Проф. Ю. Л. ПОМОРСКИЙ

# ВАРИАЦИОННАЯ СТАТИСТИКА

## ЧАСТЬ I

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

2-ое ИЗДАНИЕ
ЗНАЧИТЕЛЬНО ДОПОЛНЕННОЕ

Издание Детского Обследовательного Института
имени проф. А. С. Грибоедова
ЛЕНИНГРАД
1929

ТИПОГРАФИЯ
«КРАСНОЙ ГАЗЕТЫ»
ИМЕНИ ВОЛОДАРСКОГО
ЛЕНИНГРАД
Фонтанка, 57

Ленинградский Областлит № 5185. Заказ № 3273. Тираж 5150.

## ОТ АВТОРА.

Быстрое распространение первого издания этой книги, отзывы печати и многочисленные письма читателей с весьма ценными указаниями, пожеланиями и советами побудили меня к выпуску в свет второго издания, несколько переработанного, исправленного в деталях и дополненного по возможности во всех тех частях и отделах, на пробелы в которых обратили свое внимание мои корреспонденты.

Так, во второе издание включена глава о вариационном коэффициенте, показателе точности и коэффициентах регрессии, значительно дополнена глава о статистическом анализе качественных признаков (изложением вопроса о корреляции между несколькими качественными признаками), дано описание приближенных способов вычисления средних ошибок и вероятностей с помощью специального графика, приложенного в конце книги, указаны и разъяснены некоторые формулы, сокращающие вычислительную работу в различных частных случаях, дан список всех буквенных обозначений и формул, встречающихся в тексте, а также приведены две таблицы: для вычисления средней ошибки вариационного коэффициента и для определения наименьшего числа измерений (наблюдений), при котором коэффициент корреляции будет превышать свою среднюю ошибку по крайней мере в 3 раза.

Настоящее руководство попрежнему вовсе не затрогивает философской, логической и математической стороны статистического метода, равно как и не имеет в виду приводить описание тех или иных специфических особенностей в его приложениях к отдельным отраслям знания. Задача здесь значительно у́же и скромнее, но зато и жизненнее.

Автор имеет в виду дать исчерпывающее по ясности и подробности чисто техническое описание вычислительных приемов, применяемых при обработке массовых измерений и наблюдений, не пугая читателей излишней математической символикой и сложными логическими доказательствами, по существу все же мало понятными не математикам, составляющим большинство читателей этой книги.

Опыт общения на лекциях и при консультациях со специалистами различных категорий, применяющими математический метод в своей исследовательской работе, показал, что целесообразность этого пути ими осознана вполне, а потому и доказательство необходимости применения методов математической статистики является уже ненужным. Больше того, основные идеи метода в настоящий момент получили настолько широкое распространение, что подробно развивать их в чисто-техническом руководстве представляется совершенно излишним.

Остается лишь вопрос о рационализации самого вычислительного процесса, о возможных его упрощениях без ущерба, однако, для точности окончательных результатов, об экономии времени и труда при производстве большого количества вычислительных операций и о непрерывной проверке всех вычислений в самом процессе этой работы — вопросы, которым почти совершенно не уделяется должного внимания в существующих на русском языке руководствах по математической статистике.

Читатели, заинтересовавшиеся принципиальной стороной метода, его математическим обоснованием, а также теми или иными специальными приложениями, найдут, без сомнения, весьма ценные и интересные разъяснения в известных трудах Чупрова, Митропольского, Филипченко, Романовского, Кауфмана и др., но исчерпывающих технических указаний по вопросу о производстве соответствующих вычислений упрощенным способом они получат там все же немного или попросту не разберутся в сложных выкладках, символах и формулах, не имея для этого необходимой математической установки и соответствующей специальной тренировки. Вместе с тем овладение этой техникой необходимо всякому исследователю, применяющему в своей работе статистический метод. Нерациональная организация вычислительного



процесса, применение громоздких приемов и отсутствие контроля приводит к полной невозможности практического использования данного метода при сколько-нибудь широком масштабе работы, в то время как применение этой упрощенной техники создает условия, благоприятные для широкого его приложения в самых разнообразных отраслях знания.

Порядок расположения материала в этом руководстве определяется его целями.

Первые пять глав предназначены для читателей, не предполагающих изучать вычислительную технику. В этих главах приводится лишь краткое описание основных статистических постоянных, наглядно разъясняется их смысл и принцип применения при статистической обработке массовых материалов. Для бо̀льшей ясности способ вычисления всех этих постоянных (констант) показан здесь исключительно на примере вариационных рядов, не сведенных в разряды (классы) и при том только на очень ограниченном количестве вариант. Ознакомление с этим отделом поможет читателю уяснить особенности и границы приложимости статистического метода в области его специальности и позволит правильно интерпретировать конечные результаты математического анализа, произведенного другим лицом.

В следующих десяти главах (до XV включительно) подробно изложены технические приемы основных статистических вычислений, безусловно необходимых и всегда применяемых при математической обработке массовых числовых материалов. Этот отдел предназначается уже для лиц, выполняющих счетную работу самостоятельно.

Последние четыре главы содержат описание вычислительных приемов для более редких случаев статистической практики.

В добавлениях и приложениях собран справочный материал и образцы пособий, имеющихся в отдельной продаже.

Исключительно популярный по изложению и строго практический по содержанию характер этой книги делает ее, с моей точки зрения, особенно удобной в качестве руководства для техников-вычислителей, работающих по указаниям и под непосредственным наблюдением сведующего лица. Имея в виду эту, так сказать, педагогическую задачу, я сделал все возможное, чтобы облегчить труд наглядной передачи соответствующих сведений такому техническому сотруднику, разбив для этого все числовые примеры на составляющие их элементы и подробно описав последовательность всех промежуточных операций.

С этой же целью в книге подобраны примеры, искусственно составленные главным образом из небольших „круглых" чисел, дающих при делении и извлечении квадратного корня конечный числовой результат чаще всего также в виде целого числа. Благодаря таким приемам изложения, внимание читателя не отвлекается в сторону производства громоздких вычислений, а фиксируется по преимуществу на усвоении и запоминании общего порядка и последовательности выполняемых действий.

Настоящее руководство представляет собой, так сказать, первый концентр краткого курса вариационной статистики, включая в свою программу лишь минимальное количество необходимых в этой области предварительных сведений, с которыми, однако, представляется уже возможность оперировать на практике, разрешая ряд типовых вопросов и проблем, встречающихся во всякой научно-исследовательской работе, проводимой на массовом материале. Расширение и углубление круга включенных сюда понятий в связи с описанием более редких случаев детального статистического анализа составит содержание второй части этого руководства, обещанный выпуск которого в настоящее время несколько задерживается по независящим от меня техническим причинам, но, надо полагать, осуществится втечение ближайших же месяцев.

Читателям этой книги буду весьма признателен за всякие указания на недостаточно ясно изложенные в ней места, равно как и на желательность включения во вторую часть тех или иных вопросов и отделов, надобность в которых ощущается на практике. Заранее благодаря за эти указания, прошу по прежнему направлять их по моему адресу: Ленинград, Б. Зеленина ул., д. 26[б], кв. 42, или по телефону 546-60.

*Ю. П.*

## I. Среднее арифметическое (M).

При изучении какого-либо свойства или признака на большом количестве однородных объектов исследования, напр., при измерении роста целой группы детей определенного пола и возраста, всегда приходится сталкиваться с большей или меньшей изменчивостью этого признака (роста) у отдельных инвалидов. Линейные размеры и вес тела, сила мышц, число ударов пульса в секунду, время реакции на тот или иной раздражитель, количество правильно решенных задач—все эти и подобные им числовые результаты иссследования обычно являются крайне разнообразными, варьирующими в более или менее широких пределах даже внутри вполне однородной группы объектов исследования. Измерив, например, рост 100 мальчиков 13-летнего возраста, выросших в совершенно однородных условиях городской жизни и принадлежащих к одному и тому же классу населения (напр., только детей рабочих), мы все же обнаружим некоторую, иногда очень значительную разницу в росте отдельных детей, несмотря на то, что во всех остальных отношениях группа этих детей является, повидимому, вполне однородной.

При желании сопоставить друг с другом подобные результаты измерений для нескольких разнородных групп индивидов, напр., при сравнении роста детей различного пола, различного возраста или различных социальных группировок, было бы совершенно бесполезно пытаться сравнивать между собой рост отдельных представителей этих групп, так как в силу неизбежных колебаний роста отдельных детей мы рисковали бы случайно подвергнуть сравнению низкого представителя группы высоких детей с высоким представителем низкой группы и получить таким образом ложное представление о действительном соотношении типичного роста, свойственного этим группам в целом. Единственный правильный путь такого сопоставления заключается здесь в предварительном определении некоторой общей числовой характеристики типичного роста всей группы однородных объектов исследования и в сравнении друг с другом уже таких общих суммарных результатов.

Пусть, например, требуется сравнить рост мальчиков с ростом девочек одного и того же возраста. Сравнение отдельных мальчиков с отдельными же девочками не будет убедительным, так как некоторые девочки могут оказаться выше мальчиков даже и в том случае, если в данном возрасте мальчики в общем будут выше девочек. Для того, чтобы это стало вполне очевидным, неоходимо как-то вычислить типичный рост всех мальчиков и всех девочек отдельно и эти общие характеристики роста сопоставить уже друг с другом.

Каким же способом на основании числовых данных отдельных измерений можно составить общее представление о типичной величине изучаемого свойства или признака у всей группы объектов исследования в целом?

Для этого существуют три основных приема. Первый из них заключается в выяснении величины данного признака у того индивидуума, который занимает центральное положение среди всех остальных объектов исследования. Если расставить всех детей в ряд, расположив их строго в порядке возрастания или убывания роста, то рост ребенка, стоящего как раз в середине этого ряда, и может быть принят за типичный рост ι сех детей этой группы. Величина признака, полученная этим способом, носит название „медианы". Если середина такого ряда придется а промежутке между двумя детьми (при четном их количестве), то за медиану принимается средняя величина (полусумма), вычисленная из роста их обоих. Так, напр., если бы в середине ряда оказались два ребенка ростом в 131 и 129 см, то за медиану следует принять величину $\frac{131+129}{2}=\frac{260}{2}=130$ *см.*

Второй способ определения общей числовой характеристики признака заключается в выяснении той его величины, которая встречается чаще всего. Если в нашем примере разбить всех детей на мелкие группы так, чтобы в каждой



из этих групп оказались дети приблизительно одинакового роста, то за величину типичного роста можно принять рост детей той мелкой группы, которая является самой многочисленной. Полученная этим путем типичная величина признака называется „модой" *).

Третий способ определения типичной величины признака состоит в вычислении так называемого „среднего арифметического". Для этого в нашем примере следует сложить рост всех детей и полученную сумму разделить на число этих детей.

Поясним сказанное числовым примером. Пусть при измерении роста 11 детей мы получим следующие результаты, выраженные в сантиметрах:

127, 127, 128, 128, 128, 129, 130, 131, 132, 134, 136.

Медиана этого ряда будет 129, так как это число занимает как раз центральное место в ряду всех чисел, расположенных в порядке возрастания (слева и справа от числа 129 стоит одинаковое количество других чисел: по пяти чисел с каждой стороны).

Модой будет служить число 128, так как оно здесь повторяется чаще всего (три раза, в то время как другие числа повторяются лишь по одному или по два раза).

Для вычисления среднего арифметического следует сложить все 11 чисел и сумму их разделить на 11. Так как сумма чисел равна здесь 1430, то среднее арифметическое получим, если 1430 разделим на 11, в результате чего среднее арифметическое окажется, очевидно, равным 130.

Итак, здесь медиана = 129 *см*, мода = 128 *см*, а среднее арифметическое = 130 *см.*

Для пояснения способа определения медианы при четном числе измерений приведем еще один пример. Пусть имеем рост 10 детей:

129, 129, 130, 130, 130, 131, 132, 133, 135, 136.

Здесь середина ряда приходится между числами 130 и 131. Следовательно медиана будет $\frac{130+131}{2}=\frac{261}{2}=130{,}5$ *см.*

*) Точное значение „медианы" и „моды" вычисляется более сложным способом, который будет указан во II части этого ι уководства.

Нетрудно убедиться в том, что модой будет здесь служить 130 *см*, а среднее арифметическое окажется равным $\frac{1315}{10} = 131{,}5$ *см*.

Все указанные здесь способы определения типичной величины роста могут быть, очевидно, применены и к получению общих числовых характеристик какого угодно другого свойства или признака, напр., типичного для данной формы заболевания количества лейкоцитов, содержащихся в 1 *куб. мм* крови, количества букв, цифр или каких-либо значков, запоминаемых испытуемым после показа ему их втечение определенного промежутка времени при испытании памяти, и т. д.

Из трех указанных выше способов определения средней (типичной) величины варьирующего признака в дальнейшем мы будем пользоваться исключительно „средним арифметическим", которое обозначается обычно буквою *M*.

---

## II. Среднее квадратическое отклонение ($\sigma$).

Определение только одной средней величины *M* варьирующего признака для исчерпывающей его характеристики все же не является еще вполне достаточным. При сравнении, напр., роста отдельного ребенка со средним ростом, свойственным детям его пола и возраста, чрезвычайно важно бывает установить не только самый факт соответствия или, наоборот, несоответствия его роста „нормальному", но и определить степень отличия его роста от этой „нормы".

Пусть, например, средний рост данной возрастно-половой группы детей $M = 130$ *см*, а рост двух отдельных детей этой же группы — 129 *см* и 135 *см*. Очевидно, первый ребенок ниже среднего уровня на 1 *см*, а второй — выше на 5 *см*.

Что говорят нам эти числа сами по себе? Да, в сущности, ровно ничего, так как мы не знаем здесь, что считать за большое и за малое отличие роста от средней нормы. Повидимому, 1 *см* должен быть признан за несущественную разницу, и первого ребенка следует все же считать принадлежащим к группе детей среднего роста, в то время как отличие в 5 *см* кажется уже более значительным, и второго



ребенка следовало бы признать высоким. Но такая примитивная оценка, конечно, является совершенно произвольной. Почему разница в 1 *см* еще малая, а в 5 *см* уже большая, и где же лежит граница, отделяющая эти малые различия от различий больших?

Вполне объективный критерий такой оценки может быть найден здесь только на основании учета самой величины варьирования данного признака. Вполне очевидно, что оценка разницы между какими-либо размерами тела отдельного ребенка и средними размерами всех детей данного пола и возраста зависит от изменчивости самого этого размера. Отличие от *M* на 3 *см* в поперечном диаметре головы несомненно является уже крупным, в то время как те же 3 *см* разницы в росте еще не так существенны, потому что размеры головы у отдельных детей колеблются незначительно, а рост — в сравнительно очень широких пределах. Если бы на ряду со средним арифметическим *M* мы научились определять также и среднюю величину варьирования данного признака, то эта последняя числовая его характеристика и могла бы служить объективной мерой для оценки отличия отдельных размеров от среднего (типичного) размера всей группы объектов исследования в целом.

Исходя из каких же соображений можно вычислить такую среднюю величину варьирования? Для этого существуют два способа, которые мы последовательно здесь и рассмотрим. Пусть, напр., измерение роста 5 детей дало следующие результаты, выраженные, как и ранее, в сантиметрах:

127, 129, 130, 131, 133.

Среднее арифметическое этого ряда (*M*), очевидно, равно здесь 130 *см*. Сравним рост отдельных детей со средним ростом всей группы и выясним, на скольке сантиметров каждый отдельный ребенок оказывается выше или, наоборот, ниже этого среднего роста, т.-е. насколько сильно рост его отклоняется (в ту или иную сторону) от среднего роста всей группы. Эти различия, или „отклонения", могут быть как положительными (если ребенок оказывается выше среднего роста), так и отрицательными (если он ниже среднего роста).

В нашем примере первый ребенок ростом в 127 *см* оказался ниже среднего роста *M* (130 *см*) на 3 *см*, следовательно рост его отклоняется от среднего на „минус 3 *см*" (—3). Из тех же соображений отклонение роста второго ребенка (129 *см*) будет равно „минус 1 *см*" (—1). Рост третьего ребенка (130 *см*) случайно совпал со средним ростом всей группы (*M*), и потому соответствующее отклонение равно здесь „нулю" (0). Рост четвертого ребенка (131 *см*) превышает средний на 1 *см*, следовательно отклонение здесь равно „плюс 1 *см*" (+1). Наконец, отклонение роста последнего ребенка (133 *см*), очевидно, будет равно „плюс 3 *см* (+3). Выписав эти отклонения в соответствующем порядке, получим следующий ряд отрицательных и положительных чисел.

$$-3 \qquad -1 \qquad 0 \qquad +1 \qquad +3$$

Просматривая эти числа, мы замечаем. что рост отдельных детей нашей группы отклоняется от среднего роста *M* (130 *см*) на различную величину, колеблющуюся в абсолютном своем значении от 0 до 3 *см*. Поставим себе вопрос: на сколько же сантиметров в среднем рост отдельных детей отклоняется от общего среднего роста всей группы? Это среднее отклонение, очевидно, должно оказаться меньше самых больших отклонений (+3 и —3), но в то же время больше малых отклонений (+1, 0 и —1), занимая по отношению к этим крайним пределам некоторое промежуточное место.

Такое среднее значение отклонений, или просто „среднее отклонение", мы можем найти с помощью того же самого приема, который уже был использован нами ранее для определения средней величины самого варьирующего признака, а именно: мы можем здесь вычислить среднее арифметическое из всех этих пяти отклонений, не принимая пока во внимание их знаков (+ или —). Для этого сложим абсолютную величину всех отклонений и сумму их разделим на число этих отклонений (т.-е. на 5):

$$3+1+0+1+3=8.$$

Разделив 8 на 5, получим 1,6. Эта величина (1,6 *см*) и может служить числовой характеристикой среднего отклонения роста отдельных детей от среднего роста всей группы. В самом деле, при малом размахе колебаний роста отдельных детей, т. е. когда все измерения окажутся тесно сгруппированными около *M*, отдельные отклонения, общая их сумма, а стало-быть, и среднее отклонение также окажутся малыми; при большом же размахе колебаний отдельных измерений, т.-е. при сильной разбросанности соответствующих чисел отдельные отклонения неизбежно также окажутся большими, что тотчас же отразится и на увеличении среднего отклонения. Словом, вычисленное этим способом среднее отклонение вполне точно характеризует величину изменчивости данного признака, увеличиваясь или уменьшаясь вместе с изменением в степени его варьирования.



Но в описанном здесь способе вычисления среднего отклонения все же заключается некоторый формальный компромис. Дело в том, что, складывая отдельные отклонения для вычисления из них среднего отклонения, мы не имеем права отбрасывать знаки. Складывая же эти отклонения с учетом знаков, мы всегда получим в общем итоге нуль, так как сумма всех положительных отклонений всегда окажется в точности равной сумме всех отрицательных отклонений. В самом деле, в нашем примере:

$$-3 \qquad -1 \qquad 0 \qquad +1 \qquad +3$$

сумма всех отрицательных отклонений (—3 и —1) равна —4 а сумма всех положительных отклонений (+1 и +3) равна +4. Складывая —4 и +4, получим, очевидно, 0.

Так как эту сумму (0) мы должны далее разделить на общее число всех отклонений (т.-е. на 5), то в конечном итоге получим, очевидно, также 0. Поэтому, идя указанным путем, мы, в сущности, вместо искомой величины среднего отклонения всегда будем получать одну и ту же постоянную величину, равную нулю при всяком (как большом, так и малом) варьировании изучаемого признака.

Выход из создавшегося здесь затруднения может быть найден лишь в применении особого дополнительного приема, заключающегося в искусственном превращении всех отклонений (как положительных, так и отрицательных) только в одни положительные отклонения посредством временного возвышения их в квадрат.

Выпишем еще раз все наши 5 отклонений:

$$-3 \qquad -1 \qquad 0 \qquad +1 \qquad +3$$

После возвышения их в квадрат получим:

9 1 0 1 9

Сумма этих чисел равна 20 Но это будет уже не сумма самых отклонений, как раньше, а сумма их квадратов, т.-е. в данном случае величина заведомо большая, чем та, какую мы ищем. Разделив полученную этим путем сумму квадратов отклонений (20) попрежнему на общее число всех измерений (т.-е. на 5), получим число 4, которое также выражает собою, очевидно, не среднюю величину отдельного отклонения, как раньше, а среднюю величину его квадрата. Чтобы найти, наконец, нужное нам среднее отклонение (в первой степени, а не в квадрате), следует, извлечь из полученного нами числа (4) квадратный корень, т.-е. найти такое число (в данном случае 2), которое, будучи возвышено в квадрат, как раз и дало бы это число 4. Следовательно в нашем случае искомое среднее отклонение будет равно ]/4, т.-е. равно 2 *см.*

Вычисленное этим путем среднее отклонение, в отличие от ранее полученного (1,6), называется „средним квадратическим отклонением" и обозначается обычно греческой буквой a (сигма). Среднее квадратическое отклонение a всегда оказывается несколько больше простого среднего отклонения (у нас a = 2, а простое среднее отклонение = 1,6). Это объясняется тем, что при возвышении отдельных отклонений в квадрат особенно сильно возрастают большие отклонения, увеличивая собой общую сумму их квадратов, а стало быть и конечный результат вычислений.

В дальнейшем для характеристики средней величины варьирования мы будем пользоваться исключительно средним квадратическим отклонением o, а не простым средним отклонением.

Так как при вычислении a в конечном итоге приходится извлекать квадратный корень, который, как известно из алгебры, всегда имеет двойной знак (плюс-минус), то и самое среднее квадратическое отклонение мы будем писать с двумя знаками, напр. o = + 2. Это имеет и реальный смысл, так как данный, например, результат представляет собой некоторое среднее отличие роста отдельных детей от



типичного роста всей группы, отличие, возможное в сторону как больших, так и меньших его значений.

Величина среднего квадратического отклонения o, показывающая средний размах колебаний отдельных значений какого-либо признака, и может служить мерой изменчивости или варьирования этого признака, являясь той границей, которая отделяет малые отклонения от отклонений значительных. В нашем примере с ростом к высоким и низким мы должны, следовательно, отнести лишь тех детей, рост которых отличается от среднего роста *M* более, чем на величину o, а при меньшем различии — считать их рост „нормальным". Если бы, напр., при среднем росте *M = 130 см* среднее квадратическое отклонение o оказалось = ± 5 *см,* то высокими мы должны были бы признать лишь тех детей, рост которых превысил бы 135 *см,* а низкими — детей, рост которых оказался бы меньше 125 *см.* Рост же в пределах от 125 до 135 *см* должен был бы считаться для детей данной группы „нормальным" *).

Описанный способ вычисления o относится, собственно, к тому случаю, когда число отдельных измерений чрезвычайно велико (напр., когда a определяется из нескольких сотен и даже из нескольких тысяч отдельных измерений). В случае же, когда число измерений незначительно, как в нашем последнем примере, когда o пришлось вычислять всего лишь из 5 отдельных измерений, следует внести в этот способ некоторую поправку. Поправка эта заключается в том, что ряда чисел $M = 11{,}5$. Вычислим теперь отклонения отдельных чисел от этого среднего 11,5, т.-е. выясним, насколько каждое из них больше или меньше $M$. Получим следующие результаты:

$$+0{,}3 \quad -1{,}3 \quad -0{,}4 \quad +1{,}1 \quad +0{,}3.$$

Возвысив все эти отклонения в квадрат, будем иметь:

$$0{,}09 \quad 1{,}69 \quad 0{,}16 \quad 1{,}21 \quad 0{,}09$$

Сумма квадратов отклонений равна здесь 3,24. Эту сумму нам нужно разделить уже не на 5 (как это мы делали раньше), а на другое число, на единицу меньшее, т.-е. в данном случае на 4. В результате получим 0,81. Корень из этого числа равен 0,9 (так как 0,9 в квадрате равняется как раз 0,81). Следовательно в данном случае $\sigma = \pm 0{,}9$.

В дальнейшем мы всегда будем вычислять $\sigma$ этим последним способом (т.-е. с поправкой на малое число измерений).

Вполне очевидно, что указанный способ вычисления $\sigma$ применим к любому ряду чисел, в каких бы единицах они ни выражали результат измерения варьирующего признака: в сантиметрах, в килограммах, в секундах, в градусах, в числе экземпляров, в процентах и т. д. Таким образом $\sigma$ также как и $M$ представляет собою именованное число.

---

## III. Средняя ошибка ($m$).

Определив по указанным выше правилам среднее арифметическое $M$ для какой-либо однородной группы объектов исследования, мы, тем не менее, никогда не можем быть вполне уверены в том, что полученный нами данный частный результат совершенно точно характеризует величину этого признака и у всех других подобных же объектов, нами пока еще не исследованных. Вычислив, напр., средний рост одиннадцатилетних мальчиков на основании 100 отдельных измерений, мы не имеем никакого права утверждать, что и вообще все мальчики этого возраста будут иметь средний рост, вполне точно совпадающий со средним ростом этих 100 детей.



В самом деле, на величину среднего арифметического $M$ всегда влияет тот или иной подбор объектов исследования. Если в измеряемой группе случайно попадется несколько очень высоких или, наоборот, очень низких детей, то это обстоятельство тотчас же в большей или меньшей степени отразится и на средней величине роста всей группы, а так как состав такой группы всегда подвержен влиянию случайности, то и ручаться за абсолютную достоверность среднего арифметического $M$ представляется совершенно невозможным.

Примирившись с мыслью о некоторой неточности всякого среднего арифметического, вычисленного из ограниченного количества отдельных измерений, мы, однако, должны найти способ в каждом отдельном случае определять также и величину той ошибки, которая была допущена нами при вычислении $M$.

Основной принцип этого способа мы разъясним на следующем числовом примере.

Пусть для вычисления среднего роста детей какого-либо определенного пола и возраста мы имеем возможность производить измерения в нескольких различных школах. Допустим, далее, что, отобрав в первой школе 9 детей интересующего нас пола и возраста и измерив рост каждого из них в отдельности, мы получили следующие результаты, выраженные в сантиметрах:

137 129 131 133 125 133 121 129 141.

Для вычисления среднего роста всей группы детей, измеренных в первой школе, сложим все эти 9 чисел. Сумма их равна здесь 1179. Разделив эту сумму на число измерений (т.-е. на 9), получим 131. Следовательно искомое среднее арифметическое, которое мы обозначим через $M_1$, будет равно 131 *см.*

О чем говорит нам полученный здесь результат? Только о том, что рост э т и х девяти детей оказался в среднем равным 131 *см.* Можем ли мы на этом основании утверждать, что и во всех других школах средний рост детей данного пола и возраста будет в точности равен 131 *см*?

Очевидно, нет. Подбор девяти детей этой первой группы был все же до известной степени случайным, и у нас нет никакой гарантии в том, что среди них не попалось несколько исключительно высоких или исключительно низких ребят. И рассуждения, и опыт убеждают нас в том, что в других школах мы можем получить результаты, несколько не совпадающие с данным.

Для того, чтобы узнать, на сколько же сантиметров наше среднее арифметическое $M_1$ (131 *см*) может отличаться от таких же средних арифметических, вычисленных на основании измерения роста детей того же самого пола и возраста в других школах, нам нужно фактически определить эти средние хотя бы еще для нескольких школ.

Допустим, что после аналогичных вычислений мы получили следующие данные:

| | | | | |
|---|---|---|---|---|
| средний рост детей | второй | школы | $M_2 = 129$ | *см* |
| " " " | третьей | " | $M_3 = 128$ | " |
| " " " | четвертой | " | $M_4 = 133$ | " |
| " " " | пятой | " | $M_5 = 129$ | " |

Каким же способом на основании этих результатов можно определить среднюю величину ошибки, которую мы допустили, приняв за истинный рост ту заведомо неточную величину $M_1$ (131 *см*), которая получилась у нас на основании измерения роста детей только в первой школе? Другими словами, на сколько сантиметров мы ошиблись здесь в среднем, приняв за истинный рост величину в 131 *см*, в то время, как в других, вполне аналогичных условиях мы могли бы получить этот рост равным 129 *см*, 128 *см*, 133 *см* и 129 *см*?

Очевидно, величина нашей ошибки зависит от того, насколько сильно варьируют (не совпадают друг с другом) все эти отдельные результаты. Если бы все они оказались совершенно одинаковыми, в точности равными друг другу и нашей первой величине $M_1$, то ошибки в ее определении не было бы никакой. Чем больше эти результаты варьируют, тем, очевидно, меньше можно доверять и каждому из них в отдельности, а в том числе и нашему первому результату $M_1$.



Словом, искомая средняя ошибка нашего среднего арифметического $M_1$ определяется средней величиной варьирования всех пяти средних арифметических, вычисленных для отдельных школ.

Как же мы можем численно охарактеризовать эту среднюю величину варьирования? Очевидно, с помощью обычного приема вычисления среднего квадратического отклонения, которое вместо $\sigma$ мы обозначим здесь буквою $m$.

Итак, для определения средней ошибки $m$ любого среднего арифметического (а стало-быть, в частности, и нашего среднего $M_1$, вычисленного только для первой школы) мы должны выполнить все действия, которые были указаны ранее в применении к вычислению среднего квадратического отклонения $\sigma$, с той лишь разницей, что вместо отдельных (индивидуальных) измерений здесь будут фигурировать 5 средних арифметических, вычисленных для каждой школы отдельно.

Проделаем же все эти операции с числами нашего примера:

131 129 128 133 129.

Общее для всех пяти школ среднее арифметическое $M$ найдем, если сложим все эти числа и сумму их разделим на 5. Сумма этих чисел равна здесь 650. Следовательно $M = 130$ *см*.

Выпишем, как и раньше, отклонения отдельных чисел нашего ряда от общего их среднего арифметического 130 *см*:

$+1 \quad -1 \quad -2 \quad +3 \quad -1.$

Возвысив эти отклонения в квадрат, будем иметь:

1 1 4 9 1.

Сумма квадратов равна здесь 16. Разделив 16 на общее количество этих чисел без единицы, т. е. на 4, получим 4. Так как корень квадратный из 4 равен 2, то искомая средняя ошибка $m = \pm 2$ *см*.

Этот результат показывает, что рост детей данного пола и возраста, вычисленный только на основании 9 измерений, произведенных в первой школе, и оказавшийся равным 131 *см*, заключает в себе некоторую возможную ошибку, равную в среднем 2 *см*.

Другими словами, истинный рост детей интересующего нас пола и возраста заключается где-то в средних пределах от 129 *см* ($M_1 - m$) до 133 *см* ($M_1 + m$).

Но это только средние пределы допущенной нами неточности, так как величина $m$, подобно всякому среднему квадратическому отклонению, указывает лишь на средний размах колебаний отдельных средних арифметических, вычисленных для различных школ. Максимальные же пределы варьирования, как это доказывается в теории математической (вариационной) статистики, должны быть расширены приблизительно до трех средних квадратических отклонений, а в данном случае, следовательно, до трех средних ошибок $m$ (т.-е. до 6 *см*).

Поэтому действительный рост этих детей заключается в несколько более широких пределах, а именно: от 125 *см* ($M_1 - 3\ m$) до 137 *см* ($M_1 + 3m$), что мы имеем право утверждать уже с почти абсолютной достоверностью.

Описанный здесь способ определения средней ошибки $m$ был указан нами исключительно лишь с целью разъяснения самого понятия об этой числовой характеристике неточности всякого среднего арифметического. Нам важно было установить, что ошибка $m$ представляет собой не что иное, как средний размах возможных колебаний нескольких $M$, вычисленных для вполне однородного материала, но при различной его группировке.

В действительности, однако, нет надобности прибегать к столь сложному способу вычисления средней ошибки. Теория этого вопроса устанавливает некоторую определенную зависимость между варьированием отдельных измерений внутри одной и той же группы объектов исследования и варьированием самих средних арифметических, вычисленных для нескольких таких групп, т.-е. между $\sigma$ и искомой средней ошибкой $m$, благодаря чему получается возможность определить $m$, не прибегая к делению всех объектов исследования на те или иные мелкие группы, а прямо вычисляя эту ошибку $m$ с помощью $\sigma$.

В чем же заключается эта зависимость, и как можно воспользоваться ею для определения $m$?

Дело в том, что отдельные измерения всегда варьируют в более широких пределах, чем вычисленные из них средние



арифметические, которые, таким образом, обладают как бы большей устойчивостью (постоянством). При этом варьирование средних арифметических оказывается тем меньше, чем из большего количества измерений они вычислены. В теории математической статистики доказывается, что величина варьирования средних арифметических обратно пропорциональна корню квадратному из общего числа измерений. Это надо понимать так: если отдельные измерения (напр., рост отдельных детей) колеблются в среднем на величину $\sigma = \pm 8$ *см*, то средние арифметические, вычисленные для нескольких групп этих детей, скажем, по 16 человек в каждой, будут колебаться уже со средним размахом в 4 раза меньшим (так как $\sqrt{16} = 4$), т.-е. всего лишь на величину $= \pm 2$ *см* (а не на 8, как отдельные измерения). Таким образом для определения средней ошибки $m$ с помощью $\sigma$ можно просто разделить $\sigma$ на корень квадратный из числа всех измерений, для которых были вычислены $M$ и $\sigma$. В разобранном выше примере мы могли бы, следовательно, взамен определения нескольких средних арифметических для различных школ просто вычислить $\sigma$ для девяти измерений первой школы и, разделив эту $\sigma$ на корень квадратный из общего числа измерений (т.-е. на $\sqrt{9} = 3$), прямо получить искомую величину средней ошибки $m$.

Проделаем же эти вычисления в действительности. Ранее мы имели следующие результаты измерения роста 9-ти детей в первой школе:

137 129 131 133 125 133 121 129 141.

Так как средний рост этой группы $M_1 = 131$ *см*, то соответствующие отклонения выразятся следующими числами:

$$+6 \quad -2 \quad 0 \quad +2 \quad -6 \quad +2 \quad -10 \quad -2 \quad +10.$$

После возвышения их в квадрат получим:

36 4 0 4 36 4 100 4 100.

Сумма квадратов отклонений равна 288. Разделив эту сумму на число всех измерений без единицы, т.-е. на 8, получим 36. Так как корень из 36 равен 6, то средне квадратическое отклонение этого ряда измерений $\sigma = \pm 6$.

Согласно установленному выше правилу, для вычисления средней ошибки $m$ мы должны будем разделить полученную нами $\sigma$ (т.-е. 6) на корень из числа всех измерений т.-е. на $\sqrt{9} = 3$. Разделив 6 на 3, получим 2. Таким образом искомая средняя ошибка $m$ нашего среднего арифметического $M_1$ (131 *см*) оказалась равной $\pm 2$ *см*.

Этот результат обычно записывается так: $M_1 = 131 \pm 2$ *см*, где второе число (2), соединенное с первым (131) двойным знаком (плюс-минус), и обозначает среднюю ошибку $m$.

В дальнейшем мы всегда будем иметь в виду исключительно этот последний способ определения средней ошибки $m$, которая, таким образом, вычисляется посредством деления $\sigma$ на квадратный корень из общего числа всех измерений.

Указанная зависимость между $\sigma$, $m$ и числом измерений дает возможность проанализировать те условия, от которых зависит точность вычисления всякого среднего арифметического. Так как точность эта определяется величиной средней ошибки $m$, а эта последняя, в свою очередь, зависит от $\sigma$ и от числа измерений, то можно высказать следующее общее правило: точность определения $M$ будет тем больше, чем меньше варьирует данный материал (т.-е. чем меньше $\sigma$), и чем из большего количества измерений вычислено это среднее арифметическое $M$.

Так как та или иная величина варьирования опытного материала является присущим ему свойством и повлиять на увеличение или на уменьшение $\sigma$ не представляется возможным, то точность определения $M$ может быть повышена лишь с помощью увеличения числа измерений. В самом деле, представим себе, что данный материал варьирует в среднем на величину $\sigma = \pm 12$. Если мы вычислим $M$ из 9 измерений, то средняя ошибка $m$ будет равна этой $\sigma$ (т.-е. 12), деленной на $\sqrt{9} = 3$. Следовательно при 9 измерениях $m = \pm 4$. Если же, при той же самой величине варьирования этого материала (т.-е. при $\sigma = \pm 12$), среднее арифметическое $M$ было вычислено из 16 измерений, то средняя ошибка $m$ найдется уже посредством деления $\sigma$ на $\sqrt{16}$, т.-е. на 4 или, окончательно, $m = \pm 3$.



Если бы мы захотели еще более повысить точность определения $M$ или, что то же, уменьшить величину средней ошибки $m$, то для этого стоило бы только еще увеличить число измерений. Напр., при 36 измерениях средняя ошибка $m$ определится посредством деления 12 на $\sqrt{36}$, т.-е. на 6, в результате чего получим $m = \pm 2$.

Из этого числового примера видно, что увеличивая произвольно общее количество измерений, мы можем довести среднюю ошибку $m$ до какой угодно малой величины и тем самым повысить точность определения $M$ до любого желаемого предела.

## IV. Коэффициент корреляции ($r$).

Среднее арифметическое $M$, среднее квадратическое отклонение $\sigma$ и средняя ошибка $m$ дают в своей совокупности характеристику лишь одного варьирующего признака. Вычислив, напр., $M$, $\sigma$ и $m$ отдельно для роста или веса детей какого-либо определенного пола и возраста, мы тем самым вполне точно определим количественно все нужные нам особенности этого признака: его типичную величину, свойственную данной группе объектов исследования, степень изменчивости этого признака и ту точность, с которой представилось возможным определить $M$ при заданном числе измерений.

Но, кроме изолированного изучения отдельных варьирующих признаков, иногда представляется необходимым исследовать также и связь (или соотношение) между двумя различными признаками, т.-е. выяснить, существует или не существует какая-либо зависимость между этими признаками у отдельных объектов исследования. Наглядным примером такой связи двух признаков, или, как это принято называть, корреляции между ними, может служить зависимость между ростом и весом у отдельных детей одного и того же пола и возраста.

Вопрос здесь ставится так: существует или не существует какая-либо связь между ростом каждого отдельного ребенка и его же весом, т.-е. являются ли высокие дети в то же время и самыми тяжелыми и наоборот, или между ростом и весом нет никакой зависимости, т.-е. одинаково вероятно встретить высокого—тяжелого, высокого-легкого, низкого-тяжелого и низкого-легкого ребенка? Мыслим, очевидно, и такой тип связи роста с весом, при котором самые высокие дети оказались бы в то же время и самыми легкими, а низкие — самыми тяжелыми. Такое обратное соотношение в применении к ряду признаков является вполне возможным, и мы должны иметь его в виду, как особый вид корреляции.

Итак, вообще говоря, мы можем встретиться со следующими тремя основными формами корреляции:

а) полная прямая корреляция, когда с увеличением одного из двух признаков всегда наблюдается также и увеличение второго и притом без всяких единичных исключений,

б) отсутствие корреляции, когда оба признака варьируют у различных объектов исследования совершенно независимо друг от друга, так что никакой связи между этими признаками обнаружить не удается,

в) полная обратная корреляция, когда с увеличением одного признака второй, наоборот, всегда уменьшается, а с уменьшением первого второй возрастает, и притом эта зависимость наблюдается также у всех без исключения объектов исследования.

На ряду с этими крайними случаями корреляций мыслимы различные промежуточные степени зависимости между двумя признаками, при которых либо прямая, либо обратная связь между ними существует, но обнаруживается лишь частично и не у всех объектов исследования. Такую неполную зависимость мы будем называть частичной корреляцией, прямой или обратной, смотря по тому, обнаружена ли нами тенденция к одновременному увеличению и одновременному же уменьшению обоих признаков (прямая частичная корреляция), или с увеличением одного признака второй, чаще всего, оказывается количественно уменьшенным (обратная частичная корреляция).

Поясним эти возможные случаи корреляции на каких-либо искусственно подобранных числовых примерах.



1. Полная прямая корреляция. Пусть измерение роста четырех детей дало следующие результаты, выраженные в сантиметрах:

140 142 146 148

а измерение их веса в килограммах — такие результаты:

34 35 37 38

Сопоставляя эти данные, мы можем записать их в две параллельные строчки, причем под ростом каждого отдельного ребенка будет записан его вес:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 34 | 35 | 37 | 38 |

Эту запись, следовательно, надо понимать так: первый ребенок ростом в 140 *см* весит 34 *кг*, второй ребенок ростом в 142 *см* весит 35 *кг* и т. д.

Просматривая эти данные, мы замечаем, что здесь имеется полная прямая зависимость между ростом и весом. В самом деле, самый низкий ребенок (140 *см*) является в то же время и самым легким (34 *кг*); следующий по росту (142 *см*) оказывается уже немного тяжелее (35 *кг*); третий по счету, еще более высокий ребенок (146 *см*) весит еще больше (37 *кг*) и, наконец, самый высокий (148 *см*) в то же время оказывается и самым тяжелым (38 *кг*).

Этот случай зависимости между двумя признаками (ростом и весом) мы и условились выше называть „полной прямой корреляцией“.

2. Частичная прямая корреляция. Переставим в предыдущем примере вес у 2-го и 3-го ребенка, сохранив их рост неизменным. Тогда будем иметь следующую картину соотношения роста с весом:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 34 | 37 | 35 | 38 |

Здесь, попрежнему, самый низкий ребенок (140 *см* является в то же время и самым легким (34 *кг*), а самый высокий (148 *см*) – самым тяжелым (38 *кг*), но в середине ряда эта закономерность несколько нарушена. Так, второй ребенок (142 *см*), будучи ниже третьего (146 *см*), оказывается, тем не менее, несколько тяжелее его (второй весит 37 *кг*, а третий — только 35 *кг*).

Этот случай зависимости мы условились называть „частичной прямой корреляцией".

3. Отсутствие корреляции. Пусть соотношение между ростом и весом выражается у нас следующими двумя рядами чисел:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 37 | 34 | 38 | 35 |

Здесь самый низкий ребенок (140 *см*) является почти самым тяжелым (37 *кг*), следующий по росту (142 *см*) — самым легким (34 *кг*), третий по росту (146 *см*) — самым тяжелым (38 *кг*), а самый высокий (148 *см*) — почти самым легким (35 *кг*). Следовательно здесь не обнаруживается никакой определенной закономерности между двумя этими признаками, которые таким образом распределяются между отдельными объектами исследования в полном беспорядке.

Такой случай связи мы сговорились именовать „отсутствием корреляции".

4. Частичная обратная корреляция. Представим себе, что измерения роста и веса расположились в следующие два ряда чисел:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 38 | 35 | 37 | 34 |

Обнаружив такую картину зависимости между этими признаками, мы должны будем констатировать здесь наличие между ними уже обратной связи, так как самый низкий ребенок (140 *см*) обладает, наоборот, самым большим весом (38 *кг*), а самый высокий (148 *см*) — самым малым весом (34 *кг*). Но эта закономерность здесь все же не полная, ибо два средних по росту ребенка в отношении своего веса несколько нарушают общую картину обратной зависимости, так как сравнительно низкий второй по росту ребенок (142 *см*) оказывается несколько отстающим также по весу (35 *кг*), а более высокий третий ребенок (146 *см*) является вместе с тем и более тяжелым (37 *кг*).



Этот случай зависимости и назван был выше „частичной обратной корреляцией".

5. Полная обратная корреляция. Представим себе, наконец, зависимость между ростом и весом, выраженную такими двумя рядами чисел:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 38 | 37 | 35 | 34 |

Здесь самый низкий ребенок (140 *см*) является в то же время самым тяжелым (38 *кг*); следующий по росту (142 *см*) — уже немного легче (37 *кг*); третий еще более высокий ребенок (146 *см*) весит еще меньше (35 *кг*) и, наконец, самый высокий (148 *см*) в то же время оказывается, наоборот, самым легким (34 *кг*).

Так как обратная зависимость между ростом и весом здесь не имеет никаких исключений, то данный случай можно обозначить наименованием „полной обратной корреляции".

Во всех пяти приведенных здесь примерах характер связи между двумя признаками легко было установить непосредственно путем простого сопоставления роста с весом у каждого отдельного ребенка. Но такой примитивный способ выяснения типа корреляции оказался здесь возможным лишь благодаря чрезвычайно малому количеству объектов исследования (всего лишь четыре ребенка). Нетрудно понять, что при большем числе объектов проследить связь между двумя признаками у каждого из них в отдельности и установить наличие или отсутствие какой-либо закономерности в этой связи представляет задачу уже значительно более сложную, а при очень большом количестве объектов исследования — и просто невыполнимую.

Эти соображения заставляют нас установить какой-либо объективный числовой критерий соотношения двух признаков в виде некоторого числа (или, иначе говоря, „коэффициента"), вычисленного по определенным правилам и величиной своей прямо указывающего как на тип, так и на степень обнаруженной корреляции. Этот искусственно составленный числовой показатель связи двух признаков носит название „коэффициента корреляции" и обозначается обычно буквою $r$.

Поясним принцип вычисления $r$ на разобранных выше числовых примерах, воспользовавшись для этого сперва случаем полной прямой корреляции:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 34 | 35 | 37 | 38 |

Для этого прежде всего выразим отдельные измерения роста и веса в их отклонениях от соответствующих средних арифметических. Сложив все четыре измерения роста, получим в сумме 576 *см*. Разделив эту сумму на 4, найдем, что средний рост этих четырех детей равен 144 *см*. Точно так же найдем, что средний вес их равен 36 *кг*. Вычтя из всех отдельных измерений роста общий средний рост всех детей 144 *см*, а из всех измерений веса — средний вес 36 *кг*, получим следующее соотношение между этими признаками, выраженное уже в их отклонениях от средних арифметических:

| | | | |
|---|---|---|---|
| —4 | —2 | +2 | +4 |
| —2 | —1 | +1 | +2 |

Если мы теперь перемножим каждую пару соответствующих друг другу отклонений роста и веса (напр., —4 на —2, —2 на —1 и т. д.), то получим новый ряд чисел, составленный из произведений всех отклонений, взятых попарно. При этом, очевидно, необходимо соблюдать известное правило знаков: разные знаки (+ и — или — и +) при перемножении дают минус, а одинаковые (+ и + или — и —) в произведении дают плюс. В нашем примере рост первого ребенка отклоняется от среднего роста всех детей (144 *см*) на — 4 *см*, а его вес отклоняется от среднего веса (36 *кг*) на — 2 *кг*. Помножим — 4 на — 2, получим + 8. Вторая пара отклонений — 2 и — 1 при перемножении даст + 2. Отклонения роста и веса третьего ребенка + 2 и + 1 в произведении окажутся равными + 2. Наконец, последняя пара отклонений + 4 и + 2 при перемножении даст + 8. Итак, мы получили следующий ряд произведений:

+8 +2 +2 +8

Сложив все эти произведения, получим в сумме + 20 Это число, представляющее сумму всех произведений, взятых



попарно, как своей величиной, так и знаком (в данном случае +) вполне точно характеризует вид и степень изучаемой связи между двумя признаками, в чем легко убедиться из рассмотрения последующих числовых примеров.

Возьмем, скажем, случай частичной прямой корреляции

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 34 | 37 | 35 | 38 |

Выразив эти измерения попрежнему в их отклонениях от среднего роста (144 *см*) и среднего веса (36 *кг*), получим

| | | | |
|---|---|---|---|
| —4 | —2 | +2 | +4 |
| —2 | +1 | —1 | +2 |

Перемножив все эти отклонения попарно, будем иметь следующий ряд произведений:

+8 —2 —2 +8

(средние числа — 2 и — 2 получились отрицательными, так как мы имели здесь дело с разными знаками: — 2 множили на + 1, а + 2 на — 1).

Сумма произведений равна здесь уже не + 20, как в предыдущем примере полной прямой корреляции, а только + 12 (средние отрицательные числа — 2 и — 2 в данном случае несколько уменьшили общую их сумму).

Из этого примера видно, что с нарушением связи двух признаков тотчас же уменьшается и сумма произведений их отклонений.

Посмотрим теперь, что случится с этой суммой произведений, когда связи между двумя признаками не окажется никакой.

Воспользуемся для этого нашим примером отсутствия корреляции:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 36 | 34 | 38 | 35 |

Выразив эти измерения в их отклонениях от средних арифметических, получаем:

| | | | |
|---|---|---|---|
| —4 | —2 | +2 | +4 |
| +1 | —2 | +2 | —1 |

Произведения этих отклонений дадут здесь следующий ряд чисел:

$$-4 \qquad +4 \qquad +4 \qquad -4,$$

а сумма всех произведений будет равна 0, так как сумма всех положительных чисел (+4 и +4) оказывается в точности равной сумме всех отрицательных (— 4 и — 4).

Итак, при отсутствии корреляции сумма произведений оказывается равной 0.

Нетрудно убедиться в том, что при наличии обратной связи двух признаков сумма произведений всех отклонений сделается отрицательной (т.-е. окажется со знаком —). В самом деле, возьмем пример частичной обратной корреляции:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 38 | 35 | 37 | 34 |

или, в отклонениях от средних арифметических роста и веса:

| | | | |
|---|---|---|---|
| —4 | —2 | +2 | +4 |
| +2 | —1 | +1 | —2 |

Произведения этих отклонений выразятся здесь следующим рядом чисел:

$$-8 \qquad +2 \qquad +2 \qquad -8$$

а сумма произведений будет равна — 12.

Разберем, наконец, последний случай полной обратной корреляции:

| | | | |
|---|---|---|---|
| 140 | 142 | 146 | 148 |
| 38 | 37 | 35 | 34 |

Выразив эти измерения в их отклонениях от средних арифметических, получим:

| | | | |
|---|---|---|---|
| —4 | —2 | +2 | +4 |
| +2 | +1 | —1 | —2 |

Произведения отклонений дают следующий ряд чисел:

$$-8 \qquad -2 \qquad -2 \qquad -8$$

сумма произведений оказалась в данном случае равной — 20 (больше по своей абсолютной величине, чем в предыдущем примере частичной обратной корреляции).



Итак, в примерах прямой корреляции сумма произведений всех отклонений, взятых попарно, оказалась величиной положительной (+20 и +12), при отсутствии корреляции сумма эта обратилась в 0, а при обратной корреляции она стала уже величиной отрицательной (— 12 и — 20). Таким образом знак этой суммы (+ или —) прямо указывает нам на тип имеющейся зависимости между двумя признаками, а абсолютная ее числовая величина — на степень (величину) этой связи. В самом деле, при полной (прямой или обратной) корреляции сумма произведений оказалась наибольшей, равной +20 и — 20, а при частичной — всего только +12 и — 12.

Заметив такое чрезвычайно удобное свойство этой суммы произведений, мы могли бы прямо воспользоваться ею в качестве окончательной характеристики изучаемых корреляций. Но в таком упрощенном виде способ этот представляет все же некоторое неудобство.

Дело в том, что, складывая отдельные произведения, мы накапливаем их сумму, иногда большую, иногда меньшую, в зависимости от того, много или мало отдельных объектов исследования принимается нами в расчет. В наших пяти примерах самая бо̀льшая абсолютная величина этой суммы, соответствующая случаю полной корреляции двух признаков, оказалась равной 20 (+ или —), но если бы мы имели не 4, а значительно бо̀льшее число измерений, то сумма эта, очевидно, могла бы получиться во много раз превышающей число 20. Как же, спрашивается, в этом случае мы все же могли бы констатировать наличие полной корреляции двух признаков, т.-е. узнать, является ли полученная нами сумма произведений наибольшей?

Очевидно, такое распознавание случая полной корреляции, а стало-быть, и количественная оценка любой частичной зависимости между двумя признаками станет возможной лишь тогда, когда мы заранее будем знать максимальное значение этой суммы произведений. Тогда, сравнив полученную нами в данном частном случае такую сумму (напр., +12 или — 12) с ее максимальным значением (20), соответствующим полной связи двух признаков, мы и сможем узнать, является ли данная изучаемая нами зависимость полной, и если нет, то сильно ли отличается она от этого предельного случая полной корреляции.

Каким же способом мы могли бы определить эту наибольшую величину суммы произведений, не прибегая к рассмотрению многих отдельных частных случаев (как это было сделано в предыдущих пяти примерах исключительно с целью выяснения различных возможных результатов группировки двух признаков)?

По соображениям, о которых, однако, здесь не представляется возможным распространяться, максимальное значение суммы произведений может быть вычислено следующим способом, который мы разъясним здесь хотя бы на первом из наших пяти числовых примеров.

Выпишем два ряда измерений (роста и веса), выразив их в отклонениях от соответствующих средних арифметических (144 *см* и 36 *кг*):

| | | | |
|---|---|---|---|
| — 4 | — 2 | + 2 | + 4 |
| — 2 | — 1 | + 1 | + 2 |

Возвысив все эти отклонения в квадрат (как это мы делали раньше при вычислении $\sigma$), получим:

| | | | |
|---|---|---|---|
| 16 | 4 | 4 | 16 |
| 4 | 1 | 1 | 4 |

Сумма квадратов отклонений первого ряда (роста) равна здесь 40, а сумма квадратов отклонений второго ряда (веса) равна 10. Если эти два числа (40 и 10) помножить друг на друга и из произведения их извлечь квадратный корень, то мы и получим нужную нам максимальную величину суммы произведений, вычисленную здесь, так сказать, окольным путем.

Помножив 40 на 10, получим 400. Квадратный корень из 400 равен 20 (так как 20 при помножении на 20 же дает как раз 400). Итак, наша максимальная сумма произведений, как и следовало ожидать, оказалась равной 20.

Заметим кстати, что полученная этим способом максимальная величина суммы произведений совершенно не зависит от порядка сочетания двух признаков у отдельных объектов исследования (т. е. от типа и степени корреляции между ними). Поэтому и для всех других наших примеров,



составленных из тех же самых отклонений роста и веса, но в иных сочетаниях друг с другом, эта величина также окажется равной 20 (в чем нетрудно убедиться, проделав необходимые вычисления в действительности).

Вычислив этим путем наибольшую сумму произведений, мы можем сравнить любое частное ее значение с этой максимальной величиной и узнать таким образом, насколько близко подходит изучаемый случай связи двух признаков к предельному типу полной корреляции. Для этого стоит лишь выяснить, какую часть наша сумма произведений составляет от максимальной суммы, т. е. просто разделить данную частную сумму на это максимальное ее значение.

Ранее мы имели следующие значения суммы произведений для пяти наших примеров:

| | | |
|---|---|---|
| для случая | полной прямой корреляции . . . . . . | + 20 |
| „ „ | частичной прямой корреляции . . . . . | + 12 |
| „ „ | отсутствия корреляции . . . . . . . . | 0 |
| „ „ | частичной обратной корреляции . . . . | — 12 |
| „ „ | полной обратной корреляции . . . . . | — 20 |

Разделив эти суммы на максимальную сумму (20), получим:

| | | |
|---|---|---|
| для случая | полной прямой корреляции . . . . . . | + 1 |
| „ „ | частичной прямой корреляции . . . . | + 0,6 |
| „ „ | отсутствия корреляции . . . . . . . . | 0 |
| „ „ | частичной обратной корреляции . . . | — 0,6 |
| „ „ | полной обратной корреляции . . . . . | — 1 |

Полученные этим путем числа и могут быть приняты за окончательную характеристику связи двух признаков, т. е. за искомые коэффициенты корреляции $r$. Так как каждый из них получается в результате деления суммы произведений всех отклонений на максимальное значение этой суммы, то для крайних случаев полной корреляции (прямой и обратной) мы должны будем разделить друг на друга две равные величины, в результате чего, очевидно, всегда получим единицу (+ 1 или — 1). Таким образом величина коэффициента корреляции $r$, равная + 1, будет свидетельствовать о наличии полной прямой корреляции двух признаков, а при $r = -1$ корреляция будет тоже полной, но уже обратной.

