими признаками. В действительности же здесь зависимость несомненно имеется, но носит несколько более сложный и необычный характер, который и определяется формой заданной эмпирической линии регрессии ABC .



Черт. 115.

В подобных случаях связь между двумя признаками удобнее всего характеризовать не коэффициентом корреляции r , а особым показателем η (эта), носящим название „корреляционного отношения“.

Покажем способ вычисления корреляционного отношения η на числовых примерах, разобранных в предыдущей главе.