Во всех случаях частичной корреляции для вычисления соответствующего коэффициента $r$ нам придется делить малую сумму произведений на максимальную, т.-е. всегда меньшую величину на бо̀льшую. Поэтому в данном случае коэффициент корреляции окажется непременно величиной дробной (меньше 1), что мы и получили в наших примерах частичной прямой корреляции, когда $r = +0{,}6$, и частичной обратной корреляции, давшей $r = -0{,}6$.

Нетрудно убедиться в том, что при отсутствии корреляции, когда сумма произведений всех отклонений оказывается равной 0, коэффициент корреляции $r$ также всегда будет равен 0.

Таким образом коэффициент корреляции может принимать различные значения в пределах от $+1$ до $-1$, в частности оказываясь иногда равным 0 (или близким к 0). Больше $+1$ и меньше $-1$ коэффициент корреляции $r$ получиться никогда не может. При этом, если $r$ оказывается величиной положительной, то корреляция будет прямой, если же $r$ получается со знаком —, то это свидетельствует уже о корреляции обратной. При $r = 0$ всегда наблюдается отсутствие простейшей закономерной связи между двумя признаками.

Там, где коэффициент корреляции $r$ по своей абсолютной величине получился бо̀льшим, там и связь двух признаков будет более тесной. Напр., если в одном случае коэффициент корреляции $r_1 = +0{,}85$, а в другом $r_2 = +0{,}32$, то первый из этих результатов будет свидетельствовать о более строгой связи двух признаков, т.-е. о том, что количество отдельных нарушений обнаруженной закономерности в первом случае будет меньшим.

Чтобы покончить с вопросом о вычислении коэффициента корреляции, дадим еще один пример его определения, на этот раз из несколько большего количества отдельных измерений.

Пусть попрежнему измерения двух признаков записаны нами в две параллельные строчки, при чем числа, стоящие друг под другом, соответствуют измерениям этих признаков у одних и тех же объектов исследования.



Предположим, что результаты этих измерений расположились в следующие два ряда чисел:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 13 | 16 | 11 | 18 | 14 | 17 | 15 | 19 | 12 |
| 18 | 10 | 14 | 6 | 16 | 2 | 12 | 4 | 8 |

При простом рассмотрении этих чисел характер зависимости между ними улавливается лишь с большим трудом. Нам невозможно охватить сразу 9 пар этих чисел и проследить закономерность в соотношении их величин. Для выяснения этой закономерности необходимо вычислить указанным выше способом соответствующий коэффициент корреляции $r$.

Для того, чтобы выразить эти измерения в их отклонениях от соответствующих средних арифметических, вычислим сперва эти последние по общим правилам.

Сумма чисел верхнего ряда равна здесь 135. Разделив эту сумму на общее количество всех чисел данного ряда, т.-е. на 9, получим 15. Следовательно среднее арифметическое верхнего ряда чисел равно 15.

Так как сумма чисел нижнего ряда равна 90, а количество этих чисел здесь то же самое (т.-е. 9), то среднее арифметическое этого ряда будет равно 10.

Сравнив отдельные числа верхнего ряда с их средним арифметическим (15), а числа нижнего ряда — со вторым средним (10), получим следующие две строчки отклонений:

$$\begin{array}{ccccccccc} -2 & +1 & -4 & +3 & -1 & +2 & 0 & +4 & -3 \\ +8 & 0 & +4 & -4 & +6 & -8 & +2 & -6 & -2 \end{array}$$

После перемножения каждой пары этих отклонений будем иметь следующие их произведения:

$$-16 \quad 0 \quad -16 \quad -12 \quad -6 \quad -16 \quad 0 \quad -24 \quad +6$$

Сумма этих произведений равна здесь $-84$, т.-е. оказалась величиной отрицательной. Следовательно в данном случае мы имеем дело с обратной корреляцией заданных чисел, т.-е. с такой зависимостью между ними, при которой с увеличением чисел верхнего ряда нижние числа чаще всего, наоборот, оказываются уменьшенными, и обратно: малые числа верхнего ряда в большинстве случаев соответствуют большим числам нижнего ряда.

Но с помощью суммы произведений (— 84) мы можем определить лишь основной тип существующей здесь корреляции, величину же ее мы пока еще не знаем. В самом деле, быть-может, абсолютная величина этой суммы (— 84) является уже максимальной, и тогда наша корреляция

Не останавливаясь здесь на разъяснении самого принципа определения средней ошибки $m_r$, укажем лишь технический прием ее вычисления хотя бы на последнем нашем примере. Вспомним, что мы получили там $r = -0{,}7$, и что коэффициент этот был вычислен на основании измерения двух признаков у 9 объектов исследования. Так вот, для определения средней ошибки $m_r$ этого коэффициента корреляции $r$ мы должны возвысить его в квадрат, полученную величину вычесть из единицы и результат разделить на корень квадратный из общего числа измерений.

Проделаем все эти операции с числами нашего примера. Возвысив в квадрат коэффициент корреляции $r = -0{,}7$, получим 0,49 (знак — не ставим так как при возвышении в квадрат все числа делаются положительными). Вычтя полученную величину квадрата коэффициента корреляции (т.-е. 0,49) из единицы, получим 0,51. Это число нам нужно, далее, разделить на корень квадратный из общего количества измерений, т.-е. на $\sqrt{9} = 3$. Разделив 0,51 на 3, получим 0,17. Следовательно при данных условиях ошибка коэффициента корреляции равна $\pm 0{,}17$. Двойной знак (плюс-минус) указывает здесь на то, что средняя ошибка $m_r$ представляет собой среднюю неточность коэффициента корреляции $r$, одинаково возможную как в сторону его увеличения $(+)$, так, наоборот, и в сторону уменьшения $(-)$.

Полученный результат можно записать в обычной символической форме:

$$r = -0{,}7 \pm 0{,}17,$$

где второе число (0,17), соединенное с первым (—0,7) двойным знаком, и представляет собой среднюю ошибку коэффициента корреляции.

Данный числовой результат следует понимать так: полученная нами величина коэффициента корреляции $r$, оказавшаяся равной —0,7, не является вполне точной и заключает в себе некоторую возможную ошибку. Средняя величина этой ошибки равна здесь 0,17. Следовательно истинное значение коэффициента корреляции (которое нам неизвестно) находится где-то в средних пределах от —0,87 $(r - m_r)$ до —0,53 $(r + m_r)$. Максимальная же ошибка этого коэффициента здесь, как и в применении к среднему арифметическому $M$, будет приблизительно в т р и р а з а б о л ь ш е с р е д н е й о ш и б к и $m_r$, т.-е. в данном случае равна $\pm 0{,}51$ (а не $\pm 0{,}17$).



Так как для вычисления $m_r$ приходится квадрат коэффициента корреляции вычитать из единицы, то с увеличением самого коэффициента $r$ ошибка его будет всегда уменьшаться. Следовательно большие коэффициенты корреляции (близкие к единице) оказываются всегда точнее малых коэффициентов (близких к 0).

Так как разницу эту (между единицей и квадратом коэффициента корреляции) приходится затем делить на корень квадратный из числа измерений, то с увеличением общего количества объектов исследования ошибка $m_r$ всегда будет уменьшаться и, следовательно, точность самого коэффицента $r$ возрастать.

Поэтому при исследовании корреляций, обещающих дать большие значения коэффициента $r$, можно ограничиться сравнительно малым количеством отдельных измерений, в то время как для определения наличия малых корреляций необходимо констатировать их на возможно большем количестве объектов исследования.

## V. Оценка достоверности окончательных выводов.

Описанные выше приемы статистической обработки опытных материалов дают возможность получить различные числовые характеристики отдельных варьирующих признаков, количественно определяющие их основные свойства: типичную величину ($M$), изменчивость ($\sigma$) и связь друг с другом ($r$). Получив эти характеристики с той или иной степенью точности, мы можем, далее, делать различные выводы, касающиеся тех или иных свойств и особенностей изучаемого материала, просто путем сопоставления этих общих числовых характеристик друг с другом.

Все выводы, которые этим способом могут быть сделаны, естественно распадаются на два основных типа. К первому из них принадлежит установление той или иной с в я з и между д в у м я отдельными признаками на основании исследования вполне *однородной* группы объектов изучения, а ко второму — доказательство *разницы* в величине *одного* и того же признака у *двух* различных групп этих объектов. Примером первого типа выводов может служить, напр., заключение о наличии прямой корреляции между ростом и весом у детей одного и того же пола и возраста, а примером второго типа — констатирование разницы в росте мальчиков и девочек какого-либо определенного возраста, напр., в сторону большего роста мальчиков. Для доказательства первого из этих выводов мы должны, очевидно, вычислить коэффициент корреляции $r$ и убедиться в том, что он оказался величиной положительной (т.-е. больше 0), а для второго — сравнить средний рост всех мальчиков со средним ростом всех девочек и показать, что первая из этих величин превышает вторую.

Но каждый такой вывод сам по себе дает нам указание лишь на наличие обнаруженного соотношения только у той группы объектов исследования, которая послужила нам материалом для данного заключения. Распространять же этот вывод и на все другие объекты, нами еще не исследованные, мы, очевидно, не имеем никакого права. Выяснив, напр., что рост и вес находятся в прямой корреляции друг с другом у 100 обследованных нами детей, мы не можем еще утверждать, что такая же корреляция этих признаков непременно будет наблюдаться и у всех вообще детей этого пола и возраста, точно также, как, обнаружив больший рост мальчиков по сравнению с девочками у детей какой-либо одной школы, мы не можем распространять этот вывод и на другие школы, если для этого у нас не будет каких-либо особых дополнительных оснований.

Эта невозможность обобщения отдельных частных выводов путем безоговорочного распространения их на все аналогичные случаи изучаемых явлений и фактов объясняется здесь неизбежной неточностью, с которой мы всегда получаем отдельные числовые характеристики варьирующих признаков. Если бы в наших примерах мы могли ручаться за абсолютную точность коэффициента корреляции или отдельных средних арифметических, то никаких сомнений в общности обоих наших выводов, очевидно, не могло бы и возникнуть. Но так как в действительности мы всегда получаем



эти числовые характеристики с теми или иными неизбежными ошибками, зависящими от общего количества произведенных измерений, то и полагаться на эти выводы безоговорочно, всегда считая их вполне достоверными, мы, конечно, не вправе.

Однако, при известном соотношении между величиной самих числовых характеристик и величиной их средних ошибок, мы можем иногда делать более или менее вероятные предположения о том, что данный вывод будет оправдываться и в большинстве других аналогичных случаев, нами не испытанных в действительности, т.-е. другими словами, мы можем иногда оценивать достоверность обобщения этих частных выводов. В самом деле, если, напр., прямая корреляция роста и веса или больший средний рост мальчиков по сравнению с девочками наблюдались нами в одном частном случае, то почему бы этому же явлению не повториться и в некоторых других случаях, а может-быть, и в большинстве из них? В этом у нас может не быть полной уверенности, но некоторая доля вероятности подобного повторения наблюдаемых фактов все же, конечно, имеется всегда.

Как же, спрашивается, мы можем разобраться в вопросе о том, насколько достоверным является то или иное отдельное заключение, и можем ли мы ожидать подтверждения его и в большинстве других случаев?

Способ ориентировки в этом вопросе мы разъясним здесь отдельно в применении к обоим установленным выше основным типам возможных выводов.

Пусть, напр., при исследовании корреляции между ростом и весом мы получили коэффициент $r = +0,5$ с ошибкой $m_r = \pm 0,1$. Эти числовые результаты исследования говорят нам о том, что полученный в данном случае положительный коэффициент $r$ свидетельствует здесь о наличии прямой связи роста с весом и что при других повторных исследованиях коэффициент этот мог бы оказаться либо немного бóльшим, либо немного меньшим, причем варьирование различных значений $r$ возможно здесь в обе стороны от него в среднем на величину $m_r = \pm 0,1$. Но это только среднее отличие других коэффициентов корреляции от нашего $r$.

Максимальная же его ошибка приблизительно равняется утроенной средней ошибке $m_r$, т.-е. в данном случае может достичь величины $\pm 0{,}3$. Следовательно самое малое из всех других значений этого коэффициента будет меньше нашего $r$ на величину 0,3, т.-е. в самом крайнем случае коэффициент этот может оказаться равным $+0{,}2$ $(0{,}5 - 0{,}3 = 0{,}2)$.

Итак, при всех своих возможных колебаниях, наш коэффициент корреляции $r$ никогда не может измениться настолько, чтобы сделаться меньше $+0{,}2$, т.-е. в данном случае он всегда будет величиной положительной.

Убедившись таким образом в неизменности положительных значений коэффициента корреляции, мы можем предполагать, что прямая зависимость между ростом и весом будет наблюдаться и при всех других аналогичных повторных исследованиях, несмотря на то, что выражающий эту зависимость коэффициент $r$ и не всегда будет сохранять ту же самую величину (0,5). Для нашего окончательного вывода некоторая неустойчивость этого коэффициента в сущности и не имеет особого значении. Будет ли $r$ всегда равен 0,5 или же в некоторых случаях он немного уменьшится — это не существенно. Здесь важно было установить лишь, что при всех своих возможных изменениях коэффициент корреляции всегда останется величиной положительной, т.-е. всегда будет указывать на наличие прямой зависимости между ростом и весом. В этом мы убедились, и потому наше первоначальное заключение приобрело для нас характер достаточной достоверности.

Но совершенно иное отношение к тому же самому выводу установилось бы у нас в том случае, если бы, напр., коэффициент корреляции $r$ оказался попрежнему равным $+0{,}5$, но с ошибкой $\pm 0{,}3$. В этом случае наибольшая его неточность могла бы достичь уже величины 0,9 (в три раза большей, чем 0,3), и следовательно при максимальном своем уменьшении наш коэффициент $r$ мог бы оказаться на 0,9 меньше полученного нами (0,5), т.-е. в конечном итоге равным $-0{,}4$ $(0{,}5 - 0{,}9 = -0{,}4)$. Этот результат свидетельствует здесь о том, что некоторые из возможных значений $r$ могут, в частности, сделаться даже числами отрицательными и, стало-быть, указывать уже на наличие обратной зависимости между ростом и весом. Но если в числе других возможных значений многих варьирующих коэффициентов корреляции могут попадаться такие, которые противоречат нашему первоначальному заключению, то и надежность этого заключения будет, очевидно, поколеблена.



Получив подобные результаты, мы не сможем уже ручаться за все другие аналогичные случаи корреляций между ростом и весом и должны будем воздержаться от обобщения нашего первоначального вывода в его категорической форме.

Какое же общее правило для суждения о достоверности подобных заключений вытекает из рассмотрения этих частных примеров?

Очевидно, гарантией достоверности каждого отдельного вывода о наличии определенной связи двух признаков будет служить такая величина коэффициента корреляции $r$, при которой максимальная его ошибка (равная утроенной средней ошибке $m_r$) все же не сможет еще уменьшить этот коэффициент настолько, чтобы он сделался величиной отрицательной. Другими словами, сам коэффициент корреляции $r$ должен быть по крайней мере в 3 раза больше своей средней ошибки $m_r$.

В этом и заключается условие достоверности и общности того вывода, который может быть сделан на основании рассмотрения полученного в каждом частном случае коэффициента корреляции $r$.

Итак, чтобы убедиться в справедливости окончательного вывода о наличии прямой или обратной корреляции двух признаков, необходимо выяснить, во сколько же раз сам коэффициент корреляции $r$ превышает свою ошибку $m_r$, т.-е. попросту разделить $r$ на $m_r$. Если в результате этого деления мы получим величину, равную или бо́льшую 3, то наше первоначальное заключение может считаться действительным и для подавляющего большинства других аналогичных случаев корреляции этих признаков; если же при делении $r$ на $m_r$ получается величина, меньшая 3-х, то заключение это придется признать справедливым лишь для данного частного случая и воздержаться от распространения его и на все остальные.

Совершенно такой же критерий оценки достоверности окончательных заключений мы можем применять и к выводам второго типа.

Пусть, напр., средний рост мальчиков, который мы обозначим через $M_1$, оказался равным 137 *см*, а средний рост девочек того же возраста $M_2 = 135$ *см*, причем средние ошибки этих величин получились: $m_1 = \pm 0{,}4$ *см* (для мальчиков) и $m_2 = \pm 0{,}3$ *см* (для девочек).

Простое сравнение средних арифметических $M_1$ и $M_2$ без учета их ошибок $m_1$ и $m_2$ указывает на больший средний рост мальчиков, превышающий в данном частном случае средний рост девочек на целых 2 *см*. Спрашивается, можем ли мы на основании этих результатов считать, что и вообще все мальчики данного возраста всегда будут в среднем выше своих сверстниц-девочек, или же данный вывод может быть сделан только в отношении той группы детей, которая послужила здесь непосредственным материалом для подобного заключения?

Разрешить этот вопрос можно лишь на основании сопоставления величины полученных средних арифметических $M_1$ и $M_2$ с величиной их средних ошибок $m_1$ и $m_2$.

Будем рассуждать так: на основании измерений только одной группы детей мы получили превышение среднего роста мальчиков над средним ростом девочек на 2 *см*. Повторив эти же измерения в других группах детей того же самого пола и возраста, мы, очевидно, могли бы получить уже не 2 *см* разницы, а может быть, 1,5, или 3, или 2,5 и т. д. Если все эти разницы в среднем росте мальчиков и девочек при всех своих колебаниях все же будут оставаться величинами положительными, то это укажет нам на постоянство подмеченного ранее соотношения, т.-е. на то, что и вообще все мальчики данного возраста будут в среднем выше девочек. В самом деле, первую нашу разницу в 2 *см* мы получили путем вычитания из среднего роста мальчиков (137 *см*) среднего роста девочек (135 *см*). Эта разница оказалась у нас положительной, так как из большей величины (137 *см*) мы вычитали здесь меньшую величину (135 *см*). Отрицательная же величина этой разницы могла бы получиться у нас лишь в том случае, если бы, наоборот, средний рост мальчиков оказался меньше среднего роста девочек. Тогда, вычитая из меньшей величины большую, мы и получили бы числовой результат со знаком минус.



Итак, если бы некоторые из разниц между средним ростом мальчиков и средним ростом девочек получились отрицательными, то это свидетельствовало бы о том, что в других группах мальчики в среднем оказались ниже девочек. Поэтому наше первоначальное заключение о большем среднем росте мальчиков мы можем считать доказанным лишь в том случае, если у нас будет достаточная гарантия в получении исключительно положительных разниц и в полном отсутствии разниц отрицательных.

Что же может нас в этом убедить?

Очевидно, в данном случае, также как и при оценке достоверности коэффициента корреляции $r$, мы должны будем показать, что полученная нами в данном частном случае разница между средним ростом мальчиков и девочек (2 *см*) оказывается по крайней мере в 3 раза больше своей средней ошибки. Стало-быть, здесь задача сводится к нахождению средней ошибки этой разницы.

В теории математической статистики доказывается, что средняя ошибка разницы между двумя величинами $M_1$ и $M_2$ зависит от средних ошибок обоих средних арифметических, т.-е. от $m_1$ и $m_2$, а именно она р а в н а к о р н ю к в а д р а т н о м у и з с у м м ы к в а д р а т о в э т и х о ш и б о к[1]). В нашем примере $m_1 = \pm 0{,}4$, а $m_2 = \pm 0{,}3$. Возвысив эти ошибки в квадрат, будем иметь: 0,16 и 0,09. Сложив эти числа, найдем, что сумма квадратов ошибок равна 0,25. Так как корень из 0,25 равен 0,5, то искомая средняя ошибка нашей разницы оказывается здесь равной $\pm 0{,}5$ *см*.

Для окончательного суждения о достоверности разбираемого здесь вывода мы должны теперь величину нашей разности (2 *см*) разделить на только-что полученную ее среднюю ошибку (0,5). Разделив 2 на 0,5, окончательно получим 4. Так как в этом примере сама разность оказалась в 4 раза больше своей средней ошибки (а не в 3 раза, как это было бы уже достаточно), то заключение наше о большем росте мальчиков следует признать здесь вполне достоверным.

---

1) См. главу XVII, стр. 225.

Итак, для доказательства достоверности и общности выводов второго типа, т.-е. для получения уверенности в том, что обнаруженная в частном случае разница между двумя числовыми характеристиками варьирующего признака будет наблюдаться и в большинстве других аналогичных случаев, мы должны разделить данную разницу на корень из суммы квадратов двух ошибок этих числовых характеристик. Если полученное в результате такого деления число окажется равным или бо́льшим 3-х, то существование данной разницы и во всех других случаях может считаться доказанным; при невыполнении же этого условия в общности и абсолютной достоверности нашего первоначального вывода следует сомневаться.

Так как извлечение квадратного корня всегда представляет задачу довольно хлопотливую, то при оценке достоверности вывода второго типа можно несколько упростить технику вычислений, возвысив все величины, с которыми здесь приходится иметь дело, в квадрат. Тогда взамен деления самой разницы на корень из суммы квадратов ошибок мы должны будем разделить квадрат этой разницы просто уже на сумму квадратов ошибок (не извлекая корня). Но при этом нам придется, очевидно, возвысить в квадрат также и число 3 и ориентироваться уже, следова́тельно, по его квадрату, т.-е. по числу 9. Поясним этот упрощенный способ вычисления на прежнем примере.

Так как $M_1 = 137$ *см*, а $M_2 = 135$ *см*, то разница $M_1 - M_2 = 2$ *см*. Эту разницу следует возвысить в квадрат, в результате чего получим 4. Далее, каждую из двух ошибок (0,4 и 0,3) следует также возвысить в квадрат. Получим попрежнему: 0,16 и 0,09. Сложив эти квадраты, найдем, что сумма их равна 0,25. Теперь квадрат нашей разницы (4) следует разделить на эту сумму квадратов (0,25). В результате получим, очевидно, 16. Так как это число оказалось больше 9, то вывод наш можно считать достоверным.

В дальнейшем мы будем пользоваться чаще всего этим последним способом оценки достоверности выводов второго типа, способом, который по существу совершенно равноценен предыдущему и отличается от него лишь большей простотой вычисления.



Кроме выяснения разницы в средней величине двух варьирующих признаков, этим же способом можно определить также и достоверность того или иного различия в степени изменчивости этих признаков. Для этого стоит лишь сравнить друг с другом два средних квадратических отклонения ($\sigma$) и, учтя величину их ошибок, применить здесь описанный выше критерий оценки достоверности.

Прежде, чем приводить числовой пример такого сравнения двух средних квадратических отклонений, укажем способ вычисления их ошибок. В теории математической статистики доказывается, что среднюю ошибку $\sigma$ (сигмы) можно вычислить с помощью того же самого приема, который уже применялся нами ранее для определения ошибки $m$ среднего арифметического $M$. Для этого следует разделить $\sigma$ на квадратный корень из у д в о е н н о г о количества измерений. Если, напр., при каких-либо вычислениях $\sigma$ оказалась равной $\pm 2$ и мы вывели ее из 50 отдельных измерений, то среднюю ошибку этой $\sigma$, которую мы обозначим через $m$ со знаком $\sigma$, т.-е. символом $m_\sigma$, можно определить посредством следующих простых вычислений. Помножив общее число измерений (т.-е. 50) на 2, получим 100. Так как корень квадратный из 100 равен 10, то, разделив нашу $\sigma$ (равную 2) на 10, окончательно получим 0,2. Следовательно в данном случае искомая средняя ошибка сигмы $m_\sigma = \pm 0{,}2$.

Из описания этого способа видно, что ошибка сигмы ($m_\sigma$) всегда оказывается несколько меньше ошибки среднего арифметического ($m$). Поэтому, при всех прочих равных условиях, $\sigma$ всегда определяется с большею точностью, чем среднее арифметическое $M$.

Средняя ошибка $m_\sigma$, как и всякая средняя ошибка, дает указание на некоторый средний размах возможных колебаний сигмы при многих повторных ее вычислениях из вполне однородного материала, но на других объектах исследования. Получив, напр., $m_\sigma = \pm 0{,}2$, мы можем утверждать, что при других аналогичных исследованиях, нами фактически не выполненных, величина $\sigma$ могла бы получиться и иная, отличная от данной, но что все эти отдельные сигмы отличались бы от нашей $\sigma$ в среднем как раз на величину этой ошибки $m_\sigma = \pm 0{,}2$.

Научившись вычислять среднюю ошибку сигмы и уяснив себе ее значение, разберем какой-нибудь числовой пример сравнения степени изменчивости двух варьирующих признаков. Пусть в одном случае мы получили $\sigma_1 = \pm 3{,}6$, а в другом $\sigma_2 = \pm 2{,}3$, причем ошибка первой сигмы $m_{\sigma_1} = \pm 0{,}9$, а ошибка второй $m_{\sigma_2} = \pm 0{,}7$. Спрашивается, существует ли здесь какая-либо разница в величине этих сигм, и если существует, то насколько этому заключению можно доверять?

Первая часть этого вопроса разрешается, очевидно, чрезвычайно просто. Так как $\sigma_1$ (3,6) оказалась здесь больше $\sigma_2$ (2,3), то, очевидно, разница между ними в данном частном случае существует. Для ответа на вторую часть вопроса мы должны будем квадрат данной разницы разделить на сумму квадратов двух средних ошибок этих сигм. Если результат такого деления окажется больше 9, то достоверность нашего заключения может считаться доказанной; в противном случае обнаруженную нами разницу придется признать лишь случайной.

Так как $\sigma_1 = \pm 3{,}6$, а $\sigma_2 = \pm 2{,}3$, то разница между абсолютными их значениями в данном случае будет 1,3. Возвысив эту разницу в квадрат, получим 1,69. Теперь возвысим в квадрат обе средние ошибки этих сигм $m_{\sigma_1} = \pm 0{,}9$ и $m_{\sigma_2} = \pm 0{,}7$. Квадраты их будут соответственно равны 0,81 и 0,49. Сложив эти числа, найдем, что сумма квадратов ошибок равна здесь 1,3.

Разделив квадрат разницы (1,69) на сумму квадратов ошибок (1,3), окончательно получим 1,3. Так как это число оказалось значительно меньше 9, то наш первоначальный вывод не может еще считаться вполне достоверным. Это значит, что, повторив аналогичные вычисления на других объектах исследования, мы далеко не всегда обнаружим точно такую же разницу между двумя сигмами. Поэтому более значительное варьирование первого признака не может быть признано здесь явлением общим.

Указанный способ оценки достоверности вывода можно применить и к сравнению двух коэффициентов корреляции $r_1$ и $r_2$, вычисленных со средними ошибками $m_{r_1}$ и $m_{r_2}$.

Пусть, напр., $r_1 = +0{,}9$, $r_2 = +0{,}4$, $m_{r_1} = \pm 0{,}1$, $m_{r_2} = \pm 0{,}2$. В данном частном случае больший коэффициент $r_1$ указывает на бо́льшую связь двух признаков. Чтобы узнать, можно ли ожидать подтверждения этого вывода и в других аналогичных случаях, выполним все указанные выше операции с числами нашего примера.



Разница между двумя коэффициентами корреляции равна здесь 0,5. Квадрат этой разницы будет 0,25. Квадраты двух ошибок (0,1 и 0,2) соответственно равны 0,01 и 0,04. Сумма этих квадратов выразится числом 0,05. Разделив квадрат разницы (0,25) на эту сумму квадратов ошибок (0,05), получим 5. Так как 5 меньше 9, то за полную достоверность нашего первоначального вывода ручаться здесь также нельзя.

Указанными выше способами оценки достоверности окончательных выводов завершается статистическая обработка всяких массовых измерений. Так как с помощью отдельных числовых характеристик варьирующего материала мы можем выяснить лишь те его свойства и особенности, которые наблюдаются только у данной ограниченной группы объектов исследования, то для суждения о наличии этих свойств и у всех других подобных же объектов нам необходимо каждое такое обобщение доказать особо. Не выполнив этой заключительной и самой важной операции в процессе количественного анализа изучаемых явлений и фактов, мы всегда рискуем какую-либо случайную зависимость или различие принять за общее свойство, присущее всем без исключения подобным же объектам исследования и получить, стало-быть, заведомо ложный вывод.

Всякое научное исследование в конечном счете стремится к установлению, обоснованию и анализу общих законов. Подметить эти законы можно, конечно, и на очень небольшом количестве отдельных фактов, но убедиться в абсолютной достоверности той или иной научной гипотезы мы можем либо путем фактической проверки ее на всем имеющемся в природе материале (что практически неосуществимо), либо посредством чисто-логического доказательства, построенного на количественном анализе исследуемых явлений и фактов. Этот последний, наиболее рациональный путь доказательства общих законов на массовом материале и заключается в применении методов вариационной статистики.

## VI. Способ непосредственного вычисления.

Введем некоторые названия и обозначения, необходимые нам для использования их в дальнейшем изложении.

Полученный из опыта ряд чисел, изображающих отдельные измерения варьирующего признака, назовем „вариационным рядом“, а отдельные члены этого ряда (числовые результаты отдельных измерений) — „вариантами“. Эти варианты мы будем в дальнейшем изображать буквами $V$, а сумму их — условным символом $\Sigma$ (знак суммирования). Таким образом для обозначения суммы всех вариант данного вариационного ряда мы можем воспользоваться такой записью: $\Sigma V$ (выговаривается „сумма $V$“).

„Отклонения“ отдельных вариант от среднего арифметического $M$ мы будем обозначать буквами $x$ (икс), квадраты этих отклонений — через $x^2$, а сумму отклонений и сумму их квадратов — с помощью символов $\Sigma x$ и $\Sigma x^2$.

Если, наконец, общее число всех измерений (вариант) обозначить через $n$, то для вычисления $M$, $\sigma$ и $m$ могут служить следующие формулы:

$$M = \frac{\Sigma V}{n}, \qquad \sigma = \pm\sqrt{\frac{\Sigma x^2}{n-1}} \qquad \text{и} \quad m = \frac{\pm\sigma}{\sqrt{n}}.$$

Первая из этих формул показывает, что для вычисления среднего арифметического $M$ следует сложить все варианты $V$ и полученную сумму их ($\Sigma V$) разделить на общее количество измерений $n$.

Вторая формула показывает, что для вычисления среднего квадратического отклонения $\sigma$ следует определить отклонения всех вариант $V$ от среднего арифметического $M$, каждое такое отклонение ($x$) возвысить в квадрат (получим $x^2$), сумму этих квадратов ($\Sigma x^2$) разделить на число измерений без единицы (т. е. на $n-1$) и из полученного результата извлечь квадратный корень.

Третья формула показывает, что для вычисления средней ошибки $m$ среднего арифметического $M$ следует среднее квадратическое отклонение $\sigma$ разделить на квадратный корень из общего количества измерений $n$.

При малом количестве вариант и при условии, что все они выражаются исключительно в целых числах, вычисление $M$,



$\sigma$ и $m$ может быть произведено с помощью точного выполнения всех операций, указанных в этих формулах. Такой путь определения числовых значений $M$, $\sigma$ и $m$ мы назовем „способом непосредственного вычисления“.

Для удобства выполнения промежуточных операций, связанных с вычислением суммы всех вариант ($\Sigma V$), отдельных отклонений ($x$), их квадратов ($x^2$) и суммы этих квадратов ($\Sigma x^2$), полезно записывать отдельные результаты этих вычислений не в беспорядке, а располагая их по клеткам особой „расчетной решетки“.

ТАБЛИЦА 1.

| $V$ | $x$ | $x^2$ |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

Пусть, напр., имеем вариационный ряд, состоящий из следующих 16 вариант, **выражающих результат измерения какого-либо** варьирующего **признака** в соответствующих мерах (в сантиметрах, в килограммах, в секундах, в градусах, в процентах и т. д):

10, 8, 7, 9, 11, 5, 10, 8, 6,
12, 7, 9, 11, 9, 6, 8.

Начертим расчетную решетку (см. таблицу 1), состоящую из трех вертикальных столбцов и нескольких горизонтальных строчек. Число этих строчек определяется числом имеющихся в нашем распоряжении вариант (в данном случае 16), при чем сверху и снизу этой решетки необходимо прибавить еще по одной горизонтальной строчке, отделив их от остальных двойной линией. В верхней дополнительной строчке мы поместим буквенные обозначения соответствующих вертикальных столбцов $V$, $x$ и $x^2$. Эти заголовки указывают на то, что в первом слева вертикальном столбце записываются заданные варианты ($V$), во втором, среднем — их отклонения ($x$), а в третьем, правом — квадраты этих отклонений ($x^2$). В нижней дополнительной строчке мы будем записывать суммы чисел, подсчитанных отдельно по каждому из трех вертикальных столбцов.

Запишем теперь все наши 16 вариант в клетки вертикального столбца, озаглавленного буквою $V$ (в любом произвольном порядке), и для вычисления среднего арифметического $M$ подсчитаем сумму чисел данного столбца (см. табл. 2). Эту сумму ($\Sigma V$) запишем внизу, в последней горизонтальной строчке под числами столбца $V$. В нашем примере $\Sigma V$ оказалась равной 136. Для вычисления $M$ нужно $\Sigma V$ разделить на число измерений $n$ (в нашем случае на 16). Все эти действия запишем отдельно от таблицы в таком виде:

ТАБЛИЦА 2.

| $V$ | $x$ | $x^2$ |
|---|---|---|
| 10 | | |
| 8 | | |
| 7 | | |
| 9 | | |
| 11 | | |
| 5 | | |
| 10 | | |
| 8 | | |
| 6 | | |
| 12 | | |
| 7 | | |
| 9 | | |
| 11 | | |
| 9 | | |
| 6 | | |
| 8 | | |
| 136 | | |

$$M = \frac{\Sigma V}{n} = \frac{136}{16} = 8{,}5.$$

Итак, в нашем примере среднее арифметическое всех 16-ти заданных вариант оказалось равным 8,5.

После определения среднего арифметического $M$ следует вычислить отклонения ($x$) каждой отдельной варианты $V$ от этого среднего арифметического $M = 8{,}5$ и записать результаты с соответствующими знаками ($+$ или $-$) в средний столбец расчетной решетки, озаглавленный буквою $x$ (см. табл. 3). Для проверки вычислений мы можем сложить все числа этого столбца, принимая во внимание их знаки (т.-е. сложить отдельно все положительные числа и отдельно же—все отрицательные). Если при этом, как в данном примере, суммы положительных и отрицательных отклонений окажутся по своей абсолютной величине равными друг другу ($+13$ и $-13$), т.-е. дадут в общем итоге нуль, мы можем считать наши вычисления правильными. В противном случае следует искать ошибку или в вычислении самого $M$, или где-либо в определении отдельных отклонений $x$.



Вычислив и проверив все отклонения $x$, следует каждое из них возвысить в квадрат и результаты записать в клетки правого вертикального столбца. При этом знаки мы можем уже не ставить, так как при возвышении в квадрат все

ТАБЛИЦА 3.

| $V$ | $x$ | $x^2$ |
|---|---|---|
| 10 | +1,5 | |
| 8 | −0,5 | |
| 7 | −1,5 | |
| 9 | +0,5 | |
| 11 | +2,5 | |
| 5 | −3,5 | |
| 10 | +1 5 | |
| 8 | −0,5 | |
| 6 | −2,5 | |
| 12 | +3,5 | |
| 7 | −1,5 | |
| 9 | +0,5 | |
| 11 | +2,5 | |
| 9 | +0,5 | |
| 6 | −2,5 | |
| 8 | −0,5 | |
| 136 | 0 | |

ТАБЛИЦА 4.

| $V$ | $x$ | $x^2$ |
|---|---|---|
| 10 | +1,5 | 2,25 |
| 8 | −0,5 | 0,25 |
| 7 | −1,5 | 2,25 |
| 9 | +0,5 | 0,25 |
| 11 | +2,5 | 6,25 |
| 5 | −3,5 | 12,25 |
| 10 | +1,5 | 2,25 |
| 8 | −0,5 | 0,25 |
| 6 | −2,5 | 6,25 |
| 12 | +3,5 | 12,25 |
| 7 | −1,5 | 2,25 |
| 9 | +0,5 | 0,25 |
| 11 | +2,5 | 6,25 |
| 9 | +0,5 | 0,25 |
| 6 | −2,5 | 6,25 |
| 8 | −0,5 | 0,25 |
| 136 | 0 | 60,00 |

числа (как положительные, так и отрицательные) будут всегда иметь знак $+$.

Для получения суммы квадратов отклонений ($\Sigma x^2$) нам остается лишь сложить все числа правого столбца и подписать их сумму в соответствующей клетке нижней дополнительной строчки. В нашем примере $\Sigma x^2$ оказалась равной 60.

После всех этих вычислений и записей наша решетка примет окончательный вид, указанный на табл. 4.

Теперь, согласно указаниям формулы для вычисления среднего квадратического отклонения:

$$\sigma = \pm \sqrt{\frac{\Sigma x^2}{n-1}}$$

мы должны полученную из расчетной решетки $\Sigma x^2$ (т.-е. 60) разделить на число всех измерений без единицы (т.-е. в нашем примере на $16-1=15$). Разделив 60 на 15, получим 4. Из этого числа, далее, мы должны извлечь квадратный корень. Так как $\sqrt{4} = \pm 2$, то среднее квадратическое отклонение всех 16-ти заданных вариант

$$\sigma = \pm 2.$$

Все эти вычисления можно расположить в следующую строчку:

$$\sigma = \pm \sqrt{\frac{\Sigma x^2}{n-1}} = \pm \sqrt{\frac{60}{16-1}} = \pm \sqrt{\frac{60}{15}} = \pm \sqrt{4} = \pm 2.$$

Средняя ошибка $m$ может быть получена путем деления только-что вычисленной $\sigma$ на корень из числа измерений $n$. Так как в нашем примере $\sigma = \pm 2$, а $n = 16$, то

$$m = \frac{\pm \sigma}{\sqrt{n}} = \frac{\pm 2}{\sqrt{16}} = \frac{\pm 2}{4} = \pm 0{,}5.$$

Итак, для разобранного нами числового примера

$$M = 8{,}5 \pm 0{,}5, \quad \text{а} \quad \sigma = \pm 2.$$

Сделаем же теперь краткую сводку правил определения $M$, $\sigma$ и $m$ по способу непосредственного вычисления. Порядок этих вычислений, следовательно, таков:

1. Составить расчетную решетку, записать в нее все $n$ отдельных вариант $V$ и сложить их (получим $\Sigma V$).
2. Вычислить среднее арифметическое по формуле:

$$M = \frac{\Sigma V}{n}.$$

3. Вычислить отклонения $x$ для всех вариант, записать их в решетку и проверить сложением, убедившись в том, что

$$\Sigma x = 0.$$



4. Возвысить все отклонения в квадрат, записать их в решетку и сложить (получим $\Sigma x^2$).
5. Вычислить среднее квадратическое отклонение по формуле:

$$\sigma = \pm \sqrt{\frac{\Sigma x^2}{n-1}}.$$

6. Вычислить среднюю ошибку по формуле:

$$m = \frac{\pm \sigma}{\sqrt{n}}.$$

Как видно из приведенного выше примера, вычисление $M$, $\sigma$ и $m$ описанным здесь способом технически очень несложно. Единственным моментом, могущим представить некоторое затруднение, может явиться лишь извлечение квадратного корня. В нашем примере корень извлекался непосредственно и легко, потому что число, из которого требовалось извлечь этот корень (4 или 16), представляло собой полный квадрат искомой величины (2 или 4). В общем же случае мы, очевидно, не будем иметь такой благоприятной для нас случайности, и корень квадратный придется извлекать иными, более сложными приемами, описанию которых будет посвящена особая глава (X-я).

---

## VII. Способ округления.

Описанный выше способ „непосредственного вычисления“ $M$, $\sigma$ и $m$ не всегда, однако, является удобным в практическом отношении. Когда вариационный ряд состоит из большого количества отдельных вариант и среднее арифметическое $M$ оказывается дробным числом с несколькими десятичными знаками, вычисление всех отклонений ($x$) представляет уже чрезвычайно кропотливую работу, потому что каждое такое отклонение, в свою очередь, оказывается дробным числом. Кроме того, большое затруднение представляет и возвышение всех этих дробных отклонений в квадрат, а также сложение многих дробных чисел в столбцах $x$ и $x^2$.

Чтобы избежать этих затруднений, можно воспользоваться другим приемом вычисления этих величин, приемом, который мы назовем „способом округления“.

Для того, чтобы показать полное совпадение конечных результатов вычисления при пользовании обоими этими способами, разберем их на одном и том же числовом примере. Пусть попрежнему имеем вариационный ряд, состоящий из тех же самых 16 вариант:

10, 8, 7, 9, 11, 5, 10, 8, 6, 12, 7, 9, 11, 9, 6, 8.

ТАБЛИЦА 5.

| $V$ | $x_c$ | $x_c^2$ |
|---|---|---|
| 10 | | |
| 8 | | |
| 7 | | |
| 9 | | |
| 11 | | |
| 5 | | |
| 10 | | |
| 8 | | |
| 6 | | |
| 12 | | |
| 7 | | |
| 9 | | |
| 11 | | |
| 9 | | |
| 6 | | |
| 8 | | |
| 136 | | |

Составим расчетную решетку по форме, указанной в прошлой главе, но заголовки вертикальных столбцов поставим в ней уже иные (см. табл. 5). Первый столбец попрежнему озаглавим буквою $V$, над вторым поставим букву $x$ со значком $c$, т.-е. $x_c$, а рядом напишем $x_c^2$. Значение двух последних обозначений выяснится из описания техники дальнейших расчетов.

Теперь попрежнему запишем в столбец $V$ все варианты нашего ряда, сложим их и получим $\Sigma V$, которая, очевидно, как и раньше, окажется равной 136.

Следовательно среднее арифметическое нашего вариационного ряда

$$M = \frac{\Sigma V}{n} = \frac{136}{16} = 8{,}5.$$

Чтобы при вычислении отдельных отклонений избежать вычитания из целых вариант $V$ дробного среднего арифметического $M$ ($= 8{,}5$), округлим его до ближайшего целого числа (отсюда и название способа). Таким округленным средним арифметическим может служить здесь или число 8 (если отбросить 5 десятых), или число 9 (если прибавить к $M$ недостающую половину). Для применения этого способа совершенно безразлично, увеличится или уменьшится при округлении наше среднее арифметическое $M$, поэтому



мы можем воспользоваться здесь с одинаковым правом как числом 8, так и числом 9.

Примем за округленное среднее большее число 9 и обозначим его буквою $C$. Следовательно у нас:

$$C = 9.$$

Теперь будем вычислять „отклонения“ отдельных вариант $V$, но уже не от дробного числа $M = 8{,}5$ а от целого $C = 9$, получая в результате, очевидно, также целые числа. Эти новые отклонения мы назовем „вспомогательными“, в отличие от прежних „центральных“ отклонений. Вспомогательные отклонения мы и запишем в клетки среднего столбца расчетной решетки (см. табл. 6). Заголовок этого столбца ($x_c$) указывает, следовательно, на то, что в нем записаны „вспомогательные“ отклонения, отсчитанные от округленного среднего $C$.

ТАБЛИЦА 6.

| $V$ | $x_c$ | $x_c^2$ |
|---|---|---|
| 10 | + 1 | |
| 8 | − 1 | |
| 7 | − 2 | |
| 9 | 0 | |
| 11 | + 2 | |
| 5 | − 4 | |
| 10 | + 1 | |
| 8 | − 1 | |
| 6 | − 3 | |
| 12 | + 3 | |
| 7 | − 2 | |
| 9 | 0 | |
| 11 | + 2 | |
| 9 | 0 | |
| 6 | − 3 | |
| 8 | − 1 | |
| 136 | − 8 | |

Если теперь сложить все эти вспомогательные отклонения $x_c$, то в сумме мы уже не получим нуля, так как этим свойством обладают только „центральные“ отклонения $x$, отсчитанные от среднего арифметического $M$. В самом деле, сумма всех вспомогательных отклонений со знаком „плюс“ равна здесь 9, и со знаком „минус“ — 17. Вычитая из большей суммы (17) меньшую (9), получим 8. Так как большая сумма оказалась здесь отрицательной, то и разность (8) будет также иметь знак „минус“.

Итак, общая сумма всех вспомогательных отклонений равна — 8 (а не 0, как сумма „центральных“ отклонений, полученная в предыдущем примере при описании способа непосредственного вычисления). Эту сумму вспомогательных отклонений $(-8)$ мы обозначим буквою $S$ со значком 1, т.-е. символом $S_1$. Следовательно в данном случае

$$S_1 = -8.$$

Этот же результат мы могли бы предугадать и заранее, вычислив значение величины $S_1$ по следующей контрольной формуле:

$$S_1 = \Sigma V - nC.$$

Формула эта показывает, что для вычисления $S_1$ мы должны число $C$ (округленное среднее арифметическое, в данном случае 9) помножить на общее количество измерений $n$ (т.-е. на 16) и полученный результат вычесть из суммы всех вариант $\Sigma V$ (т.-е. из 136). Помножив 9 на 16, получим 144. Вычтя из 136 это произведение 144, очевидно, как раз и получим $-8$. Все эти выкладки удобно записать в следующую строчку

$$S_1 = \Sigma V - nC = 136 - 16 . 9 = 136 - 144 = -8.$$

Возможность предугадать заранее значение величины $S_1$ позволяет проверить вычисление всех вспомогательных отклонений $x_c$. Если фактическая сумма их $S_1$, подсчитанная путем сложения всех чисел среднего вертикального столбца расчетной решетки, окажется в точности равной тому числу, которое определяется по указанной выше контрольной формуле, мы можем считать наши расчеты правильными; в противном случае следует искать ошибку где-либо в вычислении отдельных вспомогательных отклонений $x_c$.

После проверки вспомогательных отклонений следует, как и ранее, возвысить каждое из них в квадрат, результаты записать в крайний правый столбец расчетной решетки и просуммировать (см. табл. 7).

Полученная этим путем сумма квадратов вспомогательных отклонений, которую мы обозначим буквою $S$ со значком 2, т.-е. через $S_2$, очевидно, не совпадет с суммой квадратов центральных отклонений $\Sigma x^2$, вычисленной по предыдущему способу. В самом деле, ранее применяя способ непосредственного вычисления, мы имели:

$$\Sigma x^2 = 60.$$



Теперь же при помощи способа округления получили:

$$S_2 = 64.$$

Итак, пользуясь способом округления, мы значительно упростили технику всех промежуточных вычислений, производимых внутри расчетной решетки (вместо дробных чисел все время имели дело с целыми), но зато в конечном итоге получили не искомую сумму квадратов центральных отклонений $\Sigma x^2 = 60$, а другую бо̀льшую сумму квадратов вспомогательных отклонений $S^2 = 64$.

ТАБЛИЦА 7.

| $V$ | $x_c$ | $x_c^2$ |
|---|---|---|
| 10 | $+1$ | 1 |
| 8 | $-1$ | 1 |
| 7 | $-2$ | 4 |
| 9 | 0 | 0 |
| 11 | $+2$ | 4 |
| 5 | $-4$ | 16 |
| 10 | $+1$ | 1 |
| 8 | $-1$ | 1 |
| 6 | $-3$ | 9 |
| 12 | $+3$ | 9 |
| 7 | $-2$ | 4 |
| 9 | 0 | 0 |
| 11 | $+2$ | 4 |
| 9 | 0 | 0 |
| 6 | $-3$ | 9 |
| 8 | $-1$ | 1 |
| 136 | $-8$ | 64 |

Но этим смущаться не следует, так как существует способ вычисления нужной нам суммы $\Sigma x^2$ при помощи двух полученных из расчетной решетки вспомогательных сумм $S_1$ и $S_2$. Способ этот заключается в применении следующей формулы:

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

Эта формула показывает, что для вычисления суммы $\Sigma x^2$ следует возвысить в квадрат сумму $S_1$, результат разделить на число измерений $n$ и полученную таким образом поправочную величину $\left(\frac{S_1^2}{n}\right)$ вычесть из суммы $S_2$.

В нашем примере $S_1 = -8$, $S_2 = 64$ и $n = 16$. Возвысив $-8$ в квадрат, получим 64. Разделив 64 на 16, получим 4. Вычтя из 64 эту поправочную величину 4, найдем, что искомая сумма $\Sigma x^2 = 60$. Все эти действия удобно расположить в следующую строчку:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 64 - \frac{(-8)^2}{16} = 64 - \frac{64}{16} = 64 - 4 = 60.$$

Как и следовало ожидать, этот результат в точности совпал с тем значением $\Sigma x^2$, которое ранее было получено нами прямо из расчетной решетки.

Определив по указанной формуле значение суммы квадратов центральных отклонений $\Sigma x^2$, мы заканчиваем далее наши вычисления совершенно так же, как и в случае применения предыдущего способа. А именно:

$$\sigma = \pm \sqrt{\frac{\Sigma x^2}{n-1}} = \pm \sqrt{\frac{60}{16-1}} = \pm \sqrt{\frac{60}{15}} = \pm \sqrt{4} = \pm 2.$$

$$m = \frac{\pm \sigma}{\sqrt{n}} = \frac{\pm 2}{\sqrt{16}} = \frac{\pm 2}{4} = \pm 0{,}5.$$

Итак, при производстве вычислений по способу округления мы должны выполнить следующие операции:

1. Составить расчетную решетку, занести в нее все варианты $V$ в произвольном порядке и сложить их (получим $\Sigma V$).

2. Вычислить среднее арифметическое $M$ по формуле:

$$M = \frac{\Sigma V}{n}.$$

3. Округлить полученное значение $M$ до ближайшего целого числа $C$ (все равно большего или меньшего).

4. Вычислить „вспомогательные“ отклонения $x_c$ каждой варианты $V$ от округленного среднего $C$, записать их в расчетную решетку и сложить (получим сумму вспомогательных отклонений $S_1$).

5. Проверить вычисления по контрольной формуле:

$$S_1 = \Sigma V - nC.$$

6. Возвысить каждое вспомогательное отклонение в квадрат, результаты ($x_c{}^2$) записать в расчетную решетку и сложить (получим сумму квадратов вспомогательных отклонений $S_2$).

7. Определить сумму квадратов „центральных“ отклонений (от среднего арифметического $M$) по формуле:

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



8. Вычислить среднее квадратическое отклонение по формуле:

$$\sigma = \pm \sqrt{\frac{\Sigma x^2}{n-1}}.$$

9. Вычислить среднюю ошибку по формуле:

$$m = \frac{\pm \sigma}{\sqrt{n}}.$$

ТАБЛИЦА 8.

| $V$ | $x_c$ | $x_c{}^2$ |
|---|---|---|
| 10 | + 3 | 9 |
| 8 | + 1 | 1 |
| 7 | 0 | 0 |
| 9 | + 2 | 4 |
| 11 | + 4 | 16 |
| 5 | − 2 | 4 |
| 10 | + 3 | 9 |
| 8 | + 1 | 1 |
| 6 | − 1 | 1 |
| 12 | + 5 | 25 |
| 7 | 0 | 0 |
| 9 | + 2 | 4 |
| 11 | + 4 | 16 |
| 9 | + 2 | 4 |
| 6 | − 1 | 1 |
| 8 | + 1 | 1 |
| 136 | + 24 | 96 |

В приведенном выше примере мы совершенно произвольно приняли за начало отсчетов округленное среднее $C = 9$ (вместо самого среднего арифметического $M = 8{,}5$). Заметим, что без всякого ущерба для точности вычисления окончательных результатов мы могли бы принять за $C$ какое угодно другое число, например, 10, 6, 12, 5 и т. д. При этом мы в каждом отдельном случае, очевидно, получали бы различные пары значений вспомогательных сумм $S_1$ и $S_2$, но вычисленная из этих величин сумма квадратов центральных отклонений $\Sigma x^2$ всегда оказалась бы в точности равной 60, как и в предыдущем примере.

Проделаем эти вычисления еще раз, приняв за $C$, напр., число 7 (см. табл. 8). Определив все вспомогательные отклонения, запишем их в решетку и сложим. Так как сумма всех положительных отклонений равна здесь 28, а сумма отрицательных — 4, то общая сумма всех вспомогательных отклонений $S_1 = +24$. Для проверки этих вычислений воспользуемся контрольной формулой:

$$S_1 = \Sigma V - nC.$$

Так как у нас $\Sigma V = 136$, $n = 16$, а $C = 7$, то

$$S_1 = 136 - 16.7 = 136 - 112 = +24.$$

Следовательно все вспомогательные отклонения $x_c$ и их сумма $S_1$ вычислены правильно.

Теперь возвысим эти отклонения в квадрат, запишем в расчетную решетку и сложим. Получим сумму квадратов вспомогательных отклонений $S_2 = 96$.

Чтобы с помощью двух вспомогательных сумм $S_1 = +24$ а $S_2 = 96$ найти нужную нам сумму квадратов центральных отклонений $\Sigma x^2$, воспользуемся формулой:

$$\Sigma x^2 = S_2 - \frac{S_1{}^2}{n} = 96 - \frac{(24)^2}{16} = 96 - \frac{576}{16} = 96 - 36 = 60.$$

Как и следовало ожидать, мы в данном случае получили то же самое значение $\Sigma x^2$, несмотря на то, что обе вспомогательные суммы $S_1$ и $S_2$ (зависящие от выбора числа $C$) оказались иными, чем в предыдущем примере.

Читателю, в виде самостоятельного упражнения, предлагается убедиться в том, что

$$\left.\begin{array}{llll} \text{при} & C = 11 & S_1 = -40 & S_2 = 160 \\ \text{»} & C = 10 & S_1 = -24 & S_2 = 96 \\ \text{»} & C = 9 & S_1 = -8 & S_2 = 64 \\ \text{»} & C = 8 & S_1 = +8 & S_2 = 64 \\ \text{»} & C = 7 & S_1 = +24 & S_2 = 96 \\ \text{»} & C = 6 & S_1 = +40 & S_2 = 160 \end{array}\right\} \Sigma x^2 = 60.$$

Из этой таблицы видно, что с удалением $C$ от среднего арифметического $M$ $(= 8{,}5)$ тотчас же увеличиваются и обе вспомогательные суммы $S_1$ и $S_2$, причем, если за $C$ принять число большее $M$, то $S_1$ оказывается величиной отрицательной, а если $C$ меньше $M$, то $S_1$ получается со знаком плюс. Величина же $S_2$ оказывается всегда больше суммы квадратов центральных отклонений $\Sigma x^2$ $(= 60)$.

Так как производить вычисления удобнее с малыми и притом положительными числами, то за $C$ выгодно принимать ближайшее к $M$ меньшее число. В нашем примере $M = 8{,}5$. Следовательно за $C$ удобнее всего было бы принять число 8, т.-е. просто отбросить у $M$ все десятичные знаки (мы приняли ранее $C = 9$ лишь с целью показать способ вычисления $\Sigma x^2$ при отрицательном значении $S_1$).

Заметим кстати, что для вычисления $\Sigma x^2$ по способу округления нам вовсе не нужно определять предварительно



точное значение среднего арифметического $M$. В тех случаях когда почему-либо требуется вычислить только $\sigma$ и $m$, можно выбрать произвольное значение $C$ и, определив соответствующую величину вспомогательных сумм $S_1$ и $S_2$, прямо найти с их помощью сумму квадратов центральных отклонений $\Sigma x^2$. Вычислив $\Sigma x^2$, мы без труда определим и $\sigma$, а с помощью $\sigma$ — среднюю ошибку $m$ (без предварительного вычисления среднего арифметического $M$).

Если же нам нужно вычислить только среднюю ошибку $m$, минуя $\sigma$, то для этого можно воспользоваться следующей формулой:

$$m = \pm \sqrt{\frac{\Sigma x^2}{n(n-1)}}.$$

Эта формула показывает, что для определения средней ошибки $m$ следует сумму квадратов центральных отклонений $\Sigma x^2$ разделить на произведение чисел $n$ и $(n-1)$ и из полученного частного извлечь квадратный корень. В нашем последнем примере $\Sigma x^2 = 60$, а $n = 16$. Помножив 16 ($n$) на 15 ($n-1$), получим 240. Разделив 60 на это произведение 240, получим 0,25. Так как корень из $0{,}25 = \pm 0{,}5$, то искомая средняя ошибка равна $\pm 0{,}5$. Эти вычисления можно расположить в следующую строчку:

$$m = \pm \sqrt{\frac{\Sigma x^2}{n(n-1)}} = \pm \sqrt{\frac{60}{16(16-1)}} = \pm \sqrt{\frac{60}{16 \cdot 15}} =$$
$$= \pm \sqrt{\frac{60}{240}} = \pm \sqrt{0{,}25} = \pm 0{,}5.$$

Как и следовало ожидать, этот результат в точности совпал с полученным ранее по формуле:

$$m = \frac{\pm \sigma}{\sqrt{n}}$$

разница лишь в том, что мы определили здесь $m$ без предварительного вычисления $\sigma$.

---

## VIII. Способ произведений.

Проделаем еще раз все вычисления, описанные в предыдущей главе, на том же самом числовом примере, но на этот раз расположим отдельные варианты $V$ уже не в произвольном порядке, как раньше, а в строгой последовательности их убывания (сверху большие, снизу малые варианты). Приняв за $C$ число 8, вычислим все вспомогательные отклонения $x_c$ и их квадраты $x_c^2$. После записи этих результатов в соответствующие вертикальные столбцы расчетной решетки (см. табл. 9) и сложения всех чисел в среднем и в правом столбце получим:

ТАБЛИЦА 9.

| $V$ | $x_c$ | $x_c^2$ |
|---|---|---|
| 12 | +4 | 16 |
| 11 | +3 | 9 |
| 11 | +3 | 9 |
| 10 | +2 | 4 |
| 10 | +2 | 4 |
| 9 | +1 | 1 |
| 9 | +1 | 1 |
| 9 | +1 | 1 |
| 8 | 0 | 0 |
| 8 | 0 | 0 |
| 8 | 0 | 0 |
| 7 | −1 | 1 |
| 7 | −1 | 1 |
| 6 | −2 | 4 |
| 6 | −2 | 4 |
| 5 | −3 | 9 |
| 136 | +8 | 64 |

$$S_1 = +8 \qquad S_2 = 64.$$

Просматривая эти записи, мы замечаем, что отдельные варианты $V$, их отклонения $x_c$ и квадраты этих отклонений $x_c^2$ повторяются здесь иногда по нескольку раз. Так, напр., варианта 11, имеющая отклонение $+3$, повторяется два раза, варианта 10 с отклонением $+2$ повторяется тоже два раза, варианта 9 с отклонением $+1$ — три раза и т. д.

Обнаружив наличие таких повторений, мы можем при суммировании чисел в столбцах $V$, $x_c$ и $x_c^2$ избежать сложения нескольких одинаковых величин, заменив это сложение умножением каждого повторяющегося числа на соответствующее количество его повторений. Так, например, взамен сложения двух одинаковых отклонений $+3$ и $+3$, соответствующих повторяющейся дважды варианте 11, мы можем помножить это общее их отклонение $(+3)$ на число повторений (2), в результате чего получим, очевидно, $+6$.

Для удобства выполнения всех этих перемножений начертим новую расчетную решетку, в которую запишем каждую варианту $V$ лишь по одному разу, но зато прибавим лишний вертикальный столбец справа и в этом столбце пометим, сколько раз данная варианта повторяется (см. табл. 10).



Эти новые числа, указывающие на количество повторений каждого отдельного значения варианты $V$ (а стало-быть, и на количество повторений соответствующих вспомогательных отклонений $x_c$ и их квадратов $x_c^2$), мы будем называть „частотами“ и обозначать их буквою $p$. Таким образом частоты $p$ отдельных вариант показывают, как часто (отсюда и название) варианты эти встречаются в данном вариационном ряду. В нашем примере, следовательно, частотами будут служить следующие числа:

ТАБЛИЦА 10.

| $V$ | $x_c$ | $x_c^2$ | $p$ |
|---|---|---|---|
| 12 | | | 1 |
| 11 | | | 2 |
| 10 | | | 2 |
| 9 | | | 3 |
| 8 | | | 3 |
| 7 | | | 2 |
| 6 | | | 2 |
| 5 | | | 1 |

1, 2, 2, 3, 3, 2, 2, 1.

Если теперь для вычисления среднего арифметического $M$ нам нужно подсчитать сумму всех вариант $\Sigma V$, то сделать это можно будет на основании следующих рассуждений.

ТАБЛИЦА 11.

| $V$ | $x_c$ | $x_c^2$ | $p$ | $pV$ |
|---|---|---|---|---|
| 12 | | | 1 | 12 |
| 11 | | | 2 | 22 |
| 10 | | | 2 | 20 |
| 9 | | | 3 | 27 |
| 8 | | | 3 | 24 |
| 7 | | | 2 | 14 |
| 6 | | | 2 | 12 |
| 5 | | | 1 | 5 |

Первая (самая бо́льшая) варианта 12 встречается в заданном ряду всего только один раз ($p = 1$). Следовательно в общую сумму всех вариант она войдет слагаемым тоже лишь один раз. Этот результат мы запишем справа в особом (пятом) вертикальном столбце, озаглавленном буквами $pV$ (см. табл. 11).

Следующая варианта 11 встречается уже два раза ($p=2$). Поэтому при сложении всех вариант мы должны будем учесть ее дважды, т.-е. помножить варианту 11 на частоту 2. В результате получим, очевидно, 22, что и запишем в следующей (второй сверху) клетке крайнего справа столбца $pV$.

Помножив, далее, варианту 10 на ее частоту 2, запишем полученное произведение 20 в третьей клетке столбца $pV$ и т. д.

Поступая таким же образом со всеми 8-ю повторяющимися значениями вариант $V$, т.-е. перемножая их на соответствующие частоты $p$, мы получим ряд произведений, каждое из которых представляет собой не что иное, как

ТАБЛИЦА 12.

| $V$ | $x_e$ | $x_e^2$ | $p$ | $pV$ | $px_e$ |
|---|---|---|---|---|---|
| 12 | +4 | | 1 | 12 | +4 |
| 11 | +3 | | 2 | 22 | +6 |
| 10 | +2 | | 2 | 20 | +4 |
| 9 | +1 | | 3 | 27 | +3 |
| 8 | 0 | | 3 | 24 | 0 |
| 7 | −1 | | 2 | 14 | −2 |
| 6 | −2 | | 2 | 12 | −4 |
| 5 | −3 | | 1 | 5 | −3 |

сумму одинаковых, повторяющихся вариант. Очевидно, общая сумма всех вариант $\Sigma V$ будет равна сумме этих произведений, и для вычисления ее нам придется сложить все числа столбца $pV$. Сложив эти числа в действительности, найдем, что $\Sigma V = 136$, как это мы имели и ранее.

С помощью того же самого приема можно получить и сумму вспомогательных отклонений $S_1$. Для этого следует перемножить попарно каждое вспомогательное отклонение $x_e$ на соответствующую его частоту $p$ и результаты записать в особый (шестой) столбец, озаглавленный буквами $px_e$ (см. табл. 12). Сложив числа этого столбца с учетом их знаков (отдельно все плюсы и отдельно все минусы), получим попрежнему $S_1 = +8$.



Наконец, этот же прием можно применить и к вычислению суммы квадратов вспомогательных отклонений $S_2$.

Перемножив числа столбцов $x_e^2$ и $p$, запишем результаты в последний (седьмой) столбец, озаглавленный буквами $px_e^2$ (см. табл. 13). Сумма всех чисел этого столбца и даст нужную нам величину $S_2$. Как и следовало ожидать, в данном случае.

$$S_2 = 64.$$

Все полученные в этой решетке числовые итоги ($\Sigma V = 136$, $S_1 = +8$ и $S_2 = 64$) удобно поместить в дополнительной

ТАБЛИЦА 13.

| $V$ | $x_e$ | $x_e^2$ | $p$ | $pV$ | $px_e$ | $px_e^2$ |
|---|---|---|---|---|---|---|
| 12 | +4 | 16 | 1 | 12 | +4 | 16 |
| 11 | +3 | 9 | 2 | 22 | +6 | 18 |
| 10 | +2 | 4 | 2 | 20 | +4 | 8 |
| 9 | +1 | 1 | 3 | 27 | +3 | 3 |
| 8 | 0 | 0 | 3 | 24 | 0 | 0 |
| 7 | −1 | 1 | 2 | 14 | −2 | 2 |
| 6 | −2 | 4 | 2 | 12 | −4 | 8 |
| 5 | −3 | 9 | 1 | 5 | −3 | 9 |
| — | — | — | 16 | 136 | +8 | 64 |

нижней строчке, как это показано на табл. 13. В этой же строчке под числами столбца $p$ полезно записать сумму частот, которая, очевидно, представляет собой не что иное, как общее количество всех измерений (вариант) $n$.

Получив из расчетной решетки суммы $\Sigma V$, $S_1$, и $S_2$, мы заканчиваем наши вычисления с помощью применения обычных формул:

$$M = \frac{\Sigma V}{n} = \frac{136}{16} = 8{,}5$$

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 64 - \frac{(8)^2}{16} = 60$$

$$\sigma = \pm\sqrt{\frac{\Sigma x^2}{n-1}} = \pm\sqrt{\frac{60}{16-1}} = \pm 2$$

$$m = \frac{\pm\sigma}{\sqrt{n}} = \frac{\pm 2}{\sqrt{16}} = \pm 0{,}5$$

или

$$m = \pm\sqrt{\frac{\Sigma x^2}{n(n-1)}} = \pm\sqrt{\frac{60}{16(16-1)}} = \pm 0{,}5$$

Итак, при производстве вычислений по способу произведений следует:

1. Составить расчетную решетку из 7-ми вертикальных столбцов, озаглавив их буквами: $V$, $x_c$, $x_c^2$, $p$, $pV$, $px_c$ и $px_c^2$.

2. Записать отдельные значения вариант в крайний слева столбец $V$ непременно в порядке их убывания, а в средний столбец $p$ — частоты этих вариант.

3. Приняв одно из значений варианты $V$ (все равно, какое именно) за округленное среднее $C$, вычислить для каждого значения $V$ его отклонение от этого числа $C$ и записать эти вспомогательные отклонения с соответствующими знаками во второй вертикальный столбец $x_c$.

4. Возвысить все вспомогательные отклонения в квадрат и результаты записать в третий вертикальный столбец $x_c^2$.

5. Перемножить попарно числа столбцов $V$ и $p$ и полученные результаты записать в столбец, озаглавленный буквами $pV$.

6. Перемножить попарно числа столбцов $x_c$ и $p$ (сохраняя знаки чисел $x_c$) и записать результаты в столбец $px_c$.

7. Перемножить числа столбцов $x_c^2$ и $p$ (не проставляя знаков, так как все эти числа будут положительными) и результат записать в столбец $px_c^2$.

8. Сложить числа последних 4-х столбцов, причем при суммировании чисел столбца $px_c$ следует принять во внимание их знаки (сложить отдельно все плюсы и отдельно же все минусы, из большей суммы вычесть меньшую и поставить знак большей суммы). Получим нужные для дальнейших вычислений величины: $n$, $\Sigma V$, $S_1$ и $S_2$.

9. Проверить сумму $S_1$ по контрольной формуле:

$$S_1 = \Sigma V - nC$$

10. Вычислить сумму квадратов центральных отклонений $\Sigma x^2$ по формуле:

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



11. Вычислить среднее арифметическое отклонение $M$ по формуле:

$$M = \frac{\Sigma V}{n}$$

12. Вычислить среднее квадратическое отклонение $\sigma$ по формуле:

$$\sigma = \pm\sqrt{\frac{\Sigma x^2}{n-1}}$$

13. Вычислить среднюю ошибку $m$ по формулам:

$$m = \frac{\pm\sigma}{\sqrt{n}} \quad \text{или} \quad m = \pm\sqrt{\frac{\Sigma x^2}{n(n-1)}}$$

ТАБЛИЦА 14.

| $V$ | $x$ | $x_c^2$ | $p$ | $pV$ | $px_c$ | $px_c^2$ |
|---|---|---|---|---|---|---|
| 12 | + 2 | 4 | 1 | 12 | + 2 | 4 |
| 11 | + 1 | 1 | 2 | 22 | + 2 | 2 |
| 10 | 0 | 0 | 2 | 20 | 0 | 0 |
| 9 | — 1 | 1 | 3 | 27 | — 3 | 3 |
| 8 | — 2 | 4 | 3 | 24 | — 6 | 12 |
| 7 | — 3 | 9 | 2 | 14 | — 6 | 18 |
| 6 | — 4 | 16 | 2 | 12 | — 8 | 32 |
| 5 | — 5 | 25 | 1 | 5 | — 5 | 25 |
| — | — | — | 16 | 136 | —24 | 96 |

В целях закрепления в памяти указанного здесь порядка вычисления $S_1$ и $S_2$ по способу произведений, проделаем эти вычисления на том же самом примере еще раз, приняв за $C$ хотя бы варианту 10.

Для этого в столбцах $x_c$ и $x_c^2$ против значения $V=10=C$ поставим два нуля, так как отклонение данной варианты от равного ей значения $C$ и квадрат этого отклонения очевидно равны нулю. Выше этой строчки все отклонения $x_c$ будут положительными а ниже — отрицательными. По величине же своей они, очевидно, выразятся последовательными числами: 1, 2, 3, 4 и т. д., идущими кверху и книзу от нуля, увеличиваясь с каждой клеткой ровно на единицу (см. табл. 14).

Возвысив все эти отклонения в квадрат, получим числа столбца $x_c^2$.

Перемножив попарно числа первых трех столбцов ($V$, $x_c$ и $x_c^2$) на высоты $p$, записав эти произведения в соответствующие столбцы ($pV$, $px_c$ и $px_c^2$) и просуммировав 4 последние столбца, получим:

$$n = 16 \qquad \Sigma V = 136 \qquad S_1 = -24$$
$$S_2 = 96$$

Отсюда по обычным формулам найдем:

$$\Sigma x^2 = 60 \qquad M = 8{,}5 \qquad \sigma = \pm 2$$
$$m = \pm 0{,}5$$

Сравнивая процесс вычисления $S_1$ и $S_2$ по „способу округления" с порядком получения тех же величин „способом произведений", мы убеждаемся в том, что при большом количестве вариант этот последний способ дает существенную экономию времени. В самом деле, ранее мы всегда имели дело с каждой вариантой в отдельности, вычисляя ее отклонение от числа $C$, возвышая каждое такое отклонение в квадрат и складывая все эти числа для получения $S_1$ и $S_2$. При производстве же этих расчетов по способу произведений мы оперируем уже со значительно меньшим количеством отдельных чисел, так как одинаковые варианты соединяются нами в группы, число которых, очевидно, всегда будет ограниченным.

## IX. Способ сумм.

При пользовании способом произведений вычисление вспомогательных сумм $S_1$ и $S_2$ производилось путем перемножения отдельных отклонений $x_c$ и их квадратов $x_c^2$ на соответствующие частоты $p$ или, что все равно, наоборот, посредством перемножения частот $p$ на соответствующие отклонения и их квадраты. Так как вспомогательные отклонения здесь представляют собой целые числа, увеличивающиеся в обе стороны от 0 ровно на единицу, то наша задача сводится, следовательно, к перемножению отдельных частот $p$ на ряд натуральных чисел: 1, 2, 3, 4, 5 и т. д. и на квадраты этих чисел: 1, 4, 9, 16, 25 и т. д.,



с последующим сложением всех полученных произведений для получения сумм $S_1$ и $S_2$

Покажем, что тот же результат может быть достигнут и другим способом, путем простого сложения только одних частот $p$. Разберем этот способ на прежнем числовом примере, обработав его для сравнения сперва еще раз по способу произведений, приняв за $C$ самую малую варианту 5 (см. табл. 15). Тогда в расчетную решетку против этой варианты (5) во втором и в третьем столбце следует поставить два нуля, а кверху от этих нулей написать во втором

ТАБЛИЦА 15.

| $V$ | $x_c$ | $x_c^2$ | $p$ | $pV$ | $px_c$ | $px_c^2$ |
|---|---|---|---|---|---|---|
| 12 | +7 | 49 | 1 | | +7 | 49 |
| 11 | +6 | 36 | 2 | | +12 | 72 |
| 10 | +5 | 25 | 2 | | +10 | 50 |
| 9 | +4 | 16 | 3 | | +12 | 48 |
| 8 | +3 | 9 | 3 | | +9 | 27 |
| 7 | +2 | 4 | 2 | | +4 | 8 |
| 6 | +1 | 1 | 2 | | +2 | 2 |
| 5 | 0 | 0 | 1 | | 0 | 0 |
| — | — | — | 16 | | +56 | 256 |

столбце — отклонения: 1, 2, 3, 4 и т., все со знаком плюс, в третьем же столбце — квадраты этих отклонений: 1, 4, 9, 16 и т. д. уже без всяких знаков (так как числа эти всегда положительны).

Пропустив пятый столбец $pV$ (он нам здесь неинтересен), выполним все перемножения, нужные для заполнения двух крайних столбцов $px_c$ и $px_c^2$. В столбце $px_c$ получим произведения наших частот $p$ на ряд натуральных чисел: 1, 2, 3, 4 и т. д., а в столбце $px_c^2$ — произведения тех же частот $p$ на квадраты этих чисел: 1, 4, 9, 16 и т. д. Подсчитав суммы всех чисел этих столбцов, найдем, что

$$S_1 = +56, \qquad S_2 = 256,$$

откуда сумма квадратов центральных отклонений

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 256 - \frac{(56)^2}{16} = 256 - \frac{3136}{16} = 256 - 196 = 60,$$

что мы имели и раньше.

Покажем, как можно получить те же самые вспомогательные суммы $S_1$ и $S_2$ без всяких перемножений.

Выпишем для этого наши частоты $p$ в виде отдельного вертикального столбца (см. табл. 16), а рядом с каждой из этих частот, во втором вертикальном столбце, будем записывать сумму всех предыдущих чисел этого второго столбца, сложенную с ближайшим следующим числом первого столбца.

ТАБЛИЦА 16.

| | |
|---|---|
| 1 | 1 |
| 2 | 3 |
| 2 | 5 |
| 3 | 8 |
| 3 | 11 |
| 2 | 13 |
| 2 | 15 |
| 1 | — |

ТАБЛИЦА 17.

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 3 | 4 |
| 2 | 5 | 9 |
| 3 | 8 | 17 |
| 3 | 11 | 28 |
| 2 | 13 | 41 |
| 2 | 15 | — |
| 1 | — | — |

Первый раз рядом с самой верхней частотой (1) запишем ее же (т.-е. тоже 1). Под ней следует поместить сумму этой единицы и второй частоты (2), т.-е. 3. Далее сложить только-что полученное число (3) со следующей частотой (2) и сумму их (5) записать на третьем месте второго столбца. Затем сложить 5 и 3 и сумму их (8) записать во втором столбце рядом с частотой 3 первого столбца и т. д. В результате получим ряд возрастающих чисел второго столбца, которое мы доведем лишь до предпоследнего места, кончая числом 15, записанным рядом с предпоследней частотой 2. На последнем же месте, взамен цифры, мы поставим черточку (см. табл. 16).

Получив числа второго столбца, проделаем ту же самую операцию и с ними, составив в результате по тому же способу числа третьего столбца (см. табл. 17).

Для этого на первом месте третьего столбца так же, как и ранее, запишем самую верхнюю частоту (1), под ней



сумму этой единицы и второго числа среднего столбца (3) т.-е. 4, на третьем месте – сумму только что полученного числа (4) и следующего числа среднего столбца (5), т.-е. 9, и т. д. Числа правого столбца мы также доведем не до конца предыдущего (среднего), а остановимся на одно место выше, т.-е. ограничимся лишь записью числа 41 на третьем месте снизу. Два же свободные места снизу заполним черточками.

Подсчитаем теперь отдельно суммы чисел среднего и правого столбца, обозначив их буквами $a_1$ и $a_2$. В данном случае сумма чисел среднего столбца $a_1 = 56$, а сумма чисел правого столбца $a_2 = 100$. Так вот, оба эти итога ($a_1$ и $a_2$) и могут послужить нам для вычисления искомых вспомогательных сумм $S_1$ и $S_2$. Первая из них $S_1$ как раз и будет здесь равняться первому итогу $a_1$ (т.-е. в данном случае 56), а для получения суммы $S_2$ следует второй итог $a_2$ помножить на 2 и сложить с первым итогом $a_1$. В нашем примере $a_2 = 100$. Помножив это число на 2, получим 200. Сложив 200 с первым итогом $a_1$ ($= 56$), получим искомую сумму $S_2 = 256$.

ТАБЛИЦА 18.

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 3 | 4 |
| 2 | 5 | 9 |
| 3 | 8 | 17 |
| 3 | 11 | 28 |
| 2 | 13 | **41** |
| 2 | **15** | — |
| 1 | — | — |

Итак, произведя ряд последовательных сложений по указанной выше схеме, мы без всяких перемножений частот на отклонения и их квадраты получили те же самые значения вспомогательных сумм:

$$S_1 = 56 \qquad S_2 = 256$$

В целях еще большего упрощения этих операций можно сумму чисел среднего столбца (т.-е. итог $a_1$) получить несколько иначе. Для этого (см. табл. 18) стоит лишь сложить два числа, стоящие непосредственно над черточками. Числа эти для ясности отпечатаны жирным шрифтом. Складывая 15 и 41, получим в сумме 56, что мы имели и раньше, суммируя все числа среднего столбца (т.-е. $1 + 3 + 5 + 8 + 11 + 13 + 15 = 56$).

В разобранном примере все отклонения $x_c$ оказались положительными, так как за $C$ была принята самая малая варианта 5. Покажем, что этот же способ вычисления $S_1$

и $S_2$ может быть применен и в том случае, когда за начало отсчетов $C$ принята какая-либо промежуточная варианта $V$ и следовательно отклонения $x_c$ кверху от $C$ будут положительными, а книзу — отрицательными. Разберем этот случай на том же самом числовом примере, приняв за $C$ на этот раз хотя бы варианту 8. Вспомним, что в этом случае при $C = 8$ мы получили $S_1 = +8$, а $S_2 = 64$.

Заготовим расчетную решетку, состоящую из трех вертикальных столбцов и стольких горизонтальных строчек в своей главной части, сколько отдельных значений вариант $V$ оказалось в нашем ряду после группировки и объединения повторяющихся вариант (в данном случае, следовательно, из 8 строчек). Снизу и сверху следует прибавить еще по одной горизонтальной строчке, отделив их от остальных двойной линией (см. табл. 19). В этих дополнительных строчках мы будем записывать итоги сложения чисел в отдельных столбцах. В крайней слева клетке верхней дополнительной строчки поставим букву $p$. Этот заголовок указывает здесь на то, что в левом вертикальном столбце будут записаны частоты $p$ нашего вариационного ряда.

ТАБЛИЦА 19.

| | | | |
|---|---|---|---|
| | $p$ | | |
| 12 | | | |
| 11 | | | |
| 10 | | | |
| 9 | | | |
| 8 | | | |
| 7 | | | |
| 6 | | | |
| 5 | | | |
| | | | |

Для большей наглядности описания дальнейших операций мы поместим временно на полях расчетной решетки также и значения отдельных вариант $V$, как всегда, в убывающем их порядке (сверху большие, снизу малые). Эти пометки мы сохраним до конца описания данного способа вычислений, но будем помнить, что, вообще говоря, при практическом применении этого способа они ненужны.

Заготовив таким образом расчетную решетку, запишем в левом вертикальном ее столбце все частоты нашего ряда и сложим их (см. табл. 20).

Получим сумму всех частот $p$ или, что то же, общее число измерений $n$. В данном случае $n = 16$. Это число (16) мы запишем в нижней дополнительной строчке под столбцом частот $p$. Для ясности, на полях решетки под этой клеткой поместим букву $n$, которая, таким образом, указывает здесь на то, что в этой клетке всегда записывается общее число измерений (вариант) $n$.



Так как за начало отсчетов мы решили принять $C = 8$, то против частоты 3, соответствующей этому значению варианты, поставим четыре черточки, как это показано на табл. 21. Эти черточки образуют фигуру, напоминающую положенный на бок равнобедренный треугольник, упирающийся своей вершиной в ту клетку левого столбца, которая соответствует значению варианты, принятой нами за начало отсчетов $C$ (варианта эта для ясности отпечатана жирным шрифтом). Значение этих черточек ясно. Они поставлены здесь для того, чтобы при заполнении клеток среднего и правого столбца мы оставляли эти клетки пустыми так же, как и в предыдущем примере.

ТАБЛИЦА 20.

| | | | |
|---|---|---|---|
| | $p$ | | |
| 12 | 1 | | |
| 11 | 2 | | |
| 10 | 2 | | |
| 9 | 3 | | |
| 8 | 3 | | |
| 7 | 2 | | |
| 6 | 2 | | |
| 5 | 1 | | |
| | 16 | | |
| | $n$ | | |

ТАБЛИЦА 21.

| | | | |
|---|---|---|---|
| | $p$ | | |
| 12 | 1 | 1 | 1 |
| 11 | 2 | | |
| 10 | 2 | | |
| 9 | 3 | | — |
| **8** | 3 | — | — |
| 7 | 2 | | — |
| 6 | 2 | | |
| 5 | 1 | 1 | 1 |
| | 16 | | |
| | $n$ | | |

Расположив в расчетной решетке эту фигуру из 4-х черточек, мы должны далее записать крайние частоты (верхнюю и нижнюю) в соответствующих горизонтальных строчках до конца решетки. В данном примере обе крайние частоты равны единице. Стало-быть, эти единицы нужно записать во всех трех клетках сверху и снизу главной части расчетной решетки, как это показано на табл. 21.

Проделав все эти подготовительные операции, можно приступить к заполнению оставшихся пустыми клеток нашей решетки, начиная, напр., с верхней ее части (считая от черточек).

Для этого, как и в предыдущем примере, в каждой клетке мы должны записать сумму чисел, стоящих сверху и слева. Начинать эти операции следует с угловой клетки, отмеченной для ясности на табл. 22 жирной цифрой. В этом месте следует записать сумму двух чисел: верхнего 1 и левого 2, т.-е. число 3. Под ним мы запишем сумму этого числа (3) и стоящей слева частоты 2, т.-е. 5. Под числом 5 — сумму $5+3=8$. Ниже стоят уже черточки, и следовательно суммирование можно прекратить (см. табл. 22).

ТАБЛИЦА 22.

| | $p$ | | |
|---|---|---|---|
| 12 | 1 | 1 | 1 |
| 11 | 2 | **3** | 4 |
| 10 | 2 | 5 | 9 |
| 9 | 3 | 8 | — |
| 8 | 3 | — | — |
| 7 | 2 | | — |
| 6 | 2 | | |
| 5 | 1 | 1 | 1 |
| | 16 | | |
| | $n$ | | |

Точно так же следует заполнить клетки верхней части и правого столбца, складывая верхние числа с левыми в направлении сверху вниз до клетки, занятой черточкой. Для этого под 1 рядом с числом 3, отмеченным на табл. 22 жирной цифрой, запишем 4 (так как $1+3=4$), а ниже — сумму $4+5=9$. Далее идут уже черточки, и суммирование можно прекратить (см. табл. 22).

Заполнив все клетки верхней части расчетной решетки, проделаем те же операции и с числами нижней ее части. Для этого, начиная с угловой клетки, отмеченной для ясности на табл. 23 жирной цифрой, будем суммировать нижние числа с левыми, идя в направлении снизу вверх. В среднем столбце нам придется записать только два числа: над 1 поставить 3 (так как $1+2=3$), а выше 5 (так как $3+2=5$). В следующей сверху клетке стоит уже черточка.

В правом столбце имеется всего лишь одна свободная клетка. В ней мы запишем число 4, так как снизу стоит 1, а слева 3 ($1+3=4$). Выше идут уже черточки (см. таблицу 23).



После того, как все клетки расчетной решетки будут заполнены, следует сложить отдельно верхние и отдельно же нижние числа среднего и правого столбца. В нашем примере сумма верхних чисел среднего столбца

$$a_1 = 1+3+5+8 = 17.$$

Вспомним, что этот же итог $a_1$ мы могли бы получить и более простым способом, не прибегая к действительному суммированию всех чисел этого столбца, ограничившись сложением лишь двух чисел, непосредственно примыкающих к черточкам. В нашем случае самым нижним числом среднего столбца оказалось 8, а самым нижним числом правого столбца 9. Сложив только эти два числа, мы и получим тот же самый итог

ТАБЛИЦА 23.

| | $p$ | | |
|---|---|---|---|
| 12 | 1 | 1 | 1 |
| 11 | 2 | 3 | 4 |
| 10 | 2 | 5 | 9 |
| 9 | 3 | 8 | — |
| 8 | 3 | — | — |
| 7 | 2 | 5 | — |
| 6 | 2 | **3** | 4 |
| 5 | 1 | 1 | 1 |
| | 16 | | |
| | $n$ | | |

ТАБЛИЦА 24.

| | | $a_1$ | $a_2$ |
|---|---|---|---|
| | $p$ | 17 | 14 |
| 12 | 1 | 1 | 1 |
| 11 | 2 | 3 | 4 |
| 10 | 2 | 5 | 9 |
| 9 | 3 | 8 | — |
| 8 | 3 | — | — |
| 7 | 2 | 5 | — |
| 6 | 2 | 3 | 4 |
| 5 | 1 | 1 | 1 |
| | 16 | 9 | 5 |
| | $n$ | $b_1$ | $b_2$ |

$$a_1 = 8+9 = 17.$$

Итог $a_1$ мы запишем в верхней дополнительной строчке над средним столбцом, поставив для ясности выше этой клетки соответствующий значок $a_1$ (см. табл. 24).

После этого сложим верхние числа в крайнем правом столбце. Получим второй итог:

$$a_2 = 1+4+9 = 14,$$

который запишем в крайней клетке верхней дополнительной строчки, поставив над этой клеткой значок $a_2$.

По аналогии с итогами сложения верхних чисел, суммы их в нижней части решетки обозначим через $b_1$ и $b_2$. В нашем примере, следовательно,

$$b_1 = 5 + 3 + 1 = 9$$
$$b_2 = 4 + 1 = 5$$

Заметим, что итог $b_1$ (так же, как и итог $a_1$) можно получить и без действительного суммирования чисел нижней части среднего столбца. Для этого, как и ранее, следует сложить два числа, непосредственно примыкающие к черточкам снизу (в среднем и в правом столбце). В данном случае, следовательно:

$$b_1 = 5 + 4 = 9\ ^{*})$$

Оба нижние итога $b_1$ и $b_2$ мы запишем в соответствующие клетки нижней дополнительной строчки, пометив под ними, для ясности, соответствующие значки: $b_1$ и $b_2$ (см. табл. 24).

Получив таким образом все 4 итога: $a_1$, $a_2$, $b_1$ и $b_2$, мы можем вычислить с их помощью обе нужные нам суммы: сумму вспомогательных отклонений $S_1$ и сумму их квадратов $S_2$ по формулам:

$$S_1 = a_1 - b_1$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2.$$

Первая из этих формул показывает, что для вычисления суммы $S_1$ следует из итога $a_1$ вычесть итог $b_1$. Так как в нашем примере $a_1 = 17$, а $b_1 = 9$, то

$$S_1 = 17 - 9 = +8.$$

Вторая формула показывает, что для вычисления $S_2$ следует сложить: итог $a_1$, итог $b_1$, удвоенный итог $a_2$ и

*) Аналогичный прием можно применить и при подсчете частот. Взамен действительного их суммирования достаточно сложить лишь три числа: то, которое в столбце $p$ написано слева от центральной черточки (в нашем примере 3), то, которое стоит над этой черточкой (у нас 8), и то, которое расположено под черточкой (т.-е. 5). В самом деле:

$$3 + 8 + 5 = 16,$$

что мы имели и раньше.



удвоенный же итог $b_2$. Так как в нашем примере $a_1 = 17$ $b_1 = 9$, $a_2 = 14$ и $b_2 = 5$, то

$$S_2 = 17 + 9 + 2.14 + 2.5 = 17 + 9 + 28 + 10 = 64.$$

Итак, с помощью этих формул мы получили:

$$S_1 = +8 \qquad S_2 = 64$$

т.-е те же самые результаты, как и при пользовании способом произведений.

Получив вспомогательные суммы $S_1$ и $S_2$, мы заканчиваем наши вычисления с помощью обычных формул:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 64 - \frac{(8)^2}{16} = 60.$$

$$\sigma = \pm\sqrt{\frac{\Sigma x^2}{n-1}} = \pm\sqrt{\frac{60}{16-1}} = \pm 2$$

$$m = \frac{\pm\sigma}{\sqrt{n}} = \frac{\pm 2}{\sqrt{16}} = \pm 0{,}5 \text{ или}$$

$$m = \pm\sqrt{\frac{\Sigma x_2}{n(n-1)}} = \pm\sqrt{\frac{60}{16(16-1)}} = \pm 0{,}5.$$

Единственным затруднением является здесь вычисление среднего арифметического $M$, которое раньше мы определяли по формуле:

$$M = \frac{\Sigma V}{n}.$$

В данном случае пользование этой формулой неудобно, так как сумма всех вариант ($\Sigma V$) непосредственно из расчетной решетки не выясняется (мы оперируем здесь не с самими вариантами $V$, а только с их частотами $p$). Поэтому для вычисления $M$ мы должны здесь воспользоваться другой формулой:

$$M = C + \frac{S_1}{n}.$$

Формула эта показывает, что для определения $M$ следует вспомогательную сумму $S_1$ разделить на число измерений $n$ и результат сложить с приближенным средним $C$. В нашем примере $S_1 = +8$, $n = 16$, а $C = 8$. Следовательно:

$$M = C + \frac{S_1}{n} = 8 + \frac{8}{16} = 8 + 0{,}5 = 8{,}5.$$

Как и следовало ожидать, данный результат совершенно точно совпал с тем значением $M$, которое ранее было вычислено другим способом.

Итак, при вычислении $M$, $\sigma$ и $m$ по описанному здесь способу, который можно назвать „способом сумм“, мы должны выполнить следующие операции:

1. Расположить заданный вариационный ряд в порядке убывающих значений отдельных вариант и указать частотами $p$ количество их повторений.

2. Составить расчетную решетку и занести в нее эти частоты (частоты бо̀льших вариант сверху, а частоты меньших — снизу).

3. Приняв одно из значений вариант $V$ за начало отсчетов $C$, начертить против соответствующей частоты фигуру из 4-х черточек.

4. Крайние частоты (верхнюю и нижнюю) записать в горизонтальных строчках до конца решетки.

5. Заполнить клетки верхней части расчетной решетки путем сложения верхних чисел с левыми.

6. Заполнить клетки нижней части решетки путем сложения нижних чисел с левыми.

7. Сложить верхние и нижние числа среднего и правого столбца (взамен суммирования всех чисел среднего столбца можно сложить лишь два числа, примыкающие к черточкам в среднем и в крайнем столбце) и соответствующие итоги: $a_1$, $a_2$, $b_1$ и $b_2$ записать в дополнительных строчках расчетной решетки.

8. Вычислить обе вспомогательные суммы $S_1$ и $S_2$ по формулам:

$$S_1 = a_1 - b_1$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2.$$

9. Вычислить среднее арифметическое $M$ по формуле:

$$M = C + \frac{S_1}{n}.$$

10. Вычислить сумму квадратов центральных отклонений $\Sigma x^2$ по формуле:

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



11. Вычислить среднее квадратическое отклонение $\sigma$ по формуле:

$$\sigma = \pm\sqrt{\frac{\Sigma x^2}{n-1}}.$$

12. Вычислить среднюю ошибку $m$ по формуле:

$$m = \frac{\pm\sigma}{\sqrt{n}} \quad \text{или} \quad m = \pm\sqrt{\frac{\Sigma x^2}{n(n-1)}}.$$

В последнем нашем примере мы совершенно произвольно приняли за начало отсчетов $C$ варианту 8 и расположили в расчетной решетке фигуру из 4-х черточек на уровне частоты 3, соответствующей этой варианте 8. С тем же правом, однако, мы могли бы поместить эти черточки и в другом месте, напр., несколько выше или, наоборот, ниже. При этом мы в каждом таком случае, очевидно, получили бы различные итоги $a_1$, $a_2$, $b_1$ и $b_2$ и различные же значения **вспомогательных** сумм $S_1$ и $S_2$, но вычисленная из этих величин сумма квадратов центральных отклонений $\Sigma x^2$ **всегда** оказалась бы в точности равной 60, а среднее арифметическое $M = 8{,}5$.

ТАБЛИЦА 25.

| | | $a_1$ | $a_2$ |
| --- | --- | --- | --- |
| | $p$ | 4 | 1 |
| 12 | 1 | 1 | 1 |
| 11 | 2 | 3 | — |
| 10 | 2 | — | — |
| 9 | 3 | 11 | — |
| 8 | 3 | 8 | 17 |
| 7 | 2 | 5 | 9 |
| 6 | 2 | 3 | 4 |
| 5 | 1 | 1 | 1 |
| | 16 | 28 | 31 |
| | $n$ | $b_1$ | $b_2$ |

Примем, например, за $C$ варианту 10 и расположим в решетке фигуру из 4-х черточек против ее частоты 2 (см. табл. 25).

Тогда, после выполнения всех сложений внутри решетки и подведения итогов отдельно по верхним и по нижним частям среднего и правого столбца, получим:

$$a_1 = 4 \qquad a_2 = 1$$
$$b_1 = 28 \qquad b_2 = 31.$$

Вспомогательные суммы $S_1$ и $S_2$ найдем по формулам:

$$S_1 = a_1 - b_1 = 4 - 28 = -24$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 = 4 + 28 + 2.1 + 2.31 =$$
$$= 4 + 28 + 2 + 62 = 96.$$

Итак, в этом случае при $C = 10$ мы получили

$$S_1 = -24 \qquad S_2 = 96.$$

Откуда сумма квадратов центральных отклонений:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 96 - \frac{(-24)^2}{16} = 96 - \frac{576}{16} = 96 - 36 = 60,$$

а среднее арифметическое

$$M = C + \frac{S_1}{n} = 10 + \frac{-24}{16} = 10 - 1{,}5 = 8{,}5.$$

ТАБЛИЦА 26.

| | | $a_1$ | $a_2$ |
|---|---|---|---|
| | $p$ | 1 | 0 |
| 12 | 1 | 1 | — |
| 11 | 2 | — | — |
| 10 | 2 | 13 | — |
| 9 | 3 | 11 | 28 |
| 8 | 3 | 8 | 17 |
| 7 | 2 | 5 | 9 |
| 6 | 2 | 3 | 4 |
| 5 | 1 | 1 | 1 |
| | 16 | 41 | 59 |
| | $n$ | $b_1$ | $b_2$ |

Как и следовало ожидать, мы получили здесь те же самые значения $\Sigma x^2$ и $M$, какие имели и ранее при $C = 8$.

Если принять за $C$ варианту 11, то фигура из 4-х черточек расположится в решетке так, как это показано на таблице 26. В этом случае мы можем продолжить верхнюю частоту 1 уже не до конца решетки, как раньше, а только на одну соседнюю клетку, так как следующая занята уже черточкой. Поэтому сумма всех верхних чисел среднего столбца $a_1 = 1$, а сумма верхних чисел правого столбца $a_2 = 0$ (в верхней части правого столбца нет ни одного числа). Нижние итоги оказались здесь $b_1 = 41$, $b_2 = 59$. Следовательно вспомогательные суммы:

$$S_1 = a_1 - b_1 = 1 - 41 = -40$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 = 1 + 41 + 2.0 + 2.59 =$$
$$= 1 + 41 + 0 + 118 = 160.$$



Определив $S_1 = -40$ и $S_2 = 160$, вычислим $\Sigma x^2$ и $M$:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 160 - \frac{(-40)^2}{16} = 160 - \frac{1600}{16} = 160 - 100 = 60$$
$$M = C + \frac{S_1}{n} = 11 + \frac{-40}{16} = 11 - 2{,}5 = 8{,}5.$$

В некоторых случаях черточки приходится помещать в решетке значительно выше или ниже всего имеющегося ряда частот $p$, что соответствует принятию за $C$ такого значения вариант, которое не встречается в заданном вариационном ряду вовсе.

ТАБЛИЦА 27.

| | | $a_1$ | $a_2$ |
|---|---|---|---|
| | $p$ | 104 | 316 |
| 12 | 1 | 1 | 1 |
| 11 | 2 | 3 | 4 |
| 10 | 2 | 5 | 9 |
| 9 | 3 | 8 | 17 |
| 8 | 3 | 11 | 28 |
| 7 | 2 | 13 | 41 |
| 6 | 2 | 15 | 56 |
| 5 | 1 | 16 | 72 |
| 4 | 0 | 16 | 88 |
| 3 | 0 | 16 | — |
| 2 | 0 | — | — |
| | 16 | 0 | 0 |
| | | $b_1$ | $b_2$ |

Пусть, напр., $C = 2$. Продолжим расчетную решетку на несколько клеток ниже, как это показано на табл. 27. Так как значения вариант 4, 3 и 2 не встречаются в нашем ряду ни одного раза, то частоты этих значений будут нулями. Эти нули мы запишем в столбце $p$ и против частоты (0), соответствующей варианте 2, начертим фигуру уже из трех черточек, так как четвертая нам, очевидно, здесь не нужна.

Далее заполним по обычным правилам клетки среднего и правого столбца и подведем итоги кверху. Получим:

$$a_1 = 104 \qquad a_2 = 316.$$

Нижние итоги $b_1$ и $b_2$, очевидно, равны нулям, т.-е.

$$b_1 = 0 \quad \text{и} \quad b_2 = 0.$$

Следовательно:

$$S_1 = a_1 - b_1 = 104 - 0 = +104$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 = 104 + 0 + 2.316 + 2.0 = 104 + 632 = 736,$$

отсюда:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 736 - \frac{(104)^2}{16} = 736 - \frac{10816}{16} = 736 - 676 = 60$$

$$M = C + \frac{S_1}{n} = 2 + \frac{104}{16} = 2 + 6{,}5 = 8{,}5.$$

ТАБЛИЦА 28.

| | | $a_1$ | $a_2$ |
|---|---|---|---|
| | $p$ | 0 | 0 |
| 14 | 0 | — | — |
| 13 | 0 | 16 | — |
| 12 | 1 | 16 | 72 |
| 11 | 2 | 15 | 56 |
| 10 | 2 | 13 | 41 |
| 9 | 3 | 11 | 28 |
| 8 | 3 | 8 | 17 |
| 7 | 2 | 5 | 9 |
| 6 | 2 | 3 | 4 |
| 5 | 1 | 1 | 1 |
| | 16 | 88 | 228 |
| | $n$ | $b_1$ | $b_2$ |

Если за $C$ принять варианту 14, также не встречающуюся в заданном ряду, то три черточки придется расположить, наоборот, выше всех частот $p$, прибавив к ним недостающее количество нулей (в данном случае два нуля).

После всех подсчетов внутри расчетной решетки (см. таблицу 28) получим:

$$a_1 = 0 \qquad a_2 = 0$$
$$b_1 = 88 \qquad b_2 = 228.$$

Следовательно:

$$S_1 = a_1 - b_1 = 0 - 88 = -88$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 = 0 + + 88 + 2\cdot 0 + 2\cdot 228 = = 88 + 456 = 544$$

откуда:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 544 - \frac{(-88)^2}{16} = 544 - \frac{7744}{16} = 544 - 484 = 60$$

$$M = C + \frac{S_1}{n} = 14 + \frac{(-88)}{16} = 14 - 5{,}5 = 8{,}5.$$

## X. Пользование таблицей квадратных корней.

При пользовании указанными выше формулами для вычисления $\sigma$ и $m$:

$$\sigma = \pm\sqrt{\frac{\Sigma x^2}{n-1}}, \qquad m = \frac{\pm\sigma}{\sqrt{n}} \qquad \text{или} \qquad m = \pm\sqrt{\frac{\Sigma x^2}{n(n-1)}}$$

в конечном итоге всегда приходится извлекать квадратный корень. Все разобранные ранее примеры были искусственно составлены так, что операция эта не представляла там ни-



каких затруднений. Однако на практике чаще всего приходится извлекать корень из многозначных и дробных чисел, не являющихся полными квадратами искомых величин, что требует применения уже более сложных приемов.

В тех случаях, когда особой точности в вычислениях не требуется, для извлечения квадратных корней можно воспользоваться специальной таблицей, приложенной в конце этой книги (см. стр. 281). В этих таблицах в столбцах, озаглавленных буквою $N$, выписаны числа от 1 до 1000, а рядом с ними в столбцах с заголовком $\sqrt{N}$ приведены соответствующие значения квадратных корней из этих чисел с точностью до четырех десятичных знаков после запятой (в третьем столбе этой таблицы помещены квадраты чисел $N$, т.-е. $N^2$).

В простейшем случае при извлечении квадратного корня из числа, состоящего не более, чем из 3-х цифр, ответ находится по этой таблице непосредственно.

Так, например, желая найти значение $\sqrt{7}$, мы просто выписываем из таблицы искомое число 2,6458. Точно также прямо из таблицы находим, что $\sqrt{68} = 8{,}2462$, а $\sqrt{909} = 30{,}1496$.

При извлечении квадратных корней с помощью этой таблицы из смешанных и дробных чисел, имеющих после запятой несколько десятичных знаков, приходится довольствоваться приближенным результатом, округляя заданное дробное число до ближайшего к нему целого (более точный способ указан ниже). При этом, если следующая после запятой цифра окажется меньше 5, то при таком округлении все десятичные знаки просто отбрасываются, если же цифра эта будет 5 или больше, то при отбрасывании десятичных знаков необходимо увеличить последнюю остающуюся цифру на единицу.

Пусть, например, требуется извлечь квадратный корень из числа 28,387. Так как в данном случае следующая после запятой цифра 3 оказалась меньше 5, то все десятичные знаки (3, 8 и 7) мы просто отбросим и будем считать, что требуется извлечь квадратный корень уже из целого двузначного числа 28. По таблицам корней находим, что $\sqrt{28} = 5{,}2915$. Следовательно:

$$\sqrt{28{,}387} \text{ приблизительно} = 5{,}2915.$$

Если бы требовалось извлечь корень из числа 832,729, в котором следующая после запятой цифра 7 оказалась больше 5, то, отбросив все десятичные знаки, мы должны будем прибавить к последней остающейся цифре 2 еще одну единицу и считать ее уже за 3. Таким образом в данном случае мы округлили заданное нам число до 833. Так как корню из 833 по таблице соответствует число 28,8617, то

$$\sqrt{832,729} \text{ прибл.} = 28,8617.$$

В двух предыдущих примерах мы имели дело с такими числами, которые после округления оказывались состоящими не более чем из 3-х цифр. В тех же случаях, когда число, из которого требуется извлечь корень, состоит из большего количества цифр (до запятой), мы должны предварительно уменьшить его в 100 раз, т.-е. перенести запятую на 2 места влево, затем округлить это число по общим правилам, найти соответствующее значение корня по таблицам и исправить полученный результат, увеличив его в 10 раз переносом запятой обратно на одно место вправо.

Пусть, напр., требуется извлечь квадратный корень из числа 1348,25. Разделив это число на 100, получим 13,4825. Округлив его до ближайшего целого числа, получим 13. По таблицам находим, что $\sqrt{13} = 3,6056$. Этот результат следует исправить, увеличив его в 10 раз. Следовательно:

$$\sqrt{1348,25} \text{ прибл.} = 36,056.$$

Если после уменьшения заданного числа в 100 раз и соответствующего его округления мы все же получим число, состоящее более чем из 3-х цифр, то эту же операцию (уменьшения в 100 раз) необходимо проделать повторно, передвинув запятую всего, следовательно, на 4 места влево, а найденное по таблицам значение корня исправить увеличением уже в 100 раз (а не в 10 раз), т.-е. обратным переносом запятой на 2 места вправо.

Пусть, например, требуется извлечь корень из числа 3875291,467. Уменьшив его в 100 раз дважды (т.-е. всего, следовательно, в 10000 раз), получим 387,5291467. Так как следующая после запятой цифра 5 требует увеличения последней остающейся цифры 7 на единицу, то после округления будем иметь 388. По таблицам находим, что $\sqrt{388} = 19,6977$. Увеличив этот результат в 100 раз, найдем окончательно, что

$$\sqrt{3875291,467} \text{ прибл.} = 1969,77.$$



Заметим, что при извлечении квадратного корня из числа, состоящего всего лишь из одной цифры до запятой, полезно не сразу округлять его до целого однозначного числа, а предварительно увеличить это число в 100 раз, чтобы получилось число трехзначное.

Пусть, напр., требуется извлечь корень из числа 6,8396. Мы могли бы просто округлить его до 7 и считать, что

$$\sqrt{6,8396} \text{ прибл.} = 2,6458.$$

Но этот результат очень не точен. Правильнее будет увеличить заданное число 6,8396 в 100 раз (получим 683,96) и после этого уже округлить его до ближайшего целого числа 684. Так как $\sqrt{684} = 26,1534$, то после уменьшения этого результата в 10 раз получим окончательно:

$$\sqrt{6,8396} \text{ прибл.} = 2,61534.$$

Последний результат значительно точнее предыдущего.

При извлечении квадратного корня из дробного числа, не имеющего значащих цифр до запятой, мы должны, наоборот, увеличить его в 100 или в 10000 раз, а результат исправить уменьшением в 10 раз или в 100 раз.

Пусть, напр., имеем число 0,7385. Увеличив его в 100 раз, получим 73,85. После округления его до ближайшего целого числа получим 74. По таблице найдем, что $\sqrt{74} = 8,6023$. Этот результат следует исправить, уменьшив его в 10 раз. Поэтому

$$\sqrt{0,7385} \text{ прибл.} = 0,86023.$$

Пусть требуется извлечь корень из числа 0,03672. Увеличив его только в 100 раз, получим 3,672, или после округления 4. Так как в данном случае мы получили однозначное число, а наши таблицы дают возможность извлекать корни даже из трехзначных чисел, то выгоднее будет увеличить заданное число (0,03672) не в 100, а в 10000 раз. Получим 367,2 или после округления 367. По таблицам находим, что 1/367=19,1572. Так как заданное число мы увеличили переносом запятой на 4 места вправо, то результат этот необходимо исправить обратным переносом запятой на 2 места влево. Окончательно будем иметь:

у 0,03672 прибл. = 0,191572.

Вообще следует помнить, что при извлечении квадратных корней с помощью этих таблиц следует заданное число сделать сперва двухзначным или трехзначным, уменьшив или увеличив его для этого в 100 или 10000 раз, затем округлить до ближайшего целого числа, найти соответствующее значение квадратного корня по таблицам и результат исправить обратным увеличением или уменьшением уже в 10 _или в 100 раз (если заданное число было уменьшено или увеличено в 100 раз, то результат исправляется увеличением или уменьшением только в 10 раз, если же заданное число было разделено или умножено на 10000, то результат исправляется умножением или делением его на 100).

При таких 'операциях мы, вообще говоря, получаем значения квадратных корней лишь приближенно. В особенности сильно неточность эта обнаруживается в том случае, когда заданное число мы предварительно уменьшаем в 100 или в 10000 раз с последующим увеличением результата в 10 или в 100 раз, т.-е. при извлечении корня из многозначных чисел. При извлечении же корней из дробных чисел результат получается более точный.

Заметим, что извлечение квадратных корней, равно как и все другие действия, для наших целей чаще всего достаточно производить с точностью до 0,01, т.-е. доводить вычисления только до второго десятичного знака включительно. Для

По таблицам находим, что

$$\sqrt{N} = \sqrt{279} \text{ прибл.} = 16{,}7033.$$

Это и есть приближенное значение искомого $\sqrt{Z}$, значение, к которому следует еще прибавить поправочную величину. Эта поправочная величина представляет собой произведение двух чисел. Найдем их порознь.

Первое из этих чисел найдется, как разность между значением корня из числа $N+1$ и значением корня из $N$. Так как у нас $N=279$, то $N+1=280$. По таблице находим, что

$$\sqrt{N+1} = \sqrt{280} \text{ прибл.} = 16{,}7332.$$
$$\sqrt{N} = \sqrt{279} \text{ прибл.} = 16{,}7033.$$

Разность между этими числами $16{,}7332 - 16{,}7033 = 0{,}0299$. Эта разность (0,0299) и является одним из двух чисел, произведение которых служит искомой поправкой.

Найдем второе число (другой множитель этого произведения). Это число представляет собой разность между числами $Z$ и $N$. Так как $Z=279{,}35$, а $N=279$, то искомая разность $279{,}35 - 279 = 0{,}35$.

Итак, мы получили два числа 0,0299 и 0,35 которые теперь следует перемножить. Помножив 0,0299 на 0,35, получим 0,010465. Это и есть искомая величина поправки, которую надо прибавить к приближенному значению нашего корня, т.-е. к $\sqrt{N} = 16{,}7033$.

Сложив эти числа получим окончательно:

$$16{,}7033 + 0{,}010465 = 16{,}713765.$$

Итак, $\sqrt{279{,}35}$ прибл. $= 16{,}713765$.

Все эти действия удобно было бы расположить в таком порядке:

$$\begin{array}{ll} Z = 279{,}35 & \sqrt{N+1} = 16{,}7332 \\ N = 279 & \sqrt{N} = 16{,}7033 \\ \hline Z - N = 0{,}35 & \sqrt{N+1} - \sqrt{N} = 0{,}0299 \end{array}$$

$$\sqrt{Z} = \sqrt{N} + (\sqrt{N+1} - \sqrt{N})\,(Z-N) =$$
$$= 16{,}7033 + 0{,}0299 \,.\, 0{,}35 = 16{,}7033 + 0{,}010465 = 16{,}713765.$$



При вычислении $\sqrt{Z}$ с точностью до 2-х знаков получим:

$$\sqrt{279{,}35} \text{ прибл.} = 16{,}71.$$

Применим этот более точный способ округления смешанных и дробных чисел (с поправкой) к случаю извлечения квадратного корня из числа не встречающегося в таблицах. Пусть, напр., требуется извлечь корень из числа 1256,7. Уменьшив его в 100 раз, получим 12,567. Следовательно у нас: $Z=12{,}567$, $N=12$, а $N+1=13$. Вычисления расположим так:

$$\begin{array}{ll} Z = 12{,}567 & \sqrt{N+1} = \sqrt{13} = 3{,}6056 \\ N = 12 & \sqrt{N} = \sqrt{12} = 3{,}4641 \\ \hline Z - N = 0{,}567 & \sqrt{N+1} - \sqrt{N} = 0{,}1415 \end{array}$$

$$\sqrt{Z} = \sqrt{N} + (\sqrt{N+1} - \sqrt{N})\;(Z-N) = 3{,}4641 +$$
$$+ 0{,}1415 \,.\, 0{,}567 = 3{,}4641 + 0{,}0802305 = 3{,}5443305.$$

Так как число, из которого требовалось извлечь корень (1256,7), было предварительно уменьшено в 100 раз, то полученный здесь результат следует исправить увеличением в 10 раз. Следовательно окончательно будем иметь:

$$\sqrt{1256{,}7} \text{ прибл.} = 35{,}443305$$

или, взяв это число только с 2-мя десятичными знаками после запятой:

$$\sqrt{1256{,}7} \text{ прибл.} = 35{,}44.$$

В тех случаях, когда требуется извлечь квадратный корень еще более точно (напр., с точностью до 4 знаков и более), необходимо воспользоваться таблицами логарифмов (напр., пятизначными таблицами логарифмов Пржевальского).

Способ вычисления с помощью логарифмов, как известно, заключается в том, что взамен сложных математических действий, которые требуется выполнить над заданными числами, выполняются действия более простые над особыми величинами, называемыми „логарифмами" этих чисел.

Так, вместо перемножения двух или несколько чисел можно сделать простое сложение их логарифмов, деление чисел заменить вычитанием соответствующих логарифмов, возвышение в любую степень выполнить путем умножения логарифма на показатель этой степени, а извлечение корня, наоборот, заменить простым делением логарифма.

Порядок вычисления в общем виде таков. По специальным логарифмическим таблицам отыскивают логарифмы заданных чисел, с этими логарифмами выполняют соответствующие упрощенные действия и по полученному в результате логарифму снова, при помощи тех же таблиц, находят соответствующее число.

При извлечении квадратного корня мы должны, следовательно, прежде всего отыскать логарифм того числа, из которого требуется извлечь корень, затем разделить этот логарифм на 2 и по полученному таким образом уменьшенному вдвое логарифму снова отыскать по таблицам соответствующее ему число. Это число и будет являться квадратным корнем из первоначально заданного числа. Заметим кстати, что логарифм изображается знаком *lg*; логарифм числа 120 можно записать так: *lg* 120.

Пусть, напр., в конечном итоге произведенных вычислений мы получили:

$$\sigma = \pm\sqrt{4{,}032}$$

Для извлечения квадратного корня из числа 4,032 отыщем по таблицам Пржевальского логарифм этого числа (см. описание способа пользования этими таблицами, приведенное там на стр. 4). Мы найдем, что $lg\ 4{,}032 = 0{,}60552$.

Этот логарифм нам нужно разделить пополам:

$$\frac{lg\ 4{,}032}{2} = \frac{0{,}60552}{2} = 0{,}30276.$$

Полученный результат и является уже логарифмом искомой величины, т.-е. $lg\ \sigma = 0{,}30276$.

Нам остается теперь лишь отыскать по таблицам Пржевальского то число, которое соответствует этому логарифму. Таким числом будет 2,008.

Следовательно, $\sigma = \pm 2{,}008$ или, округляя его до двух знаков, $\sigma = \pm 2{,}01$.



Все эти действия можно записать так:

$$\sigma = \pm\sqrt{4{,}032}$$
$$lg\ 4{,}032 = 0{,}60552$$
$$lg\ \sigma = \frac{lg\ 4{,}032}{2} = \frac{0{,}60552}{2} = 0{,}30276$$
$$\sigma = \pm 2{,}01.$$

Как видно из этого примера, извлечение квадратных корней с помощью логарифмов не является особенно сложным. Некоторое затруднение может возникнуть здесь лишь при операциях с дробными числами, имеющими, как известно, отрицательные логарифмы. При делении такого логарифма на 2 приходится иногда применять особый искусственный прием, который мы и поясним здесь на нескольких числовых примерах.

Пусть требуется разделить на 2 логарифм такого вида:

$$\overline{4}{,}82678$$

Здесь знак минус, поставленный над числом 4, указывает на то, что только это число 4 является отрицательным, все же остальные десятичные знаки в этом обозначении представляют собой цифры положительной дроби. Другими словами, заданный логарифм представляет собой сумму отрицательного целого числа — 4 и положительной дроби + 0,82678.

Так как в данном случае отрицательная часть такого логарифма (— 4) является числом четным (делящимся на 2 без остатка), то выполнить это деление можно по обычным правилам, т.-е. разделить на 2 отдельно эту отрицательную часть логарифма (— 4) и отдельно же положительную его часть (0,82678). В результате, очевидно, получим:

$$\overline{2}{,}41339.$$

Иначе пришлось бы поступить в том случае, если бы отрицательная часть логарифма оказалась числом н е ч е т н ы м. Пусть, напр., имеем такой логарифм:

$$\overline{3}{,}26408.$$

В этом случае мы можем прибегнуть к следующему искусственному приему. Прибавим мысленно к отрицательной части нашего логарифма (— 3) еще одну отрицательную же единицу, т.-е. будем считать, что эта отрицательная часть логарифма у нас уже не — 3, а — 4. Чтобы не изменить общей величины всего заданного первонанально логарифма, прибавим к его положительной (дробной) части тоже одну единицу, но уже, очевидно, положительную. Тогда эта положительная часть нашего логарифма сделается равной 1,26408 (вместо 0,26408). Преобразовав, таким образом, этот логарифм, мы можем выполнить деление его на 2 совершенно так же, как и в предыдущем примере, т.-е. отдельно разделить на 2 отрицательную часть — 4 (получим — 2) и отдельно же положительную часть 1,26408. Выполнив это деление с величиной 1,26408, получим 0,63204. Следовательно окончательно будем иметь $\bar{2},63204$.

Итак, при делении на 2 такого логарифма, отрицательная часть которого представляет собой нечетное число (напр., $\bar{1}$, $\bar{3}$, $\bar{5}$ и т. д.), следует дополнить это число до следующего четного числа (напр., вместо $\bar{1}$ считать $\bar{2}$, вместо $\bar{5}$ — $\bar{6}$ и т. д.), к положительной дроби прибавить единицу, разделить на 2 обе части такого преобразованного логарифма и результаты соединить в обычном символическом изображении.

Пусть, напр., в конечном итоге вычислений оказалось, что

$$\sigma = \pm\sqrt{0,961}.$$

По таблицам логарифмов находим, что

$$lg\ 0,961 = \bar{1},98272.$$

При делении этого логарифма пополам прибавим мысленно к его отрицательной части (— 1) еще одну отрицательную же единицу (получим — 2), а к положительной дроби (0,98272) — положительную единицу (получим 1,98272). Разделив обе части этого преобразованного логарифма на 2, в отрицательной его части получим, очевидно, — 1, а в положительной — дробь 0,99136, т.-е. в конечном итоге:

$$lg\ \sigma = \bar{1},99136.$$

Этому логарифму по таблицам Пржевальского соответствует число 0,9803. Следовательно: $\sigma = \pm 0,98$.



Все эти действия можно записать так:

$$\sigma = \pm\sqrt{0,961}$$
$$lg\ 0,961 = \bar{1},98272$$
$$lg\ \sigma = \frac{lg\ 0,961}{2} = \frac{\bar{1},98272}{2} = \bar{1},99136$$
$$\sigma = \pm 0,98.$$

Так как самый способ нахождения по заданному числу соответствующего ему логарифма и, наоборот, по заданному логарифму соответствующего ему числа с достаточною ясностью изложен в описании процесса логарифмирования, приложенном к таблицам Пржевальского, то вопрос об извлечении квадратного корня с помощью логарифмических таблиц можно считать исчерпанным.

---

## XI. Разбивка вариационного ряда на классы.

Во всех разобранных ранее примерах определения $M$, $\sigma$ и $m$ по способу сумм отдельные значения вариант отличались друг от друга ровно на единицу. Так, последний раз мы имели вариационный ряд, отдельные члены которого (варианты) принимали только следующие значения:

5, 6, 7, 8, 9, 10, 11, 12.

Однако на практике приходится иногда иметь дело с такими рядами, в которых отдельные значения вариант отличаются друг от друга и на другую величину. Пусть, напр., имеем вариационный ряд, состоящий из следующих 14 вариант:

30, 15, 40, 25, 20, 10, 25, 20, 10, 30, 20, 25, 15, 20.

Замечая, что варианты эти колеблются здесь в максимальных пределах от 10 до 40 и что отдельные значения их отличаются друг от друга не на единицу, а на 5, мы можем представить этот ряд в виде следующей сводки:

| | | | |
|---|---|---|---|
| варианта 40 | встретилась в заданном ряду . . . . | 1 раз | частоты $p$. |
| " 35 | " " " " . . . . | 0 " | |
| " 30 | " " " " . . . . | 2 " | |
| " 25 | " " " " . . . . | 3 " | |
| " 20 | " " " " . . . . | 4 " | |
| " 15 | " " " " . . . . | 2 " | |
| " 10 | " " " " . . . . | 2 " | |

Этим путем мы разбили вариационный ряд на 7 отдельных групп, или „классов", при чем значения вариант двух соседних классов отличаются здесь всегда на 5 единиц. Эту величину (5) мы обозначим буквою $k$ и назовем классовым промежутком. Итак, в нашем примере классовый промежуток $k = 5$.

При вычислении $M$, $\sigma$ и $m$ для такого ряда по способу сумм мы должны проделать все обычные операции, связанные с получением вспомогательных сумм $S_1$ и $S_2$, и вычислить с их помощью сумму квадратов центральных отклонений $\Sigma x^2$ *), но вместо указанных ранее формул для определения $M$, $\sigma$ и $m$:

$$M = C + \frac{S_1}{n}; \quad \sigma = \pm\sqrt{\frac{\Sigma x^2}{n-1}} \quad \text{и} \quad m = \pm\sqrt{\frac{\Sigma x^2}{n(n-1)}}$$

воспользоваться уже другими формулами, в которые входит величина классового промежутка $k$, а именно:

$$M = C + \frac{kS_1}{n}; \quad \sigma = \pm k\sqrt{\frac{\Sigma x^2}{n-1}} \quad \text{и} \quad m = \pm k\sqrt{\frac{\Sigma x^2}{n(n-1)}}.$$

Первая из этих формул показывает, что для вычисления среднего арифметического $M$ следует полученную из расчетной решетки сумму $S_1$ помножить на классовый промежуток $k$, произведение это разделить на число измерений $n$ и результат прибавить к началу отсчетов $C$.

Вторая формула показывает, что для вычисления среднего квадратического отклонения $\sigma$ следует сумму квадратов центральных отклонений $\Sigma x^2$ разделить на число измерений без единицы (т.-е. на $n - 1$), из полученного результата

*) В сущности при $k$ не $= 1$ величины $S_1$ $S_2$ и $\Sigma x^2$ уже не представляют собою истинных значений суммы вспомогательных отклонений, суммы квадратов вспомогательных отклонений и суммы квадратов центральных отклонений. Мы сохраняем здесь эти наименования лишь условно, чтобы избежать введения лишних терминов.



извлечь квадратный корень и помножить его на классовый промежуток $k$.

Третья формула показывает, что для вычисления средней ошибки $m$ следует сумму квадратов центральных отклонений $\Sigma x^2$ разделить на число измерений $n$, помноженное предварительно на то же число измерений без единицы ($n-1$), из полученного результата извлечь квадратный корень и помножить его на классовый промежуток $k$.

Если средняя ошибка $m$ вычисляется не самостоятельно, а с помощью $\sigma$, то никакой поправки в соответствующую формулу вносить, очевидно, не нужно (так как поправка эта входит уже в формулу для вычисления самой $\sigma$). Итак, в этом случае, так же, как и ранее, среднюю ошибку $m$ можно вычислить по основной формуле:

$$m = \frac{\pm\sigma}{\sqrt{n}}.$$

Заметим, что прежние наши формулы для вычисления $M$, $\sigma$ и $m$ представляли собой лишь частный случай этих новых (общих) формул с поправкой на величину классового промежутка $k$. В самом деле, ранее отдельные значения вариант всегда отличались друг от друга ровно на единицу, следовательно классовый промежуток $k$ равнялся там 1. Если в новых формулах принять $k = 1$, то мы и получим наши прежние формулы, соответствующие этому частному случаю.

Проделаем же все эти вычисления с данными последнего примера.

Составив расчетную решетку (см. таблицу 29), занеся в нее все частоты заданного ряда и проделав обычные операции, связанные с определением итогов по верхним и нижним частям среднего и правого столбца, получим:

$$a_1 = 5 \qquad a_2 = 3$$
$$b_1 = 14 \qquad b_2 = 8$$

откуда определим и обе вспомогательные суммы $S_1$ и $S_2$:

$$S_1 = a_1 - b_1 = 5 - 14 = -9$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 = 5 + 14 + 2\cdot3 + 2\cdot8 = 5 + 14 + 6 + 16 = 41.$$

Определив суммы $S_1 = -9$ и $S_2 = 41$, найдем сумму $\Sigma x^2$:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 41 - \frac{(-9)^2}{14} = 41 - \frac{81}{14} = 41 - 5{,}79 = 35{,}21$$

откуда

$$\sigma = \pm k\sqrt{\frac{\Sigma x^2}{n-1}} = \pm 5\sqrt{\frac{35{,}21}{14-1}} = \pm 5\sqrt{\frac{35{,}21}{13}} = \pm 5\sqrt{2{,}71}$$

ТАБЛИЦА 29.

| | $p$ | $a_1$ = 5 | $a_2$ = 3 |
|---|---|---|---|
| 40 | 1 | 1 | 1 |
| 35 | 0 | 1 | 2 |
| 30 | 2 | 3 | — |
| 25 | 3 | — | — |
| 20 | 4 | 8 | — |
| 15 | 2 | 4 | 6 |
| 10 | 2 | 2 | 2 |
| | 14 | 14 | 8 |
| | $n$ | $b_1$ | $b_2$ |

так как $\sqrt{2{,}71} = 1{,}64621$, то

$$\sigma = \pm 5 \cdot 1{,}64621 = \pm 8{,}23105$$

или после округления *)

$$\sigma = \pm 8{,}23.$$

Итак,

$$\sigma = \pm 8{,}23.$$

Среднюю ошибку $m$ найдем по формуле:

$$m = \frac{\pm\sigma}{\sqrt{n}} = \frac{\pm 8{,}23}{\sqrt{14}}.$$

Так как $\sqrt{14}$ прибл. $= 3{,}74$ **), то

$$m = \frac{\pm 8{,}23}{3{,}74} = \pm\frac{823}{374} = \pm 2{,}20.$$

Итак,

$$m = \pm 2{,}20.$$

Для вычисления среднего арифметического $M$ выясним, какое значение варианты $V$ было принято нами за начало отсчетов $C$. Так как фигура из 4-х черточек была расположена в расчетной решетке против частоты 3, а эта частота соответствует варианте 25 (см. табл. 29), то за $C$ мы и должны будем принять это число 25. Итак, у нас

$$C = 25$$

*) Для бо́льшей точности это округление сделано здесь уже в самом конце вычислений, а не при извлечении корня из 2,71.

**) Здесь взято только два знака после запятой, чтобы облегчить деление 8,23 на 3,74.



поэтому

$$M = C + \frac{kS_1}{n} = 25 + \frac{5\cdot(-9)}{14} = 25 - \frac{45}{14} = 25 - 3{,}21 = 21{,}79.$$

Итак,

$$M = 21{,}79.$$

Эти конечные результаты мы можем записать в такой форме:

$$M = 21{,}79 \pm 2{,}20$$
$$\sigma = \pm 8{,}23.$$

ТАБЛИЦА 30.

| | $p$ | $a_1$ = 21 | $a_2$ = 19 |
|---|---|---|---|
| 7 | 2 | 2 | 2 |
| 6,5 | 1 | 3 | 5 |
| 6 | 4 | 7 | 12 |
| 5,5 | 2 | 9 | — |
| 5 | 5 | — | — |
| 4,5 | 2 | 6 | — |
| 4 | 3 | 4 | 6 |
| 3,5 | 0 | 1 | 2 |
| 3 | 1 | 1 | 1 |
| | 20 | 12 | 9 |
| | $n$ | $b_1$ | $b_2$ |

Приведем еще один пример вариационного ряда с классовым промежутком $k$, неравным единице. Пусть имеем следующие 20 целых и дробных вариант:

6, 5, 4½, 4, 6, 3, 5½, 5, 7, 4, 5, 4½, 5, 6, 4, 5½, 7, 5, 6½, 6.

После подсчета одинаковых (повторяющихся) вариант получим следующую сводку:

| | | | |
|---|---|---|---|
| варианта | 7 | встретилась в заданном ряду . . . . . . | 2 раза |
| " | 6½ | " " " " . . . . . . | 1 " |
| " | 6 | " " " " . . . . . . | 4 " |
| " | 5½ | " " " " . . . . . . | 2 " |
| " | 5 | " " " " . . . . . . | 5 " |
| " | 4½ | " " " " . . . . . . | 2 " |
| " | 4 | " " " " . . . . . . | 3 " |
| " | 3½ | " " " " . . . . . . | 0 " |
| " | 3 | " " " " . . . . . . | 1 " |

частоты $p$

Так как соседние значения вариант отличаются друг от друга на ½, то классовый промежуток

$$k = 0{,}5.$$

Составив расчетную решетку (см. табл. 30) и произведя в ней все необходимые вычисления, найдем, что:

$$a_1 = 21 \qquad a_2 = 19$$
$$b_1 = 12 \qquad b_2 = 9$$

откуда

$$S_1 = a_1 - b_1 = 21 - 12 = +9$$

$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 =$$

$$= 21 + 12 + 2 \cdot 19 + 2 \cdot 9 = 21 + 12 + 38 + 18 = 89.$$

Следовательно:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 89 - \frac{(+9)^2}{20} = 89 - \frac{81}{20} = 89 - 4{,}05 = 84{,}95$$

$$\sigma = \pm k \sqrt{\frac{\Sigma x^2}{n-1}} = \pm 0{,}5 \sqrt{\frac{84{,}95}{20-1}} = \pm 0{,}5 \sqrt{\frac{84{,}95}{19}}$$

$$= \pm 0{,}5 \sqrt{4{,}47}\,.$$

Так как $\sqrt{4{,}47} = 2{,}11424$, то $\sigma = \pm 0{,}5 \cdot 2{,}11424 = \pm 1{,}05712$.

Итак, $\sigma = \pm 1{,}06$.

$$m = \frac{\pm \sigma}{\sqrt{n}} = \frac{\pm 1{,}06}{\sqrt{20}}.$$ Так как $\sqrt{20}$ прибл. $= 4{,}47$, то

$$m = \frac{\pm 1{,}06}{4{,}47} = \pm \frac{106}{447} = \pm 0{,}24.$$

Итак, $m = \pm 0{,}24$.

Так как за начало отсчетов $C$ принята варианта 5 (см. табл. 30), то

$$M = C + \frac{kS_1}{n} = 5 + \frac{0{,}5\,(+9)}{20} = 5 + \frac{4{,}5}{20} = 5 + 0{,}23 = 5{,}23.$$

Итак, $M = 5{,}23$.

Следовательно:

$$M = 5{,}23 \pm 0{,}24 \qquad \sigma = \pm 1{,}06$$

В тех случаях, когда после подсчета одинаковых вариант общее количество классов оказывается слишком большим, а классовый промежуток $k$, наоборот, слишком малым, полезно бывает объединить несколько соседних мелких классов в более крупные группы.

Пусть, напр., после подсчета повторяющихся вариант какого-либо ряда мы получили следующую первичную их сводку, представленную на табл. 31. Здесь в столбце $V$ указаны отдельные значения этих вариант, а рядом с ними в столбце $p$ — количество повторений этих значений (т.-е. их частоты).



ТАБЛИЦА 31.

| | V | p | |
|---|---|---|---|
| 23 | 24 | 1 | 4 |
| | 23 | 1 | |
| | 22 | 2 | |
| 20 | 21 | 2 | 10 |
| | 20 | 3 | |
| | 19 | 5 | |
| 17 | 18 | 7 | 24 |
| | 17 | 11 | |
| | 16 | 6 | |
| 14 | 15 | 4 | 9 |
| | 14 | 2 | |
| | 13 | 3 | |
| 11 | 12 | 0 | 3 |
| | 11 | 1 | |
| | 10 | 2 | |

Желая уменьшить общее количество классов (15) за счет увеличения классового п р о м е ж у т к а $k$ (равного здесь 1), мы можем сгруппировать эти классы, напр., по три, получив в результате уже не 15, а всего только 5 таких объединенных (крупных) классов. Очевидно, классовый промежуток $k$ в этом случае также утроится и будет равен уже не 1, а 3.

Результаты такой группировки мелких классов в более крупные представлены на табл. 32, где в клетках столбца $V$ перечислены отдельные значения вариант, вошедшие в каждый из пяти классов, а в соседнем столбце $p$ выписаны и суммированы по отдельным (новым) классам частоты этих значений. Таким образом частоты $p$ этих пяти новых классов выразятся числами: 4, 10, 24, 9, 3

ТАБЛИЦА 32.

| V | p |
|---|---|
| 24, 23, 22 | $1 + 1 + 2 = 4$ |
| 21, 20, 19 | $2 + 3 + 5 = 10$ |
| 18, 17, 16 | $7 + 11 + 6 = 24$ |
| 15. 14, 13 | $4 + 2 + 3 = 9$ |
| 12, 11, 10 | $0 + 1 + 2 = 3$ |

В дальнейшем при таком объединении нескольких мелких классов в более крупные мы будем считать, что в каждом из них заключаются н е р а з л и ч н ы е значения вариант $V$ (как это было в действительности), а совершенно одинаковые, в точности равные среднему значению вариант данного объединенного класса. Так, напр., в нашем примере первый (сверху) класс состоит в действительности из трех различных значений вариант: 24, 23 и 22. Отныне мы будем считать, что все варианты, вошедшие в этот класс, являются уже одинаковыми, в точности равными среднему значению этого класса, т.-е. 23.

Это среднее значение класса мы будем определять, как полусумму крайних его значений:

$$\frac{24+22}{2}=\frac{46}{12}=23.$$

Точно также все варианты следующего (второго сверху) класса мы будем считать равным среднему значению этого класса:

$$\frac{21+19}{2}=\frac{40}{2}=20 \text{ и т. д.}$$

Такое искусственное уравнивание различных значений вариант внутри каждого объединенного класса является, конечно, совершенно произвольным и делается лишь для удобства дальнейших вычислений в ущерб их точности. Однако теория и практика подобных операций показывает, что при достаточно большом количестве измерений ($n$) неточность эта бывает обычно настолько незначительной, что ею просто можно пренебречь.

ТАБЛИЦА 33.

| Границы классов (от — до) | Среднее значение классов | Частоты $p$ |
|---|---|---|
| 24 — 22 . . . | 23 | 4 |
| 21 — 19 . . . | 20 | 10 |
| 18 — 16 . . . | 17 | 24 |
| 15 — 13 . . | 14 | 9 |
| 12 — 10 . . . | 11 | 3 |



Итак, после преобразования нашего вариационного ряда мы можем представить его в окончательной форме, указанной на табл. 33. Здесь в левом столбце выписаны границы всех пяти классов (от — до), в следующем столбце — средние значения этих классов, а в правом — частоты этих значений. Таблицу эту следует понимать так: в числе различных значений наших вариант встретились 4 варианты, заключающиеся в пределах от 24 до 22 включительно, т.-е. по величине своей близкие к среднему значению этого класса 23; далее, 10 вариант оказались заключенными в границах от 21 до 19, т.-е. примыкающими к среднему значению 20, и т. д.

Приняв середины всех классов за отдельные значения повторяющихся вариант, а разницу между ними — за классовый промежуток $k$, мы можем составить обычного типа расчетную решетку (см. табл. 34) и вычислить для этого ряда $M$, $\sigma$ и $m$ по общим правилам.

ТАБЛИЦА 34.

| | | $a_1$ | $a_2$ |
|---|---|---|---|
| | $p$ | 18 | 4 |
| 23 | 4 | 4 | 4 |
| 20 | 10 | 14 | — |
| 17 | 24 | — | — |
| 14 | 9 | 12 | — |
| 11 | 3 | 3 | 3 |
| | 50 | 15 | 3 |
| | $n$ | $b_1$ | $b_2$ |

Проделав все необходимые вычисления с началом отсчетов $C=17$, получим:

$$a_1=18 \qquad a_2=4$$
$$b_1=15 \qquad b_2=3$$

откуда:

$$S_1=a_1-b_1=18-15=+3$$
$$S_2=a_1+b_1+2a_2+2b_2=18+15+2\cdot4+2.3=$$
$$=18+15+8+6=47.$$

Следовательно:

$$\Sigma x^2=S_2-\frac{S_1^2}{n}=47-\frac{(+3)^2}{50}=47-\frac{9}{50}=47-0{,}18=46{,}82$$

$$\sigma=\pm k\sqrt{\frac{\Sigma x^2}{n-1}}=\pm3\sqrt{\frac{46{,}82}{50-1}}=\pm3\sqrt{\frac{46{,}82}{49}}=\pm3\sqrt{0{,}96}$$

Так как $\sqrt{0{,}96}=0{,}9798$, то $\sigma=\pm3\cdot0{,}9798=\pm2{,}94$.

Итак, $\sigma = \pm 2{,}94$

$$m = \frac{\pm \sigma}{\sqrt{n}} = \frac{\pm 2.94}{\sqrt{50}}.$$ Так как $\sqrt{50}$ прибл. $= 7{,}07$, то

$$m = \frac{\pm 2{,}94}{7{,}07} = \frac{\pm 294}{707} = \pm 0{,}42.$$

Итак, $m = \pm 0{,}42$

$$M = C + \frac{kS_1}{n} = 17 + \frac{3 \cdot (+3)}{50} = 17 + \frac{9}{50} = 17 + 0{,}18 = 17{,}18.$$

Итак,

$$M = 17{,}18.$$

Следовательно:

$$M = 17{,}18 \pm 0{,}42 \qquad \sigma = \pm 2{,}94.$$

В только что разобранном примере границы всех пяти объединенных классов были установлены в расчете только на целые значения отдельных вариант $V$. Так, к первому из этих классов мы отнесли варианты: 24, 23, 22, ко второму — 21, 20, 19 и т. д., установив для этих классов следующие границы:

от 24 до 22 включительно
„ 21 — 19 „
„ 18 — 16 „
„ 15 — 13 „
„ 12 — 10 „

Однако кроме таких целых вариант в нашем ряду, вообще говоря, могли бы встретиться также и дробные числа, напр., с половинами (17,5 или 23,5 и т. д.). Куда же следовало бы отнести их при разбивке на классы?

Тут возможны два случая. Если такая дробная варианта по величине своей окажется в границах какого-либо класса, то ее и следует, очевидно, отнести к этому классу. Напр., варианту 23,5 следует отнести к первому классу, так как она находится в его границах от 24 до 22.

Но возможен и другой случай, когда такая дробная варианта по величине своей придется между двумя соседними классами. Пусть, напр., имеем варианту 21,5. Так как варианта эта меньше 22, то отнести ее к первому классу нельзя (класс этот заключает в себе варианты от 24 до 22,



т.-е. бо́льшие 22). Но ее нельзя отнести также и к следующему (второму) классу, так как в этом классе, по условию, заключаются лишь варианты, меньшие 21 (от 21 до 19). Поэтому в данном случае варианта 21,5 попадает как бы в промежуток между двумя соседними классами, первым и вторым.

То же самое произошло бы и с вариантой 18,5, которая также оказывается в промежутке между вторым и третьим классом, и т. д.

Чтобы избежать подобных затруднений при распределении вариант по классам, можно расширить установленные ранее их границы, включив в них также и эти промежуточные дробные значения. В данном случае, расчитывая на дробные значения вариант только с половинами (а не с десятыми долями или еще мельче), можно изменить, наприм., нижние границы всех классов, уменьшив их на 0,5. Тогда в каждый объединенный класс мы должны будем отнести уже не по 3 целых, а по 6 таких дробных значений вариант $V$, что можно представить в виде следующей наглядной сводки (см. табл. 35).

ТАБЛИЦА 35.

| Отдельные значения вариант $V$, вошедшие в каждый из 5-ти классов | Границы классов (от — до) | Средние значения классов | Частоты $p$ |
|---|---|---|---|
| 24, 23½, 23, 22½, 22, 21½ . . | 24 — 21,5 | 22,75 | 4 |
| 21, 20½, 20, 19½, 19, 18½ . . | 21 — 18,5 | 19,75 | 10 |
| 18, 17½, 17, 16½, 16, 15½ . . | 18 — 15,5 | 16,75 | 24 |
| 15, 14½, 14, 13½, 13, 12½ . . | 15 — 12,5 | 13,75 | 9 |
| 12, 11½, 11, 10½, 10, 9½ . . | 12 — 9,5 | 10,75 | 3 |

Для определения средних значений каждого такого класса мы должны попрежнему вычислить полусумму его границ. Так, напр., в первом классе заключаются теперь все варианты.

от 24 до 21,5 включительно. Сложив обе эти крайние его границы, получим $24 + 21{,}5 = 45{,}5$. Разделив эту сумму (45,5) на 2, получим 22,75. Следовательно за среднее значение первого класса (от 24 до 21,5) мы должны принять величину 22,75 (см. табл. 35). Таким же образом получим и среднее значение следующего (второго) класса:

$$\frac{21 + 18{,}5}{2} = \frac{39{,}5}{2} = 19{,}75 \text{ и т. д.}$$

Если бы среди отдельных вариант нашего ряда попадались также и дробные числа с различными десятичными знаками (а не только с половинами), напр., 19,3 или 28,8 и т. д., то некоторые из этих значений все же могли бы попасть при такой разбивке в промежуток между двумя соседними классами. Напр., куда следовало бы отнести варианту 21,2? К первому классу она не подходит, так как нижняя граница этого класса = 21,5 (т.-е, больше 21,2), а следующий, второй класс начинается уже со значения 21, в то время как наша варианта 21,2 больше этой верхней его границы. Словом, такая варианта и в данном случае не попадет ни в один из этих классов, т.-е. окажется в промежутке между ними.

В расчете на такой случай мы можем еще больше расширить границы наших классов, хотя бы опять за счет уменьшения их нижних границ. Для этого стоит лишь приблизить эти границы почти вплотную к верхним границам следующих (меньших) классов, доведя эту разницу в данном случае уже до 0,1. Тогда границы отдельных классов и их средние значения определятся следующим образом:

| | | |
|---|---|---|
| от 24 до 21,1 | включительно, со средним значением . . . . . . . . . | 22,55 |
| „ 21 — 18,1 | „ „ „ „ . . . . . . . . . . | 19,55 |
| „ 18 — 15,1 | „ „ „ „ . . . . . . . . . . | 16,55 |
| „ 15 — 12,1 | „ „ „ „ . . . . . . . . . . | 13,55 |
| „ 12 — 9,1 | „ „ „ „ . . . . . . . . . . | 10,55 |

При таких классовых границах уже ни одна дробная варианта не окажется в промежутке между двумя соседними классами и при разбивке непременно войдет в какой-либо определенный класс. Средние значения этих классов получены здесь, как и ранее, путем сложения двух границ и деления этой суммы пополам. В самом деле, границами первого класса служат здесь числа 24 и 21,1. Сумма их $24 + 21{,}1 = 45{,}1$. Разделив эту сумму на 2, получим 22,55. Это и есть среднее значение первого класса. Так же найдем средние значения и остальных классов.



Если бы различие между отдельными вариантами было еще меньше и выражалось, скажем, в сотых долях, т.-е. варианты эти имели бы, напр., такой вид: 14,73 или 18,06 и т. д., то нижние границы всех классов пришлось бы подвести еще ближе к верхним границам следующих классов, доведя эту разницу уже до 0,01. В этом случае границы классов и их средние значения можно было бы представить в таком виде:

| | | |
|---|---|---|
| от 24 до 21,01 | включительно, со средним значением . . . . . . . . | 22,505 |
| „ 21 — 18,01 | „ „ „ „ . . . . . . . . | 19,505 |
| „ 18 — 15,01 | „ „ „ „ . . . . . . . . | 16,505 |
| „ 15 — 12,01 | „ „ „ „ . . . . . . . . | 13,505 |
| „ 12 — 9,01 | „ „ „ „ . . . . . . . . | 10,505 |

Итак, при установлении классовых границ мы должны наметить верхние их значения в круглых числах (у нас верхними границами служили числа: 24, 21, 18, 15 и 12, отличающиеся друг от друга ровно на величину классового промежутка $k = 3$), затем, приняв во внимание точность самих измерений (вариант $V$), зафиксировать соответствующие нижние границы этих классов, приблизив их к верхним границам следующих (меньших) классов на такую величину, на которую минимально могут отличаться друг от друга отдельные варианты $V$ (т.-е. на 0,5, или 0,1, или 0,01 и т. д.). Средние же значения отдельных классов следует определить путем сложения двух этих границ (верхней и нижней) с последующим делением этой суммы на 2.

Заметим, что в действительности совершенно излишне определять все средние значения отдельных классов. Для производства вычислений по способу сумм необходимо знать лишь среднее значение того класса, против частоты которого ($p$) в расчетной решетке будут расположены четыре черточки. Среднее значение этого класса мы и должны будем принять за начало отсчетов $C$ (в данном случае, следовательно, это $C$ может оказаться и дробным числом). Таким образом установление точных границ отдельных классов является необходимым лишь для разбивки по этим классам отдельных вариант заданного ряда, а при производстве вычислений внутри самой расчетной решетки взамен этих границ и их средних значений нам нужно будет знать лишь величину классового промежутка $k$ и начало отчетов $C$.

После этих предварительных замечаний разберем конкретный числовой пример обработки вариационного ряда, состоящего из сравнительно большого количества целых и дробных вариант (самый общий случай).

Пусть, напр., при каких-либо измерениях мы получили следующие числовые их результаты, выраженные в см, в сек., $^0/_0$ $^0/_0$ или других мерах:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12,2 | 8,9 | 15,3 | 15,8 | 10,9 | 15,0 | 13,7 | 13,1 |
| 9,9 | 7,9 | 13,3 | 9,8 | 8,8 | 18,6 | 10,6 | 11,3 |
| 6,0 | 12,4 | 7,1 | 10,4 | 11,3 | 4,6 | 10,3 | 14,5 |
| 10,8 | 3,2 | 12,0 | 11,7 | 10,0 | 7,5 | 10,4 | 13,5 |
| 16,4 | 10,5 | 6,4 | 9,1 | 14,6 | 4,9 | 8,1 | 7,7 |
| 19,4 | 11,0 | 13.0 | 6,8 | 8,4 | 14,0 | 12,7 | 11,8 |

Мы имеем здесь вариационный ряд, состоящий из 48 отдельных вариант ($n = 48$). Для разбивки его на классы нам нужно прежде всего выяснить наибольший размах колебаний всех этих чисел, т.-е. определить приблизительную разницу между самой большой и самой малой вариантой. В данном случае самая большая варианта, которую мы обозначим символом $max\ V$ (максимум $V$), оказалась равной 19,4, а самая малая, $min\ V$ (минимум $V$), 3,2.

$$max\ V = 19{,}4$$
$$min\ V = 3{,}2$$

Следовательно отдельные варианты нашего ряда по своей величине колеблются здесь в пределах от 19,4 до 3,2. Эти крайние пределы колебаний мы можем округлить до ближайших к ним целых или хотя бы дробных с половинами чисел, но так, чтобы промежуток между ними во всяком случае не уменьшился. Для этого самое большое значение варианты (19,4) при округлении мы еще увеличим до 19,5, а самое малое значение (3,2), наоборот, уменьшим до 3.

Тогда приблизительно будем иметь следующие округленные пределы:

$$max\ V \text{ прибл.} = 19{,}5$$
$$min\ V \text{ прибл.} = 3.$$



Приблизительная же величина размаха колебаний всех наших варианта будет равна, очевидно, разности этих округленных пределов, т.-е.

$$19{,}5 - 3 = 16{,}5$$

Все эти предварительные вычисления можно записать так:

$$max\ V = 19{,}4 \text{ или прибл.} = 19{,}5$$
$$min\ V = 3{,}2 \text{ или прибл.} = 3{,}0$$
$$\text{разница прибл.} = 16{,}5$$

Определив величину наибольшего размаха колебаний отдельных вариант нашего ряда, следует решить вопрос о том, на какое же число классов желательно в данном случае разбить этот ряд? Тут общих правил не существует. Чем большее количество классов мы установим, тем точнее будут выполнены вычисления, но зато тем кропотливее будет и наша работа. Во всяком случае нецелесообразно разбивать вариационный ряд менее, чем на 8—10 классов. Наибольшее их количество обыкновенно бывает 20—25.

Предположим, что в данном случае мы остановились на 10 классах. Тогда весь промежуток между наибольшей и наименьшей вариантой (т.-е. только-что вычисленная разность 16,5) разобьется на 10 равных частей. Разделив 16,5 на 10, получим 1,65. Это и есть приблизительная величина классового промежутка $k$ при разбивке ряда на 10 классов. Так как нам не обязательно разбивать этот ряд непременно на 10 классов (мы можем установить и 11 или 9 классов), то полезно будет округлить полученную приближенную величину классового промежутка (1,65), приняв ее, напр., равной 1,5. Итак, у нас

$$k = 1{,}5$$

Теперь следует наметить границы классов, начиная с их верхних пределов. Так как наибольшая варианта $max\ V$ приблиз. $= 19{,}5$, а классовый промежуток $k = 1{,}5$, то, вычитая последовательно из 19,5 несколько раз под-ряд эту величину 1,5, получим все верхние границы наших классов. Нижние же их границы найдем, прибавляя к верхним границам следующих классов величину 0,1, так как отдельные варианты в нашем примере отличаются друг от друга минимально на эту величину 0,1.

Если пронумеровать все классы, начиная от больших к меньшим (т.-е. сверху вниз), то получим следующую картину.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Класс | № 1 | заключает в себе варианты | от | 19,5 | до | 18,1 | включительно |
| „ | № 2 | „ „ „ „ | „ | 18 | — | 16,6 | „ |
| „ | № 3 | „ „ „ „ | „ | 16,5 | — | 15,1 | „ |
| „ | № 4 | „ „ „ „ | „ | 15 | — | 13,6 | „ |
| „ | № 5 | „ „ „ „ | „ | 13,5 | — | 12,1 | „ |
| „ | № 6 | „ „ „ „ | „ | 12, | — | 10,6 | „ |
| „ | № 7 | „ „ „ „ | „ | 10,5 | — | 9,1 | „ |
| „ | № 8 | „ „ „ „ | „ | 9 | — | 7,6 | „ |
| „ | № 9 | „ „ „ „ | „ | 7,5 | — | 6,1 | „ |
| „ | № 10 | „ „ „ „ | „ | 6 | — | 4,6 | „ |
| „ | № 11 | „ „ „ „ | „ | 4,5 | — | 3,1 | „ |

Здесь №№ классов проставлены лишь временно, для удобства описания способа распределения отдельных вариант по этим классам. В действительности же без этих №№ можно, очевидно, и обойтись.

Установив таким образом границы всех классов (которых в данном случае оказалось 11), следует выяснить, сколько же отдельных вариант нашего ряда войдет в каждый из этих классов?

Для этого удобнее всего воспользоваться особой „разносной решеткой", придав ей форму, указанную на табл. 36.

Начертив такую решетку, следует пометить в ней каким-либо условным значком (напр., вертикальной черточкой) каждую отдельную варианту в строчке того класса, в границах которого она окажется. Так, напр., первая в нашем списке варианта 12,2 должна быть отнесена к классу № 5, так как этот класс заключает в себе все варианты, меньшие 13,5 и бóльшие 12,1, а наша варианта (12,2) как раз и удовлетворяет этому условию. Поэтому в широком столбце разносной решетки, озаглавленном словом „пометки", мы поставим одну черточку против границ класса № 5 (см. табл. 36). Этот условный значок указывает, следовательно, на то, что в числе отдельных вариант нашего ряда имеется одна такая, величина которой заключается в пределах от 13,5 до 12,1.

Следующая по списку варианта 9,9 должна быть отнесена к классу № 7, так как класс этот заключает в себе



ТАБЛИЦА 36.

| №№ классов | Границы классов (от — до) | Пометки | Частоты p |
|---|---|---|---|
| 1 | 19,5 — 18,1 | | |
| 2 | 18 — 16,6 | | |
| 3 | 16,5 — 15,1 | | |
| 4 | 15 — 13,6 | | |
| 5 | 13,5 — 12,1 | / | |
| 6 | 12 — 10,6 | | |
| 7 | 10,5 — 9,1 | / | |
| 8 | 9 — 7,6 | | |
| 9 | 7,5 — 6,1 | | |
| 10 | 6 — 4,6 | / | |
| 11 | 4,5 — 3,1 | | |

варианты от 10,5 до 9,1, т.-е. в том числе и нашу варианту 9,9. В соответствующей строчке разносной решетки мы и поставим одну черточку.

ТАБЛИЦА 37.

| Границы классов (от — до) | Пометки | Частоты p |
|---|---|---|
| 19,5 — 18,1 | // | 2 |
| 18 — 16,6 | | 0 |
| 16,5 — 15,1 | /// | 3 |
| 15 — 13,6 | ///// | 5 |
| 13,5 — 12,1 | /////// | 7 |
| 12 — 10,6 | ///////// | 9 |
| 10,5 — 9,1 | //////// | 8 |
| 9 — 7,6 | ////// | 6 |
| 7,5 — 6,1 | //// | 4 |
| 6 — 4,6 | /// | 3 |
| 4,5 — 3,1 | / | 1 |

Третью по списку варианту 6 по аналогичным соображениям мы должны будем отнести к классу № 10 (от 6 до 4,6) и в строчке этого класса поставить также одну черточку (см. табл. 36).

Поступая так же со всеми 48 вариантами заданного ряда, в конечном итоге получим 48 черточек, разнесенных по отдельным строчкам нашей решетки (см. табл. 37). Количество этих черточек в отдельных строчках укажет нам на число вариант, вошедших в каждый из 11 классов. Подсчитав это количество черточек в каждой строчке *), мы запишем результаты такого подсчета в крайний правый столбец, озаглавленный буквой $p$. Числа этого столбца и будут служить частотами нашего вариационного ряда.

ТАБЛИЦА 38.

| | $a_1$ | $a_2$ |
|---|---|---|
| $p$ | 36 | 34 |
| 2 | 2 | 2 |
| 0 | 2 | 4 |
| 3 | 5 | 9 |
| 5 | 10 | 19 |
| 7 | 17 | — |
| 9 | — | — |
| 8 | 22 | - |
| 6 | 14 | 27 |
| 4 | 8 | 13 |
| 3 | 4 | 5 |
| 1 | 1 | 1 |
| 48 | 49 | 46 |
| $n$ | $b_1$ | $b_2$ |

Получив частоты $p$ заданного вариационного ряда, составим обычного типа расчетную решетку (см. табл. 38), запишем в нее эти частоты и где-либо посередине решетки расположим фигуру из 4-х черточек. Так как черточки эти разместились здесь на уровне частоты 9, соответствующей классу с границами от 12 до 10,6 (см. табл. 38 и 37), то середину этого класса мы и должны будем принять за начало отсчетов $C$. Величину $C$ определим, как полусумму границ этого класса:

$$C = \frac{12 + 10{,}6}{2} = \frac{22{,}6}{2} = 11{,}3.$$

После обычных вычислений получим итоги:

$$a_1 = 36 \qquad a_2 = 34$$
$$b_1 = 49 \qquad b_2 = 46.$$

*) Для удобства подсчета этих черточек полезно располагать их группами по пяткам. Для этого можно, например, над каждыми четырьмя вертикальными черточками проводить одну горизонтальную.



Следовательно:

$$S_1 = a_1 - b_1 = 36 - 49 = -13$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 = 36 + 49 + 2 \cdot 34 + 2 \cdot 46 =$$
$$= 36 + 49 + 68 + 92 = 245$$

откуда:

$$\Sigma x^2 = S_2 - \frac{S_1^2}{n} = 245 - \frac{(-13)^2}{48} = 245 - \frac{169}{48} = 245 - 3{,}52 = 241{,}48$$

При

$$C = 11{,}3 \quad \text{и} \quad k = 1{,}5$$
$$M = C + \frac{kS_1}{n} = 11{,}3 + \frac{1{,}5(-13)}{48} = 11{,}3 - \frac{19{,}5}{48} =$$
$$= 11{,}3 - 0{,}41 = 10{,}89$$
$$\sigma = \pm k\sqrt{\frac{\Sigma x^2}{n-1}} = \pm 1{,}5\sqrt{\frac{241{,}48}{48-1}} = \pm 1{,}5\sqrt{\frac{241{,}48}{47}} =$$
$$= \pm 1{,}5\sqrt{5{,}14}.$$

Так как $\sqrt{5{,}14} = 2{,}26716$, то $\sigma = \pm 1{,}5 \cdot 2{,}26716 = \pm 3{,}40074$ или, округляя, $\sigma = \pm 3{,}40$

$$m = \frac{\pm\sigma}{\sqrt{n}} = \frac{\pm 3{,}40}{\sqrt{48}}.$$ Так как $\sqrt{48}$ прибл. $= 6{,}93$, то

$$m = \frac{\pm 3{,}40}{6{,}93} = \pm\frac{340}{693} = \pm 0{,}49.$$

Итак, для заданного ряда:

$$M = 10{,}89 \pm 0{,}49, \quad \text{а} \quad \sigma = \pm 3{,}40.$$

Сделаем же сводку правил, определяющих порядок вычисления $M$, $\sigma$ и $m$ по способу сумм в общем случае, когда заданный вариационный ряд состоит из большого количества членов и отдельные варианты представляют собой не только целые, но и дробные числа. Порядок этот, следовательно, таков:

1. Найти $max\ V$ и $min\ V$, округлить их и путем вычитания $min\ V$ из $max\ V$ определить приблизительную величину наибольшего размаха колебаний отдельных вариант заданного ряда.

2. Произвольно задаться числом классов (от 8 до 25) и определить приблизительную величину классового промежутка $k$ путем деления разницы между $max\ V$ и $min\ V$ на принятое число классов. Эту приблизительную величину

классового промежутка полезно также округлить (или до целого числа, или до ближайшего дробного с половиной).

3. Определить верхние границы всех классов путем последовательного вычитания несколько раз под ряд классового промежутка $k$ из наибольшей варианты $max\ V$.

4. Определить нижние границы этих классов путем прибавления к верхним границам соседних (меньших) классов той величины, на которую минимально отличаются друг от друга отдельные варианты заданного ряда.

5. Составить разносную решетку и путем постановки черточек распределить в ней все варианты по отдельным классам *).

6. Определить частоты $p$ отдельных классов путем подсчета черточек в каждой отдельной строчке разносной решетки и сложить эти частоты (получим общее количество измерений $n$).

7. Составить „расчетную решетку" обычного типа и, произведя в ней все необходимые действия, получить итоги: $a_1$, $a_2$, $b_1$ и $b_2$.

8. С помощью этих итогов определить обе вспомогательные суммы $S_1$ и $S_2$ по обычным формулам:

$$S_1 = a_1 - b_1$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2.$$

9. Определить начало отсчетов $C$, как полусумму границ того класса, против частоты которого $p$ в расчетной решетке расположена фигура из 4-х черточек.

10. Вычислить среднее арифметическое $M$ по формуле:

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

11. Вычислить сумму квадратов центральных отклонений $\Sigma x^2$ по формуле:

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

*) При разноске по классам большого количества вариант полезно воспользоваться специальными карточками, распределяемыми по отделениям особого раскладочного ящика. Такие карточки и ящики имеются в продаже на складе кооперативного товарищества „Современник" (Ленинград, пр. Володарского, 27).



12. Вычислить среднее квадратическое отклонение $\sigma$ по формуле:

$$\sigma = \pm \sqrt{\frac{\Sigma x^2}{n-1}}.$$

13. Вычислить среднюю ошибку $m$ по формуле:

$$m = \frac{\pm\sigma}{\sqrt{n}} \text{ или } m = \pm k \sqrt{\frac{\Sigma x^2}{n(n-1)}} \text{ (помимо } \sigma\text{).}$$

Величины $M$, $\sigma$ и $m$, вычисленные для вариационного ряда, предварительно разбитого на классы, являются, очевидно, не вполне точными. В самом деле, при обработке такого ряда действительные значения отдельных вариант заменяются здесь средними значениями соответствующих классов и потому являются приближенными и, конечно, неточными числами. Поэтому и $M$, $\sigma$ и $m$, вычисленные для таких приближенных вариант, будут неизбежно содержать в себе некоторую ошибку и притом тем большую, чем грубее было сделано объединение отдельных вариант в классе. Поэтому степень точности $M$, $\sigma$ и $m$ будет зависеть от величины классового промежутка $k$. Чем меньше $k$ (и чем больше, следовательно, общее количество всех классов), тем точнее будут и величины $M$, $\sigma$ и $m$.

Итак, в отдельных случаях мы имеем возможность пользоваться тремя основными способами вычисления $M$, $\sigma$ и $m$ в зависимости от состава и характера подлежащих обработке вариационных рядов: способом непосредственного вычисления, способом округления и способом сумм (способ произведений в счет не идет, так как способ сумм его вполне заменяет *).

Если вариационный ряд состоит из очень небольшого количества вариант (напр., не более 5 или 6), или когда среднее арифметическое $M$ случайно оказывается целым числом, можно применять способ непосредственного вычисления.

Когда вариационный ряд состоит из небольшого количества вариант (напр., 20—30) и среднее арифметическое $M$

*) Способ произведений является незаменимым лишь в том случае, когда вариационный ряд приходится разбивать на классы с неодинаковым классовым промежутком.

оказывается числом дробным, полезно воспользоваться способом округления.

Если же вариационный ряд состоит из большого количества вариант (больше 30—40) и притом варианты эти оказываются не только целыми, но и дробными числами, приходится разбивать такой ряд на классы и обрабатывать его по способу сумм. При этом желательно разбивать такой ряд на возможно большее количество классов, что при пользовании способом сумм не может особенно усложнить технику вычислений.

## XII. Одновременная обработка нескольких вариационных рядов со сводкой.

В тех случаях, когда $M$, $\sigma$ и $m$ приходится определять не для одного, а для нескольких вариационных рядов (что чаще всего и случается), чрезвычайно удобно бывает производить все промежуточные вычисления в одной общей расчетной решетке. При таком способе вычислений, как увидим дальше, получается возможность проверять все итоги ($a_1$, $a_2$, $b_1$, $b_2$) и обе вспомогательные суммы ($S_1$ и $S_2$), определяемые для каждого ряда в отдельности, с помощью очень несложного приема, сводящегося к простому сложению соответствующих чисел. В возможности такого контроля и заключается одно из главных преимуществ способа сумм.

Покажем применение этого способа проверки на числовом примере. Пусть имеем 3 вариационных ряда, состоящих из следующих вариант:

| I ряд: | | II ряд: | | | | III ряд: | |
|---|---|---|---|---|---|---|---|
| 7,5 | 9,5 | 13,5 | 11,0 | 12,5 | 13,5 | 19,5 | 19,0 |
| 3,5 | 5,5 | 7,0 | 7,5 | 16,0 | 15,0 | 22,0 | 20,5 |
| 5,5 | 2,5 | 14,5 | 21,5 | 14,5 | 12,5 | 18,5 | 21,5 |
| 8,0 | 6,0 | 11,5 | 11,0 | 15,0 | 14,0 | 24,0 | 21,0 |
| 6,5 | 4,5 | 18,0 | 10,5 | 17,5 | 15,5 | 23,5 | 20,5 |

Разобьем эти ряды на классы, установив для них о б щ и е границы с о д и н а к о в ы м классовым промежутком $k$. Для этого прежде всего необходимо отыскать наибольшую и наименьшую варианту ($max\ V$ и $min\ V$) вне зависимости от



того, в каком из трех рядов они окажутся. Просматривая варианты всех этих рядов, мы находим, что наибольшая из них 24 оказалась в III ряду, а наименьшая 2,5—в I ряду. Так как обе эти крайние варианты уже представляют собой сравнительно „круглые“ числа (без мелких десятых и сотых долей), то округлять их, очевидно, не следует. Итак,

$$max\ V = 24{,}0$$
$$min\ V = \ \ 2{,}5$$
$$\text{разница} = 22{,}5.$$

ТАБЛИЦА 39.

| Границы классов (от—до) | I ряд | | II ряд | | III ряд | |
|---|---|---|---|---|---|---|
| | Пометки | $p$ | Пометки | $p$ | Пометки | $p$ |
| 24 — 22,5 | | | | | // | 2 |
| 22 — 20,5 | | | / | 1 | ///// | 5 |
| 20 — 18,5 | | | | 0 | /// | 3 |
| 18 — 16,5 | | | // | 2 | | |
| 16 — 14,5 | | | ////// | 6 | | |
| 14 — 12,5 | | | ///// | 5 | | |
| 12 — 10,5 | | | //// | 4 | | |
| 10 — 8,5 | / | 1 | | 0 | | |
| 8 — 6,5 | /// | 3 | // | 2 | | |
| 6 — 4,5 | //// | 4 | | | | |
| 4 — 2,5 | // | 2 | | | | |

Этот наибольший размах колебаний всех наших вариант (22,5) мы можем разбить на произвольное число классов. Остановимся для простоты на числе классов, равном приблизительно 10 (в действительности, конечно, рекомендуется разбивать такие ряды на значительно большее количество классов). Тогда, разделив 22,5 на 10, найдем приблизительную величину классового промежутка $k$, которая в данном случае окажется равной 2,25. Округлив это значение до ближайшего целого числа, окончательно найдем, что

$$k = 2.$$

Так как наибольшая варианта $max\ V = 24$, то, вычитая последовательно величину $k = 2$ из числа 24, определим все верхние границы наших классов. Нижние же их границы нам придется приблизить к верхним границам соседних (меньших) классов на 0,5, так как отдельные варианты всех заданных рядов отличаются друг от друга минимально на эту величину 0,5. В результате получим границы всех классов, помещенные на табл. 39 в крайнем слева столбце разносной решетки. Решетка эта составлена здесь для рас-

ТАБЛИЦА 40.

| I ряд | | | II ряд | | | III ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *p* | | | *p* | | | *p* | | | *p* | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

пределения вариант по классам внутри каждого ряда отдельно.

После разноски с помощью черточек всех вариант каждого ряда по соответствующим строчкам этой решетки подсчитаем количество черточек, оказавшихся в каждом классе, и результаты запишем в столбце *p*. Это и будут частоты всех трех наших рядов.

Для обработки этих рядов по способу сумм составим общую расчетную решетку, придав ей форму, указанную на табл. 40.



Решетка эта представляет собой не что иное, как простое соединение четырех расчетных решеток обычного типа. Первые три части этой решетки предназначаются для отдельной обработки трех заданных вариационных рядов порознь, а последняя, озаглавленная словом „сводка“, — для обработки нового (4-го) ряда, составленного из всех 40 вариант, независимо от того, к которому из отдельных рядов они принадлежат. Из данных этой сводки можно тоже вычислить $M$, $\sigma$ и $m$, относящиеся уже ко всем 40 вариан-

ТАБЛИЦА 41.

| Границы классов | I ряд | | | II ряд | | | III ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *p* | | | *p* | | | *p* | | | *p* | | |
| 24 — 22,5 | | | | | | | 2 | | | | | |
| 22 — 20,5 | | | | 1 | | | 5 | | | | | |
| 20 — 18,5 | | | | 0 | | | 3 | | | | | |
| 18 — 16,5 | | | | 2 | | | | | | | | |
| 16 — 14,5 | | | | 6 | | | | | | | | |
| 14 — 12,5 | | | | 5 | | | | | | | | |
| 12 — 10,5 | | | | 4 | | | | | | | | |
| 10 — 8,5 | 1 | | | 0 | | | | | | | | |
| 8 — 6,5 | 3 | | | 2 | | | | | | | | |
| 6 — 4,5 | 4 | | | | | | | | | | | |
| 4 — 2,5 | 2 | | | | | | | | | | | |
| | | | | | | | | | | | | |

там в целом, и кроме того сводка эта, как увидим сейчас, будет служить средством проверки всех промежуточных вычислений, производимых внутри первых трех частей этой решетки.

В составленную таким образом комбинированную расчетную решетку следует записать частоты всех трех рядов, заполнив ими вертикальные столбцы, озаглавленные буквами *p*. Для удобства разноски частот по горизонтальным строчкам пометим временно на полях решетки границы всех 11 классов. После этих записей наша решетка примет вид, указанный на таблице 41.

Далее следует заполнить клетки столбца $p$ в последней (четвертой) части решетки, озаглавленной словом „сводка“, путем сложения всех частот в трех рядах по горизонтальным строчкам и затем подытожить все четыре столбца $p$. В результате получим количество измерений ($n$) в каждом из трех рядов и общее количество всех измерений (40) в сводке. Очевидно, сумма всех трех чисел $n$, получаемая

ТАБЛИЦА 42.

| Границы классов | I ряд | | | II ряд | | | III ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $p$ | | | $p$ | | | $p$ | | | $p$ | | |
| 24 — 22,5 | | | | | | | 2 | | | 2 | | |
| 22 — 20,5 | | | | 1 | | | 5 | | | 6 | | |
| 20 — 18,5 | | | | 0 | | | 3 | | | 3 | | |
| 18 — 16,5 | | | | 2 | | | | | | 2 | | |
| 16 — 14,5 | | | | 6 | | | | | | 6 | | |
| 14 — 12,5 | | | | 5 | | | | | | 5 | | |
| 12 — 10,5 | | | | 4 | | | | | | 4 | | |
| 10 — 8,5 | 1 | | | 0 | | | | | | 1 | | |
| 8 — 6,5 | 3 | | | 2 | | | | | | 5 | | |
| 6 — 4,5 | 4 | | | | | | | | | 4 | | |
| 4 — 2,5 | 2 | | | | | | | | | 2 | | |
| | 10 | | | 20 | | | 10 | | | 40 | | |

в первых трех частях решетки, должна равняться этому же числу $n$ в сводке. В самом деле:

$$10 + 20 + 10 = 40.$$

После всех этих подсчетов и записей наша решетка примет вид, указанный на таблице 42.

После разноски и подытоживания всех частот следует в каждой части этой решетки расположить фигуру из 4-х черточек и притом во всех трех рядах и в сводке непременно на одном и том же общем уровне, как это показано на табл. 43.

Высота расположения этих черточек произвольна. Их можно было бы поместить как выше, так и ниже, но удоб-



нее располагать их где-либо посередине решетки, так как при этом условии значительно упрощаются все вычисления (уменьшаются обе вспомогательные суммы $S_1$ и $S_2$).

Так как черточки эти мы расположили здесь против той строчки, которая соответствует классу с границами от 14 до 12, 5, то полусумму этих границ мы и должны будем принять за начало отсчетов $C$.

ТАБЛИЦА 43.

| | I ряд | | | II ряд | | | III ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $p$ | | | $p$ | | | $p$ | | | $p$ | | |
| 24 — 22,5 | | | | | | | 2 | | | 2 | | |
| 22 — 20,5 | | | | 1 | | | 5 | | | 6 | | |
| 20 — 18,5 | | | | 0 | | | 3 | | | 3 | | |
| 18 — 16,5 | | | | 2 | | | | | | 2 | | |
| 16 — 14,5 | | | — | 6 | | — | | | — | 6 | | — |
| 14 — 12,5 | | — | — | 5 | — | — | | — | — | 5 | — | — |
| 12 — 10,5 | | | — | 4 | | — | | | — | 4 | | — |
| 10 — 8,5 | 1 | | | 0 | | | | | | 1 | | |
| 8 — 6,5 | 3 | | | 2 | | | | | | 5 | | |
| 6 — 4,5 | 4 | | | | | | | | | 4 | | |
| 4 — 2,5 | 2 | | | | | | | | | 2 | | |
| | 10 | | | 20 | | | 10 | | | 40 | | |

Следовательно при данном расположении черточек

$$C = \frac{14 + 12{,}5}{2} = \frac{26{,}5}{2} = 13{,}25.$$

Итак, для всех трех рядов и для сводки будем иметь одинаковую величину начала отсчетов $C$ и одинаковый же классовый промежуток $k$, а именно:

$$C = 13{,}25$$
$$k = 2.$$

После того, как в расчетной решетке будут начерчены все фигуры из 4-х черточек, следует в неполных столбцах $p$ прибавить столько нулей, сколько это понадобится для заполнения ими столбцов до середины черточек. В данном случае для этого придется прибавить два нуля сверху к частотам I ряда и три нуля снизу к частотам III ряда. В результате получим картину, представленную на таблице 44.

Здесь расчетная решетка изображена уже без добавочных обозначений классовых границ, которые ранее были записаны для памяти слева на полях решетки. Надобность в обозначении этих границ отпала с того момента, когда

ТАБЛИЦА 44.

| I ряд | | | II ряд | | | III ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | | | $p$ | | | $p$ | | | $p$ | | |
| | | | | | | 2 | | | 2 | | |
| | | | 1 | | | 5 | | | 6 | | |
| | | | 0 | | | 3 | | | 3 | | |
| | | | 2 | | | 0 | | | 2 | | |
| | | | 6 | | — | 0 | | - | 6 | | |
| 0 | — | — | 5 | — | — | 0 | — | — | 5 | — | — |
| 0 | | — | 4 | | — | | | | 4 | | — |
| 1 | | | 0 | | | | | | 1 | | |
| 3 | | | 2 | | | | | | 5 | | |
| 4 | | | | | | | | | 4 | | |
| 2 | | | | | | | | | 2 | | |
| 10 | | | 20 | | | 10 | | | 40 | | |

было установлено начало отсчетов $C$ и классовый промежуток $k$.

После этих подготовительных операций следует обычным приемом продолжить крайние частоты каждого ряда и сводки до конца соответствующей части расчетной решетки и заполнить оставшиеся свободными клетки в верхней ее части путем сложения верхних чисел с левыми, а в нижней — нижних чисел с левыми.

После сложения всех чисел в верхних и нижних частях решетки отдельно по средним и правым столбцам получим для каждого ряда и сводки итоги: $a_1$, $a_2$, $b_1$ и $b_2$. Итоги эти мы запишем в обычных местах, после чего наша решетка примет вид, указанный на таблице 45.



В этом этапе работы мы имеем возможность сделать первую проверку произведенных вычислений. Для этого стоит лишь сложить одноименные итоги в первых трех частях расчетной решетки и убедиться в том, что сумма их будет в точности равняться соответствующему итогу сводки.

Проделаем эту проверку для всех четырех итогов $a_1$, $a_2$, $b_1$ и $b_2$ отдельно.

ТАБЛИЦА 45.

| I ряд | | | II ряд | | | III ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0 | 0 | $p$ | 14 | 8 | $p$ | 39 | 59 | $p$ | 53 | 67 |
| | | | | | | 2 | 2 | 2 | 2 | 2 | 2 |
| | | | 1 | 1 | 1 | 5 | 7 | 9 | 6 | 8 | 10 |
| | | | 0 | 1 | 2 | 3 | 10 | 19 | 3 | 11 | 21 |
| | | | 2 | 3 | 5 | 0 | 10 | 29 | 2 | 13 | 34 |
| | | — | 6 | 9 | — | 0 | 10 | — | 6 | 19 | — |
| 0 | — | — | 5 | — | — | 0 | — | — | 5 | — | — |
| 0 | 10 | — | 4 | 6 | — | | | — | 4 | 16 | |
| 1 | 10 | 27 | 0 | 2 | 4 | | | | 1 | 12 | 31 |
| 3 | 9 | 17 | 2 | 2 | 2 | | | | 5 | 11 | 19 |
| 4 | 6 | 8 | | | | | | | 4 | 6 | 8 |
| 2 | 2 | 2 | | | | | | | 2 | 2 | 2 |
| 10 | 37 | 54 | 20 | 10 | 6 | 10 | 0 | 0 | 40 | 47 | 60 |

Итоги $a_1$ в различных частях нашей решетки оказались следующими:

Для I ряда $a_1 = 0$
„ II „ $a_1 = 14$ Для сводки $a_1 = 53$.
„ III „ $a_1 = 39$

Сумма их $= 53$

Итоги $a_2$ оказались следующими:

Для I ряда $a_2 = 0$
„ II „ $a_2 = 8$ Для сводки $a_2 = 67$.
„ III „ $a_2 = 59$

Сумма их $= 67$

Итоги $b_1$ получились следующие:

| | | |
|---|---|---|
| Для I ряда | $b_1 = 37$ | |
| „ II „ | $b_1 = 10$ | Для сводки $b_1 = 47$. |
| „ III „ | $b_1 = 0$ | |
| Сумма их | $= 47$ | |

Итоги $b_2$ оказались равными:

| | | |
|---|---|---|
| Для I ряда | $b_2 = 54$ | |
| „ II „ | $b_2 = 6$ | Для сводки $b_2 = 60$. |
| „ III „ | $b_2 = 0$ | |
| Сумма их | $= 60$ | |

Убедившись в том, что суммы эти совпадают, мы можем считать все выполненные вычисления правильными; в противном случае следовало бы искать ошибку в промежуточных выкладках.

После проверки итогов $a_1$, $a_2$, $b_1$ и $b_2$ следует для каждого ряда и для сводки вычислить обе вспомогательные суммы $S_1$ и $S_2$ по обычным формулам:

$$S_1 = a_1 - b_1$$
$$S_2 = a_1 + b_1 + 2a_2 + 2b_2.$$

В результате этих вычислений получим:

| Для I ряда: | Для II ряда: | Для III ряда: | Для сводки: |
|---|---|---|---|
| $S_1 = -\ 37$ | $S_1 = +\ 4$ | $S_1 = +\ 39$ | $S_1 = +\ 6$ |
| $S_2 = 145$ | $S_2 = 52$ | $S_2 = 157$ | $S_2 = 354$ |

После определения вспомогательных сумм $S_1$ и $S_2$ можно сделать вторую проверку правильности произведенных вычислений. Проверка эта, так же, как и первая, заключается в сложении одноименных значений $S$ в первых трех частях расчетной решетки, в результате чего сумма этих значений должна оказаться в точности равной соответствующему значению $S$ для сводки.

В нашем примере мы имеем следующие значения $S_1$:

| | | |
|---|---|---|
| Для I ряда | $S_1 = -37$ | |
| „ II „ | $S_1 = +\ 4$ | Для сводки $S_1 = +6$. |
| „ III „ | $S_1 = +39$ | |
| Сумма их | $= +\ 6$ | |



Значения $S_2$ нами были получены следующие:

| | | |
|---|---|---|
| Для I ряда | $S_2 = 145$ | |
| „ II „ | $S_2 = 52$ | Для сводки $S_2 = 354$. |
| „ III „ | $S_2 = 157$ | |
| Сумма их | $= 354$ | |

Следовательно все вычисления были произведены правильно.

После вычисления и проверки обеих вспомогательных сумм $S_1$ и $S_2$ следует для каждого из трех заданных рядов отдельно определить сумму квадратов центральных отклонений $\Sigma x^2$ по обычной формуле:

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

После соответствующих вычислений получим:

| | |
|---|---|
| Для I ряда | $\Sigma x^2 = 8{,}1$ |
| „ II „ | $\Sigma x^2 = 51{,}2$ |
| „ III „ | $\Sigma x^2 = 4{,}9.$ |

Заметим, что если бы мы таким же образом вычислили $\Sigma x^2$ и для сводки, то в данном случае сумма этих величин, взятая для отдельных рядов, уже не сошлась бы с величиной $\Sigma x^2$, вычисленной для сводки, как это было в применении к итогам $a_1$, $b_1$, $a_2$ и $b_2$, а также в применении к вспомогательным суммам $S_1$ и $S_2$. Таким образом суммы квадратов центральных отклонений $\Sigma x^2$ для проверки вычислений этим способом служить не могут.

Далее, для каждого из трех рядов отдельно следует вычислить $M$, $\sigma$ и $m$ по обычным формулам:

$$M = C + \frac{kS_1}{n};\quad \sigma = \pm k\sqrt{\frac{\Sigma x^2}{n-1}} \text{ и } m = \pm\frac{\sigma}{\sqrt{n}}$$

$$\text{или } m = \pm k\sqrt{\frac{\Sigma x^2}{n(n-1)}}.$$

Так как в нашем примере $C = 13{,}25$, а $k = 2$, то

| | | | |
|---|---|---|---|
| Для I ряда: | $M = 5{,}85$ | $\sigma = \pm 1{,}90$ | $m = \pm 0{,}60$ |
| „ II „ | $M = 13{,}29$ | $\sigma = \pm 3{,}28$ | $m = \pm 0{,}73$ |
| „ III „ | $M = 21{,}05$ | $\sigma = \pm 1{,}47$ | $m = \pm 0{,}47.$ |

Эти окончательные результаты вычислений, равно как и промежуточные величины $S_1$, $S_2$ и $\Sigma x^2$, мы можем записать в нижних частях расчетной решетки, которая после этого примет окончательный вид, указанный на таблице 46.

ТАБЛИЦА 46 *).

| I ряд | | | II ряд | | | III ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0 | 0 | $p$ | 14 | 8 | $p$ | 39 | 59 | $p$ | 53 | 67 |
| | | | | | | 2 | 2 | 2 | 2 | 2 | 2 |
| | | | 1 | 1 | 1 | 5 | 7 | 9 | 6 | 8 | 10 |
| | | | 0 | 1 | 2 | 3 | 10 | 19 | 3 | 11 | 21 |
| | | | 2 | 3 | 5 | 0 | 10 | 29 | 2 | 13 | 34 |
| | | — | 6 | 9 | — | 0 | 10 | — | 6 | 19 | — |
| 0 | — | — | 5 | — | — | 0 | — | — | 5 | — | — |
| 0 | 10 | — | 4 | 6 | — | | | — | 4 | 16 | — |
| 1 | 10 | 27 | 0 | 2 | 4 | | | | 1 | 12 | 31 |
| 3 | 9 | 17 | 2 | 2 | 2 | | | | 5 | 11 | 19 |
| 4 | 6 | 8 | | | | | | | 4 | 6 | 8 |
| 2 | 2 | 2 | | | | | | | 2 | 2 | 4 |
| 10 | 37 | 54 | 20 | 10 | 6 | 10 | 0 | 0 | 40 | 47 | 60 |
| $S_1 = -\ 37$ | | | $S_1 = +\ 4$ | | | $S_1 = +\ 39$ | | | $S_1 = +\ 6$ | | |
| $S_2 = 145$ | | | $S_2 = 52$ | | | $S_2 = 157$ | | | $S_2 = 354$ | | |
| $\Sigma x^2 = 8{,}1$ | | | $\Sigma x^2 = 51{,}2$ | | | $\Sigma x^2 = 4{,}9$ | | | — | | |
| $M = 5{,}85$ | | | $M = 13{,}29$ | | | $M = 21{,}05$ | | | | | |
| $\sigma = \pm 1{,}90$ | | | $\sigma = \pm 3{,}28$ | | | $\sigma = \pm 1{,}47$ | | | | | |
| $m = \pm 0{,}60$ | | | $m = \pm 0{,}73$ | | | $m = \pm 0{,}47$ | | | | | |

*) При большом количестве вычислений полезно заготовить такие расчетные решетки типографским способом. Готовые решетки с формулами (форма № 1) имеются в книжном складе кооперативного товарищества „Современник" в Ленинграде (пр. Володарского. 27).



Если по самому характеру вопроса нам пришлось бы вычислить $M$, $\sigma$ и $m$ тоже для сводки, то при помощи тех же формул мы, очевидно, получили бы:

$$\Sigma x^2 = 353{,}1 \qquad M = 13{,}55 \qquad \sigma = \pm 6{,}02 \qquad m = \pm 0{,}95.$$

Итак, при одновременной обработке нескольких вариационных рядов, разбитых на классы с одинаковым классовым промежутком $k$ и с одинаковыми же границами этих классов, мы должны последовательно выполнить следующие отдельные операции:

1. Составить общую расчетную решетку, занести в нее частоты всех рядов, составить частоты сводки, просуммировать все столбцы $p$ (получим общее количество измерений $n$ в каждом ряду и в сводке) и убедиться в том, что сумма всех чисел $n$ для отдельных рядов совпадает с числом $n$ для сводки.

2. Где-либо по середине расчетной решетки на одном общем уровне для всех рядов и сводки начертить фигуры из 4-х черточек и выяснить значение начала отсчетов $C$, определив его как полусумму границ того класса, против частоты которого проставлены эти черточки.

3. Проделать все вычисления, необходимые для получения итогов: $a_1$, $a_2$, $b_1$ и $b_2$.

4. Проверить правильность произведенных вычислений путем сложения одноименных итогов для всех рядов, в результате чего суммы эти должны оказаться равными соответствующим итогам сводки.

5. Для каждого ряда и сводки вычислить вспомогательные суммы $S_1$ и $S_2$.

6. Проверить правильность их вычисления, убедившись в том, что сумма всех $S_1$ для отдельных рядов в точности равняется $S_1$ для сводки, а сумма всех $S_2$ для отдельных рядов равна $S_2$ для сводки.

7. Для всех рядов (а если нужно, то и для сводки) вычислить по обычным формулам $\Sigma x^2$, $M$, $\sigma$ и $m$.

В разобранном выше примере сводка могла иметь вполне реальное значение. Она представляла собой совокупность всех 40 вариант, разбитых на те же самые классы, на какие были разбиты и три заданные ряда, и отражала в себе общую картину распределения этих вариант по отдельным классам независимо от того, к какому из трех рядов они относятся. Поэтому в данном случае определение $M$, $\sigma$ и $m$ для сводки могло иметь вполне определенный смысл, а в некоторых случаях даже вызываться необходимостью.

Но совершенно иную картину мы получим в том случае, когда отдельные вариационные ряды будут разбиты на классы, не совпадающие по своей величине, границам и наименованию. При этих условиях сводка не будет иметь уже никакого реального значения, и надобность в получении для нее общих числовых характеристик $M$, $\sigma$ и $m$ совершенно отпадет (у сводки останутся тогда лишь контрольные функции).

Поясним эту разницу на каком-либо конкретном примере. Пусть, в первом случае, нас интересует рост детей различного возраста. После разбивки всего материала на одинаковые по своей величине и границам классы (в *см* или их долях) мы получим несколько вариационных рядов, характеризующих распределение детей каждого отдельного возраста по установленным таким образом классам. Сводка же, составленная из этих данных, будет, очевидно, указывать общее распределение всех измеренных детей по этим же самым классам независимо от возраста отдельных детей. В конечном итоге мы получим несколько комплектов числовых характеристик $M$, $\sigma$ и $m$ для каждого возраста отдельно и один такой же комплект этих величин, общий для всей группы детей в целом.

Но мыслим и такой случай: допустим, что мы измеряем отдельно рост, вес, емкость легких, мышечную силу, число ударов пульса в секунду и т. д. Для каждого из этих признаков мы должны будем установить свою собственную скалу распределения по классам и указать их границы уже в различных единицах измерения (в *см*, *кг*, *куб. см* и т. д.). В результате разбивки этих данных по классам мы также получим несколько вариационных рядов, но ряды эти окажутся, очевидно, уже не однородными, хотя и относящимися к одним и тем же объектам исследования. При одновременной обработке подобных вариационных рядов мы также имеем возможность составить сводку из их частот $p$, но сводкой этой придется воспользоваться уже только



в целях проверки промежуточных вычислений, определять же для нее $M$, $\sigma$ и $m$, очевидно, не нужно.

Покажем на числовом примере, как следует поступать в подобных случаях.

Пусть требуется обработать одновременно два следующих вариационных ряда, разбитых уже на классы:

| | | | | | | |
|---|---|---|---|---|---|---|
| I ряд. | От 100 | до | 96 | заключается | 3 | варианты |
| | „ 95 | „ | 91 | „ | 8 | „ |
| | „ 90 | „ | 86 | „ | 22 | „ |
| | „ 85 | „ | 81 | „ | 31 | „ |
| | „ 80 | „ | 76 | „ | 18 | „ |
| | „ 75 | „ | 71 | „ | 6 | „ |
| | „ 70 | „ | 66 | „ | 2 | „ |
| | „ 65 | „ | 61 | „ | 2 | „ |
| II ряд. | От 3 | до | 2,51 | заключается | 2 | варианты |
| | „ 2,50 | „ | 2,01 | „ | 5 | „ |
| | „ 2 | „ | 1,51 | „ | 12 | „ |
| | „ 1,50 | „ | 1,01 | „ | 10 | „ |
| | „ 1 | „ | 0,51 | „ | 7 | „ |
| | „ 0,50 | „ | 0,01 | „ | 4 | „ |

Прежде всего установим в каждом ряду отдельно те классы, против соответствующих частот которых в расчетной решетке мы расположим фигуру из 4-х черточек. Эти классы мы выберем с тем расчетом, чтобы они оказались приблизительно в середине ряда. В данном случае для I ряда удобно, напр., остановиться на классе с границами от 85 до 81 с частотой $p = 31$, а для II ряда — на классе от 2 до 1,51 с частотой $p = 12$.

Установив эти исходные классы, мы сразу же можем вычислить и начало отсчетов $C$ для каждого ряда отдельно.

Для I ряда $C = \frac{85 + 81}{2} = \frac{166}{2} = 83.$

Для II ряда $C = \frac{2 + 1{,}51}{2} = \frac{3{,}51}{2} = 1{,}755$ или округляя 1,76.

Классовые промежутки $k$ здесь, очевидно, равны 5 и 0,5. Итак,

для I ряда $C = 83$ $k = 5$

„ II „ $C = 1{,}76$ $k = 0{,}5$

После этого составим общую расчетную решетку, проставим во всех трех ее частях фигуру из 4-х черточек на одном и том же уровне, а снизу для памяти запишем только-что вычисленные значения $C$ и $k$ для каждого ряда в отдельности. В этом виде наша решетка представлена на таблице 47.

Далее следует записать в эту решетку частоты обоих рядов, следя за тем, чтобы частоты выбранных ранее

ТАБЛИЦА 47.

| I ряд | | | II ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|
| $p$ | | | $p$ | | | $p$ | | |
| | | | | | | | | |
| | | | | | | | | |
| | | — | | | — | | | — |
| | — | — | | — | — | | — | — |
| | | — | | | — | | | — |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| $C=83$ $k=5$ <br> $S_1=$ <br> $S_2=$ <br> $\Sigma x^2=$ | | | $C=1{,}76$ $k=0{,}5$ <br> $S_1=$ <br> $S_2=$ <br> $\Sigma x^2=$ | | | — <br> $S_1=$ <br> $S_2=$ <br> — | | |
| $M=$ <br> $\sigma=\pm$ <br> $m=\pm$ | | | $M=$ <br> $\sigma=\pm$ <br> $m=\pm$ | | | | | |

средних классов (31 и 12) оказались как раз на уровне середины черточек. После этого нужно составить частоты сводки, подытожить числа во всех столбцах $p$ и проверить эти вычисления обычным способом сходимости чисел $n$ в двух рядах и в сводке. В результате получим картину, представленную на таблице 48.



В данном случае частоты сводки составлены правильно, так как число $n$ для I ряда (92) в сумме с числом $n$ для II ряда (40) в точности равняется числу $n$ для сводки (132).

После этих записей и проверки частот сводки следует заполнить свободные клетки во всех трех частях расчетной

ТАБЛИЦА 48.

| I ряд | | | II ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|
| $p$ | | | $p$ | | | $p$ | | |
| 3 | | | | | | 3 | | |
| 8 | | | 2 | | | 10 | | |
| 22 | | — | 5 | | — | 27 | | — |
| 31 | — | — | 12 | — | — | 43 | — | — |
| 18 | | — | 10 | | — | 28 | | — |
| 6 | | | 7 | | | 13 | | |
| 2 | | | 4 | | | 6 | | |
| 2 | | | | | | 2 | | |
| 92 | | | 40 | | | 132 | | |
| $C=83$ $k=5$ <br> $S_1=$ <br> $S_2=$ <br> $\Sigma x^2=$ | | | $C=1{,}76$ $k=0{,}5$ <br> $S_1=$ <br> $S_2=$ <br> $\Sigma x^2=$ | | | — <br> $S_1=$ <br> $S_2=$ <br> — | | |
| $M=$ <br> $\sigma=\pm$ <br> $m=\pm$ | | | $M=$ <br> $\sigma=\pm$ <br> $m=\pm$ | | | | | |

решетки по общим правилам и подвести итоги: $a_1$, $a_2$, $b_1$ и $b_2$.

В результате получим:

Для I ряда: $a_1=47$ $a_2=17$ $b_1=44$ $b_2=24$
" II " $a_1=9$ $a_2=2$ $b_1=36$ $b_2=19$
" сводки $a_1=56$ $a_2=19$ $b_1=80$ $b_2=43$

Все эти итоги запишем на обычных местах. Тогда расчетная решетка примет следующий вид (см. таблицу 49).

После определения итогов $a_1$, $a_2$, $b_1$ и $b_2$ необходимо их проверить, убедившись в том, что, сумма одноименных итогов в отдельных рядах в точности равна соответствующим итогам в сводке.

ТАБЛИЦА 49.

| I ряд | | | II ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|
| $p$ | 47 | 17 | $p$ | 9 | 2 | $p$ | 56 | 19 |
| 3 | 3 | 3 | | | | 3 | 3 | 3 |
| 8 | 11 | 14 | 2 | 2 | 2 | 10 | 13 | 16 |
| 22 | 33 | — | 5 | 7 | — | 27 | 40 | — |
| 31 | — | — | 12 | — | — | 43 | — | — |
| 18 | 28 | — | 10 | 21 | — | 28 | 49 | — |
| 6 | 10 | 16 | 7 | 11 | 15 | 13 | 21 | 31 |
| 2 | 4 | 6 | 4 | 4 | 4 | 6 | 8 | 10 |
| 2 | 2 | 2 | | | | 2 | 2 | 2 |
| 92 | 44 | 24 | 40 | 36 | 19 | 132 | 80 | 43 |
| $C=83$ $k=5$ | | | $C=1{,}76$ $k=0{,}5$ | | | — | | |
| $S_1=$ | | | $S_1=$ | | | $S_1=$ | | |
| $S_2=$ | | | $S_2=$ | | | $S_2=$ | | |
| $\Sigma x^2=$ | | | $\Sigma x^2=$ | | | — | | |
| $M=$ | | | $M=$ | | | | | |
| $\sigma=\pm$ | | | $\sigma=\pm$ | | | | | |
| $m=\pm$ | | | $m=\pm$ | | | | | |

Проверяя итоги $a_1$, получим:

Для I ряда $a_1=47$
„ II „ $a_1=9$ Для сводки $a_1=56$

Сумма их $=56$



Проверяя итоги $a_2$, получим:

Для I ряда $a_2=17$
„ II „ $a_2=2$ Для сводки $a_2=19$

Сумма их $=19$

Проверяя итоги $b_1$, получим:

Для I ряда $b_1=44$
„ II „ $b_1=36$ Для сводки $b_1=80$

Сумма их $=80$

Проверяя итоги $b_2$, получим:

Для I ряда $b_2=24$
„ II „ $b_2=19$ Для сводки $b_2=43$.

Сумма их $=43$

Следовательно все вычисления произведены правильно.

Теперь вычислим для каждого ряда и для сводки обе вспомогательные суммы $S_1$ и $S_2$. Суммы эти окажутся здесь равными:

| Для I ряда: | Для II ряда: | Для сводки: |
|---|---|---|
| $S_1=+3$ | $S_1=-27$ | $S_1=-24$ |
| $S_2=173$ | $S_2=87$ | $S_2=260$ |

Эти суммы необходимо также проверить.

Проверяя суммы $S_1$, получим:

Для I ряда $S_1=+3$
„ II „ $S_1=-27$ Для сводки $S_1=-24$.

Сумма их $=-24$

Проверяя суммы $S_2$, получим:

Для I ряда $S_2=173$
„ II „ $S_2=87$ Для сводки $S_2=260$.

Сумма их $=260$

Следовательно и вспомогательные суммы $S_1$ и $S_2$ вычислены правильно.

После определения сумм $S_1$ и $S_2$ вычислим для каждого ряда сумму квадратов центральных отклонений $\Sigma x^2$ (для сводки вычислять $\Sigma x^2$ уже не нужно). В результате получим для I ряда: $\Sigma x^2 = 172{,}90$; для II ряда: $\Sigma x^2 = 68{,}77$.

С помощью найденных величин $C$, $k$, $n$, $S_1$, $S_2$ и $\Sigma x^2$ можно определить и все числовые характеристики обоих рядов. После соответствующих вычислений получим:

| | | | |
|---|---|---|---|
| Для I ряда: | $M = 83{,}16$ | $\sigma = \pm 6{,}89$ | $m = \pm 0{,}72$ |
| „ II „ | $M = 1{,}42$ | $\sigma = \pm 0{,}66$ | $m = \pm 0{,}11$ |

Эти результаты мы можем записать в расчетную решетку против соответствующих буквенных обозначений, после чего решетка эта примет следующий окончательный вид (см. таблицу 50).

ТАБЛИЦА 50.

| I ряд | | | II ряд | | | Сводка | | |
|---|---|---|---|---|---|---|---|---|
| $p$ | 47 | 17 | $p$ | 9 | 2 | $p$ | 56 | 19 |
| 3 | 3 | 3 | | | | 3 | 3 | 3 |
| 8 | 11 | 14 | 2 | 2 | 2 | 10 | 13 | 16 |
| 22 | 33 | — | 5 | 7 | — | 27 | 40 | — |
| 31 | — | — | 12 | — | — | 43 | — | — |
| 18 | 28 | — | 10 | 21 | — | 28 | 49 | — |
| 6 | 10 | 16 | 7 | 11 | 15 | 13 | 21 | 31 |
| 2 | 4 | 6 | 4 | 4 | 4 | 6 | 8 | 10 |
| 2 | 2 | 2 | | | | 2 | 2 | 2 |
| 92 | 44 | 24 | 40 | 36 | 19 | 132 | 80 | 43 |
| $C = 83$<br>$K = 5$<br>$S_1 = +3$<br>$S_2 = 173$<br>$\Sigma x^2 = 172{,}90$ | | | $C = 1{,}76$<br>$K = 0{,}5$<br>$S_1 = -27$<br>$S_2 = 87$<br>$\Sigma x^2 = 68{,}77$ | | | —<br>—<br>$S_1 = -24$<br>$S_2 = 260$<br>— | | |
| $M = 83{,}16$<br>$\sigma = \pm 6{,}89$<br>$m = \pm 0{,}72$ | | | $M = 1{,}42$<br>$\sigma = \pm 0{,}66$<br>$m = \pm 0{,}11$ | | | —<br>—<br>— | | |



Таким образом, при одновременной обработке нескольких вариационных рядов, разбитых на классы, отличающиеся в каждом отдельном ряду по своим размерам, границам и наименованию, следует поступать так:

1. Установить для каждого ряда тот класс, частоту которого следует поместить в расчетной решетке против середины 4-х черточек, и определить для всех рядов начало отсчетов $C$, как полусумму границ этих средних классов.

2. Составить расчетную решетку и занести в нее частоты заданных рядов, следя за тем, чтобы частоты выбранных средних классов оказались как раз на уровне 4-х черточек.

3. Составить частоты сводки, проссумировать все столбцы $p$ (получим числа $n$) и проверить их по обычному правилу сходимости.

4. Проделать все вычисления, связанные с определением итогов $a_1$, $a_2$, $b_1$, $b_2$ для всех рядов и для сводки, и проверить эти итоги тем же способом.

5. Вычислить для всех рядов и для сводки обе вспомогательные суммы $S_1$ и $S_2$ и проверить их.

6. Вычислить $\Sigma x^2$, $M$, $\sigma$ и $m$ по обычным формулам только для отдельных рядов (но не для сводки).

## XIII. Дополнительный учет отдельных вариант с большими отклонениями.

Способ сумм дает тем большую точность, чем меньше был взят классовый промежуток $k$ и чем, следовательно, на большее количество классов был предварительно разбит данный вариационный ряд. Однако при чрезмерном уменьшении классового промежутка могут возникнуть чисто технические затруднения при вычерчивании и заполнении расчетной решетки, которая в этом случае может просто не поместиться на листе бумаги обычного формата. Слишком большое количество классов в особенности может затруднить вычислительную работу тогда, когда отдельные, сравнительно редко встречающиеся значения вариант оказываются далеко отброшенными от среднего арифметического $M$. Поясним этот случай числовым примером, приняв сперва для простоты классовый промежуток $k = 1$.

1. Пусть имеем следующий „основной“ ряд:

варианта 20 встречалась 3 раза
„ 19 „ 12 „
„ 18 „ 20 „
„ 17 „ 9 „
„ 16 „ 6 „

Пусть, далее, кроме этих 50-ти „основных“ вариант, сгруппировавшихся в сравнительно узком интервале (от 20 до 16), в данном ряду встречаются еще три „дополнительные“ их значения: 25, 23 и 23 (всего, следовательно, в заданном ряду будет 53 варианты).

При обработке такого ряда мы можем итти двумя путями: или учесть одновременно все 53 варианты, продолжив наш ряд кверху до наибольшего значения 25 и обозначив частотами, равными нулю, отсутствие промежуточных значений (24, 22 и 21), или же обработать только выписанные выше 50 „основных“ вариант и внести соответствующую поправку в наши вычисления уже в конце этой работы.

Проделаем эти вычисления обоими способами и убедимся в том, что конечные результаты у нас получаются совершенно одинаковыми. Начнем с учета всех 53 вариант. В этом случае наш ряд представится в следующем виде:

варианта 25 встретилась 1 раз
„ 24 „ 0 „
„ 23 „ 2 „
„ 22 „ 0 „
„ 21 „ 0 „
„ 20 „ 3 „
„ 19 „ 12 „
„ 18 „ 20 „
„ 17 „ 9 „
„ 16 „ 6 „

Начертив расчетную решетку и произведя в ней все необходимые выкладки (при $C=18$), получим картину, представленную на табл. 51.

Табл. 51.

| Знач. вар. | $p$ | 35 | 44 |
|---|---|---|---|
| 25 | 1 | 1 | 1 |
| 24 | 0 | 1 | 2 |
| 23 | 2 | 3 | 5 |
| 22 | 0 | 3 | 8 |
| 21 | 0 | 3 | 11 |
| 20 | 3 | 6 | 17 |
| 19 | 12 | 18 | — |
| 18 | 20 | — | — |
| 17 | 9 | 15 | — |
| 16 | 6 | 6 | 6 |
| | 53 | 21 | 6 |



Здесь:

$$S_1=+14$$
$$S_2=156.$$

Следовательно $\Sigma x^2=152{,}30$. При $k=1$ окончательно получим:

$$M=18{,}26$$
$$\sigma=\pm 1{,}71.$$

Но можно итти и другим путем, воспользовавшись указанными выше свойствами контрольной сводки.

Вспомним, что в сводке вспомогательные величины $n$, $S_1$ и $S_2$ слагаются из сумм соответствующих вспомогательных величин отдельных рядов, из которых эта сводка составлена. Если наш ряд мы разобьем на две части, на „основную“ и „дополнительную“, и для каждой из них порознь определим все три вспомогательные величины $n$, $S_1$ и $S_2$, то, сложив их затем попарно, очевидно и получим те же самые значения $n$, $S_1$ и $S_2$, которые только что были вычислены непосредственно, т. е.

$$n=53 \qquad S_1=+14 \qquad S_2=156.$$

Табл. 52.

| Знач. вар. | $p$ | 18 | 3 |
|---|---|---|---|
| 20 | 3 | 3 | 3 |
| 19 | 12 | 15 | — |
| 18 | 20 | — | — |
| 17 | 9 | 15 | — |
| 16 | 6 | 6 | 6 |
| | 50 | 21 | 6 |

Проделаем эти вычисления в действительности, начав с „основного“ ряда (без вариант 25, 23 и 23). В этом случае наша решетка примет вид указанный на табл. 52.

Обозначив общее число измерений, сумму вспомогательных отклонений и сумму их квадратов через $n$, $S_1$ и $S_2$ с дополнительными значками сверху в виде штриха ('), т.-е. через $n'$, $S_1'$ и $S_2'$, и приняв за $C$ попрежнему число 18, для данного ряда получим:

$$n'=50 \qquad S_1'=-3 \qquad S_2'=57.$$

Эти данные относятся к „основному“ ряду, в котором не учтены еще 3 редко встречающиеся варианты с большими отклонениями (25, 23 и 23). Составив из них особый „дополнительный“ ряд, вычислим и для него все три вспомогательные величины $n$, $S_1$ и $S_2$, обозначив их для отличия двумя штрихами ($''$), т.-е. приняв обозначения $n''$, $S_1''$ и $S_2''$. В виду того, что количество этих вариант крайне незначительно ($n''=3$), в данном случае вычисление удобно произвести по способу округления. За начало отсчетов мы должны взять непременно то же самое число $C$, которое было принято и при обработке „основного ряда“, т.-е. $C=18$.

Вычтя из всех „дополнительных“ вариант (25, 23 и 23) это число $C=18$, получим вспомогательные отклонения $x_c(+7,\ +5$ и $+5)$, а возвысив эти отклонения во вторую степень, найдем и квадраты вспомогательных отклонений $x_c^2$ (49, 25 и 25). Все эти данные следует записать в соответствующие столбцы расчетной решетки и просуммировать. В результате получим:

Табл. 53.

| $V$ | $x_c$ | $x_c^2$ |
|---|---|---|
| 25 | $+7$ | 49 |
| 23 | $+5$ | 25 |
| 23 | $+5$ | 25 |
| — | $+17$ | 99 |

$$S_1''=17 \qquad S_2''=99.$$

Общее же количество этих „дополнительных“ вариант

$$n''=3.$$

Итак, для „основного“ ряда:

$$n'=50 \qquad S_1'=-3 \qquad S_2'=57$$

а для „дополнительного“:

$$n''=3 \qquad S_1''=+17 \qquad S_2''=99.$$

Для получения всех трех вспомогательных величин $n$, $S_1$ и $S_2$ (без штриха), относящихся ко всем 53 вариантам (что необходимо нам для вычисления $M$ и $\sigma$), следует сложить попарно $n'+n''$, $S_1'+S_1''$ и $S_2'+S_2''$. В результате получим:

$$n=n'+n''=50+3=53$$
$$S_1=S_1'+S_1''=-3+17=+14$$
$$S_2=S_2'+S_2''=57+99=156.$$

Как и следовало ожидать, найденные этим путем величины $n=53$, $S_1=+14$ и $S_2=156$ в точности совпали с теми их значениями, которые ранее были получены прямо из расчетной решетки (см. выше).



В разобранном только что примере классовый промежуток $k$ равнялся 1. В общем же случае, когда $k$ оказывается либо больше, либо меньше 1, способ этот требует внесения некоторых поправок, которые удобнее всего разъяснить также на числовом примере. Пусть имеем следующий „основной“ ряд:

В интервале от 60 до 56 заключается 4 варианты
„ „ „ 55 „ 51 „ 14 „
„ „ „ 50 „ 46 „ 23 „
„ „ „ 45 „ 41 „ 26 „
„ „ „ 40 „ 36 „ 18 „
„ „ „ 35 „ 31 „ 5 „

Кроме того пусть имеется еще 5 „дополнительных“ вариант, невошедших в приведенное выше распределение по классам: 93, 78, 68, 23 и 18. Здесь классовый промежуток

$$k=5.$$

Чтобы учесть все варианты (включая и 5 „дополнительных“), следует или увеличить количество классов до 16 и обработать эти данные в общей расчетной решетке обычным способом или же поступить так, как это мы сделали в предыдущем примере, т.-е. обработать сперва „основной“ ряд, состоящий всего лишь из 6 классов, а затем внести соответствующую поправку на включение 5 „дополнительных“ вариант.

Для сравнения результатов проделаем эти вычисления обоими способами, начав с обработки всех вариант в одной общей решетке (см. табл. 54).

Табл. 54.

| Гран. клас. | $p$ | 85 | 102 |
|---|---|---|---|
| 95—91 | 1 | 1 | 1 |
| 90—86 | 0 | 1 | 2 |
| 85—81 | 0 | 1 | 3 |
| 80—76 | 1 | 2 | 5 |
| 75—71 | 0 | 2 | 7 |
| 70—66 | 1 | 3 | 10 |
| 65—61 | 0 | 3 | 13 |
| 60—56 | 4 | 7 | 20 |
| 55—51 | 14 | 21 | 41 |
| 50—46 | 23 | 44 | — |
| 45—41 | 26 | — | — |
| 40—36 | 18 | 25 | — |
| 35—31 | 5 | 7 | 12 |
| 30—26 | 0 | 2 | 5 |
| 25—21 | 1 | 2 | 3 |
| 20—16 | 1 | 1 | 1 |
| | 95 | 37 | 21 |

В этой решетке пять „дополнительных" вариант разместятся в следующих классах:

| | | | | | | |
|---|---|---|---|---|---|---|
| варианта | 93 | войдет в класс | от | 95 | до | 91 |
| „ | 78 | „ „ „ | „ | 80 | „ | 76 |
| „ | 68 | „ „ „ | „ | 70 | „ | 66 |
| „ | 23 | „ „ „ | „ | 25 | „ | 21 |
| „ | 18 | „ „ „ | „ | 20 | „ | 16 |

Так как каждая из этих вариант встречается лишь по одному разу, то в вертикальном столбце $p$ расчетной решетки против границ соответствующих классов следует поставить по 1. После этого в ту же решетку запишем частоты „основного" ряда: 4, 14, 23, 26, 18 и 5, а против границ тех классов, в которых не оказалось ни одной варианты, проставим нули. Приняв за начало отсчетов $C$ середину классов от 45 до 41, т.-е. полусумму его границ.

Табл. 55.

| Гран. клас. | $p$ | 63 | 26 |
|---|---|---|---|
| 60—56 | 4 | 4 | 4 |
| 55—51 | 14 | 18 | 22 |
| 50—46 | 23 | 41 | — |
| 45—41 | 26 | — | — |
| 40—36 | 18 | 23 | — |
| 35—31 | 5 | 5 | 5 |
| | 90 | 28 | 5 |

$$\frac{45+41}{2}=\frac{86}{2}=43$$

и проделав обычные вычисления, получим (при $C=43$)

$$n=95 \qquad S_1=+48 \qquad S_2=368.$$

Посмотрим, к каким результатам мы придем, если разобьем наш ряд на „основной" и „дополнительный".

Вычисление вспомогательных величин $n'$, $S_1'$ и $S_2'$ для „основного" ряда (при $C=43$) представлено на табл. 55.

В данном случае

$$n'=90 \qquad S_1'=+35 \qquad S_2'=153.$$

При вычислении же вспомогательных величин $n''$, $S_1''$ и $S_2''$ для „дополнительных" вариант (по способу округления) придется внести в этот процесс некоторую поправку, в связи с тем, что классовый промежуток $k$ здесь не равен уже еди-



нице, как в предыдущем примере (теперь $k=5$). Приняв попрежнему $C=43$, получим расчет, представленный на табл. 56.

Сумма чисел среднего столбца ($x_c$) равна здесь $+65$, а сумма чисел правого столбца ($x_c^2$) равна 5375. Но эти суммы в данном случае еще не представляют собой искомых вспомогательных величин $S_1''$ и $S_2''$ „дополнительного" ряда, как это имело место в предыдущем примере, при $k=1$. Для получения из этих чисел ($+65$ и 5375) искомых сумм $S_1''$ и $S_2''$ следует первое число ($+65$) разделить еще на классовый промежуток $k$, а второе число (5375) — на квадрат этой величины, т.-е. на $k^2$. Так как в нашем примере $k=5$, и, следовательно, $k^2=25$, то искомые вспомогательные величины $S_1''$ и $S_2''$ найдутся после следующих несложных вычислений:

Табл. 56.

| $V$ | $x_c$ | $x_c^2$ |
|---|---|---|
| 93 | $+50$ | 2500 |
| 78 | $+35$ | 1225 |
| 68 | $+25$ | 625 |
| 23 | $-20$ | 400 |
| 18 | $-25$ | 625 |
| — | $+65$ | 5375 |

$$S_1''=\frac{+65}{5}=+13$$

$$S_2''=\frac{5375}{25}=215.$$

Общее же количество дополнительных вариант $n''$ очевидно без всяких изменений равно 5. Итак, для „основного" ряда: $n'=90$, $S_1'=+35$, $S_2'=153$, а для „дополнительного": $n''=5$, $S_1''=+13$, $S_2''=215$.

Сложив эти числа попарно ($n'+n''$, $S_1'+S_1''$ и $S_2'+S_2''$), получим значения вспомогательных величин $n$, $S_1$ и $S_2$, относящихся ко всем 95 вариантам. В конечном итоге, очевидно, будем иметь:

$$n=95 \qquad S_1=+48 \qquad S_2=368.$$

Эти данные мы имели и раньше, но должны были вычислять их в громоздкой решетке, состоящей из 16 строчек, в то время, как здесь мы получили их путем сравнительно сокращенных операций. Эта экономия работы в связи с упрощением и сокращением отдельных вычислений становится особенно заметной в том случае, когда „дополнительные" варианты очень сильно отклоняются от среднего арифметического $M$. Если такой ряд обрабатывать по общим правилам, то расчетную решетку придется делать необычайно больших размеров, что технически представляет, очевидно, крупное неудобство. Выполняя же эти вычисления в два приема, мы в обоих случаях пользуемся расчетными решетками ограниченного формата и, стало быть, в значительной степени упрощаем этот процесс.

Разобранный в последнем примере случай все же не является еще общим. Дело в том, что 5 „дополнительных" вариант (93, 78, 68, 23 и 18) были искусственно подобраны с тем расчетом, чтобы числовые значения их пришлись как раз посредине соответствующих классов. Так, например, варианта 93 равна здесь полусумме границ класса от 95 до 91, так как

$$\frac{95+91}{2}=\frac{186}{2}=93,$$

и стало быть число 93 занимает в этом классе как раз центральное положение. Точно также „дополнительная" варианта 78 расположена в центре класса с границами от 80 до 76 и т. д.

Благодаря такому искусственному подбору этих вариант, конечные результаты вычисления без разбивки и с разбивкой ряда на „основной" и „дополнительный" оказались равными друг другу в точности. В общем же случае, при произвольных значениях „дополнительных" вариант, совпадение конечных результатов вычисления будет уже не полным, хотя практически и вполне достаточным. Поясним сказанное примером.

Пусть имеем следующий „основной" ряд, разбитый на 15 классов, с классовым промежутком $k=0{,}5$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| В границах | от | 43 | до | 42,6 | заключается | 2 | варианты |
| „ | „ | 42,5 | „ | 42,1 | „ | 3 | „ |
| „ | „ | 42 | „ | 41,6 | „ | 6 | „ |
| „ | „ | 41,5 | „ | 41,1 | „ | 12 | „ |
| „ | „ | 41 | „ | 40,6 | „ | 21 | „ |
| „ | „ | 40,5 | „ | 40,1 | „ | 32 | „ |
| „ | „ | 40 | „ | 39,6 | „ | 42 | „ |
| „ | „ | 39,5 | „ | 39,1 | „ | 36 | „ |
| „ | „ | 39 | „ | 38,6 | „ | 25 | „ |
| „ | „ | 38,5 | „ | 38,1 | „ | 17 | „ |



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| В границах | от | 38 | до | 37,6 | заключается | 9 | вариант |
| „ | „ | 37,5 | „ | 37,1 | „ | 4 | „ |
| „ | „ | 37 | „ | 36,6 | „ | 2 | „ |
| „ | „ | 36,5 | „ | 36,1 | „ | 2 | „ |
| „ | „ | 36 | „ | 35,6 | „ | 1 | „ |

и кроме того 10 „дополнительных" вариант, невошедших в эту разбивку по классам: 49,4, 47,2, 45,5, 44,0, 35,1, 33,6, 31,3, 29,0, 26,4, 25,9.

Для обработки всего этого ряда в целом (т.-е. включая и 10 „дополнительных" вариант) пришлось бы увеличить число классов до 48 (читателям предлагается выполнить эту разбивку и произвести соответствующее вычисление самостоятельно).

Взяв за начало отсчетов $C$ средину класса с границами от 40 до 39,6, т.-е. приняв $C=39{,}8$ и выполнив обычные вычисления по способу сумм, в конечном итоге получим:

$$n=224 \qquad S_1=-128$$
$$S_2=4516.$$

Табл. 57.

| Гран. клас. | $p$ | 161 | 153 |
|---|---|---|---|
| 43 —42,6 | 2 | 2 | 2 |
| 42,5—42,1 | 3 | 5 | 7 |
| 42 —41,6 | 6 | 11 | 18 |
| 41,5—41,1 | 12 | 23 | 41 |
| 41 —40,6 | 21 | 44 | 85 |
| 40,5—40,1 | 32 | 76 | — |
| 40 —39,6 | 42 | — | — |
| 39,5—39,1 | 36 | 96 | — |
| 39 —38,6 | 25 | 60 | 131 |
| 38,5—38,1 | 17 | 35 | 71 |
| 38 —37,6 | 9 | 18 | 36 |
| 37,5—37,1 | 4 | 9 | 18 |
| 37 —36,6 | 2 | 5 | 9 |
| 36,5—36,1 | 2 | 3 | 4 |
| 36 —35,6 | 1 | 1 | 1 |
| | 214 | 227 | 270 |

Посмотрим, к каким результатам мы придем, если обработаем сперва „основную" часть заданного ряда и потом уже внесем поправку на включение 10 „дополнительных" вариант.

Составив расчетную решетку для 15 классов (см. табл. 57) и произведя в ней обычные выкладки при $C=39{,}8$, в конечном итоге получим:

$$n'=214$$
$$S_1'=-66$$
$$S_2'=1234.$$

Эти данные относятся к „основному“ ряду. Для вычисления величин $n''$, $S_1''$, $S_2''$, относящихся к 10 „дополнительным“ вариантам, составим расчетную решетку, указанную на табл. 58.

Табл 58.

| $V$ | $x_c$ | $x_c^2$ |
|---|---|---|
| 49,4 | + 9,6 | 92,16 |
| 47,2 | + 7,4 | 54,76 |
| 45,5 | + 5,7 | 32,49 |
| 44,0 | + 4,2 | 17,64 |
| 35,1 | — 4,7 | 22,09 |
| 33,6 | — 6,2 | 38,44 |
| 31,3 | — 8,5 | 72,25 |
| 29,0 | — 10,8 | 116,64 |
| 26,4 | — 13,4 | 179,56 |
| 25,9 | — 13,9 | 193,21 |
| — | — 30,60 | 819,24 |

Выполнив в этой решетке обычные вычисления при $C=39{,}8$ и разделив сумму чисел, полученных в среднем столбце, на $k=0{,}5$, а сумму чисел, полученных в правом столбце, на $k^2=0{,}25$, найдем искомые значения $S_1''$ и $S_2''$

$$S_1''=\frac{-30{,}60}{0{,}5}=-61{,}20$$

$$S_2''=\frac{819{,}24}{0{,}25}=3276{,}96.$$

Общее же количество „дополнительных“ вариант $n''$ очевидно $=10$. Итак:

| | | | |
|---|---|---|---|
| для основного ряда | $n'=214$ | $S_1'=-66$ | $S'_2=1234$ |
| для „дополн.“ вар. | $n''=10$ | $S_1''=-61{,}2$ | $S_2''=3276{,}96$ |
| суммы этих величин: | $n=224$ | $S_1=-127{,}2$ | $S_2=4510{,}96$ |

Округлив дробные значения $S_1$ и $S_2$ до ближайших к ним целых чисел, получим окончательно:

$$n=224 \qquad S_1=-127 \qquad S_2=4511.$$

Вспомним, что при обработке этого ряда без дробления его на „основную“ и „дополнительную“ часть (при разбивке на 48 классов) мы имели следующие значения тех же вспомогательных величин:

$$n=224 \qquad S_1=-128 \qquad S_2=4516.$$

Как видно из этого сопоставления, результаты вычисления в обоих случаях несколько отличаются друг от друга (для $S_1$ на 1, а для $S_2$ на 5), что впрочем на величине $M$ и $\sigma$ обычно или совершенно не отражается или же сказывается в несовпадении лишь последних десятичных знаков.



Так, определив $M$ и $\sigma$ с помощью вспомогательных величин $S_1=-127$ и $S_2=4511$ (при $n=224$, $C=39{,}8$ и $=0{,}5$), в конечном итоге будем иметь:

$$M=39{,}52$$
$$\sigma=\pm 2{,}23.$$

Воспользовавшись же ранее выведенными значениями $S_1=-128$ и $S_2=4516$ (при тех же значениях $n$ и $C$ и $k$), получим:

$$M=39{,}51$$
$$\sigma=\pm 2{,}23.$$

В данном случае различие в конечных результатах этих вычислений сказалось лишь в последнем десятичном (втором после запятой) знаке у числа $M$, а оба значения $\sigma$ в точности совпали друг с другом.

Описанный здесь способ дополнительного учета вариант большими отклонениями удобно применять, например, в том случае, когда в процессе разбивки первичных числовых данных по заранее зафиксированным классам некоторые из вариант с большими отклонениями по случайному недосмотру или по каким-либо иным причинам окажутся выходящими далеко за пределы установленных классов и, вследствие этого, не поместятся в расчетной решетке. Тогда, вместо составления новой решетки, расчитанной на большое количество классов, эти несколько неуместившихся вариант можно пометить где-либо на полях как „дополнительные“ и в конце вычислений (уже после проверки вычислений с помощью контрольной сводки) включить их в общий расчет путем соответствующей поправки.

С помощью аналогичного приема можно выполнить и обратную задачу (выключение отдельных вариант из заданного ряда). Пусть, например, имеем следующий вариационный ряд, разбитый на 14 классов, с классовым промежутком $k=0{,}1$ (см. табл. 59). Приняв $C=2{,}2$, получим: $S_1=+15$ и $S_2=533$, откуда при $n=100$ и $k=0{,}1$ $M=2{,}22$ и $\sigma=\pm 0{,}23$.

В заданном ряду большинство вариант (96) сгруппировано в сравнительно узких пределах от 2,6 до 1,9, и лишь 4 варианты, записанные сверху (3,2, два раза по 3,1 и 3,0), оказались сильно отклоняющимися от $M$ в сторону больших значений. Пусть по каким либо соображениям мы пожелали бы теперь выключить эти 4 варианты из заданного ряда и заново вычислить $M$ и $\sigma$ для оставшихся 96 вариант. Для этого составим новую расчетную решетку (см. табл. 60) и по способу округления вычислим $S_1$ и $S_2$ для 4-х выключаемых вариант (при том же $C = 2{,}2$).

Вычтя из всех вариант это постоянное число $C$, запишем разности в столбец $x_c$, а квадраты соответствующих разностей в соседний столбец $x_c^2$. Просуммировав числа обоих столбцов, получим $+3{,}6$ и $3{,}26$. Разделив первую из этих сумм на $k = 0{,}1$, а вторую на $k^2 = 0{,}01$, получим:

$$S_1'' = \frac{+3{,}6}{0{,}1} = +36$$

$$S_2'' = \frac{3{,}26}{0{,}01} = 326.$$

Таблица 59.

| $V$ | $p$ | 79 | 167 |
|---|---|---|---|
| 3,2 | 1 | 1 | 1 |
| 3,1 | 2 | 3 | 4 |
| 3,0 | 1 | 4 | 8 |
| 2,9 | 0 | 4 | 12 |
| 2,8 | 0 | 4 | 16 |
| 2,7 | 0 | 4 | 20 |
| 2,6 | 1 | 5 | 25 |
| 2,5 | 3 | 8 | 33 |
| 2,4 | 7 | 15 | 48 |
| 2,3 | 16 | 31 | — |
| 2,2 | 28 | — | — |
| 2,1 | 23 | 41 | — |
| 2,0 | 13 | 18 | 23 |
| 1,9 | 5 | 5 | 5 |
| | 100 | 64 | 28 |

Таблица 60

| $V$ | $x_c$ | $x_c^2$ |
|---|---|---|
| 3,2 | +1,0 | 1,00 |
| 3,1 | +0,9 | 0,81 |
| 3,1 | +0,9 | 0,81 |
| 3,0 | +0,8 | 0,64 |
| — | +3,6 | 3,26 |

Эта пара вспомогательных величин $S_1''$ и $S_2''$ относится к „дополнительной“ части заданного ряда, подлежащей выключению. Чтобы найти суммы $S_1'$ и $S_2'$, относящиеся только к „основной“ части нашего ряда (т.-е. к 96 вариантам), следует из полученных ранее значений $S_1 = +15$ и $S_2 = 533$ вычесть эти выключаемые значения $S_1'' = +36$ и $S_2'' = 326$.

$$S_1' = S_1 - S_1'' = (+15) - (+36) = -21$$
$$S_2' = S_2 - S_2'' = 533 - 326 = 207.$$



Итак, для основной части заданного ряда (оставшейся после выключения 4-х вариант) вспомогательные величины:

$$n' = 96 \qquad S_1' = -21 \qquad \text{и} \qquad S_2' = 207.$$

Следовательно:

$$M = C + \frac{kS_1'}{n} = 2{,}2 + \frac{0{,}1(-21)}{96} = 2{,}2 - \frac{2{,}1}{96} = 2{,}2 - 0{,}02 = 2{,}18.$$

$$\Sigma x^2 = S_2 - \frac{(S_1')^2}{n} = 207 - \frac{(-21)^2}{96} = 207 - \frac{441}{96} = 207 - 4{,}59 = 202{,}41$$

$$\sigma = k\sqrt{\frac{\Sigma x^2}{n-1}} = 0{,}1\sqrt{\frac{202{,}41}{95}} = 0{,}1\sqrt{2{,}13} = 0{,}1 \cdot 1{,}46 = 0{,}15.$$

Итак, для оставшейся после выключения „основной“ части ряда (состоящей из 96 вариант):

$$M = 2{,}18 \quad \text{и} \quad \sigma = \pm 0{,}15.$$

Как и следовало ожидать, выключение вариант с большими отклонениями несколько изменило в соответствующем направлении $M$ (вместо 2,22 стало 2,18) и значительно уменьшило $\sigma$ (вместо 0,23 стало 0,15).

## XIV. Вычисление коэффициента корреляции при малом числе измерений.

Описанные в предыдущих главах облегченные способы вычисления $M$, $\sigma$ и $m$ дают возможность легко и быстро определять сумму квадратов центральных отклонений $\Sigma x^2$ для любого ряда варьирующих чисел. В этом и заключается главное преимущество этих облегченных способов по сравнению с непосредственным вычислением $\Sigma x^2$, так как дальнейшие операции, связанные с определением $M$, $\sigma$ и $m$ по соответствующим формулам, не представляют уже особых затруднений.

Так как при определении коэффициента корреляции $r$ приходится вычислять суммы квадратов центральных отклонений для обоих коррелирующих рядов, то в данном случае мы также имеем возможность воспользоваться как способом округления, так и способом сумм.

Покажем, как это можно выполнить.

Если центральные отклонения вариант одного из двух коррелирующих рядов обозначить через $x$, а второго — через $y$, то формула для определения коэффициента $r$ примет следующий вид (см. гл. IV):

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

Формула эта показывает, что для определения $r$ следует сумму произведений всех отклонений, взятых попарно ($\Sigma xy$), разделить на квадратный корень из произведения суммы квадратов центральных отклонений первого ряда ($\Sigma x^2$) на сумму квадратов центральных отклонений второго ряда ($\Sigma y^2$).

Разберем какой-либо конкректный случай применения этой формулы на числовом примере, начав с обработки его сперва способом непосредственного вычисления.

Пусть имеем следующие 16 пар вариант двух коррелирующих рядов:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 4 | 2 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 3 | 2 | 3 | 3 | 5 | 4 |
| 8 | 6 | 8 | 7 | 7 | 6 | 5 | 9 | 7 | 8 | 9 | 7 | 5 | 8 | 10 | 10 |

Здесь так же, как и раньше (см. главу IV), под числами первого верхнего ряда записаны соответствующие им числа второго (нижнего) ряда. Следовательно запись эту нужно понимать так: у первого объекта исследования (крайняя слева пара чисел) первый признак при измерении оказался равным 3, в то время, как второй признак (у него же) оказался равным 8; у второго объекта первый признак = 4, а второй = 6; у третьего объекта первый признак = 2, а второй = 8 и т. д. Признаки эти, очевидно, могут быть измерены и выражены здесь в различных (неодинаковых) мерах (напр., один из них в *см*, а другой в *кг* и т. д.).

Для определения (по способу непосредственного вычисления) коэффициента корреляции $r$, показывающего наличие, характер и степень связи между двумя этими признаками, начертим расчетную решетку, состоящую из 7-ми вертикальных столбцов и 16-ти горизонтальных строчек (по числу коррелирующих пар), не считая двух дополнительных строчек сверху и снизу этой решетки для заголовков и записи сумм (см. табл. 61).



Первый из этих столбцов (крайний слева) озаглавим буквою $V$ со знаком $x$, т.-е. символом $V_x$. Этот заголовок указывает на то, что в данном столбце мы будем записывать варианты первого из двух коррелирующих рядов, центральные отклонения которого в дальнейшем будут обозначены через $x$ (ряд этот записан в нашем задании в верхней строчке).

ТАБЛИЦА 61.

| $V_x$ | $V_y$ | $x$ | $x^2$ | $y$ | $y^2$ | $xy$ |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

В следующем столбце, озаглавленном тоже буквою $V$, но уже со знаком $y$, т.-е. символом $V_y$, мы будем записывать варианты другого ряда, центральные отклонения которого в дальнейшем будут изображены с помощью буквы $y$ (этот ряд в задании был записан во второй строчке).

Заголовки остальных столбцов понятны: в столбце $x$ будут записаны центральные отклонения первого ряда, в следующем столбце $x^2$ — их квадраты, в столбце $y$ — центральные отклонения второго ряда, в столбце $y^2$ — квадраты этих отклонений, а в последнем (справа) столбце $xy$ — произведения каждого $x$ на соответствующий ему (парный) $y$.

ТАБЛИЦА 62.

| $V_x$ | $V_y$ | $x$ | $x^2$ | $y$ | $y^2$ | $xy$ |
|---|---|---|---|---|---|---|
| 3 | 8 | | | | | |
| 4 | 6 | | | | | |
| 2 | 8 | | | | | |
| 2 | 7 | | | | | |
| 4 | 7 | | | | | |
| 3 | 6 | | | | | |
| 1 | 5 | | | | | |
| 4 | 9 | | | | | |
| 3 | 7 | | | | | |
| 2 | 8 | | | | | |
| 3 | 9 | | | | | |
| 2 | 7 | | | | | |
| 3 | 5 | | | | | |
| 3 | 8 | | | | | |
| 5 | 10 | | | | | |
| 4 | 10 | | | | | |
| 48 | 120 | | | | | |

Для определения центральных отклонений $x$ и $y$ нам нужно предварительно вычислить средние арифметические обоих рядов, а это, в свою очередь, требует определения суммы всех вариант по каждому ряду отдельно. Записав варианты заданных рядов в столбцы $V_x$ и $V_y$, сложим числа этих столбцов и суммы их $\Sigma V_x$ и $\Sigma V_y$ запишем в нижней дополнительной строчке (см. табл. 62).

В данном случае сумма вариант первого ряда

$$\Sigma V_x = 48,$$



а сумма вариант второго ряда

$$\Sigma V_y = 120.$$

Так как общее число измерений для обоих рядов одинаково ($n = 16$), то среднее арифметическое для первого ряда, которое мы обозначим через $M_x$, найдется по обычной формуле:

$$M_x = \frac{\Sigma V_x}{n} = \frac{48}{16} = 3,$$

а среднее арифметическое для второго ряда:

$$M_y = \frac{\Sigma V_y}{n} = \frac{120}{16} = 7{,}5.$$

Определив оба средних арифметических $M_x$ и $M_y$, следует вычислить отклонения $x$ всех вариант первого ряда и отклонения $y$ вариант второго ряда. Эти отклонения мы запишем в соответствующие столбцы $x$ и $y$ расчетной решетки и проверим вычисления, убедившись в том, что обе суммы этих чисел равны 0, т.-е. что

$$\Sigma x = 0 \quad \text{и} \quad \Sigma y = 0$$

После этого нужно каждый $x$ помножить на соответствующий ему $y$ и полученное произведение $xy$ записать в последней (правый) столбец расчетной решетки. При этом необходимо соблюдать правило знаков (если оба отклонения имеют одинаковые знаки, то произведение их окажется с $+$, если же отклонения эти имеют разные знаки, то произведение $xy$ будет иметь знак $-$).

Далее следует сложить числа столбца $xy$ с учетом их знаков (т.-е. сложить отдельно все $+$ и отдельно же все $-$, из большей суммы вычесть меньшую и поставить знак большей). В результате получим сумму произведений всех отклонений, взятых попарно, $\Sigma xy$. Этот результат

$$\Sigma xy = +12$$

мы запишем в правом нижнем углу расчетной решетки, которая после этого примет вид, указанный на таблице 63.

Так как $\Sigma xy$ получилась здесь со знаком $+$, то корреляция окажется прямой (см. главу IV).

Теперь следует каждое отклонение $x$ и каждое отклонение $y$ возвысить в квадрат, результаты записать в столбцах $x^2$ и $y^2$ и просуммировать их. Получим суммы квадратов центральных отклонений $\Sigma x^2$ и $\Sigma y^2$.

В данном случае для I ряда $\Sigma x^2 = 16$,
а „ II „ $\Sigma y^2 = 36$.

ТАБЛИЦА 63.

| $V_x$ | $V_y$ | $x$ | $x^2$ | $y$ | $y^2$ | $xy$ |
|---|---|---|---|---|---|---|
| 3 | 8 | 0 | | +0,5 | | 0 |
| 4 | 6 | +1 | | −1,5 | | −1,5 |
| 2 | 8 | −1 | | +0,5 | | −0,5 |
| 2 | 7 | −1 | | −0,5 | | +0,5 |
| 4 | 7 | +1 | | −0,5 | | −0,5 |
| 3 | 6 | 0 | | −1,5 | | 0 |
| 1 | 5 | −2 | | −2,5 | | +5,0 |
| 4 | 9 | +1 | | +1,5 | | +1,5 |
| 3 | 7 | 0 | | −0,5 | | 0 |
| 2 | 8 | −1 | | +0,5 | | −0,5 |
| 3 | 9 | 0 | | +1,5 | | 0 |
| 2 | 7 | −1 | | −0,5 | | +0,5 |
| 3 | 5 | 0 | | −2,5 | | 0 |
| 3 | 8 | 0 | | +0,5 | | 0 |
| 5 | 10 | +2 | | +2,5 | | +5.0 |
| 4 | 10 | +1 | | +2,5 | | +2,5 |
| 48 | 120 | 0 | | 0 | | +12,0 |

Эти суммы мы запишем в нижней дополнительной строчке под соответствующими столбцами $x^2$ и $y^2$, после чего наша решетка примет окончательный вид, представленный на таблице 64.

Итак, в результате всех вычислений мы получим следующие три величины, нужные нам для вычисления с их помощью коэффициента корреляции $r$, а именно:

$$\Sigma x^2 = 16 \qquad \Sigma y^2 = 36 \qquad \Sigma xy = +12$$



Подставляя их в формулу для коэффициента корреляции получим:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{+12}{\sqrt{16 \cdot 36}} = \frac{+12}{\sqrt{576}} = \frac{+12}{24} = +0,5$$

После вычисления коэффициента корреляции $r$ необходимо определить также и его среднюю ошибку $m_r$. Ошибка эта (см. главу IV) найдется по формуле:

ТАБЛИЦА 64.

| $V_x$ | $V_y$ | $x$ | $x^2$ | $y$ | $y^2$ | $xy$ |
|---|---|---|---|---|---|---|
| 3 | 8 | 0 | 0 | +0,5 | 0,25 | 0 |
| 4 | 6 | +1 | 1 | −1,5 | 2,25 | −1,5 |
| 2 | 8 | −1 | 1 | +0,5 | 0,25 | −0,5 |
| 2 | 7 | −1 | 1 | −0,5 | 0,25 | +0,5 |
| 4 | 7 | +1 | 1 | −0,5 | 0.25 | −0,5 |
| 3 | 6 | 0 | 0 | −1,5 | 2,25 | 0 |
| 1 | 5 | −2 | 4 | −2,5 | 6,25 | +5,0 |
| 4 | 9 | +1 | 1 | +1,5 | 2,25 | +1,5 |
| 3 | 7 | 0 | 0 | −0,5 | 0,25 | 0 |
| 2 | 8 | −1 | 1 | +0,5 | 0,25 | −0,5 |
| 3 | 9 | 0 | 0 | +1,5 | 2,25 | 0 |
| 2 | 7 | −1 | 1 | −0,5 | 0,25 | +0,5 |
| 3 | | 0 | 0 | −2,5 | 6,25 | 0 |
| 3 | 8 | 0 | 0 | +0,5 | 0,25 | 0 |
| | 10 | +2 | 4 | +2,5 | 6,25 | +5,0 |
| 4 | 10 | +1 | 1 | +2,5 | 6,25 | +2,5 |
| 48 | 120 | 0 | 16 | 0 | 36,00 | +12,0 |

$$m_r = \pm \frac{1 - r^2}{\sqrt{n}}.$$

Эта формула показывает, что для определения величины $m_r$ следует квадрат коэффициента корреляции ($r^2$) вычесть из единицы и результат разделить на квадратный корень из общего количества измерений ($n$).

В нашем примере $r = +0{,}5$, а $n = 16$. Следовательно

$$m_r = \pm \frac{1-(0{,}5)^2}{\sqrt{16}} = \pm \frac{1-0{,}25}{4} = \pm \frac{0{,}75}{4} = \pm 0{,}19.$$

Эти конечные результаты вычисления можно записать в такой форме:

$$r = +0{,}50 \pm 0{,}19.$$

Полученный здесь коэффициент корреляции $r = +0{,}50$ указывает на наличие довольно значительной прямой зависимости между двумя заданными рядами чисел. Чтобы судить о достоверности этого утверждения, следует (см. главу V) разделить абсолютную величину коэффициента корреляции $r$ (т.-е. 0,50) на абсолютную величину его средней ошибки (т.-е. на 0,19).

$$\frac{r}{m_r} = \frac{0{,}50}{0{,}19} = 2{,}63.$$

Так как в данном случае коэффициент корреляции $r$ превосходит свою ошибку $m_r$ всего лишь в 2,63 раза (а не в 3 раза, как это требуется), то высказанное выше утверждение (о наличии прямой связи двух признаков) не может еще считаться доказанным. Это значит, что при иных условиях этот коэффициент мог бы оказаться и близким к 0, а может-быть, даже сделаться и величиной отрицательной (что указывало бы уже на наличие обратной зависимости между двумя заданными рядами чисел). *)

Итак, при определении коэффициента корреляциа $r$ и его ошибки $m_r$ по способу непосредственного вычисления следует:

1. Составить расчетную решетку на 7 вертикальных столбцов, записать в нее варианты $V_x$ первого ряда в произвольном порядке, а рядом с ними соответствующие варианты $V_y$ второго ряда и сложить их. Получим суммы вариант $\Sigma V_x$ и $\Sigma V_y$.

*) Взамен вычисления средней ошибки $m_r$ можно определить то минимальное число измерений $n$, при котором данный коэффициент корреляции $r$ будет превышать свою ошибку $m_r$ по крайней мере в 3 раза. Соответствующая таблица чисел $n$ для последовательных значений $r$ от 0 до 0,75 помещена в конце книги (см. стр. 289).



2. Вычислить оба средних арифметических $M_x$ и $M_y$ по формулам:

$$M_x = \frac{\Sigma V_x}{n} \qquad M_y = \frac{\Sigma V_y}{n}$$

3. Вычислить центральные отклонения $x$ вариант первого ряда и центральные отклонения $y$ вариант второго ряда, записать их в расчетную решетку и проверить сложением, убедившись в том, что

$$\Sigma x = 0 \quad \text{и} \quad \Sigma y = 0.$$

4. Перемножить каждый $x$ на соответствующий ему $y$ (с учетом знаков), результаты записать в столбец $xy$ и сложить. Получим сумму произведений этих отклонений $\Sigma xy$.

5. Возвысить все $x$ и все $y$ в квадрат, результаты записать в решетку и сложить. Получим суммы квадратов центральных отклонений $\Sigma x^2$ и $\Sigma y^2$.

6. Вычислить коэффициент корреляции $r$ по формуле:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}.$$

7. Вычислить среднюю ошибку коэффициента корреляции $m_r$ по формуле:

$$m_c = \pm \frac{1-r^2}{\sqrt{n}}.$$

8. Оценить достоверность наличия обнаруженной зависимости, убедившись в том, что коэффициент корреляции $r$ превосходит свою среднюю ошибку $m_r$ по крайней мере в 3 раза.

Желая избавиться от дробных отклонений, можно при вычислении коэффициента корреляции $r$ воспользоваться способом округления. Покажем применение этого способа на том же самом числовом примере.

Составим расчетную решетку на этот раз из 8 вертикальных столбцов, озаглавив их так, как это показано на таблице 65. Так как отдельные отклонения в данном случае будут отсчитываться уже от округленных средних, то в отличие от центральных отклонений $x$ и $y$ мы обозначим их через $x_c$ и $y_c$. Значение двух последних заголовков $z_c$ и $z_c^2$ выяснится из описания дальнейших операций.

В столбцах $V_x$ и $V_y$ мы попрежнему запишем варианты обоих рядов, сложим их и полученные суммы $\Sigma V_x$ и $\Sigma V_y$ поместим снизу. В данном случае так же, как и раньше:

ТАБЛИЦА 65.

| $V$ | $V_y$ | $x_c$ | $x_c^2$ | $y_c$ | $y_c^2$ | $z_c$ | $z_c^2$ |
|---|---|---|---|---|---|---|---|
| 3 | 8 | | | | | | |
| 4 | 6 | | | | | | |
| 2 | 8 | | | | | | |
| 2 | 7 | | | | | | |
| 4 | 7 | | | | | | |
| 3 | 6 | | | | | | |
| 1 | 5 | | | | | | |
| 4 | 9 | | | | | | |
| 3 | 7 | | | | | | |
| 2 | 8 | | | | | | |
| 3 | 9 | | | | | | |
| 2 | 7 | | | | | | |
| 3 | 5 | | | | | | |
| 3 | 8 | | | | | | |
| 5 | 10 | | | | | | |
| 4 | 10 | | | | | | |
| 48 | 120 | | | | | | |

$$\Sigma V_x = 48$$
$$\Sigma V_y = 120.$$

Средние арифметические $M_x$ и $M_y$ найдутся по обычным формулам:

$$M_x = \frac{\Sigma V_x}{n} = \frac{48}{16} = 3$$

$$M_y = \frac{\Sigma V_y}{n} = \frac{120}{16} = 7{,}5.$$



Так как первое из этих средних $M_x$ случайно оказалось целым числом ($M_x = 3$), то округлять его, очевидно, и не нужно. Второе среднее ($M_y = 7{,}5$) мы можем округлить или до 7, или до 8. Примем за округленное среднее второго ряда число 7. Обозначив эти округленные средние буквами $C$ со значками $x$ и $y$, будем иметь:

$$C_x = 3 \qquad C_y = 7.$$

Теперь вычислим вспомогательные отклонения $x_c$ и $y_c$ отдельных вариант двух наших рядов от соответствующих округленных средних $C_x$ и $C_y$. Эти вспомогательные отклонения мы запишем в столбцы $x_c$ и $y_c$ и просуммируем (см. табл. 66).

Так как округленное среднее $C_x$ случайно совпало с точным значением самого среднего арифметического $M_x = 3$, то сумма вспомогательных отклонений этого ряда окажется равной нулю. Но это не является общим правилом. Сумма вспомогательных отклонений второго ряда равна здесь уже $+8$. Обе эти суммы (0 и $+8$) мы попрежнему обозначим через $S_1$, указав соответствующею буквою, поставленною в скобках, к какому ряду это обозначение относится. Итак, в нашем примере сумма вспомогательных отклонений первого ряда

$$S_1(x) = 0,$$

а сумма вспомогательных отклонений второго ряда

$$S_1(y) = +8.$$

Обе эти суммы можно проверить с помощью того же самого приема, каким при вычислении $\sigma$ по способу округления мы проверяли единственную получавшуюся там сумму $S_1$, т. е. по контрольной формуле:

$$S_1 = \Sigma V - nC.$$

В данном случае:

$$S_1(x) = \Sigma V_x - nC_x = 48 - 16 \cdot 3 = 48 - 48 = 0.$$
$$S_1(y) = \Sigma V_y - nC_y = 120 - 16 \cdot 7 = 120 - 112 = +8.$$

Следовательно обе суммы $S_1(x)$ и $S_1(y)$ у нас вычислены правильно.

Теперь внесем в порядок вычисления $r$ еще одно изменение, которое имеет целью упростить все наши выкладки и в особенности будет полезно при дальнейшем применении способа сумм. Изменение это заключается в том, что взамен умножения каждого вспомогательного отклонения $x_c$ на со-

ТАБЛИЦА 66.

| $V_x$ | $V_y$ | $x_a$ | $x_c^2$ | $y_c$ | $y_c^2$ | $z_c$ | $z_c^2$ |
|---|---|---|---|---|---|---|---|
| 3 | 8 | 0 | | +1 | | +1 | |
| 4 | 6 | +1 | | −1 | | 0 | |
| 2 | 8 | −1 | | +1 | | 0 | |
| 2 | 7 | −1 | | 0 | | −1 | |
| 4 | 7 | +1 | | 0 | | +1 | |
| 3 | 6 | 0 | | −1 | | −1 | |
| 1 | 5 | −2 | | −2 | | −4 | |
| 4 | 9 | +1 | | +2 | | +3 | |
| 3 | 7 | 0 | | 0 | | 0 | |
| 2 | 8 | −1 | | +1 | | 0 | |
| 3 | 9 | 0 | | +2 | | +2 | |
| 2 | 7 | −1 | | 0 | | −1 | |
| 3 | 5 | 0 | | −2 | | −2 | |
| 3 | 8 | 0 | | +1 | | +1 | |
| 5 | 10 | +2 | | +3 | | +5 | |
| 4 | 10 | +1 | | +3 | | +4 | |
| 48 | 120 | 0 | | +8 | | +8 | |

ответствующее ему вспомогательное отклонение $y_c$ мы будем просто складывать эти отклонения, записывая соответствующие суммы в столбец $z_c$. Следовательно числа этого столбца будут представлять собою суммы вспомогательных отклонений, взятых попарно. Очевидно, при сложении $x_c$ и $y_c$ следует учитывать их знаки (т.-е. в случае, когда $x_c$ и $y_c$ окажутся с разными знаками, из большего отклонения нужно вычесть меньшее и поставить знак большего). Итак, каждый

$$z_c = x_c + y_c.$$



Записав эти суммы в предпоследний столбец расчетной решетки и просуммировав числа этого столбца с учетом их знаков, получим картину, представленную на той же таблице 66. Здесь сумма чисел $z_c$, которую мы по аналогии с двумя предыдущими суммами $S_1(x)$ и $S_1(y)$ обозначим символом $S_1(z)$, оказалась равной $+8$. Эту сумму мы запишем в нижней дополнительной строчке расчетной решетки под соответсвующим вертикальным столбцом.

Итак, у нас:

$$S_1(z) = +8.$$

Сумму $S_1\ (z)$ мы всегда можем проверить с помощью следующей контрольной формулы:

$$S_1(z) = S_1(x) + S_1(y).$$

Действительно, в нашем случае $S_1(x) = 0$, а $S_1(y) = +8$. Следовательно:

$$S_1(z) = 0 + 8 = +8,$$

что мы и получили только-что непосредственно из расчетной решетки.

После вычисления всех трех сумм:

$$S_1(x),\ S_1(y) \text{ и } S_1(z)$$

следует возвысить в квадрат каждое из чисел, стоящих в столбцах $x_c$, $y_c$ и $z_c$, и записать результаты в соседние столбцы $x_c^2$, $y_c^2$ и $z_c^2$. Сложив числа всех этих столбцов, получим суммы квадратов вспомогательных отклонений обоих наших рядов и, по аналогии с ними, сумму квадратов чисел $z_c$.

Обозначив эти суммы через $S_2$ с соответствующими пояснительными значками в скобках, получим:

$$S_2(x) = 16 \qquad S_2(y) = 40 \qquad S_2(z) = 80.$$

Записав эти результаты в нижней строчке решетки под соответствующими стобцами, получим картину представленную на таблице 67.

С помощью суммы вспомогательных отклонений $S_1$ и суммы квадратов этих отклонений $S_2$ мы можем для каждого ряда вычислить и сумму квадратов его центральных отклонений по обычной формуле, применявшейся ранее (при определении $\sigma$), а именно:

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

В данном случае, следовательно:

$$\Sigma x^2 = S_2(x) - \frac{S_1(x)^2}{n} = 16 - \frac{0^2}{16} = 16 - \frac{0}{16} = 16 - 0 = 16$$

$$\Sigma y^2 = S_2(y) - \frac{S_1(y)^2}{n} = 40 - \frac{8^2}{16} = 40 - \frac{64}{16} = 40 - 4 = 36$$

ТАБЛИЦА 67.

| $V_x$ | $V_y$ | $x_c$ | $x_c^2$ | $y_c$ | $y_c^2$ | $z_c$ | $z_c^2$ |
|---|---|---|---|---|---|---|---|
| 3 | 8 | 0 | 0 | + 1 | 1 | + 1 | 1 |
| 4 | 6 | + 1 | 1 | − 1 | 1 | 0 | 0 |
| 2 | 8 | − 1 | 1 | + 1 | 1 | 0 | 0 |
| 2 | 7 | − 1 | 1 | 0 | 0 | − 1 | 1 |
| 4 | 7 | + 1 | 1 | 0 | 0 | + 1 | 1 |
| 3 | 6 | 0 | 0 | − 1 | 1 | − 1 | 1 |
| 1 | 5 | − 2 | 4 | − 2 | 4 | − 4 | 16 |
| 4 | 9 | + 1 | 1 | + 2 | 4 | + 3 | 9 |
| 3 | 7 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 8 | − 1 | 1 | + 1 | 1 | 0 | 0 |
| 3 | 9 | 0 | 0 | + 2 | 4 | + 2 | 4 |
| 2 | 7 | − 1 | 1 | 0 | 0 | − 1 | 1 |
| 3 | 5 | 0 | 0 | − 2 | 4 | − 2 | 4 |
| 3 | 8 | 0 | 0 | + 1 | 1 | + 1 | 1 |
| 5 | 10 | + 2 | 4 | + 3 | 9 | + 5 | 25 |
| 4 | 10 | + 1 | 1 | + 3 | 9 | + 4 | 16 |
| 48 | 120 | 0 | 16 | + 8 | 40 | + 8 | 80 |

а по аналогии с ними:

$$\Sigma z^2 = S_2(z) - \frac{S_1(z)^2}{n} = 80 - \frac{8^2}{16} = 80 - \frac{64}{16} = 80 - 4 = 76$$

Как и следовало ожидать, величины $\Sigma x^2 = 16$ и $\Sigma y^2 = 36$ в точности совпали с теми их значениями, которые ранее



были получены нами прямо из расчетной решетки (см. способ непосредственного вычисления $r$). Третья величина $\Sigma z^2$, представляющая собою сумму квадратов чисел $z_c$, как увидим сейчас, понадобится нам для вычисления суммы произведений всех отклонений, взятых попарно, т.-е. величины $\Sigma xy$, которая ранее получалась также непосредственно из решетки (у нас $\Sigma xy$ оказалась там равной $+12$). В данном случае эту сумму произведений мы можем вычислить по следующей формуле:

$$\Sigma xy = \frac{\Sigma z^2 - \Sigma y^2 - \Sigma x^2}{2}$$

Формула эта показывает, что для определения $\Sigma xy$ следует из величины $\Sigma z^2$ вычесть обе суммы квадратов центральных отклонений $\Sigma y^2$ и $\Sigma x^2$ и результат разделить на 2. В нашем примере:

$$\Sigma xy = \frac{76 - 36 - 16}{2} = \frac{76 - 52}{2} = \frac{+24}{2} = +12$$

Результат этот, как и следовало ожидать, совпал с полученным ранее прямо из расчетной решетки.

Получив все три величины $\Sigma x^2$, $\Sigma y^2$ и $\Sigma xy$, нужные для определения коэффициента корреляции $r$, подставим их в формулу для вычисления этого коэффициента:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 . \Sigma y^2}} = \frac{+12}{\sqrt{16 . 36}} = \frac{+12}{\sqrt{576}} = \frac{+12}{24} = +0,5$$

Определив $r$, найдем и его среднюю ошибку $m_r$:

$$m_r = \pm \frac{1 - r^2}{\sqrt{n}} = \pm \frac{1 - (0,5)^2}{\sqrt{16}} = \pm \frac{1 - 0,25}{4} = \pm \frac{0,75}{4} = \pm 0,19$$

Следовательно здесь, как и раньше:

$$r = +0,50 \pm 0,19.$$

Итак, при определении коэффициента корреляции $r$ и его средней ошибки $m_r$ по способу округления следует:

1. Составить расчетную решетку из 8 вертикальных столбцов, записать в нее все коррелирующие пары вариант $V_x$ и $V_y$, сложить их (получим $\Sigma V_x$ и $\Sigma V_y$) и вычислить оба средних арифметических $M_x$ и $M_y$ по формулам:

$$M_x = \frac{\Sigma V_x}{n} \qquad M_y = \frac{\Sigma V_y}{n}$$

2. Округлить эти средние до ближайших к ним целых чисел $C_x$ и $C_y$.

3. Вычислить вспомогательные отклонения $x_c$ и $y_c$ отдельных вариант обоих рядов от этих округленных средних $C_x$ и $C_y$, результаты записать в расчетную решетку и сложить. Получим суммы $S_1(x)$ и $S_1(y)$.

4. Проверить правильность вычисления $S_1(x)$ и $S_1(y)$ по контрольным формулам:

$$S_1(x) = \Sigma V_x - n\,.\,C_x$$
$$S_1(y) = \Sigma V_y - n\,.\,C_y$$

5. Сложить каждое вспомогательное отклонение $x_c$ с соответствующим ему вспомогательным отклонением $y_c$, суммы их ($z_c$) записать в решетку и сложить. Получим вспомогательную сумму $S_1(z)$.

6. Проверить величину $S_1(z)$ по контрольной формуле:

$$S_1(z) = S_1(x) + S_1(y).$$

7. Возвысить в квадрат все вспомогательные отклонения $y_c$ и $x_c$, а также и числа $z_c$, результаты записать в расчетную решетку и сложить по столбцам. Получим суммы квадратов вспомогательных отклонений $S_2(x)$ и $S_2(y)$ и, по аналогии с ними, вспомогательную сумму $S_2(z)$.

8. Для каждого из двух коррелирующих рядов и для третьего фиктивного ряда (с отклонениями $z$) вычислить сумму квадратов их центральных отклонений по формулам:

$$\Sigma x^2 = S_2(x) - \frac{S_1(x)^2}{n} \qquad \Sigma z^2 = S_2(z) - \frac{S_1(z)^2}{n}$$
$$\Sigma y^2 = S_2(y) - \frac{S_1(y)^2}{n}$$

9. Вычислить сумму произведений всех отклонений, взятых попарно, по формуле:

$$\Sigma xy = \frac{\Sigma z^2 - \Sigma y^2 - \Sigma x^2}{2}$$

10. Вычислить коэффициент корреляции $r$ по обычной формуле:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2\,.\,\Sigma y^2}}$$



11. Вычислить среднюю ошибку коэффициента корреляции $m_r$ по обычной формуле:

$$m_r = \pm \frac{1 - r^2}{\sqrt{n}}$$

12. Оценить достоверность наличия обнаруженной зависимости, убедившись в том, что коэффициент $r$ превосходит свою среднюю ошибки $m_r$ по крайней мере в 3 раза.

Описанный здесь способ округления удобно применять к вычислению коэффициента корреляции $r$ в том случае, когда коэффициент этот определяется из сравнительно небольшого количества отдельных измерений (напр., не более 20—30).

## XV. Вычисление коэффициента корреляции при большом числе измерений.

Пользуясь возможностью заменить перемножение отдельных вспомогательных отклонений $x_c$ и $y_c$ простым их сложением (см. предыдущую главу), попробуем применить к определению коэффициента корреляции $r$ наиболее совершенный из всех существующих способов вычисления—способ сумм (см. главу IX и последующие). Покажем сперва применение этого способа на том же самом упрощенном примере, который уже был разобран ранее для определения $r$ способом непосредственного вычисления и способом округления. Пусть, попрежнему, имеем два коррелирующих ряда:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 4 | 2 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 3 | 2 | 3 | 3 | 5 | 4 |
| 8 | 6 | 8 | 7 | 7 | 6 | 5 | 9 | 7 | 8 | 9 | 7 | 5 | 8 | 10 | 10 |

Замечая, что отдельные значения вариант в каждом из этих рядов повторяются иногда по несколько раз, подсчитаем количество сочетаний каждого отдельного значения варианты $V_x$ первого ряда с каждым отдельным значением варианты $V_y$ второго ряда. Для этого начертим особую „корреляционную решетку“, состоящую из нескольких вертикальных столбцов и горизонтальных строчек (см. таблицу 68). В этой решетке вертикальные столбцы будут соответствовать отдельным значениям вариант первого ряда ($V_x$), а горизонтальные строчки — отдельным значениям вариант второго ряда ($V_y$). Так как в первом ряду встречаются значения вариант $V_x$ от 1 до 5 включительно, то вертикальных столбцов следует заготовить здесь пять, озаглавив их сверху числами: 1, 2, 3, 4 и 5 (в направлении слева направо). Так как, далее, во втором ряду встречаются варианты $V_y$ от 5 до 10 включительно, т.-е. шесть отдельных их значений, то решетка эта должна состоять из 6 горизонтальных строчек. Заголовки их мы проставим слева (сверху — бо̀льшие, снизу — меньшие). Этими заголовками будут служить здесь числа: 10, 9, 8, 7, 6, 5.

ТАБЛИЦА 68.

| $V_y$ \ $V_x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 10 | | | | | |
| 9 | | | | | |
| 8 | | | / | | |
| 7 | | | | | |
| 6 | | | | / | |
| 5 | | | | | |

Заготовив такую решетку, разнесем по ее клеткам все наши 16 коррелирующих пар. Так как для первого объекта измерений $V_x = 3$, а $V_y = 8$, то данную пару вариант мы должны будем зарегистрировать в той клетке этой решетки, которая лежит на пересечении вертикального столбца с заголовком 3 и горизонтальной строчки с заголовком 8. Поставив в эту клетку для памяти одну вертикальную че рточку мы и отметим ею существование одного такого объекта измерений, который по первому свойству оказался равным 3 и в то же самое время по второму свойству дал измерение, равное 8.

Следующий объект имел $V_x = 4$, а $V_y = 6$. Эти данные мы пометим в той клетке корреляционной решетки, которая находится на пересечении вертикального столбца с заголовком 4 и горизонтальной строчки с заголовком 6, куда



и нужно, следовательно, поставить вторую вертикальную черточку (см. таблицу 68).

Поступая так же со всеми 16 коррелирующими парами (т.-е. отмечая каждую из них в соответствующей клетке

ТАБЛИЦА 69.

| $V_y$ \ $V_x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 10 | | | | / | / |
| 9 | | | / | / | |
| 8 | | // | // | | |
| 7 | | // | / | / | |
| 6 | | | / | / | |
| 5 | / | | / | | |

корреляционной решетки постановкой вертикальной черточки), в конечном итоге получим полную картину разноски всех 16 вариант двух заданных рядов по клеткам этой решетки, что в окончательном виде и представлено на таблице 69.

Если теперь мы подсчитаем количество черточек, оказавшихся в каждой отдельной клетке, и взамен черточек

ТАБЛИЦА 70.

| $V_y$ \ $V_x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 10 | | | | 1 | 1 |
| 9 | | | 1 | 1 | |
| 8 | | 2 | 2 | | |
| 7 | | 2 | 1 | 1 | |
| 6 | | | 1 | 1 | |
| 5 | 1 | | 1 | | |

поставим в них уже эти числа повторений (т.-е. частоты $p$), то решетка наша примет вид, указанный на таблице 70. Эти частоты показывают здесь, сколько раз каждое отдельное сочетание двух значений вариант $V_x$ и $V_y$ в действительности встречалось в нашем задании. Например, частота 2, поставленная в корреляционной решетке на пересечении вертикального столбца 3 и горизонтальной строчки 8, указывает на то, что среди 16 объектов исследования встретились 2 такие, которые по первому свойству имели оценку 3, в то время, как измерение их по второму свойству оказалось равным 8 и т. д.

Так как для производства вычислений по способу сумм нам необходимо иметь частоты каждого из двух рядов в отдельности, т.-е. знать, сколько же раз повторяется каждая из вариант 1, 2, 3, 4, 5 в первом ряду и, независимо от этого, сколько раз повторяется каждая из вариант 10, 9, 8, 7, 6 и 5 во втором ряду, то для ответа на этот вопрос

ТАБЛИЦА 71.

| $V^y$ \ $V_x$ | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| 10 | | | | 1 | 1 | 2 |
| 9 | | | 1 | 1 | | 2 |
| 8 | | 2 | 2 | | | 4 |
| 7 | | 2 | 1 | 1 | | 4 |
| 6 | | | 1 | 1 | | 2 |
| 5 | 1 | | 1 | | | 2 |
| | 1 | 4 | 6 | 4 | 1 | |

нам придется подсчитать суммы частот отдельно во всех вертикальных столбцах и отдельно же во всех горизонтальных строчках (см. табл. 71).

Подсчитав суммы всех частот по вертикальным столбцам, получим частоты первого ряда:

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

(суммы эти записаны на таблице 71 снизу под решеткой). После такого же подсчета частот по горизонтальным строчкам получим частоты второго ряда:

$$2 \quad 2 \quad 4 \quad 4 \quad 2 \quad 2$$

(эти суммы на таблице 71 записаны справа).

Очевидно, суммы частот каждого ряда будут равны друг другу, так как каждая из них представляет собою общее количество всех коррелирующих пар (количество



измерений $n$). В данном случае сумма частот первого ряда (записанная снизу) равна 16 и сумма частот второго ряда (записанная справа) также равна 16.

Имея частоты обоих рядов, мы можем по способу сумм вычислить и суммы квадратов их центральных отклонений $\Sigma x^2$ и $\Sigma y^2$. Но как же определить из данных корреляционной решетки третью необходимую нам сумму $\Sigma z^2$ (см. предыдущую главу XIV)?

Вспомним, что каждое $z_c$ представляет собой сумму двух вспомогательных отклонений $x_c$ и $y_c$, т.-е.

$$z_c = x_c + y_c$$

Разберемся же в вопросе о том, какие клетки корреляционной решетки будут соответствовать **одинаковым** числам $z_c$. Для этого перепишем заново нашу корреляционную решетку и, не проставляя в ее клетках никаких частот, озаглавим все вертикальные столбцы и горизонтальные строчки уже не значениями самих вариант, как раньше, а выпишем в заголовках вспомогательные отклонения этих вариант от соответствующих округленных средних $C_x$ и $C_y$. Эти округленные средние $C_x$ и $C$ для удобства возьмем те же самые, которые были приняты нами в предыдущей главе при вычислении коэффициента корреляции $r$ по способу округления, а именно:

$$C_x = 3 \qquad C_y = 7$$

Тогда наша корреляционная решетка примет вид, указанный на таблице 72. В клетках этой решетки мы будем теперь записывать числа $z_c$, т.-е. суммы соответствующих отклонений $x_c$ и $y_c$. Так, в клетке, находящейся на пересечении вертикального столбца, озаглавленного — 1, и горизонтальной строчки с заголовком $+2$, мы должны будем записывать $+1$, так как в этом месте мы, по условию, должны зарегистрировать такую пару вариант, для которой $x_c = -1$, $y_c = +2$, а сумма их

$${}_c = -1 + 2 = +1.$$

Эту $+1$ мы и запишем в данной клетке корреляционной решетки (см. таблицу 72).

Если и во всех других клетках записать сумму чисел верхнего и левого заголовка, то получим все числа $z_c$, со-

ТАБЛИЦА 72.

| $y_c$ \ $x$ | $-2$ | $-1$ | $0$ | $+1$ | $+2$ |
|---|---|---|---|---|---|
| $+3$ | | | | | |
| $+2$ | | $+1$ | | | |
| $+1$ | | | | | |
| $0$ | | | | | |
| $-1$ | | | | | |
| $-2$ | | | | | |

ответствующие различным парам коррелирующих вариант (см. табл. 73).

Просматривая эти записи, мы замечаем, что одинаковые числа $z_c$ располагаются здесь по диагональным направлениям корреляционной решетки (от левого верхнего ее угла к правому нижнему). Следовательно для подсчета коррели-

ТАБЛИЦА 73.

| $y_c$ \ $x$ | $-2$ | $-1$ | $0$ | $+1$ | $+2$ |
|---|---|---|---|---|---|
| $+3$ | $+1$ | $+2$ | $+3$ | $+4$ | $+5$ |
| $+2$ | $0$ | $+1$ | $+2$ | $+3$ | $+4$ |
| $+1$ | $-1$ | $0$ | $+1$ | $+2$ | $+3$ |
| $0$ | $-2$ | $-1$ | $0$ | $+1$ | $+2$ |
| $-1$ | $-3$ | $-2$ | $-1$ | $0$ | $+1$ |
| $-2$ | $-4$ | $-3$ | $-2$ | $-1$ | $0$ |

рующих пар с одинаковыми числами $z_c$ мы должны будем в нашей прежней корреляционной решетке сложить все частоты, расположенные по диагональным клеткам.

Выполнив такой подсчет в действительности, получим картину, представленную на таблице 74, где сумма частот,



соответствующих одинаковым числам $z_c$ (которые здесь уже не помечены), записана справа и снизу на полях корреляционной решетки, а самые диагональные направления суммирования обозначены пунктирными линиями со стрелками на концах.

Итак, частотами третьего фиктивного ряда (ряда $z$) будут служить следующие 10 чисел:

1 1 1 1 3 4 3 1 0 1

Сумма этих частот, очевидно также равна общему количеству всех измерений $n$ (т.-е. $= 16$).

ТАБЛИЦА 74.

| $V_y$ \ $V_x$ | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| 10 | | | | 1 | 1 | |
| 9 | | | 1 | 1 | | 1 |
| 8 | | 2 | 2 | | | 1 |
| 7 | | 2 | 1 | 1 | | 1 |
| 6 | | | 1 | 1 | | 1 |
| 5 | 1 | | 1 | | | 3 |
| | | 1 | 0 | 1 | 3 | 4 |

Имея частоты всех трех рядов, мы можем обычным путем вычислить суммы их вспомогательных отклонений $S_1$ и суммы квадратов этих вспомогательных отклонений $S_2$, а с помощью этих величин — и суммы квадратов центральных отклонений $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$.

Для выполнения всех этих вычислений по способу сумм начертим обычного типа комбинированную расчетную решетку (см. главу XII), состоящую из трех основных частей, предназначенных для вычисления сумм $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ и четвертой добавочной части для контрольной сводки (см. табл. 75).

Для удобства описания дальнейших вычислений все эти ряды мы обозначим теми же буквами, какими изображаются в них отдельные отклонения, т.-е. $x$, $y$ и $z$, а по аналогии с ними четвертый (контрольный) ряд обозначим буквою $u$.

Число горизонтальных строчек в этой решетке определяется количеством отдельных частот в наиболее длинном ряду. В данном примере ряд $x$ имел лишь 5 частот (1, 4, 6, 4, 1), ряд $y$ — 6 частот (2, 2, 4, 4, 2, 2), а ряд $z$ — 10 частот (1, 1, 1, 1, 3, 4, 3, 1, 0, 1). Итак, наибольшее количество частот, которое нужно поместить в эту решетку, будет равно 10. Следовательно нам необходимо составить эту ре-

ТАБЛИЦА 75.

| Ряд $x$ | | | Ряд $y$ | | | Ряд $z$ | | | Контр. сводка (ряд $u$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | | | $p$ | | | $p$ | | | $p$ | | |
| | | | | | | | | | | | |
| | | | | | | 1 | | | | | |
| | | | | | | 1 | | | | | |
| | | | 2 | | | 1 | | | | | |
| 1 | | | 2 | | | 1 | | | | | |
| 4 | | - | 4 | | | 3 | | | | | |
| 6 | | | 4 | | | 4 | | | | | |
| 4 | | | 2 | | | 3 | | | | | |
| 1 | | | 2 | | | 1 | | | | | |
| | | | | | | 0 | | | | | |
| | | | | | | 1 | | | | | |
| | | | | | | | | | | | |

шетку из 10 горизонтальных строчек (не считая двух дополнительных сверху и снизу для записи итогов $a_1$, $a_2$, $b_1$ и $b_2$).

В столбцах $p$ всех трех основных частей этой решетки мы запишем частоты трех наших рядов, причем частоты больших значений вариант (с положительными отклонениями) поместим в решетке сверху, а частоты малых значений (с отрицательными отклонениями) — снизу. Частоты третьего (фиктивного) ряда $z$ можно расположить как угодно, в любом произвольном порядке. Высота расположения этих частот также совершенно произвольна. Выгоднее всего располагать частоты всех трех рядов так, чтобы они тяготели



к середине расчетной решетки. В результате этих записей получим картину, представленную на таблице 75.

Разместив в расчетной решетке частоты всех трех основных рядов, следует составить из них частоты сводки. Для этого, как всегда, сложим частоты, оказавшиеся в одинаковых горизонтальных строчках, и суммы их запишем в столбец $p$ последней четвертой части решетки (ряд $u$). Сложив затем числа всех четырех столбцов $p$, в первых трех частях решетки получим, очевидно, одинаковые числа $n$, а в сводке — величину в 3 раза большую (т.-е. $3n$). Подсчет

ТАБЛИЦА 76.

| Ряд $x$ | | | Ряд $y$ | | | Ряд $z$ | | | Контр. сводка (ряд $u$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | | | $p$ | | | $p$ | | | $p$ | | |
| | | | | | | | | | | | |
| | | | | | | 1 | | | 1 | | |
| | | | | | | 1 | | | 1 | | |
| | | | 2 | | | 1 | | | 3 | | |
| 1 | | | 2 | | | 1 | | | 4 | | |
| 4 | | – | 4 | | – | 3 | | — | 11 | | — |
| 6 | — | — | 4 | — | — | 4 | — | — | 14 | — | — |
| 4 | | — | 2 | | — | 3 | | — | 9 | | — |
| 1 | | | 2 | | | 1 | | | 4 | | |
| | | | | | | 0 | | | 0 | | |
| | | | | | | 1 | | | 1 | | |
| 16 | | | 16 | | | 16 | | | 48 | | |

всех частот по отдельным столбцам и сверка их с соответствующей суммой частот в контрольной сводке может служить здесь первой проверкой этих вычислений (см. табл. 76).

Далее, как всегда, проставим фигуры из 4-х черточек во всех трех рядах и в сводке на одном общем уровне (см. табл. 76).

Проделав затем все вычисления, связанные с получением итогов $a_1$, $a_2$, $b_1$, $b_2$ для трех рядов и сводки, получим картину, представленную на таблице 77.

После проверки этих итогов по общим правилам (см. главу XII) вычислим для каждого ряда и сводки обе вспомогательные суммы $S_1$ и $S_2$. Сохраняя прежние обозначения этих сумм для различных рядов (с помощью соответствующих букв, поставленных в скобках), найдем, что:

| | | |
|---|---|---|
| Для ряда $x$: | $S_1(x) = 0$ | $S_2(x) = 16$ |
| „ „ $y$: | $S_1(y) = +8$ | $S_2(y) = 40$ |
| „ „ $z$: | $S_1(z) = +8$ | $S_2(z) = 80$ |
| Для сводки $u$: | $S_1(u) = +16$ | $S_2(u) = 136$ |

ТАБЛИЦА 77.

| Ряд $x$ | | | Ряд $y$ | | | Ряд $z$ | | | Контр. сводка (ряд $u$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 6 | 1 | $p$ | 14 | 8 | $p$ | 17 | 20 | $p$ | 37 | 29 |
| | | | | | | 1 | 1 | 1 | 1 | 1 | 1 |
| | | | | | | 1 | 2 | 3 | 1 | 2 | 3 |
| | | | 2 | 2 | 2 | 1 | 3 | 6 | 3 | 5 | 8 |
| 1 | 1 | 1 | 2 | 4 | 6 | 1 | 4 | 10 | 4 | 9 | 17 |
| 4 | 5 | — | 4 | 8 | — | 3 | 7 | — | 11 | 20 | — |
| 6 | — | — | 4 | — | — | 4 | — | — | 14 | — | — |
| 4 | 5 | — | 2 | 4 | — | 3 | 5 | — | 9 | 14 | — |
| 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 4 | 4 | 5 | 7 |
| | | | | | | 0 | 1 | 2 | 0 | 1 | 2 |
| | | | | | | 1 | 1 | 1 | 1 | 1 | 1 |
| 16 | 6 | 1 | 16 | 6 | 2 | 16 | 9 | 7 | 48 | 21 | 10 |

Эти вспомогательные суммы мы также должны проверить, убедившись в том, что сумма всех $S_1$ для отдельных рядов равна $S_1$ для сводки, а сумма всех $S_2$ для отдельных рядов равна $S_2$ для сводки. В данном случае оба эти условия выполнены, следовательно все вычисления произведены правильно.

После определения и проверки всех вспомогательных сумм следует для каждого из трех основных рядов (но уже не для сводки, которая после проверки сумм $S_1$ и $S_2$ сыграла свою роль до конца) вычислить величины $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ по обычным формулам:



$$\Sigma x^2 = S_2(x) - \frac{S_1(x)^2}{n} = 16 - \frac{0^2}{16} = 16 - \frac{0}{16} = 16 - 0 = 16$$

$$\Sigma y^2 = S_2(y) - \frac{S_1(y)^2}{n} = 40 - \frac{8^2}{16} = 40 - \frac{64}{16} = 40 - 4 = 36$$

$$\Sigma z^2 = S_2(z) - \frac{S_1(z)^2}{n} = 80 - \frac{(8)^2}{16} = 80 - \frac{64}{16} = 80 - 4 = 76$$

Итак, суммы квадратов центральных отклонений, как и ранее, оказались равными

$$\Sigma x^2 = 16 \qquad \Sigma y^2 = 36 \qquad \Sigma z^2 = 76.$$

Здесь, благодаря тому, что числа $n$ для всех трех рядов всегда одинаковы (у нас $n = 16$), мы можем проверить вычисление всех трех сумм $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ по следующей контрольной формуле:

$$\Sigma x^2 + \Sigma y^2 + \Sigma z^2 = S_2(u) - \frac{\Sigma S_1^2}{n}.$$

Эта формула показывает, что сумма всех трех только-что полученных величин: $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ должна оказаться равной выражению, стоящему в правой части равенства. Для вычисления же этого выражения мы должны сложить квадраты всех трех $S_1$, полученных для рядов $x$, $y$ и $z$, сумму эту ($\Sigma S_1^2$) разделить на число измерений $n$ и результат вычесть из суммы $S_2$, полученной в контрольной сводке, т.-е. из $S_2(u)$. В нашем примере мы имели следующие значения $S_1$ для рядов $x$, $y$ и $z$:

| | | | |
|---|---|---|---|
| Для ряда $x$: | $S_1(x) = 0$, | следовательно | $S_1(x)^2 = 0$. |
| „ „ $y$: | $S_1(y) = +8$, | „ | $S_1(y)^2 = 64$. |
| „ „ $z$: | $S_1(z) = +8$, | „ | $S_1(z)^2 = 64$. |

Сумма квадратов всех трех величин $S_1$ т.-е. $\Sigma S_1^2 = 128$.

Разделив эту сумму квадратов $\Sigma S_1^2 = 128$ на число измерений $n = 16$, получим:

$$\frac{\Sigma S_1^2}{n} = \frac{128}{16} = 8.$$

Эту величину (8) нам нужно теперь вычесть из суммы $S_2$, полученной в сводке (т.-е. в четвертом ряду, обозначенном буквою $u$). У нас

$$S_2(u) = 136.$$

Вычитая из 136 только-что полученную перед тем величину 8, получим 128. Итак, на основании указанной выше контрольной формулы, сумма всех величин $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ должна оказаться равной 128. Посмотрим, случится ли это на самом деле?

Так как

$$\Sigma x^2 = 16, \qquad \Sigma y^2 = 36, \qquad \text{а} \quad \Sigma z^2 = 76,$$

то

$$\Sigma x^2 + \Sigma y^2 + \Sigma z^2 = 16 + 36 + 76 = 128.$$

Следовательно в данном случае мы можем ручаться за правильность вычисления всех трех сумм квадратов центральных отклонений $\Sigma x^2$, $\Sigma y^2$, $\Sigma z^2$.

Эти вычисления удобно было бы расположить так:

$$\Sigma x^2 + \Sigma y^2 + \Sigma z^2 = S_2(u) - \frac{\Sigma S_1{}^2}{n}$$

$$\underbrace{16 + 36 + 76} = 136 - \frac{0 + 64 + 64}{16} = 136 - \frac{128}{16} = \underbrace{136 - 8}$$

$$128 = 128.$$

После вычисления и проверки всех трех сумм квадратов центральных отклонений $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ следует с их помощью определить и сумму произведений всех отклонений, взятых попарно, т.-е. величину $\Sigma xy$ по обычной формуле:

$$\Sigma xy = \frac{\Sigma z^2 - \Sigma y^2 - \Sigma x^2}{2} = \frac{76 - 36 - 16}{2} = \frac{76 - 52}{2} = \frac{+24}{2} = +12.$$

Все эти промежуточные вычисления, необходимые в конечном итоге для определения с их помощью коэффициента корреляции $r$, полезно тотчас же, по мере их нахождения и проверки, записывать в нижних частях расчетной решетки, которая после этого примет следующий окончательный вид (см. таблицу 78).

В этой решетке суммы $S_1$ и $S_2$ помечены уже без пояснительных буквенных значков $x$, $y$, $z$ и $u$, которые ранее были поставлены в скобках. Сделано это потому, что самое размещение этих сумм $S_1$ и $S_2$ по определенным частям решетки прямо уже указывает здесь на принадлежность их к тому или иному определенному ряду.



С помощью величин $\Sigma x^2$, $\Sigma y^2$ и $\Sigma xy$ мы легко определим и коэффициент корреляции $r$ по обычной формуле:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{+12}{\sqrt{16 \cdot 36}} = +0{,}5.$$

ТАБЛИЦА 78 *).

| Ряд $x$ | | | Ряд $y$ | | | Ряд $z$ | | | Контр. сводка (ряд $u$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 6 | 1 | $p$ | 14 | 8 | $p$ | 17 | 20 | $p$ | 37 | 29 |
| | | | | | | 1 | 1 | 1 | 1 | 1 | 1 |
| | | | | | | 1 | 2 | 3 | 1 | 2 | 3 |
| | | | 2 | 2 | 2 | 1 | 3 | 6 | 3 | 5 | 8 |
| 1 | 1 | 1 | 2 | 4 | 6 | 1 | 4 | 10 | 4 | 9 | 17 |
| 4 | 5 | — | 4 | 8 | — | 3 | 7 | — | 11 | 20 | — |
| 6 | — | — | 4 | — | — | 4 | — | — | 14 | — | — |
| 4 | 5 | — | 2 | 4 | — | 3 | 5 | — | 9 | 14 | — |
| 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 4 | 4 | 5 | 7 |
| | | | | | | 0 | 1 | 2 | 0 | 1 | 2 |
| | | | | | | 1 | 1 | 1 | 1 | 1 | 1 |
| 16 | 6 | 1 | 16 | 6 | 2 | 16 | 9 | 7 | 48 | 21 | 10 |
| $S_1 = 0$ | | | $S_1 = +8$ | | | $S_1 = +8$ | | | $S_1 = +16$ | | |
| $S_2 = 16$ | | | $S_2 = 40$ | | | $S_2 = 80$ | | | $S_2 = 136$ | | |
| $S_1^2 = 0$ | | | $S_1^2 = 64$ | | | $S_1^2 = 64$ | | | $\Sigma S_1{}^2 = 128$ | | |
| $\Sigma x^2 = 16$ | | | $\Sigma y^2 = 36$ | | | $\Sigma z^2 = 76$ | | | $\Sigma xy = +12$ | | |

Средняя ошибка этого коэффициента $m_r$ найдется по обычной формуле:

$$m_r = \pm \frac{1 - r^2}{\sqrt{n}} = \pm \frac{1 - (0{,}5)^2}{\sqrt{16}} = \pm 0{,}19.$$

Здесь, как и ранее, коэффициент корреляции $r$ оказался в 2,63 раза больше своей средней ошибки $m_r$, что указы-

*) При большом количестве вычислений полезно заготовить такие расчетные решетки типографским способом. Готовые решетки с формулами (форма № 2) имеются в книжном складе кооперативного товарищества „Современник“ в Ленинграде (пр. Володарского, 27).

вает на недостаточную достоверность заключения о наличии прямой связи между двумя заданными рядами чисел.

Итак, при вычислении коэффициента корреляции $r$ по способу сумм следует:

1. Составить корреляционную решетку и распределить в ней все коррелирующие пары вариант постановкой черточек с последующим подсчетом этих черточек в каждой клетке. Получим частоты $p$.

2. Сложить эти частоты по вертикальным, горизонтальным и диагональным направлениям. Получим частоты рядов: $x$, $y$ и $z$.

3. Составить расчетную решетку для одновременной обработки трех рядов $x$, $y$ и $z$ с контрольной сводкой ($u$), записать в нее эти частоты, составить частоты сводки, сложить числа во всех четырех столбцах $p$ и проверить вычисления, убедившись в том, что суммы всех частот в первых трех частях решетки равны друг другу ($=n$), а сумма частот в сводке равна числу в 3 раза большему (т.-е. $=3n$).

4. Расположив фигуру из 4 черточек на одном общем уровне во всех четырех частях решетки, проделать все вычисления, связанные с определением итогов: $a_1$, $a_2$, $b_1$ и $b_2$ для всех трех рядов и сводки, и проверить эти итоги обычным способом (сверяя суммы одноименных итогов в отдельных рядах с соответствующими итогами сводки).

5. Для каждого ряда и сводки вычислить обе вспомогательные суммы $S_1$ и $S_2$ и проверить их тем же способом.

6. Для каждого ряда вычислить суммы квадратов центральных отклонений $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ по обычной формуле:

$$\Sigma x^2 \text{ (или } \Sigma y^2, \text{ или } \Sigma z^2) = S_2 - \frac{S_1^2}{n}.$$

7. Проверить правильность вычисления этих сумм ($\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$) по контрольной формуле:

$$\Sigma x^2 + \Sigma y^2 + \Sigma z^2 = S_2(u) - \frac{\Sigma S_1^2}{n}.$$

8. Вычислить сумму произведений всех отклонений, взятых попарно ($\Sigma xy$), по формуле:

$$\Sigma xy = \frac{\Sigma z^2 - \Sigma y^2 - \Sigma x^2}{2}.$$



9. Вычислить коэффициент корреляции $r$ по формуле:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}.$$

10. Вычислить среднюю ошибку коэффициента корреляции $m_r$ по формуле:

$$m_r = \pm \frac{1 - r^2}{\sqrt{n}}.$$

11. Оценить достоверность наличия обнаруженной зависимости, убедившись в том, что коэффициент корреляции $r$ превышает свою ошибку $m_r$ по крайней мере в 3 раза.

Разобрав на предыдущем упрощенном примере общую схему определения коэффициента корреляции $r$ по способу сумм, проделаем эти вычисления еще раз, воспользовавшись теперь уже более обширным и сложным материалом. Пусть, напр., имеем следующие 50 коррелирующих пар:

| | | | | |
|---|---|---|---|---|
| 6,4—58 | 6,2—63 | 12,6—64 | 6,8—63 | 6,7—63 |
| 3,2-56 | 11,0—64 | 5,3-71 | 16,0—55 | 14,1-65 |
| 8,1—66 | 9,9—64 | 7,9-57 | 11,2—68 | 12,3-65 |
| 7,8—62 | 13,6—65 | 5,1—58 | 11,9—60 | 11,6—64 |
| 10,1—63 | 8,7—62 | 5,9—51 | 5,4—69 | 13,7—59 |
| 9,3—67 | 9,0—61 | 9,6-53 | 4,2—69 | 8,5 58 |
| 10,9-67 | 10,4—57 | 13,2-52 | 12,8—68 | 16,7—56 |
| 5,8 66 | 10,3—59 | 7,5—70 | 9,4—70 | 14,5—59 |
| 8,3—61 | 15,8—75 | 7,3—70 | 9,7—61 | 7,2—57 |
| 11,2-63 | 10,8—73 | 11,7—60 | 10,8—63 | 11,4—56 |

В этом списке имеется 5 столбцов, в каждом из которых заключается по 10 соответствующих друг другу парных вариант. Числа, записанные во всех столбцах слева, представляют собой варианты ряда $x$ (т.-е. $V_x$), а стоящие справа — соответствующие им варианты второго ряда $y$ (т.-е. $V_y$).

Для разноски этих пар по клеткам корреляционной решетки следует прежде всего установить границы классов для ряда $x$ и, независимо от этого, для ряда $y$.

Начнем с ряда $x$ (т.-е. чисел, стоящих слева). Наибольшая варианта ($max\ V_x$) равна здесь 16,7 или, после округления, 17, а наименьшая ($min\ V_x$) $=3{,}2$ или, округляя, 3.

Итак, для первого ряда:

$$\begin{array}{l} max\ V_x \text{ прибл.} = 17 \\ min\ V_x \quad " \quad = 3 \\ \hline \text{разница прибл.} = 14. \end{array}$$

Разобьем этот ряд для простоты всего на 7 классов (в действительности желательно в этих случаях установить классов 10—15). Тогда классовый промежуток для первого ряда, который мы обозначим через $k_x$, найдется путем деления наибольшего размаха колебаний вариант этого ряда (т.-е. 14) на принятое число классов (7):

$$k = \frac{14}{7} = 2.$$

При этих условиях границами всех 7-ми классов ряда $x$ будут служить следующие числа:

от 17 до 15,1
" 15 " 13,1
" 13 " 11,1
" 11 " 9,1
" 9 " 7,1
" 7 " 5,1
" 5 " 3,1

Теперь установим таким же образом границы классов и для ряда $y$. Здесь:

$$\begin{array}{l} max\ V_y = 75 \\ min\ V_y = 51 \\ \hline \text{разница} = 24. \end{array}$$

Разобьем этот ряд, напр., на 5 классов. Тогда классовый промежуток для ряда $y$

$$k_y = \frac{24}{5} = 4{,}8 \text{ или, округляя, } k_y = 5.$$

Следовательно ряд $y$ разобьется на следующие классы:

от 75 до 71
" 70 " 66
" 65 " 61
" 60 " 56
" 55 " 51.



После этого составим разносную корреляционную решетку, придав ей форму, указанную на таблице 79.

Здесь границы классов ряда $x$ записаны сверху и идут, как всегда, увеличиваясь в направлении слева направо, а границы классов ряда $y$ помещены слева в убывающем порядке (сверху бóльшие, снизу меньшие). Этот порядок расположения классов является обязательным и, во избежание ошибок, его следует строго придерживаться (можно было бы, конечно, принять и другой способ записи этих классов, но тогда следовало бы изменить в соответствии с этим и порядок всех последующих вычислений).

ТАБЛИЦА 79.

| $V_y$ \ $V$ | 3,1—5 | 5,1—7 | 7,1—9 | 9,1 - 11 | 11,1—13 | 13,1—15 | 15,1—17 |
|---|---|---|---|---|---|---|---|
| 75—71 | | | | | | | |
| 70—66 | | | | | | | |
| 65—61 | | | | | | | |
| 60 - 56 | | / | | | | | |
| 55—51 | | | | | | | |

Составив разносную решетку, следует каждую пару коррелирующих вариант, помещенных в нашем первоначальном списке, отметить в определенной клетке этой решетки путем постановки в ней одной черточки. Первая пара вариант обозначена в списке числами 6,4 и 58. Это значит, что $V_x = 6{,}4$ соответствует $V_y = 58$. Так как варианта первого ряда $V_x (= 6{,}4)$ относится к классу с границами от 5,1 до 7, а соответствующая ей варианта второго ряда $V_y (= 58)$ — к классу с границами от 60 до 56, то данную пару этих вариант мы должны будем зарегистрировать в той клетке нашей решетки, которая лежит на пересечении второго слева вертикального столбца и четвертой сверху горизонтальной строчки. В этой клетке на табл. 79 и поставлена первая вертикальная черточка.

Отметив черточками все остальные пары наших вариант в клетках этой решетки, получим полную картину разноски, представленную на табл. 80.

Теперь подсчитаем количество черточек в каждой клетке этой решетки и полученные частоты $p$ запишем в соответствующие клетки корреляционной решетки, которую начер-

ТАБЛИЦА 80.

| $V_y$ \ $V_x$ | 3,1—5 | 5,1—7 | 7,1—9 | 9,1—11 | 11,1—13 | 13,1—15 | 15,1—17 |
|---|---|---|---|---|---|---|---|
| 75—71 | | / | | / | | | / |
| 70—66 | / | // | /// | /// | // | | |
| 65—61 | | /// | //// | ///// | //// | // | |
| 60—56 | / | // | | // | /// | // | / |
| 55—51 | | / | | / | | / | / |

тим уже без обозначения классовых границ (надобность в них отпала с момента окончания разноски). В результате получим картину, представленную на таблице 81.

После этого нам нужно подсчитать частоты ряда $x$ ряда $y$ и ряда $z$.

Для получения частот ряда $x$ следует сложить все числа, стоящие в клетках этой решетки по вертикальным столбцам. Частоты ряда $y$ найдутся, если сложить эти же числа по направлению горизонтальных строчек. Наконец, для определения частот ряда $z$ мы должны будем сложить эти же самые числа по направлению диагональных клеток (от левого верхнего угла к правому нижнему). Результаты этого подсчета мы запишем так: суммы чисел отдельных вертикальных столбцов поместим сверху, суммы чисел горизонтальных строчек—слева, а результаты диагонального сумми-

ТАБЛИЦА 81.

| | | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 1 | | | 1 |
| 1 | 2 | 3 | 3 | 2 | | |
| | 3 | 4 | 5 | 4 | 2 | |
| 1 | 2 | 3 | 2 | 3 | 2 | 1 |
| | 1 | | 1 | | 1 | 1 |

рования—снизу и справа против стрелок пунктирных линий, указывающих направление этого суммирования (см. таблицу 82).



Итак, в данном случае частотами ряда $x$ будут служить следующие числа:

2, 9, 10, 12, 9, 5 и 3;

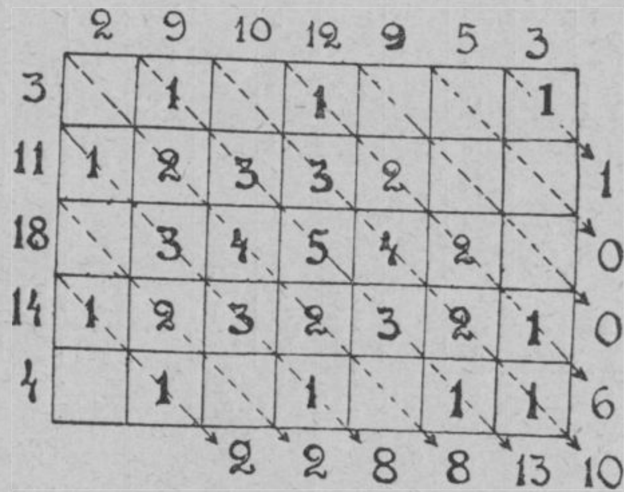
частотами ряда $y$:

3, 11, 18, 14 и 4,

а частотами ряда $z$:

1, 0, 0, 6, 10, 13, 8, 8, 2 и 2.

ТАБЛИЦА 82.




Теперь начертим расчетную решетку, запишем в нее эти частоты, составим частоты сводки, сложим числа во всех четырех столбцах $p$ и проверим вычисления, убедившись в том, что суммы частот в трех основных рядах ($x$, $y$ и $z$) равны общему числу измерений $n$ (т.-е. 50), а сумма частот в сводке равна $3n$ (т.-е. 150). После этого наша решетка примет вид, указанный на таблице 83.

Далее проделаем в расчетной решетке все вычисления, необходимые для определения итогов: $a_1$ и $a_2$, $b_1$ и $b_2$, проверим эти вычисления, сличая суммы одноименных итогов с соответствующими итогами сводки, найдем для каждого ряда и сводки обе вспомогательные суммы $S_1$ и $S_2$, проверим их тем же способом и определим три суммы квадратов центральных отклонений $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ (см табл. 84).

Эти суммы необходимо проверить по контрольной формуле:

$$\Sigma x^2 + \Sigma y^2 + \Sigma z^2 = S_2(u) - \frac{\Sigma S_1^2}{n}.$$

ТАБЛИЦА 83.

| Ряд $x$ | | | Ряд $y$ | | | Ряд $z$ | | | Контр. сводка (ряд $u$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | | | $p$ | | | $p$ | | | $p$ | | |
| | | | | | | 1 | | | 1 | | |
| | | | | | | 0 | | | 0 | | |
| 3 | | | | | | 0 | | | 3 | | |
| 5 | | | 3 | | | 6 | | | 14 | | |
| 9 | | | 11 | | | 10 | | | 30 | | |
| 12 | | | 18 | | | 13 | | | 43 | | |
| 10 | | | 14 | | | 8 | | | 32 | | |
| 9 | | | 4 | | | 8 | | | 21 | | |
| 2 | | | | | | 2 | | | 4 | | |
| | | | | | | 2 | | | 2 | | |
| 50 | | | 50 | | | 50 | | | 150 | | |
| $S_1 =$ | | | $S_1 =$ | | | $S_1 =$ | | | $S_1 =$ | | |
| $S_2 =$ | | | $S_2 =$ | | | $S_2 =$ | | | $S_2 =$ | | |
| $S_1^2 =$ | | | $S_1^2 =$ | | | $S_1^2 =$ | | | $\Sigma S_1^2 =$ | | |
| $\Sigma x^2 =$ | | | $\Sigma y^2 =$ | | | $\Sigma z^2 =$ | | | $\Sigma xy =$ | | |

В нашем примере:

$$\Sigma x^2 = 119{,}28 \qquad \Sigma y^2 = 52{,}50 \qquad \Sigma z^2 = 146{,}58$$

$S_2$ для сводки, т.-е. $S_2(u) = 322$

$$\Sigma S_1^2 = 36 + 25 + 121 = 182.$$

Подставляя эти данные в контрольную формулу получим:

$$119{,}28 + 52{,}50 + 146{,}58 = 322 - \frac{182}{50}$$

$$318{,}36 = 322 - 3{,}64 = 318{,}36.$$



Следовательно суммы $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$ вычислены правильно.

Далее определим величину $\Sigma xy$ по формуле:

$$\Sigma xy = \frac{\Sigma z^2 - \Sigma y^2 - \Sigma x^2}{2}.$$

ТАБЛИЦА 84.

| Ряд $x$ | | | Ряд $y$ | | | Ряд $z$ | | | Контр. сводка (ряд $u$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 28 | 14 | $p$ | 17 | 3 | $p$ | 27 | 16 | $p$ | 72 | 33 |
| | | | | | | 1 | 1 | 1 | 1 | 1 | 1 |
| | | | | | | 0 | 1 | **2** | 0 | 1 | 2 |
| 3 | 3 | 3 | | | | 0 | 1 | 3 | 3 | 4 | 6 |
| 5 | 8 | 11 | 3 | 3 | 3 | 6 | 7 | 10 | 14 | 18 | 24 |
| 9 | 17 | — | 11 | 14 | — | 10 | 17 | — | 30 | 48 | — |
| 12 | — | — | 18 | — | — | 13 | — | — | 43 | — | — |
| 10 | 21 | — | 14 | 18 | — | 8 | 20 | — | 32 | 59 | — |
| 9 | 11 | 13 | 4 | 4 | 4 | 8 | 12 | 18 | 21 | 27 | 35 |
| 2 | **2** | **2** | | | | **2** | 4 | 6 | 4 | 6 | 8 |
| | | | | | | 2 | 2 | 2 | **2** | 2 | 2 |
| 50 | 34 | 15 | 50 | 22 | 4 | 50 | 38 | 26 | 150 | 94 | 45 |
| $S_1 = -6$ | | | $S_1 = 5$ | | | $S_1 = -11$ | | | $S_1 = -22$ | | |
| $S_2 = 120$ | | | $S_2 = 53$ | | | $S_2 = 149$ | | | $S_2 = 322$ | | |
| $S_1^2 = 36$ | | | $S_1^2 = 25$ | | | $S_1^2 = 121$ | | | $\Sigma S_1^2 = 182$ | | |
| $\Sigma x^2 = 119{,}28$ | | | $\Sigma y^2 = 52{,}50$ | | | $\Sigma z^2 = 146{,}58$ | | | $\Sigma xy = -12{,}60$ | | |

Для нашего примера, следовательно:

$$\Sigma xy = \frac{146{,}58 - 52{,}50 - 119{,}28}{2} = \frac{146{,}58 - 171{,}78}{2} =$$

$$= \frac{-25{,}20}{2} = -12{,}6.$$

После записи всех этих результатов в расчетную решетку, последняя примет окончательный вид, указанный на табл. 84.

Получив все три величины, нужные для определения коэффициента корреляции $r$ (т.-е. $\Sigma x^2$, $\Sigma y^2$ и $\Sigma xy$), подставим их в соответствующую формулу:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{-12{,}60}{\sqrt{119{,}28 \cdot 52{,}50}} = \frac{-12{,}60}{\sqrt{6262{,}20}}$$

Так как $\sqrt{6262{,}2}$ прибл. $= 79{,}13$, то

$$r = \frac{-12{,}60}{79{,}13} = -0{,}16.$$

Ошибка $m_r$ этого коэффициента найдется по формуле:

$$m_r = \pm \frac{1 - r^2}{\sqrt{n}} = \pm \frac{1 - (-0{,}16)^2}{\sqrt{50}} = \pm \frac{1 - 0{,}0256}{7{,}07} =$$

$$= \pm \frac{0{,}9744}{7{,}07} = \pm 0{,}14.$$

Итак, зависимость между двумя заданными рядами чисел выражается здесь следующим значением коэффициента корреляции и его средней ошибки:

$$r = -0{,}16 \pm 0{,}14.$$

Так как в данном случае коэффициент корреляции $r$ превосходит свою среднюю ошибку $m_r$ всего лишь в 1,14 раза (а не в 3 раза, как это требуется), то обнаруженная здесь слабая обратная зависимость между заданными рядами чисел не может еще считаться доказанной и, повидимому, является лишь чисто-случайной.

Имея под руками таблицы логарифмов, можно значительно упростить вычисление коэффициента корреляциии в конечной части этой работы, т.-е. при пользовании формулой:

$$r = \frac{\Sigma yx}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}.$$

Для этого следует отыскать по таблице логарифмы всех трех чисел: $\Sigma x^2$, $\Sigma y^2$ и $\Sigma xy$, сложить логарифмы чисел: $\Sigma x^2$ и $\Sigma y^2$, сумму их разделить на 2, найденную величину вычесть из логарифма числа $\Sigma xy$ и по полученному в результате логарифму отыскать в этих же таблицах соответствующее число. Это число и будет служить искомым коэффициентом корреляции $r$.



Проделаем эти вычисления с данными последнего примера. Там мы имели:

$$\Sigma x^2 = 119{,}28 \qquad \Sigma y^2 = 52{,}50 \qquad \Sigma xy = -12{,}60.$$

Отыскав по таблице логарифмы этих чисел, получим:

$$\begin{aligned} lg\,\Sigma x^2 &= lg\ 119{,}28 = 2{,}07657 \\ lg\,\Sigma y^2 &= lg\ 52{,}50 = 1{,}72016 \\ lg\,\Sigma xy &= lg\ 12{,}60 = 1{,}10037. \end{aligned}$$

Число $\Sigma xy$ взято здесь без всякого знака, так как отрицательные числа, как известно, логарифмов не имеют. Соответствующую поправку мы внесем в самом конце наших вычислений, просто поставив у полученного коэффициента $r$ знак минус (знак коэффициента корреляции всегда определяется знаком величины $\Sigma xy$; если $\Sigma xy$ получилась со знаком $+$, то и $r$ будет величиной положительной; если $\Sigma xy$ оказалась отрицательной, то и коэффициент корреляции получится со знаком $-$).

Теперь нам нужно сложить логарифмы чисел $\Sigma x^2$ и $\Sigma y^2$:

$$\begin{array}{r} +\ \begin{array}{r} 2{,}07657 \\ 1{,}72016 \end{array} \\ \hline 3{,}79673 \end{array}$$

Эту сумму логарифмов следует разделить пополам

$$\frac{3{,}79673}{2} = 1{,}89836.$$

Полученный логарифм нужно, далее, вычесть из логарифма числа $\Sigma xy$, т.-е. из 1,10037.

$$\begin{array}{r} -\ \begin{array}{r} 1{,}10037 \\ 1{,}89836 \end{array} \\ \hline \bar{1}{,}20201 \end{array}$$

Последнее вычитание следует выполнить по общим правилам и лишь в самом конце действия, при вычитании целых единиц (стоящих слева от запятой), отметить получение отрицательного количества постановкой минуса над соответствующей цифрой (в данном случае над 1).

Отыскав по таблицам то число, которому соответствует найденный здесь логарифм, мы и получим коэффициент корреляции $r$.

Так как логарифму 1,20201 приблизительно соответствует число 0,1592, то искомый коэффициент корреляции

$$r = -0{,}1592$$

или ограничиваясь двумя знаками после запятой,

$$r = -0{,}16,$$

что мы имели и раньше.

Здесь знак — прибавлен на том основании, что величина $\Sigma xy$, определяющая этот знак, была отрицательной.

---

## XVI. Вариационный коэффициент ($v$), показатель точности (P) и коэффициенты регрессии (R) *).

В предыдущих главах были разобраны способы определения среднего арифметического $M$, среднего квадратического отклонения $\sigma$ и коэффициента корреляции $r$ с их средними ошибками $m$, $m_\sigma$ и $m_r$, причем для вычисления их были указаны следующие основные формулы:

$$M = \frac{\Sigma V}{n} \quad \text{или} \quad M = C + \frac{kS_1}{n}$$

$$\sigma = \pm k\sqrt{\frac{\Sigma x^2}{n-1}} \qquad r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

$$m = \pm\frac{\sigma}{\sqrt{n}} \quad \text{или} \quad m = \pm k\sqrt{\frac{\Sigma x^2}{n(n-1)}}$$

$$m_\sigma = \pm\frac{\sigma}{\sqrt{2n}} \qquad m_r = \pm\frac{1-r^2}{\sqrt{n}}$$

*) Эти три сводные характеристики искуственно объединены в одной общей главе по причине сравнительно редкого применения их на практике. Чтобы не отвлекать внимание читателей от основных методов вычисления, описанию показателя и коэффициентов ($v$, $P$ и $R$) уделена эта специальная глава.



Вспомним, что в приведенной здесь формуле для вычисления среднего квадратического отклонения ($\sigma$)

$$\sigma = \pm k\sqrt{\frac{\Sigma x^2}{n-1}}$$

сумму квадратов центральных отклонений $\Sigma x^2$ мы должны были делить под корнем на $n-1$, а не на $n$, лишь при малом числе измерений $n$. При большом же $n$ эта разница между $n-1$ и $n$ по сравнению с самим $n$ оказывается настолько незначительной, что без всякого ущерба для точности мы можем пользоваться формулой

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

или, при классовом промежутке $k = 1$, полагать:

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

Далее заметим, что среднюю ошибку сигмы ($m_\sigma$) можно определить также и с помощью ошибки $m$ среднего арифметического ($M$) по формуле:

$$m_\sigma = \frac{m}{\sqrt{2}}$$

Результаты применения этой и указанной выше формулы окажутся всегда совершенно точно совпадающими друг с другом, в чем нетрудно убедиться на любом числовом примере.

Полезно будет указать здесь также и другую формулу для вычисления коэффициента корреляции $r$, удобную в том случае, когда предварительно были определены средние квадратические отклонения обоих коррелирующих рядов (обозначим их через $\sigma_x$ и $\sigma_y$). Если, при достаточно большом числе измерений $n$, оба эти средние квадратические отклонения ($\sigma_x$ и $\sigma_y$) были вычислены без поправки на малое число измерений, т.-е. по формуле

$$\sigma = \sqrt{\frac{\Sigma x^2}{n}}, \text{ а не по формуле } \sigma = \sqrt{\frac{\Sigma x^2}{n-1}}$$

то искомый коэффициент корреляции $r$ можно определить также и по формуле

$$r = \frac{\Sigma xy}{n\,\sigma_x\,\sigma_y}$$

Пусть, например, имеем следующие два ряда чисел:

Ряд $x$: 7 9 10 11 и 13

Ряд $y$: 9 17 15 13 и 21.

Для первого из этих рядов:

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n}} = \sqrt{\frac{20}{5}} = \sqrt{4} = \pm 2,$$

а для второго:

$$\sigma_y = \sqrt{\frac{80}{5}} = \sqrt{16} = \pm 4.$$

Коэффициент корреляции, вычисленный по обычной формуле:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{+32}{\sqrt{20.80}} = \frac{+32}{\sqrt{1.600}} = \frac{+32}{40} = +0{,}8$$

По второй формуле получим те же результаты:

$$r = \frac{\Sigma xy}{n\,\sigma_x\,\sigma_y} = \frac{+32}{5 \cdot 2 \cdot 4} = \frac{+32}{40} = +0{,}8.$$

В данном случае конечные результаты совпали совершенно точно. Если же обе сигмы ($\sigma_x$ и $\sigma_y$) мы определили бы по обычной формуле (с поправкой на малое число измерений, т. е. разделив под корнем $\Sigma x^2$ на $n-1$), то, пользуясь этой формулой для вычисления $r$, мы получили бы результат лишь приближенный и тем менее точный, чем из меньшего числа измерений ($n$) пришлось бы определять $r$. При $n = 250$ и больше эта разница в конечных результатах может сказаться лишь на 3-м знаке коэффициента $r$, не отражаясь вовсе на его втором знаке после запятой.

Введем теперь некоторые дополнительные числовые характеристики отдельного вариационного ряда и связи между двумя рядами, указав попутно и формулы для их вычисления.

1. **Вариационный коэффициент** ($v$). Средний размах колебаний отдельных измерений какого-либо варьирующего



признака в обе стороны от среднего арифметического $M$, как известно, характеризуется величиной среднего квадратического отклонения ($\sigma$), и при сравнении средней амплитуды колебаний отдельных вариант в двух или нескольких рядах мы можем судить о различиях в их варьировании по величине соответствующих сигм.

Так, например, если известно, что в каком либо возрасте рост мальчиков при средней его величине $M_1 = 154$ *см* дает колебания в обе стороны от этого $M_1$ со средним размахом $\sigma_1 = \pm 9$ *см*, а у девочек этого же возраста при $M_2 = 153$ *см* $\sigma = \pm 6$ *см*, то судить о варьировании роста мальчиков и девочек в этом возрасте можно путем простого сопоставления друг с другом соответствующих сигм: $\sigma_1 = 9$ *см* и $\sigma_2 = 6$ *см*. Так как в данном случае $\sigma_1$ оказалась больше $\sigma_2$ на целых 3 *см*, то варьирование роста мальчиков для данного возраста должно быть признано более широким (достоверность этого заключения, очевидно, подлежит еще выяснению с помощью обычного критерия, учитывающего величину соответствующих средних ошибок, но для нас это в данном случае не представляет интереса).

Подобное заключение о величине варьирования на основании простого сопоставления друг с другом двух сигм ($\sigma_1$ и $\sigma_2$) мы имели право сделать здесь лишь благодаря тому, что средний рост мальчиков $M_1 = 154$ *см* мало чем отличался от среднего роста девочек $M_2 = 153$ *см*. Приблизительное равенство двух этих исходных величин позволило нам учитывать лишь абсолютную величину варьирования, выраженную с помощью сигм, не прибегая к понятию об относительном варьировании, т. е. о варьировании на ту или иную долю самого среднего роста $M$.

Рассмотрим такой отвлеченный числовой пример:

в I ряду: $M_1 = 40$ $\sigma_1 = 6$

во II ряду: $M_2 = 25$ $\sigma_2 = 5$.

Спрашивается, какой из этих рядов варьирует в более широких пределах? Абсолютно, конечно, первый, так как $\sigma_1$ больше $\sigma_2$.

Но в данном случае среднее арифметическое $M_1$ уже не равно $M_2$, и бóльшее варьирование первого ряда ($\sigma_1 = 6$) по отношению к соответствующему среднему арифметическому $M_1=40$ составляет меньшую величину, чем варьирование второго ряда ($\sigma_2=5$) по отношению к своему среднему арифметическому $M_2=25$.

Выразим для наглядности обе эти сигмы в процентах от соответствующих средних ($M$).

Для I ряда:

$$\frac{100.6}{40}=\frac{600}{40}=15\%.$$

Для II ряда:

$$\frac{100.5}{25}=\frac{500}{25}=20\%.$$

Итак, $\sigma_1=6$, характеризующая среднюю величину варьирования первого ряда, составляет от среднего арифметического $M_1$ этого ряда всего лишь $15\%$, в то время как $\sigma_2=5$, характеризующая средний размах колебаний чисел второго ряда, составляет от среднего арифметического $M_2$ этого ряда уже $20\%$. Таким образом, относительное варьирование (по сравнению с величиной соответствующих средних арифметических $M$) во втором ряду оказалось большим, в то время как абсолютное варьирование, выраженное величиной самих сигм, было, наоборот, большим в первом ряду. В тех случаях, когда по сути вопроса интерес представляет не абсолютная величина варьирования (выраженная с помощью $\sigma$), а относительная, показывающая, на какую долю самого среднего арифметического $M$ отдельные варианты колеблются в обе стороны около этого $M$, полезно вычислить процентное отношение $\sigma$ к $M$, называемое коэффициентом вариации (или вариационным коэффициентом). Обозначив этот коэффициент через $v$ (малое), будем иметь следующую формулу для его вычисления:

$$v=\frac{100\sigma}{M}\%.$$

Пусть, например, в нашем распоряжении имеются числовые результаты обработки двух вариационных рядов, относящиеся к детям одного и того же пола в каком-либо из старших возрастов.

Для роста стоя:

$$M_1=155{,}7 \qquad \sigma_1=6{,}10.$$



Для роста сидя:

$$M_2=82{,}9 \qquad \sigma_2=3{,}24.$$

Желая сравнить варьирование роста стоя этих детей с варьированием их роста сидя, мы прежде всего пытаемся разрешить эту задачу путем сопоставления друг с другом соответствующих сигм. В данном случае $\sigma_1$ почти вдвое больше $\sigma_2$. Это дает нам право заключить, что абсолютная величина варьирования роста стоя в данном случае оказывается большей. Посмотрим теперь, какую долю (какой процент) составляют эти средние размахи колебаний роста стоя ($\sigma_1$) и роста сидя ($\sigma_2$) отдельных детей от соответствующих средних арифметических $M_1$ и $M_2$. Вычислив оба вариационных коэффициента ($v$), получим:

Для роста стоя:

$$v_1=\frac{100\cdot 6{,}10}{155{,}7}=3{,}92\%.$$

Для роста сидя:

$$v_2=\frac{100\cdot 3{,}24}{82{,}9}=3{,}91\%.$$

Итак, относительное варьирование этих размеров оказалось почти одинаковым. Другими словами, рост стоя варьирует здесь в более широких пределах лишь в силу того, что и отдельные варианты этого ряда оказываются значительно больше вариант второго ряда, по сравнению же с их средней величиной размахи колебаний в обоих случаях оказываются приблизительно равными.

Вариационный коэффициент $v$, будучи отвлеченным числом (выраженным в $\%\%$), дает возможность сравнивать варьирование признаков, измеряемых в единицах различных наименований, например, варьирование роста с варьированием веса. Простое сравнение соответствующих сигм здесь явно неприменимо, так как оба средних квадратических отклонения выражены здесь в различных единицах измерения ($\sigma$ роста в *см*, а $\sigma$ веса в *кг*). Сравнение же вариационных коэффициентов и в этом случае является допустимым, так как каждый из них указывает лишь на ту долю, которую составляет сигма от соответствующего $M$ (в $\%\%$).

Пусть, например, имеем следующие данные:

Для роста:

$$M_1 = 138{,}21\ см \qquad \sigma_1 = 7{,}41\ см.$$

Для веса:

$$M_2 = 32{,}30\ кг \qquad \sigma_2 = 4{,}74\ кг.$$

В силу полной невозможности сравнения $\sigma_1 = 7{,}41$, выраженной в числе сантиметров, с $\sigma_2 = 4{,}74$, выраженной в килограммах, остается лишь вычислить вариационные коэффициенты.

Для роста:

$$v_1 = \frac{100 \cdot 7{,}41}{138{,}21} = 5{,}36\%.$$

Для веса:

$$v_2 = \frac{100 \cdot 4{,}74}{32{,}30} = 14{,}67\%.$$

Эти две отвлеченные величины: $v_1 = 5{,}36$ и $v_2 = 14{,}67$, выраженные в одинаковых единицах (в %%), мы можем уже подвергнуть сравнению. В данном случае относительное варьирование веса оказалось больше относительного варьирования роста (в весе наблюдается большее разнообразие).

При сравнении друг с другом двух величин, с целью выяснения обнаруженного между ними различия, мы всегда должны оценить также и достоверность этого различия путем применения обычного критерия, определяющего соотношение между данной разницей и ее средней ошибкой (см. главу V).

Поэтому для сопоставления двух вариационных коэффициентов $v_1$ и $v_2$ необходимо вычислить также и их средние ошибки $m_{v_1}$ и $m_{v_2}$. Ошибка вариационного коэффициента $v$ вычисляется по формуле:

$$m_v = \frac{v\sqrt{0{,}5 + \left(\frac{v}{100}\right)^2}}{\sqrt{n}},$$

показывающей, что данный коэффициент $v$ следует разделить на 100, полученное частное возвысить в квадрат, к этой величине прибавить всегда одно и то же постоянное число



0,5, из полученной суммы извлечь квадратный корень, помножить его на величину самого коэффициента $v$ и это произведение разделить на квадратный корень из числа измерений $n$, т.-е. на $\sqrt{n}$.

Пусть, например,

$$v = 40\%, \quad а \quad n = 66.$$

Разделив $v = 40$ на 100; получим 0,4. Возвысив 0,4 в квадрат, будем иметь 0,16. К этой величине нужно теперь прибавить 0,5, в результате чего, очевидно, получим 0,66. Корень из 0,66 приблизительно равен 0,8124. Помножив 0,8124 на $v = 40$, получим 32,496.

Теперь эту величину 32,496 следует разделить еще на $\sqrt{n}$, т.-е. на $\sqrt{66} = 8{,}124$. Разделив 32,496 на 8,124, получим 4. Следовательно, ошибка вариационного коэффициента

$$m_v = \pm 4\%.$$

Все эти действия можно было бы записать в следующую строчку:

$$m_v = \frac{v\sqrt{0{,}5 + \left(\frac{v}{100}\right)^2}}{\sqrt{n}} = \frac{40\sqrt{0{,}5 + \left(\frac{40}{100}\right)^2}}{\sqrt{66}} =$$

$$= \frac{40\sqrt{0{,}5 + (0{,}4)^2}}{8{,}124} = \frac{40\sqrt{0{,}5 + 0{,}16}}{8{,}124} = \frac{40\sqrt{0{,}66}}{8{,}124} = \frac{40 \cdot 0{,}8124}{8{,}124} =$$

$$= \frac{32{,}496}{8{,}124} = \pm 4\%.$$

В разобранном выше примере (варьирования роста и веса) мы получили:

Для роста:

$$v_1 = 5{,}36\%.$$

Для веса:

$$v_2 = 14{,}67\%.$$

Допустим, что оба эти вариационные коэффициента были вычислены при $n = 64$. Тогда соответствующие ошибки обоих коэффициентов $v$ определяются по формулам:

$$m_{v_1} = \frac{5{,}36\sqrt{0{,}5 + \left(\frac{5{,}36}{100}\right)^2}}{\sqrt{64}} = \frac{5{,}36\sqrt{0{,}5 + 0{,}0536^2}}{8} =$$

$$= \frac{5{,}36\sqrt{0{,}5 + 0{,}00287296}}{8} = \frac{5{,}36\sqrt{0{,}50287296}}{8} = \frac{5{,}36 \cdot 0{,}71}{8} =$$

$$= \frac{3{,}8056}{8} = \pm 0{,}48.$$

$$m_{v_2} = \frac{14{,}67\sqrt{0{,}5 + \left(\frac{14{,}67}{100}\right)^2}}{\sqrt{64}} = \frac{14{,}67\sqrt{0{,}5 + 0{,}1467^2}}{8} =$$

$$= \frac{14{,}67\sqrt{0{,}5 + 0{,}02152089}}{8} = \frac{14{,}67\sqrt{0{,}52152089}}{8} = \frac{14{,}67 \cdot 0{,}72}{8} =$$

$$= \frac{10{,}5624}{8} = \pm 1{,}32.$$

Итак, при сравнении варьирования роста с варьированием веса, мы получили:

Для роста:

$$v_1 = 5{,}36 \qquad m_{v_1} = \pm 0{,}48.$$

Для веса:

$$v_2 = 14{,}67 \qquad m_{v_2} = \pm 1{,}32.$$

Чтобы оценить достоверность различия в величине коэффициентов вариации $v_1$ и $v_2$, следует применить обычный критерий (см. главу V).

$$\frac{(v_1 - v_2)^2}{m_{v_1}^2 + m_{v_2}^2} > 9.$$

Подставляя сюда найденные значения $v_1$ $v_2$ $m_{v_1}$ $m_{v_2}$, получим:

$$\frac{(5{,}36 - 14{,}67)^2}{(0{,}48)^2 + (1{,}32)^2} = \frac{(-9{,}31)^2}{0{,}2304 + 1{,}7424} = \frac{86{,}6761}{1{,}9728} = 43{,}94.$$

Так как в данном случае величина

$$\frac{(v_1 - v_2)^2}{m_{v_1}^2 + m_{v_2}^2} = 43{,}94,$$

т.-е. оказалась значительно больше 9, мы можем признать различие между $v_1$ и $v_2$ доказанным.



При определении ошибки вариационного коэффициента по формуле

$$m_v = \frac{v\sqrt{0{,}5 + \left(\frac{v}{100}\right)^2}}{\sqrt{n}}$$

особенно кропотливыми являются выкладки при вычислении выражения стоящего в числителе этой дроби, т.-е.

$$v\sqrt{0{,}5 + \left(\frac{v}{100}\right)^2}.$$

Для облегчения и ускорения работы по вычислению этого выражения, в конце книги приложена особая таблица значений величины $\gamma$ (гамма)

$$\gamma = v\sqrt{0{,}5 + \left(\frac{v}{100}\right)^2}$$

для всех последовательных значений вариационного коэффициента от 1 до 125 включительно.

Таким образом средняя ошибка вариационного коэффициента при определении ее с помощью этой таблицы может быть вычислена по следующей упрощенной формуле:

$$m_v = \frac{\gamma}{\sqrt{n}},$$

где величина $\gamma$ для данного $v$ определяется по специальной таблице (см. стран. 288).

Пусть, например, $v = 35$, а $n = 85$. В левом столбце таблицы, озаглавленном буквою $v$, находим значение $v = 35$. Рядом с этим числом, справа, в соседнем столбце, озаглавленном буквою $\gamma$, стоит число 27,61. Следовательно для $v = 35$ величина $\gamma = 27{,}61$.

Подставляя это значение $\gamma$ в упрощенную формулу, получим:

$$m_v = \frac{\gamma}{\sqrt{n}} = \frac{27{,}61}{\sqrt{85}} = \frac{27{,}61}{9{,}22} = \pm 2{,}99.$$

Итак, при $v = 35$ и $n = 85$ ошибка вариационного коэффициента

$$m_v = \pm 2{,}99.$$

Если вариационный коэффициент $v$ не является целым числом, а представляет собой целое с дробью, то по той же таблице в рубрике $v$ следует отыскать ближайшее к нему меньшее целое число (т.-е. просто отбросить все десятичные знаки после запятой) и в графе $\gamma$ найти приближенное значение величины $\gamma$. Далее, в самом крайнем (справа) столбце, озаглавленном в таблице буквою $\Delta$ (дельта), следует найти число, стоящее рядом с только что найденным приближенным значением $\gamma$. Эта величина ($\Delta$) представляет собой $\frac{1}{10}$ той разности между данной $\gamma$ и следующей (ниже ее стоящей) $\gamma$, которая соответствует $\frac{1}{10}$ самого вариационного коэффициента $v$. К приближенному значению $\gamma$ теперь надо прибавить эту разность $\Delta$, предварительно увеличив ее во столько раз, сколько десятых было отброшено у $v$.

Пусть, например, $v = 48{,}3$, а $n$ по прежнему $= 85$. Отбросив 3 десятых (т.-е. 0,3), ищем в столбце $v$ число 48. Рядом с ним в графе $\gamma$ стоит число 41,02. Это и есть приближенная величина $\gamma$, соответствующая $v = 48$. Но наш вариационный коэффициент в действительности больше этого приближенного значения (48) на 3 десятых (т.-е. на 0,3). Поэтому и величина $\gamma$ в действительности должна быть больше найденного приближенного ее значения 41,02. Чтобы узнать, какую величину прибавить к этому приближенному числу $\gamma$ (т.-е. к 41,02), нужно найти в таблице разность $\Delta$. В данном случае $\Delta = 0{,}113$. Это значит, что каждая отброшенная у $v$ десятая доля единицы увеличивает $\gamma$ на $\Delta = 0{,}113$. У нас было отброшено 3 десятых. Следовательно к приближенному значению $\gamma$ нужно прибавить 3 $\Delta$, т.-е. $3 \cdot 0{,}113 = 0{,}339$. Прибавив к 41,02 эту поправку 0,339, получим окончательно:

$$\gamma = 41{,}02 + 0{,}339 = 41{,}359.$$

Итак, для $v = 48{,}3$ величина $\gamma = 41{,}359$.

Подставляя это исправленное значение $\gamma$ в упрощенную формулу для вычисления средней ошибки вариационного коэффициента, получим:

$$m_v = \frac{\gamma}{\sqrt{n}} = \frac{41{,}359}{\sqrt{85}} = \frac{41{,}359}{9{,}22} = \pm 4{,}49\%.$$



Если бы у $v$ было при округлении отброшено два знака, то к поправке для $\gamma$ пришлось бы прибавить еще часть $\Delta$, соответствующую отброшенным сотым долям $v$. Пусть, например, $v = 98{,}35$, а $n = 85$. Отыскивая по таблице в графе $v$ число 98, найдем приближенную величину $\gamma = 118{,}43$. Поправка $\Delta = 0{,}201$. Помножив ее на число отброшенных десятых долей единицы, т.-е. на 3, получим $3 \cdot 0{,}201 = 0{,}603$. Но этого мало. Мы не учли здесь еще 5 сотых. Если одна десятая соответствует $\Delta$, то 1 сотая будет соответствовать $\frac{\Delta}{10}$ т.-е. в данном случае 0,0201, а 5 таких сотых потребует прибавки в 5 раз больше, т.-е. $5 \cdot 0{,}0201 = 0{,}1005$. Итак, поправка на десятые доли оказалась равной 0,603, а на сотые 0,1005. Полная поправка будет, очевидно, равна их сумме, т.-е.

$$0{,}603 + 0{,}1005 = 0{,}7035.$$

Этот расчет можно было бы записать так:

| | | |
|---|---|---|
| для десятых . . . . | $3 \cdot 0{,}201$ | $= 0{,}603$ |
| для сотых . . . . . | $5 \cdot 0{,}0201$ | $= 0{,}1005$ |
| полная поправка: | | $= 0{,}7035.$ |

Прибавив к приближенному числу $\gamma$ (118,43) эту поправку (0,7035), получим:

$$118{,}43 + 0{,}7035 = 119{,}1335.$$

Итак для $v = 8{,}35$ величина $\gamma = 119{,}1335$.

Искомая ошибка найдется по формуле:

$$m_v = \frac{\gamma}{\sqrt{n}} = \frac{119{,}1335}{9{,}22} = \pm 12{,}92\%.$$

2. **Показатель точности.** ($P$). Подобно вариационному коэффициенту $v$, представляющему собой процентное отношение $\sigma$ к $M$, можно и среднюю ошибку $m$ выразить в %% от соответствующего среднего арифметического $M$. Составленный этим путем процентный показатель мы обозначим через $P$ и будем называть „показателем точности“.

Таким образом показатель точности

$$P = \frac{100\, m}{M}\ \%.$$

Если, например, в одном случае мы получили

$$M_1 = 25 \quad \text{и} \quad m = 0{,}5$$

а в другом:

$$M_2 = 10 \quad \text{и} \quad m = 0{,}2$$

то соответствующие показатели точности $P_1$ и $P_2$ окажутся равными

$$P_1 = \frac{100\, m_1}{M_1} = \frac{100 \,.\, 0{,}5}{25} = \frac{50}{25} = 2\%.$$

$$P_2 = \frac{100\, m_2}{M_2} = \frac{100 \,.\, 0{,}2}{10} = \frac{20}{10} = 2\%.$$

Этот результат показывает, что при различной абсолютной точности двух заданных средних $M_1$ и $M_2$, выраженной величиной их средних ошибок ($m_1 = 0{,}5$ и $m_2 = 0{,}2$), относительная точность этих средних $P_1$ и $P_2$ оказалась совершенно одинаковой (в обоих случаях средний предел ожидаемых колебаний этих средних арифметических составляет $2\%$ от их собственной величины).

Если бы мы имели следующие абсолютные данные:

$$M_1 = 16 \qquad m_1 = 0{,}8$$
$$M_2 = 32 \qquad m_2 = 0{,}8,$$

то соответствующие показатели точности

$$P_1 = \frac{100 \,.\, 0{,}8}{16} = \frac{80}{16} = 5\%$$

$$P_2 = \frac{100 \,.\, 0{,}8}{32} = \frac{80}{32} = 2{,}5\%.$$

Этот пример показывает, что иногда при одинаковой абсолютной точности заданных средних ($m_1 = m_2 = 0{,}8$) их относительная точность оказывается различной ($P_1 = 5\%$, $P_2 = 2{,}5\%$).

В тех случаях, когда предварительно был вычислен вариационный коэффициент $v$, для определения показателя точности $P$ может служить следующая формула:

$$P = \frac{v}{\sqrt{n}}$$

Показатель точности $P$ имеет широкое применение в опытной агрономии, характеризуя надежность результатов полевых испытаний, производимых в целях выяснения тех



или иных свойств и особенностей различных культурных растений. При этом принято считать, что достаточная надежность такого эксперимента будет обеспечена лишь в том случае, если показатель точности $P$ не превышает $5\%$ (т. е. средняя ошибка $m$ составляет не более 5 процентов от соответствующего среднего арифметического $M$).

В тех случаях, когда подобный опыт представляет интерес главным образом со стороны методологической (например, при оценке способа его постановки) сравнению друг с другом на ряду со средними арифметическими $M$ могут подлежать также и показатели точности $P$. Для выяснения вопроса о достоверности различия между двумя показателями точности $P_1$ и $P_2$ необходимо предварительно определить их средние ошибки $m_{p_1}$, $m_{p_2}$.

Эти ошибки могут быть вычислены по формуле:

$$m_p = p\sqrt{\frac{1}{2n} + \left(\frac{P}{100}\right)^2}$$

где через $n$, как всегда, обозначено общее количество измерений, а через $P$ данный показатель точности.

Пусть, например, $P = 3\%$, а $n = 50$. Тогда

$$m_p = 3\sqrt{\frac{1}{2 \,.\, 50} + \left(\frac{3}{100}\right)^2} = 3\sqrt{\frac{1}{100} + (0{,}03)^2} =$$
$$= 3\sqrt{0{,}01 + 0{,}0009} = 3\sqrt{0{,}0109}.$$

Так как $\sqrt{0{,}0109}$ приблизительно $= 0{,}1044$, то

$$m_p = 3 \,.\, 0{,}1044 = 0{,}3132.$$

Итак, окончательно, средняя ошибка данного показателя точности

$$m_p = \pm 0{,}31\%.$$

Пусть, например, испытываются два способа постановки какого-либо опыта (две методики), причем в первом случае показатель точности $P_1 = 2\%$, а во втором $P_2 = 3\%$. Пусть, далее, соответствующие ошибки этих показателей оказались равными: $m_{p_1} = \pm 0{,}2\%$, а $m_{p_2} = \pm 0{,}3\%$. Спрашивается, гарантирует ли первый из этих способов всегда большую точность результатов опыта, или же применение второй методики дало менее точные результаты лишь случайно?

Применив обычный критерий оценки достоверности,

$$\frac{(P_1 - P_2)^2}{m_{p_1}{}^2 + m_{p_2}{}^2} > 9$$

получим:

$$\frac{(2-3)^2}{(0{,}2)^2 + (0{,}3)^2} = \frac{(-1)^2}{0{,}04 + 0{,}09} = \frac{1}{0{,}13} = 7{,}69.$$

Так как в данном случае квадраты оцениваемых различий между двумя показателями точности превышают сумму квадратов двух средних ошибок всего лишь в 7,69 раза (а не в 9 раз, как это требуется), то преимущество первого способа постановки опыта перед вторым здесь не может еще считаться доказанным.

3. **Коэффициенты регрессии (R).** На ряду с вопросом о выяснении наличия прямой или обратной зависимости между двумя варьирующими признаками (например, ростом и весом) во многих случаях может представить интерес также и определение количественной стороны этой зависимости.

Расположив все объекты исследования в ряд по возрастанию одного из двух связанных друг с другом признаков, мы, при наличии прямой корреляции между ними, всегда обнаружим более или менее ясно выраженное возрастание и другого признака, а при обратной корреляции, наоборот, некоторое убывание второго признака. При этом может возникнуть вопрос о том, на какую же величину этот второй признак возрастает или убывает в связи с возрастанием первого признака?

Поясним эту мысль таким примером. Пусть речь идет попрежнему о корреляции между ростом и весом у некоторого количества детей, вполне однородных по возрасту и полу. Разобьем этих детей на группы (классы), отличающиеся друг от друга ровно на 1 см. по росту. Таким образом в первой группе мы будем иметь нескольких детей, обладающих самым низким ростом (одинаковым для всей этой группы). Во вторую группу войдут дети, ровно на 1 *см* превышающие первую группу. В следующую (третью) группу — еще более высокие дети, отличающиеся от второй группы также на 1 *см* (а от первой, следовательно, на 2 *см*) и т. д. вплоть до последней группы самых высоких детей.



Если между ростом и весом этих детей существует прямая корреляция (хотя бы и частичная), то и вес их будет возрастать от группы к группе, хотя при неполной зависимости между ростом и весом это возрастание второго признака (веса) и не будет носить здесь столь же строгий характер, как это имело место в применении к искусственно подобранному нами росту этих детей.

Если для каждой отдельной группы детей с одинаковым ростом вычислить их средний вес, то полученный ряд средних арифметических этого веса у отдельных таких групп обнаружит более или менее ярко выраженную тенденцию к возрастанию (от групп с низким ростом к группам с высоким ростом).

Такое среднее увеличение веса, соответствующее увеличению роста на 1 *см*, вообще говоря, может оказаться и не одинаковым — в одном случае несколько большим, в другом, наоборот, меньшим.

Так вот, та средняя величина возрастания веса, которая соответствует увеличению роста на 1 *см*, и называется коэффициентом регрессии (веса на рост или веса к росту).

Если бы мы разбили этих же детей на группы по возрастанию веса ровно на 1 килограмм и определили бы средний рост для каждой такой группы, то среднее увеличение роста, соответствующее увеличению веса на 1 килограмм, дало бы другой коэффициент регрессии (роста на вес или роста к весу).

Коэффициенты регрессии принято обозначать буквами $R$, при чем для отличия их друг от друга, у этих букв обычно ставят условные знаки: $y$ или $x$. Так, если за ряд $x$ принят рост, а рядом $y$ служит вес, то коэффициент регрессии веса на рост ($y$ на $x$) обозначается символически:

$$R_y\,,$$

а коэффициент регрессии роста на вес ($x$ на $y$) через:

$$R_x\,.$$

Формула для определения числовых значений этих коэффициентов по способу непосредственного вычисления или по способу округления (т. е. без разбивки рядов на классы) имеет следующий вид:

$$R_y = \frac{\Sigma xy}{\Sigma x^2} \qquad R_x = \frac{\Sigma xy}{\Sigma y^2}.$$

При вычислении же этих коэффициентов по способу сумм следует пользоваться формулами:

$$R_y = \frac{k_y}{k} \cdot \frac{\Sigma xy}{\Sigma x^2} \qquad R_x = \frac{k_x}{k_y} \cdot \frac{\Sigma xy}{\Sigma y^2},$$

где через $k_y$ обозначен классовый промежуток ряда $y$, а через $k_x$ — классовый промежуток ряда $x$.

Пусть, например, имеем следующие два ряда коррелирующих чисел (5 пар; $n = 5$).

| | | | | | | |
|---|---|---|---|---|---|---|
| Ряд $x$: | 12 | 14 | 15 | 16 | и | 18 |
| Ряд $y$: | 4 | 12 | 10 | 8 | и | 16 |

Среднее арифметическое чисел первого ряда $M_x = 15$, а для второго ряда $M_y = 10$. Выразив все заданные варианты в их отклонениях от соответствующих средних арифметических $M_x$ и $M_y$, получим:

| | | | | | |
|---|---|---|---|---|---|
| $x$: | $-3$ | $-1$ | $0$ | $+1$ | $+3$ |
| $y$: | $-6$ | $+2$ | $0$ | $-2$ | $+6$ |

Перемножив каждый $x$ на соответствующий ему $y$, получим следующий ряд произведений $xy$:

$$xy: \quad +18 \quad -2 \quad 0 \quad -2 \quad +18;$$

сумма чисел этого ряда $\Sigma xy = +32$.

Возвысив в квадрат все отклонения ($x$) первого ряда, получим:

$$x^2: \quad 9 \quad 1 \quad 0 \quad 1 \quad 9$$

Сумма этих квадратов $\Sigma x^2 = 20$.

Возвысив в квадрат отклонения ($y$) второго ряда, получим:

$$y^2: \quad 36 \quad 4 \quad 0 \quad 4 \quad 36$$

Сумма этих квадратов $\Sigma y^2 = 80$.

Итак, для нашего числового примера:

$$\Sigma xy = +32 \qquad \Sigma x^2 = 20 \qquad \Sigma y^2 = 80$$



Подставляя эти данные в формулы для вычисления коэффициентов регрессии, будем иметь:

$$R_y = \frac{\Sigma xy}{\Sigma x^2} = \frac{+32}{20} = +1{,}6$$

$$R_x = \frac{\Sigma xy}{\Sigma y^2} = \frac{+32}{80} = +0{,}4$$

Первый из этих коэффициентов регрессии ($R_y = +1{,}6$) показывает, что с возрастанием чисел первого ряда (ряда $x$) на единицу (1), числа второго ряда (ряда $y$) в среднем увеличиваются (+) на 1,6. Второй коэффициент ($R_x = +0{,}4$) показывает, что с увеличением чисел второго ряда (ряда $y$) на единицу (1) числа первого ряда (ряда $x$) в среднем также увеличивается (+), но уже только на 0,4.

Если бы между двумя заданными рядами существовала обратная корреляция, то $\Sigma xy$ и оба коэффициента регрессии оказались бы отрицательными. Это показывало бы, что с возрастанием чисел одного из рядов числа второго ряда, наоборот, в среднем уменьшаются (—).

Заметим, кстати, что между коэффициентом корреляции $r$ и обоими коэффициентами регрессии $R_y$ и $R_x$ существует зависимость, определяемая формулой:

$$r^2 = R_y \cdot R_x,$$

показывающая, что квадрат коэффициента корреляции $r$ равен произведению двух коэффициентов регрессии $R_y$ и $R_x$.

В нашем примере (при $\Sigma xy = +32$, $\Sigma x^2 = 20$ и $\Sigma y^2 = 80$) коэффициент корреляции

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}} = \frac{+32}{\sqrt{20 \cdot 80}} = \frac{+32}{\sqrt{1600}} = \frac{+32}{40} = +0{,}8$$

стало быть $r^2 = 0{,}8^2 = 0{,}64$.

Но то же значение $r^2 = 0{,}64$ мы получили и по только что указанной формуле (при $R_y = +1{,}6$ и $R_x = +04$)

$$r^2 = R_y \cdot R_x = (+1{,}6)\,(+0{,}4) = 0{,}64.$$

Эта формула может служить для проверки вычислений. Следует лишь помнить, что в тех случаях, когда приходится извлекать корень или производить деление не нацело, указанная зависимость в силу неизбежной неточности вычислений (при отбрасывании десятичных знаков) будет иметь лишь приближенный характер.

Если для каждого из двух коррелирующих рядов были уже предварительно вычислены средние квадратические отклонения $\sigma_x$ и $\sigma_y$ (по формуле без поправки на малое число измерений), то коэффициенты регрессии могут быть определены по другим формулам:

$$R_y = \frac{\sigma_y}{\sigma_x}\, r \qquad R_x = \frac{\sigma_x}{\sigma_y}\, r$$

В нашем примере для первого ряда (ряда $x$)

$$\sigma_x = \sqrt{\frac{20}{5}} = \sqrt{4} = \pm 2,$$

а для второго ряда (ряда $y$)

$$\sigma_y = \sqrt{\frac{80}{5}} = \sqrt{16} = \pm 4$$

Коэффициент же корреляции между этими рядами

$$r = +0{,}8$$

Подставляя эти данные в указанные выше формулы для вычисления $R_y$ и $R_x$, получим:

$$R_y = \frac{\sigma_y}{\sigma_x}\, r = \frac{4}{2}(+0{,}8) = \frac{+3{,}2}{2} = +1{,}6$$

$$R_x = \frac{\sigma_x}{\sigma_y}\, r = \frac{2}{4}(+0{,}8) = \frac{+1{,}6}{4} = +0{,}4$$

т.-е. те же самые результаты, которые мы имели и ранее.

Чтобы показать применение этих формул при вычислении $R_y$ и $R_x$ по способу сумм, рассмотрим следующий числовой пример корреляции двух признаков, разобранный уже нами ранее (см. главу XV).

На таблице 85 представлена корреляционная решетка с проставленными в ее клетках заданными частотами $p$.

Для этих данных (см. стр. 182) мы ранее получили:

$$\Sigma x^2 = 119{,}28 \qquad \Sigma xy = -12{,}6$$
$$\Sigma y^2 = 52{,}50 \qquad n = 50.$$



Здесь классовый промежуток ряда $x$, т.-е. $k_x = 2$, а классовый промежуток ряда $y$, т.-е. $k_y = 5$.

Подставляя эти данные в формулы для вычисления $R_y$ и $R_x$, получим:

$$R_y = \frac{k_y}{k_x} \cdot \frac{\Sigma xy}{\Sigma x^2} = \frac{5}{2} \cdot \frac{(-12{,}6)}{119{,}28} = \frac{-63}{238{,}56} = -0{,}26$$

$$R_x = \frac{k_x}{k_y} \cdot \frac{\Sigma xy}{\Sigma y^2} = \frac{2}{5} \cdot \frac{(-12{,}6)}{52.50} = \frac{-25{,}2}{262{,}5} = -0{,}10.$$

В данном случае эти коэффициенты регрессии оказались отрицательными. Первый из них $R_y = -0{,}26$ показывает, что с возрастанием числовых значений ряда $x$ на 1, соответствующие значения другого рода $y$, наоборот, уменьшаются в среднем на 0,26, а с увеличением вариант второго ряда ($y$) средние значения первого ряда ($x$) имеют также тенденцию к убыванию, в среднем на 0,10.

ТАБЛИЦА 85.

| $y$ \ $x$ | 3,1—5 | 5,1—7 | 7,1—9 | 9,1—11 | 11,1—13 | 13,1—15 | 15,1—17 |
|---|---|---|---|---|---|---|---|
| 75—71 | | 1 | | 1 | | | 1 |
| 70—66 | 1 | 2 | 3 | 3 | 2 | | |
| 65—61 | | 3 | 4 | 5 | 4 | 2 | |
| 60—56 | 1 | 2 | 3 | 2 | 3 | 2 | 1 |
| 55—51 | | 1 | | 1 | | 1 | 1 |

Заметим, кстати, что оба коэффициента регрессии, так же как и коэффициент корреляции $r$, всегда имеют одинаковые знаки (либо все три коэффициента положительные, либо все три отрицательные).

Если бы в данном примере были уже вычислены оба средние квадратические отклонения заданных рядов ($\sigma_x$ и $\sigma_y$), то нам выгоднее было бы воспользоваться другими формулами для вычисления $R_y$ и $R_x$:

$$R_y = \frac{\sigma_y}{\sigma} r \quad \text{и} \quad R_x = \frac{\sigma_x}{\sigma_y}\, r,$$

где $\sigma_x$ и $\sigma_y$ берутся всегда только со знаком $+$.

Вычислив $\sigma_x$ и $\sigma_y$ из наших данных, получим:

$$\sigma_x = 3{,}09 \quad \sigma_y = 5{,}12$$

Так как коэффициент корреляции $r = -0{,}16$, то

$$R_y = \frac{5{,}12}{3{,}09}(-0{,}16) = \frac{-0{,}8192}{3{,}09} = -0{,}26$$

$$R_x = \frac{3{,}09}{5{,}12}(-0{,}16) = \frac{-0{,}4944}{5{,}12} = -0{,}10.$$

Проверка коэффициентов регрессии по формуле:

$$r^2 = R_y \cdot R_x$$

дает здесь следующий результат:

$$r^2 = (-0{,}16)^2 = 0{,}0256$$
$$R_y \cdot R_x = (-0{,}26)\ (-0{,}10) = 0{,}026,$$

т.-е. приблизительное равенство с расхождением в 3-м знаке.

Средние ошибки коэффициентов регрессии вычисляются по следующим формулам (без учета величины классового промежутка $k_x$ и $k_y$):

для $R_y$:

$$m_y = \pm\sqrt{\frac{\Sigma y^2}{\Sigma x^2}\cdot\frac{1-r^2}{n}}$$

для $R_x$:

$$m_x = \pm\sqrt{\frac{\Sigma x^2}{\Sigma y^2}\cdot\frac{1-r^2}{n}}.$$

Если же вычисления велись по способу сумм, и оба ряда были предварительно разбиты на классы с классовым промежутком $k_x$ и $k_y$, то для определения этих ошибок могут служить следующие формулы:

$$m_y = \pm\frac{k_y}{k_x}\sqrt{\frac{\Sigma y^2}{\Sigma x^2}\cdot\frac{1-r^2}{n}}$$

$$m_x = \pm\frac{k_x}{k_y}\sqrt{\frac{\Sigma x^2}{\Sigma y^2}\cdot\frac{1-r^2}{n}}.$$

В тех же случаях, когда известными являются оба средних квадратических отклонения заданных рядов ($\sigma_x$ и $\sigma_y$), для вычисления средних ошибок коэффициентов регрессии удобно воспользоваться такими формулами:



$$m_y = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r^2}{n}}$$

$$m_x = \frac{\sigma_x}{\sigma_y}\sqrt{\frac{1-r^2}{n}}.$$

Во всех этих случаях для проверки вычисления может служить следующая контрольная формула:

$$m_y \cdot m_x = \frac{1-r^2}{n},$$

показывающая, что произведение средних ошибок обоих коэффициентов регрессии равняется выражению:

$$\frac{1-r^2}{n}.$$

В разобранном выше примере (см. стр. 202) мы получили:

$$\Sigma x^2 = 20, \quad \Sigma y^2 = 80, \quad r = +0{,}8, \quad n = 5.$$

Подставляя эти данные в первую пару указанных здесь формул для вычисления средних ошибок коэффициентов $R$, получим:

$$m_y = \sqrt{\frac{\Sigma y^2}{\Sigma x^2}\cdot\frac{1-r^2}{n}} = \sqrt{\frac{80}{20}\cdot\frac{1-(0{,}8)^2}{5}} = \sqrt{4\cdot\frac{1-0{,}64}{5}} =$$
$$= \sqrt{4\cdot\frac{0{,}36}{5}} = \sqrt{\frac{1{,}44}{5}} = \sqrt{0{,}288} = \pm 0{,}54$$

$$m_x = \sqrt{\frac{\Sigma x^2}{\Sigma y^2}\cdot\frac{1-r^2}{n}} = \sqrt{\frac{20}{80}\cdot\frac{1-(0{,}8)^2}{5}} = \sqrt{\frac{1}{4}\cdot\frac{1-0{,}64}{5}} =$$
$$= \sqrt{\frac{1}{4}\cdot\frac{0{,}36}{5}} = \sqrt{\frac{0{,}36}{20}} = \sqrt{0{,}018} = \pm 0{,}13.$$

Итак: $m_y = \pm 0{,}54$
$m_x = \pm 0{,}13.$

Чтобы проверить эти вычисления, воспользуемся контрольной формулой:

$$m_y \cdot m_x = \frac{1-r^2}{n}$$

Здесь произведение этих ошибок

$$0{,}54 \cdot 0{,}13 = 0{,}0702$$

величина же

$$\frac{1-r^2}{n} = \frac{1-(0{,}8)^2}{5} = \frac{1-0{,}64}{5} = \frac{0{,}36}{5} = 0{,}072.$$

Расхождение получилось лишь в третьем знаке.

В другом примере (см. стр. 204) мы имели:

$$\Sigma x^2 = 119{,}28 \quad k_x = 2 \quad r = -0{,}16$$
$$\Sigma y^2 = 52{,}50 \quad k_y = 5 \quad n = 50.$$

Подставив эти данные во вторую пару формул для вычисления средней ошибки коэффициентов $R$ (с учетом классовых промежутков $k_x$ и $k_y$), получим:

$$m_y = \frac{k_y}{k_x}\sqrt{\frac{\Sigma y^2}{\Sigma x^2}\cdot\frac{1-r^2}{n}} = \frac{5}{2}\sqrt{\frac{52{,}50}{119{,}28}\cdot\frac{1-(-0{,}16)^2}{50}} = \pm 0{,}23$$

$$m_x = \frac{k_x}{k_y}\sqrt{\frac{\Sigma x^2}{\Sigma y^2}\cdot\frac{1-r^2}{n}} = \frac{2}{5}\sqrt{\frac{119{,}28}{52{,}50}\cdot\frac{1-(-0{,}16)^2}{50}} = \pm 0{,}08.$$

В том же примере мы имели $\sigma_x = 3{,}10$, $\sigma_y = 5{,}10$.

Подставив эти данные в третью пару формул для вычисления средних ошибок коэффициентов $R$, получим:

$$m_y = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r^2}{n}} = \frac{5{,}12}{3{,}09}\sqrt{\frac{1-(-0{,}16)^2}{50}} = \pm 0{,}23$$

$$m_x = \frac{\sigma_x}{\sigma_y}\sqrt{\frac{1-r^2}{n}} = \frac{3{,}09}{5{,}12}\sqrt{\frac{1-(-0{,}16)^2}{50}} = \pm 0{,}08.$$

Эти результаты следует также проверить по контрольной формуле:

$$m_y \cdot m_x = \frac{1-r^2}{n}.$$

В данном случае, при

$$m_y = 0{,}23 \text{ и } m_x = 0{,}08,$$
$$m_y \cdot m_x = (0{,}23).\ (0{,}08) = 0{,}0184$$

и величина

$$\frac{1-r^2}{n} = \frac{1-(-0{,}16)^2}{50} = \frac{1-0{,}0256}{50} = \frac{0{,}9744}{50} = 0{,}0195.$$



Следовательно вычисление произведено правильно (расхождение в 3 знаке).

Применение средних ошибок коэффициентов регрессии аналогично применению средней ошибки коэффициента корреляции $r$. Так как коэффициенты регрессии определяют среднюю величину нарастания (или убывания) одного из признаков при возрастании другого признака на 1, то для суждения о достоверности этого нарастания (или убывания) необходимо убедиться в том, что абсолютная величина коэффициента регрессии превышает абсолютную величину своей средней ошибки по крайней мере в 3 раза, т. е.:

$$\frac{R_y}{m_y} > 3 \quad \text{или} \quad \frac{R_x}{m_x} > 3$$

Так в одном из разобранных выше примеров (см. стр. 203) мы получили коэффициент регрессии

$$R_y = +1{,}6,$$

а вычислив его среднюю ошибку (см. стр. 207) нашли, что

$$m_y = \pm 0{,}54.$$

В данном случае отношение

$$\frac{R_y}{m_y} = \frac{1{,}6}{0{,}54} = 2{,}96.$$

Следовательно увеличение признака $y$ в связи с увеличением признака $x$ не может при данных условиях считаться доказанным.

Сравнивать оба коэффициента регрессии друг с другом можно, очевидно, лишь в том случае, если оба коррелирующих признака выражены в одинаковых единицах измерения (например, при корреляции двух антропометрических признаков, выраженных в *см*).

То же соображение относится и к случаю сравнения двух коэффициентов регрессии, вычисленных для различных корреляционных решеток (например, при корреляции одной и той же пары антропометрических признаков у детей разного пола).

Для оценки достоверности обнаруженного различия между двумя коэффициентами регрессии $R_1$ и $R_2$, вычисленными с ошибками $m_1$ и $m_2$, применяется обычный критерий:

$$\frac{(R_1-R_2)^2}{m_1^2+m_2^2}>9.$$

Пусть, например, при корреляции двух признаков у мальчиков один из коэффициентов регрессии $R_1=+0{,}9$ со средней ошибкой $m_1=\pm 0{,}2$, а при корреляции тех же признаков у девочек соответствующий коэффициент регрессии $R_2=+0{,}5$ со средней ошибкой $m_2=\pm 0{,}1$.

Спрашивается, является ли достоверным это различие в коэффициентах регрессии $R_1$ и $R_2$, т.-е. можно ли при данных условиях считать доказанным большее среднее увеличение изучаемого признака у мальчиков? Применяя критерий оценки достоверности, получим:

$$\frac{(0{,}9-0{,}5)^2}{(0{,}2)^2+(0{,}1)^2}=\frac{(0{,}4)^2}{0{,}04+0{,}01}=\frac{0{,}16}{0{,}05}=3{,}2.$$

Так как в результате этих вычислений мы получили величину 3,2 (меньше 9), то намечавшийся вывод не может считаться доказанным.

## XVII. Некоторые частные случаи статистической обработки вариационных рядов.

В целях упрощения и сокращения работы по вычислению всех описанных выше общих числовых характеристик, как отдельного ряда, так и связи между двумя рядами признаков полезно будет ознакомиться с некоторыми свойствами и особенностями этих величин.

Прежде всего поставим себе вопрос: как изменятся числовые значения $M$ и $\sigma$, если каждую отдельную варианту увеличить или уменьшить на какую-либо постоянную величину? Разберем этот случай на числовом примере. Пусть имеем ряд, состоящий из пяти вариант:

31 29 28 33 и 29.

Для этих пяти чисел $M=30$, а $\sigma=\pm 2$, в чем легко убедиться проделав соответствующие выкладки по способу непосредственного вычисления (см. стр. 50).



Прибавим теперь к каждой варианте какую-нибудь постоянную величину, например 3, и посмотрим, как это отразится на $M$ и $\sigma$. Новые (увеличенные на 3) варианты будут теперь иметь следующие значения:

34 32 31 36 и 32.

Для этого ряда чисел $M=33$, а $\sigma=\pm 2$, в чем также нетрудно убедиться, проделав соответствующие вычисления в действительности.

Приведенный пример показывает, что от прибавления ко всем вариантам одной и той же величины (у нас 3) среднее арифметическое $M$ увеличивается на ту же самую величину (т.-е. на 3), а среднее квадратическое отклонение $\sigma$ остается без изменения ($\sigma$ была и осталась $\pm 2$).

Отняв от всех первоначально взятых нами пяти вариант какую-либо постоянную величину, например 4, получим следующие новые варианты:

27 25 24 29 и 25.

Для этого ряда $M=26$ (уменьшилось на 4 единицы), а $\sigma$ по прежнему $=\pm 2$. Следовательно при вычитании из всех вариант какой-либо постоянной величины среднее арифметическое $M$ уменьшается на ту же самую величину, а $\sigma$ остается без изменения.

Несколько иначе происходит изменение $M$ и $\sigma$ при увеличении или уменьшении всех вариант в постоянное число раз. Увеличим, например, все наши варианты: 31, 29, 28, 33 и 29, скажем, в 2 раза и посмотрим, как это отразится на $M$ и $\sigma$. После умножения на 2 получим следующие 5 новых вариант:

62, 58, 56, 66 и 58.

Для этого ряда $M=60$, и $\sigma=\pm 4$. Следовательно при умножении всех вариант на постоянное число (в нашем примере на 2), как $M$, так и $\sigma$ также увеличиваются и при том во столько же раз.

Разделив наши первоначальные варианты, например, на 10, получим:

3,1 2,9 2,8 3,3 и 2,9

Для этого ряда $M = 3$, а $\sigma = \pm 0,2$. Следовательно при делении всех вариант на постоянное число (у нас на 10) $M$ и $\sigma$ уменьшаются во столько же раз (т. е. в 10 раз).

Указанные свойства $M$ и $\sigma$ дают возможность в некоторых случаях значительно упростить способ вычисления. Так, если некоторые из вариант оказываются отрицательными числами (например, температура ниже 0), то, чтобы не иметь дела со знаком —, мы можем прибавить ко всем измерениям какую-либо постоянную величину, в результате чего все наши варианты обратятся исключительно в положительные числа. Вычислив для них $M$ и $\sigma$ и помня, что $M$ оказалось больше своего истинного значения, а $\sigma$ не изменилась, мы должны будем внести сюда конечную поправку, уменьшив $M$ на соответствующую величину.

Точно также, желая избавиться от дробных чисел, мы можем помножить все варианты на 10 или на 100, что изменит $M$ и $\sigma$ во столько же раз. В конечном итоге нам придется разделить $M$ и $\sigma$ на соответствующую величину (на 10 или на 100).

Указанным свойством $M$ и $\sigma$ удобно воспользоваться в том случае, когда варианты представляют собой процентное отношение каких-либо абсолютных измерений к одному и тому же постоянному числу (что соответствует умножению этих вариант на частное от деления 100 на это постоянное число). Тогда, взамен вычисления этих процентных отношений для каждого измерения отдельно, мы можем подвергнуть обработке только абсолютные числа и внести соответствующую поправку уже в конечные значения $M$ и $\sigma$, выразив их в $^0/_0{}^0/_0$ от заданной величины.

Аналогичные операции приходится выполнять и при переходе от одних единиц измерения к другим (например, от вершков к *см* или от пудов к килограммам), а также при исправлении постоянных неточностей в самих измерениях (например, при измерении роста людей, стоящих на подставке определенной высоты над полом, или при пользовании оборванной сантиметровой лентой, в которой недостает некоторого постоянного числа делений).

Возможен и такой случай, когда все варианты изменяются путем умножения или деления на постоянную величину с последующим прибавлением или вычитанием также постоянной величины.



Пусть, например, требуется перейти от градусов по Фаренгейту к градусам по Цельсию. Как известно, перечисление это делается путем умножения числа градусов Фаренгейта на $\frac{5}{9}$ с последующим вычитанием из этого произведения постоянного числа $\frac{160}{9}$. Так, например, 20 градусов по Фаренгейту $= -6,67$ градусов по Цельсию, т. к.

$$\frac{5}{9}\cdot 20 - \frac{160}{9} = \frac{100}{9} - \frac{160}{9} = \frac{-60}{9} = -6,67$$

Если варианты представляют собой температуру, измеренную в градусах по Фаренгейту, и для них вычислены уже $M$ и $\sigma$, то для получения $M$ и $\sigma$ в градусах Цельсия нет надобности выполнять эти вычисления для каждой варианты отдельно, т. к. соответствующую поправку мы можем внести прямо в конечные значения $M$ и $\sigma$. Т. к. $M$ изменяется и при увеличении или уменьшении вариант на постоянную величину и при изменении их в постоянное число раз, а $\sigma$ — только при умножении или делении вариант, то эта конечная поправка выразится в умножении $M$ на $\frac{5}{9}$ с последующим вычитанием $\frac{160}{9}$, а $\sigma$ придется только умножить на $\frac{5}{9}$ (не вычитая $\frac{160}{9}$).

Пусть, например, $M = 50$, а $\sigma = 9$ в градусах по Фаренгейту. Чтобы выразить $M$ в градусах Цельсия, следует 50 умножить на $\frac{5}{9}$ и из полученного произведения вычесть $\frac{160}{9}$. Таким образом искомое

$$M = \frac{5}{9}\cdot 50 - \frac{160}{9} = \frac{250}{9} - \frac{160}{9} = \frac{90}{9} = +10.$$

Среднее квадратическое отклонение $\sigma$ в градусах Цельсия получится просто путем умножения 9 на $\frac{5}{9}$

$$\sigma = \frac{5}{9}\cdot 9 = 5$$

Выяснив на разобранных выше числовых примерах закон изменения $M$ и $\sigma$ при различных изменениях всех вариант, легко представить себе и характер соответствующих изменений в величине средней ошибки $m$, вариационного коэффициента $v$ и показателя точности $P$.

Т. к. средняя ошибка ($m$) вычисляется по формуле:

$$m = \frac{\sigma}{\sqrt{n}},$$

то при постоянном числе измерений ($n$) величина $m$ во всех своих изменениях следует изменениям $\sigma$, т. е. при увеличении или уменьшении всех вариант на постоянную величину $m$ не меняется, а при увеличении или уменьшении их в постоян, ное число раз — $m$ увеличивается или уменьшается во столько же раз.

Вариационный коэффициент $v$, представляющий собой процентное отношение $\sigma$ к $M$, вычисляется по формуле:

$$v = \frac{100\,\sigma}{M}\ \%.$$

Так как при умножении или делении всех вариант на постоянное число, как $\sigma$ так и $M$ увеличиваются или уменьшаются во столько же раз, то отношение их, очевидно, не претерпевает при этом никаких изменений. Поэтому в случае изменения вариант в постоянное число раз вариационный коэффициент $v$ сохраняет свою первоначальную величину.

Так как при увеличении или уменьшении всех вариант на постоянную величину соответствующим образом меняется лишь $M$, а $\sigma$ сохраняет свое первоначальное значение, то и вариационный коэффициент $v$ в этом случае оказывается измененным. Пусть, например, имеем ряд чисел: 21, 19, 18 23, и 19, для которого $M = 20$, а $\sigma = \pm 2$. Следовательно вариационный коэффициент:

$$v = \frac{100\,\sigma}{M} = \frac{100 \cdot 2}{20} = 10\%.$$

Прибавим ко всем вариантам какую-нибудь постоянную величину, например 5. Тогда получим новый ряд:

26, 24, 23, 28, и 24,

для которого $M = 25$, а $\sigma = \pm 2$. Следовательно теперь:

$$v = \frac{100 \,.\, 2}{25} = 8\%$$



Отсюда видно, что при увеличении всех вариант на постоянную величину (у нас на 5) вариационный коэффициент $v$, наоборот, уменьшается, но уже не просто на ту же самую величину (5), а более сложно (в данном случае на 2%).

Если все первоначально заданные варианты уменьшить на постоянную величину (например на 5), то получим новый ряд

16, 14, 13, 18 и 14,

для которого $M = 15$, а $\sigma = \pm 2$. Тогда

$$v = \frac{100 \cdot 2}{15} = 13{,}3\ \%$$

Следовательно при вычитании из всех вариант постоянной величины (5) вариационный коэффициент $v$ несколько увеличивается (у нас на 3,3%).

Показатель точности $P$ представляет собой процентное отношение средней ошибки $m$ к среднему арифметическому $M$

$$P = \frac{100\,m}{M}\ \%$$

Поэтому характер его изменений при увеличении или уменьшении отдельных вариант вполне совпадает с только что разобранным случаем изменения вариационного коэффициента $v$: при умножении или делении вариант на постоянное число — $P$ не меняется, при увеличении же их на постоянную величину — $P$ уменьшается, а при уменьшении увеличивается (как и $v$).

Посмотрим теперь, как будет изменяться коэффициент корреляции $r$ при увеличении или уменьшении всех вариант на постоянную величину или в постоянное число раз. Пусть имеем два ряда коррелирующих друг с другом парных вариант (подписанных друг под другом):

7, 9, 11 и 13
9, 17, 13 и 21.

Если для этих рядов мы определим коэффициент корреляции $r$ по способу непосредственного вычисления (см. стр. 35), то получим

$$r = +0{,}8$$

Изменим как нибудь оба ряда вариант, например, уменьшив числа верхнего ряда на 3 единицы и помножив числа нижнего ряда на 2. Тогда получим новую пару рядов:

4, 6, 8 и 10
18, 34, 26, 42.

Для этих рядов коэффициент корреляции $r = +0{,}8$.

Следовательно от каких-бы то ни было одинаковых (в пределах каждого ряда) изменений всех вариант коэффициент корреляции $r$ не претерпевает изменений, сохраняя свою первоначальную величину. Это свойство коэффициента корреляции дает возможность прибавлять к вариантам каждого ряда любую постоянную величину или увеличивать их в произвольное постоянное число раз (или обратно вычитать и делить их), вовсе не меняя вариант другого ряда или же изменяя их независимо от первого ряда. Все эти операции совершенно не отражаются на величине коэфициента корреляции (чем между прочим и объясняется отсутствие в формуле для $r$ обозначения величины классового промежутка $k$).

Иначе реагируют на изменения вариант оба коэффициента регрессии: $R_y$ и $R_x$. Пусть имеем:

Ряд $x$ : 12 16 20 и 24
Ряд $y$ : 20 36 28 и 44.

Для этой пары коррелирующих рядов (при $M_x = 18$, а $M_y = 32$)

$$\Sigma x^2 = 80 \qquad \Sigma y^2 = 320 \qquad \Sigma xy = +128$$

Поэтому коэффициент регрессии $y$ на $x$ (показывающий на сколько в среднем изменяются варианты ряда $y$ соответственно увеличению вариант ряда $x$ на единицу)

$$R_y = \frac{\Sigma xy}{\Sigma x^2} = \frac{+128}{80} = +1{,}6.$$

Коэффициент же регрессии $x$ на $y$ (показывающий на сколько в среднем изменяются варианты ряда $x$ соответственно увеличению вариант ряда $y$ на единицу)

$$R_x = \frac{\Sigma xy}{\Sigma y^2} = \frac{+128}{320} = +0{,}4$$



Если все варианты ряда $x$ и, независимо от этого, все варианты ряда $y$ мы увеличим или уменьшим на какую-либо постоянную величину, то ни один из коэффициентов регрессии не изменится (также как и коэффициент корреляции $r$). В самом деле вычтем, например, из чисел первого ряда 2 и прибавим к числам второго ряда по 1. Тогда будем иметь новые ряды:

Ряд $x$ : 10 14 18 и 22
Ряд $y$ : 21 37 29 и 45

Для этих рядов $M_x = 16$, а $M_y = 33$, но попрежнему:

$$\Sigma x^2 = 80 \qquad \Sigma y^2 = 320 \quad \text{и} \quad \Sigma xy = +128$$

Следовательно и коэффициенты регрессии останутся без изменения

$$R_y = +1{,}6 \quad \text{и} \quad R_x = +0{,}4.$$

Теперь попробуем изменить варианты только одного ряда, например ряда $x$, умножив их на какое-либо постоянное число, например на 2, и посмотрим, как это отразится на величине коэффициентов регрессии:

Ряд $x$ : 24 32 40 и 48
Ряд $y$ : 20 36 28 и 44.

Для этих новых рядов: $M_x = 36$, а $M_y = 32$. Следовательно

$$\Sigma x^2 = 320 \qquad \Sigma y^2 = 320 \quad \text{и} \quad \Sigma xy = +256$$

откуда:

$$R_y = \frac{+256}{320} = +0{,}8 \quad \text{и} \quad R_x = \frac{+256}{320} = +0{,}8.$$

Итак, при умножении вариант ряда $x$ на постоянную величину (у нас на 2) соответствующий коэффициент регрессии $x$ на $y$ (т. е. $R_x$) увеличивается во столько же раз (был $= 0{,}4$, а стал $= 0{,}8$), в то время как другой коэффициент ($R_y$), наоборот, уменьшается во столько же раз (был $= 1{,}6$, а стал $= 0{,}8$).

Если этот же ряд (т. е. ряд $x$) уменьшить вдвое, не изменяя другого ряда ($y$), то получим следующие числа:

Ряд $x$ : 6 8 10 и 12
Ряд $y$ : 20 36 28 и 44.

Для этих рядов: $R_y = +3{,}2$ (вместо $+1{,}6$), а $R_x = +0{,}2$ (вместо $+0{,}4$).

Следовательно при уменьшении вариант какого-либо ряда данный коэффициент регрессии уменьшается во столько же раз, а другой коэффициент, наоборот, оказывается соответственно увеличенным.

Нетрудно убедиться из тех же числовых примеров, что при умножении второго ряда ($y$) на постоянную величину коэффициент регрессии $R_y$ увеличивается во столько же раз, а другой коэффициент $R_x$ уменьшается, и наоборот: при делении чисел ряда $y$ на постоянную величину коэффициент $R_y$ уменьшается, а коэффициент $R_x$ увеличивается во столько же раз.

Таким образом при умножении и делении вариант коэффициент $R_y$ изменяется одинаково с рядом $y$, а коэффициент $R_x$ — одинаково с рядом $x$.

Отсюда становится понятной и та роль, которую играют в этих формулах классовые промежутки обоих рядов $k_x$ и $k_y$, входящие в состав этих формул при вычислениях, ведущихся по способу сумм (см. стр. 202).

Выше мы рассмотрели случай изменения всех вариант на некоторую постоянную величину. Посмотрим теперь, как отразится на величине $M$ и $\sigma$ прибавление к вариантам переменных чисел. Случай этот можно иллюстрировать следующим примером. Пусть имеем два ряда варьирующих чисел:

| | | | | | |
|---|---|---|---|---|---|
| I ряд: | 14 | 18 | 20 | 22 и | 26 |
| II ряд: | 7 | 11 | 10 | 9 и | 13. |

Прибавим к каждому из верхних чисел соответствующее ему (записанное под ним) нижнее число и рассмотрим новый ряд, составленный из вариант двух заданных рядов, просуммированных попарно:

III ряд: 21 29 30 31 и 39.

Вычислив среднее арифметическое для всех трех рядов, получим:

для I ряда: $M_1 = 20$

для II ряда: $M_2 = 10$

для III ряда: $M_3 = 30$.



Этот результат указывает на существование определенной зависимости между средним арифметическим нового ряда ($M_3$) и двумя средними арифметическими обоих заданных рядов ($M_1$ и $M_2$). Зависимость эта выражается формулой:

$$M_3 = M_1 + M_2,$$

показывающей, что среднее арифметическое нового ряда, составленного из сумм вариант двух заданных рядов, равно сумме их средних арифметических.

Зададимся теперь целью определить этим же способом среднее квадратическое отклонение $\sigma_3$ нового ряда с помощью двух средних квадратических отклонений $\sigma_1$ и $\sigma_2$ обоих заданных рядов. Но при этом будем вычислять все сигмы не по указанной ранее формуле (с поправкой на малое число измерений):

$$\sigma = \sqrt{\frac{\Sigma x^2}{n-1}},$$

а по другой формуле, в которой под корнем $\Sigma x^2$ делят не на $n-1$, а на $n$:

$$\sigma = \sqrt{\frac{\Sigma x^2}{n}}$$

Эту замену в формуле ($n-1$ на $n$) мы вправе сделать при достаточно большом количестве измерений ($n$), т. к. в этом случае разница между $n-1$ и $n$ по сравнению с самим числом $n$ оказывается ничтожной. Тогда для нашего примера средние квадратические отклонения двух заданных рядов примут следующие значения:

$$\text{Для I ряда } \sigma_1 = \sqrt{\frac{80}{5}} = \sqrt{16} = \pm 4$$

$$\text{Для II ряда } \sigma_2 = \sqrt{\frac{20}{5}} = \sqrt{4} = \pm 2,$$

а для нового (третьяго) ряда, варианты которого составлены из сумм вариант двух заданных рядов,

$$\sigma_3 = \sqrt{\frac{164}{5}} = \sqrt{32{,}8} = \pm 5{,}73$$

Оказывается, что этот последний результат ($\sigma_3 = \pm 5{,}73$) мы могли бы получить и не прибегая к непосредственной обработке соответствующего ряда чисел, пользуясь следующей формулой:

$$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2 + 2r\sigma_1\sigma_2},$$

где через $r$, как всегда, обозначен коэффициент корреляции между двумя заданными рядами чисел. В нашем примере $r = +0{,}8$.

Подставив в указанную формулу числовые значения:

$$\sigma_1 = 4 \qquad \sigma_2 = 2 \qquad r = +0{,}8,$$

получим:

$$\sigma_3 = \sqrt{4^2 + 2^2 + 2\,.\,0{,}8\,.\,4\,.\,2} = \sqrt{16 + 4 + 1{,}6\,.\,8} = \\ = \sqrt{20 + 12{,}8} = \sqrt{32{,}8} = \pm 5{,}73,$$

т. е. тот же самый результат, который мы имели и раньше.

Указанная формула для вычисления $\sigma_3$ по величине $\sigma_1$, $\sigma_2$ и $r$ дает абсолютно точные результаты в том случае, если все сигмы вычисляются без поправки на малое число измерений. При обыкновенном же способе вычислений $\sigma$ данная формула имеет лишь приближенный характер и тем более точный, чем с большим числом вариант мы оперируем.

Таким образом, $\sigma$ ряда, составленного из сумм, зависит не только от соответствующих сигм двух заданных рядов, но также и от величины коэффициента корреляции между этими рядами. Чем больше $r$, тем, при прочих равных условиях, больше и $\sigma$ нового ряда и наоборот.

В частном случае при полной корреляции между числами заданных рядов (т. е. при $r = +1$) искомая сигма окажется равной сумме двух сигм этих рядов:

$$\sigma_3 = \sigma_1 + \sigma_2$$

При полной обратной корреляции (т. е. при $r = -1$) она равна разности соответствующих сигм:

$$\sigma_3 = \sigma_1 - \sigma_2.$$

При отсутствии же корреляции (т. е. при $r = 0$) искомая сигма равна квадратному корню из суммы квадратов двух заданных сигм:

$$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2}.$$



Покажем эти соотношения на числовых примерах. Пусть имеем следующие два ряда чисел:

| | | | | | |
|---|---|---|---|---|---|
| I ряд: | 14 | 18 | 20 | 22 | 26. |
| II ряд: | 7 | 9 | 10 | 11 | 13. |

В данном случае $\sigma_1 = 4$, $\sigma_2 = 2$, а коэффициент корреляции $r = +1$. Составим новый ряд, путем сложения этих чисел попарно:

III ряд: 21 27 30 33 39.

Для этого ряда $\sigma_3 = 6$ (при непосредственном ее вычислении из данного ряда).

Тот же результат мы получим и по указанной выше формуле, соответствующей случаю: $r = +1$.

$$\sigma_3 = \sigma_1 + \sigma_2 = 4 + 2 = 6.$$

Возьмем теперь те же числа, но переставим нижний их ряд в обратном порядке:

| | | | | | |
|---|---|---|---|---|---|
| I ряд: | 14 | 18 | 20 | 22 | 26. |
| II ряд: | 13 | 11 | 10 | 9 | 7. |

Здесь $\sigma_1 = 4$, $\sigma_2 = 2$ и $r = -1$.

III ряд: 27 29 30 31 33.

Для данного ряда $\sigma_3 = 2$. Но и по формуле (при $r = -1$):

$$\sigma_3 = \sigma_1 - \sigma_2 = 4 - 2 = 2.$$

Наконец, рассмотрим следующие два ряда чисел:

| | | | | | |
|---|---|---|---|---|---|
| I ряд: | 14 | 18 | 20 | 22 | 26. |
| II ряд: | 9 | 13 | 10 | 7 | 11. |

здесь $\sigma_1 = 4 \qquad \sigma_2 = 2 \qquad r = 0.$

III ряд: 23 31 30 29 37.

Для этого ряда $\sigma_3 = 4{,}47$ (если ее определять прямо из 5 заданных чисел). Но и по формуле, соответствующей случаю $r = 0$, мы будем иметь

$$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{4^2 + 2^2} = \sqrt{16 + 4} = \sqrt{20} = 4{,}47.$$

Изложенные правила вычисления $M$ и $\sigma$, относящиеся к вариантам, составленным из сумм соответствующих друг другу вариант двух других рядов, могут иметь применение, например, в том случае, когда нам известны $M$ и $\sigma$ для отдельных частей каких-либо объектов исследования и по ним требуется вычислить $M$ и $\sigma$ для этих объектов в целом. Пусть, в какой-либо группе детей был измерен рост сидя и длина ног, при чем оказалось, что

для роста сидя: $M_1 = 69{,}83 \qquad \sigma_1 = 2{,}81$
для длины ног: $M_2 = 59{,}74 \qquad \sigma_2 = 3{,}72.$

Кроме того, пусть известно, что коэффициент корреляции между ростом сидя и длиной ног:

$$r = +0{,}62.$$

Так как полная длина тела (рост стоя) каждого ребенка слагается из его роста сидя (верхней части тела) и длины ног (нижней части тела), то в данном случае мы имеем возможность при вычислении $M_3$ и $\sigma_3$, относящихся к полной длине тела всех детей, избежать измерения роста стоя у каждого отдельного ребенка, вычислив эти $M_3$ и $\sigma_3$ прямо по формулам:

$$M_3 = M_1 + M_2$$

$$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2 + 2r\sigma_1\sigma_2}\,.$$

В нашем примере, следовательно, при

$$M_1 = 69{,}83$$

и

$$M_2 = 59{,}74$$

$$M_3 = 69{,}83 + 59{,}74 = 129{,}57,$$

и т. к.

$$\sigma_1 = 2{,}81 \quad \sigma_2 = 3{,}72 \quad \text{и} \quad r = +0{,}62,$$

то

$$\sigma_3 = \sqrt{(2{,}81)^2 + (3{,}72)^2 + 2(0{,}62)\ (2{,}81)\ (3{,}72)} = \pm 5{,}89.$$

Таким образом для роста стоя

$$M_3 = 129{,}57 \text{ см} \quad \text{и} \quad \sigma_3 = \pm 5{,}89 \text{ см}.$$

Приведенная выше формула удобна в том случае, когда известен коэффициент корреляции $r$ между двумя рядами суммируемых чисел (в нашем примере между ростом сидя и длиной ног). Если же, наоборот, известной является сигма ряда, составленного из сумм (т. е. $\sigma_3$), то коэффициент корреляции $r$, в свою очередь, может быть вычислен по обратной формуле:



$$r = \frac{\sigma_3^2 - \sigma_1^2 - \sigma_2^2}{2\sigma_1\sigma_2}\,.$$

В нашем примере следовательно коэффициент корреляции между ростом сидя и длиной ног может быть вычислен по этой формуле, если известны все три сигмы, характеризующие варьирование роста стоя ($\sigma_3$), роста сидя ($\sigma_1$) и длины ног ($\sigma_2$). Подставив в нашу формулу эти числовые данные:

$$\sigma_3 = 5{,}89 \qquad \sigma_1 = 2{,}81 \qquad \text{и} \ \sigma_2 = 3{,}72$$

получим:

$$r = \frac{(5{,}89)^2 - (2{,}81)^2 - (3{,}72)^2}{2(2{,}81)\ (3{,}72)} = +0{,}62.$$

Итак, коэффициент корреляции для роста сидя и длины ног

$$r = +0{,}62.$$

В некоторых случаях может представить интерес производный (III) ряд, составленный не из сумм, а из разностей вариант двух заданных рядов. Например:

| | | | | | |
|---|---|---|---|---|---|
| I ряд: | 14 | 18 | 20 | 22 | 26 |
| II ряд: | 7 | 11 | 10 | 9 | 13 |
| III ряд: | 7 | 7 | 10 | 13 | 13 |

Между средними арифметическими $M_3$ и средним квадратическим отклонением $\sigma_3$ в этих рядах существует следующая зависимость:

$$M_3 = M_1 - M_2 \qquad \sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2}$$

в чем не трудно убедиться на приведенных выше числовых примерах. В частном случае:

при $r = +1 \qquad \sigma_3 = \sigma_1 - \sigma_2$
при $r = -1 \qquad \sigma_3 = \sigma_1 + \sigma_2$
при $r = 0 \qquad \sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2}$

Практически эти формулы могут быть использованы, например, в том случае, когда известными являются $M$ и $\sigma$ для полной длины тела (роста стоя) и для роста сидя, причем требуется по этим данным вычислить $M$ и $\sigma$ для длины ног.

Если коэффициент корреляции между ростом стоя и ростом сидя

$$r = +0{,}87$$

Для роста стоя

$$M_1 = 129{,}57 \qquad \sigma_1 = 5{,}89.$$

Для роста сидя

$$M_2 = 69{,}83 \qquad \sigma_2 = 2{,}81,$$

то $M_3$ и $\sigma_3$ для длины ног можно найти по формулам:

$$M_3 = M_1 - M_2 \qquad \sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2}.$$

Подставив сюда числовые данные нашего примера, получим:

$$M_3 = 129{,}57 - 69{,}83 = 59{,}74$$

$$\sigma_3 = \sqrt{(5{,}89)^2 + (2{,}81)^2 - 2(0{,}87)\,(5{,}89)\,(2{,}81)} = \pm 3{,}71.$$

Если же, наоборот, известной является $\sigma_3 = 3{,}71$, а искомым — коэффициент корреляции $r$ между ростом стоя и ростом сидя, то для определения $r$ по нашим данным можно воспользоваться обратной формулой, которая в применении к данному случаю будет иметь такой вид:

$$r = \frac{\sigma_1^2 + \sigma_2^2 - \sigma_3^2}{2\sigma_1\sigma_2}.$$

Подставив сюда числовые значения $\sigma_1$, $\sigma_2$ и $\sigma_3$, получим:

$$r = \frac{(5{,}89)^2 + (2{,}81)^2 - (3{,}71)^2}{2(5{,}89)\,(2{,}81)} = +0{,}87.$$

Эту же зависимость можно распространить и на случай суммации многих признаков. Пусть нам известны следующие $M$ и $\sigma$, вычисленные, например, для трех частей руки, составляющих в общей сложности полную ее длину.

| | | |
|---|---|---|
| Для | плеча | $M_1$ и $\sigma_1$ |
| „ | предплечья | $M_2$ и $\sigma_2$ |
| „ | кисти | $M_3$ и $\sigma_3$. |

Пусть, кроме того, известны коэффициенты корреляции между этими размерами:

1) Коэффициент корреляции между длиной плеча и предплечья: $r_{12}$.



2) Коэффициент корреляции между длиной плеча и кистью: $r_{13}$.

3) Коэффициент корреляции между длиной предплечья и кистью: $r_{23}$.

По этим данным можем вычислить $M$ и $\sigma$ для полной длины руки, пользуясь формулами:

$$M = M_1 + M_2 + M_3$$

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2r_{12}\sigma_1\sigma_2 + 2r_{13}\sigma_1\sigma_3 + 2r_{23}\sigma_2\sigma_3}$$

Зависимость между сигмами частей какого-либо полного размера и его общей сигмой имеет место и в применении к соответствующим средним ошибкам ($m$) этих размеров, так как при постоянном числе измерений ($n$) средняя ошибка зависит лишь от соответствующей сигмы, увеличиваясь и уменьшаясь пропорционально величине последней, что и определяется формулой:

$$m = \frac{\sigma}{\sqrt{n}}.$$

Поэтому все то, что нами было установлено в применении к средним квадратическим отклонениям ($\sigma$), можно безоговорочно распространить и на средние ошибки $m$.

Так, если новый (третий) ряд составляется из сумм вариант двух заданных рядов (первого и второго), то между средней ошибкой ($m_3$) нового ряда и средними ошибками ($m_1$ и $m_2$) двух заданных рядов существует зависимость, определяемая формулой:

$$m_3 = \sqrt{m_1^2 + m_2^2 + 2r\,m_1m_2},$$

где через $r$ обозначен коэффициент корреляции между двумя заданными рядами суммируемых вариант. В частности:

| | | |
|---|---|---|
| при | $r = +1$ | $m_3 = m_1 + m_2$ |
| „ | $r = -1$ | $m_3 = m_1 - m_2$ |
| „ | $r = 0$ | $m_3 = \sqrt{m_1^2 + m_2^2}$. |

Если же новый ряд составляется не из сумм, а из разностей, то

$$m_3 = \sqrt{m_1^2 + m_2^2 - 2r\,m_1m_2}$$

в частности:

$$
\begin{array}{lll}
\text{при} & r = +1 & m_3 = m_1 - m_2 \\
\text{„} & r = -1 & m_3 = m_1 + m_2 \\
\text{„} & r = 0 & m_3 = \sqrt{m_1^2 + m_2^2}.
\end{array}
$$

В правильности этих соотношений легко убедиться из рассмотрения тех же числовых примеров, на которых была выяснена зависимость между соответствующими сигмами (см. стр. 221).

Особый интерес может для нас представить случай составления нового ряда, посредством вычитания из вариант первого ряда соответствующих им вариант второго ряда. При этих условиях средняя ошибка нового ряда может быть, как было уже указано выше, вычислена по формуле:

$$m_3 = \sqrt{m_1^2 + m_2^2 - 2r\, m_1 m_2}.$$

Приведем числовой пример использования этой формулы. Пусть числа первого ряда представляют собой результат взвешивания группы детей при поступлении их в санаторию, а числа второго ряда — вес этих же детей в конце какого-либо периода пребывания их в этой санатории:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| I ряд (до): | 29 | 32 | 34 | 27 | 33 | 35 | 31 | 28 | 30 |
| II ряд (после): | 35 | 36 | 38 | 31 | 33 | 37 | 34 | 30 | 32 |

Пусть, далее, нам требуется выяснить результаты пребывания этих детей в санатории в отношении влияния соответствующего режима на их вес (на сколько килограмм они в среднем прибавили в весе, и достоверным ли является наше утверждение, что эта прибавка не носит случайного характера).

Идя обычным путем, мы должны были бы вычислить средний вес этих детей при поступлении в санаторию ($M_1$), средний их вес в конце периода санаторного лечения ($M_2$) и обе средние ошибки ($m_1$ и $m_2$) этих средних арифметических.

Далее, применив обычный критерий достоверности (см. стран. 45), мы должны были бы убедиться в том, что

$$\frac{M_2 - M_1}{\sqrt{m_2^2 + m_1^2}} > 3.$$

В нашем примере:



Для I ряда:

$$M_1 = 31 \text{ кг} \quad m_1 = 0{,}86 \text{ кг}$$

Для II ряда:

$$M_2 = 34 \text{ кг} \quad m_2 = 0{,}86 \text{ кг}$$

Применяя этот критерий, получим:

$$\frac{M_2 - M_1}{\sqrt{m_2^2 + m_1^2}} = \frac{34 - 31}{\sqrt{(0{,}86)^2 + (0{,}86)^2}} = \frac{3}{\sqrt{0{,}74 + 0{,}74}} = \frac{3}{\sqrt{1{,}48}} = \frac{3}{1{,}22} = 2{,}46.$$

Итак, в данном случае, разница между $M_2$ и $M_1$, т. е. средняя прибавка в весе, не может еще считаться достоверной, т. к. абсолютная величина этой разницы (т. е. 3) превышает свою среднюю ошибку (1,22) только в 2,46 раза, а не в 3 раза, как это требуется. Невозможность доказать наличие достоверного увеличения в весе этих 9 детей объясняется здесь чрезмерно большой средней ошибкой данной разницы. В самом деле, среднюю ошибку прибавки веса мы определяли здесь как

$$\sqrt{m_1^2 + m_2^2},$$

т. е. по формуле, соответствующей тому случаю, когда коэффициент корреляции между двумя заданными рядами чисел

$$r = 0.$$

Ничего иного мы и не имели права предположить, т. к., рассматривая изолированно вес детей до поступления в санаторию и вес их после проделанного курса лечения, мы не обладали никакими сведениями о корреляции этих данных, почему, из осторожности, и должны были считать эту корреляцию отсутствующей (т. е. $r = 0$). Если бы речь шла здесь о разных детях, из которых 9 человек были взвешены при приеме, а другие 9 при выпуске, то о корреляции между этими данными и вообще вопрос бы не поднимался, т. к. совершенно неизвестно, каких-же детей из первой группы следует сопоставить с определенными детьми из второй группы при выяснении корреляции между их весами. Но в нашем примере варианты двух заданных рядов распадаются уже на естественные пары, что и дает право на выяснение вопроса о корреляции между этими рядами. В самом деле, каждый ребенок имел два измерения своего веса (до и после), следовательно каждая варианта первого ряда (вес ребенка при поступлении в санаторию) имела свою парную варианту во втором ряду (вес того же ребенка после лечения).

Вычислив коэффициент корреляции между двумя рядами заданных чисел, получим:

$$r = +0{,}8.$$

На этом основании среднюю ошибку разностей вариант этих рядов (т. е. средней прибавки в весе) мы можем теперь вычислить по формуле:

$$m = \sqrt{m_1^2 + m_2^2 - 2\,r m_1 m_2}.$$

Подставив сюда наши числовые значения:

$$m_1,\ m_2 \text{ и } r,$$

получим

$$m = \pm 0{,}54.$$

Это новое значение средней ошибки разности, вычисленное с учетом корреляции между весом до и весом после, оказалось несколько меньше полученного ранее (без учета данной корреляции, т. е. в предположении, что $r = 0$).

В данном случае абсолютная величина самой разности (3) превышает свою среднюю ошибку (0,54) уже более чем в 3 раза

$$\frac{3}{0{,}55} = 5{,}45.$$

Следовательно, приняв во внимание корреляцию между заданными рядами чисел, мы получили результат сравнения уже достаточно достоверный.

Практически, выгоднее будет, взамен вычисления коэффициента корреляции $r$, несколько видоизменить самый ход обработки первичного материала, подвергнув этой обработке не оба заданные ряда вариант порознь, а сразу же вычислив разности между весом до и после для каждого из 9 детей отдельно. Этим путем мы получим, очевидно, те же самые результаты. В самом деле:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| I ряд (вес до): | 29 | 32 | 34 | 27 | 33 | 35 | 31 | 28 | 30 |
| II ряд (вес после): | 35 | 36 | 38 | 31 | 33 | 37 | 34 | 30 | 32 |
| III ряд (прибавка в весе): | 6 | 4 | 4 | 4 | 0 | 2 | 3 | 2 | 2. |



Для третьего ряда

$$M = 3 \qquad m = \pm 0{,}54$$

что мы имели и раньше, применяя формулы

$$M = M_2 - M_1 \qquad m = \sqrt{m_1^2 + m_2^2 - 2r\,m_1 m_2}.$$

Итак, во всех случаях, когда предметом конечного исследования является средняя величина разности между двумя заданными рядами вариант, и варианты эти могут быть сгруппированы в естественные пары, выгоднее бывает подвергать обработке не отдельные ряды (в изолированном друг от друга виде), а сразу же составить новый ряд из разностей соответствующих пар вариант и, вычислив среднее арифметическое $M$ и среднюю ошибку $m$ для этого нового ряда, применить обычный критерий оценки достоверности, требующий, чтобы средняя величина разности (т. е. в данном случае само $M$) превышала свою ошибку $m$ по крайней мере в 3 раза

$$\frac{M}{m} > 3.$$

При сравнении, например, среднего роста мальчиков ($M_1$) со средним ростом девочек ($M_2$) мы вынуждены итти обычным путем изолированной обработки каждого ряда порознь и применять критерий оценки достоверности:

$$\frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} > 3.$$

Но, если речь идет не о мальчиках и девочках вообще, а, скажем, о братьях и сестрах, то нам выгоднее будет вычислить сперва разности между ростом каждого брата и ростом его сестры, после чего подвергнуть обработке уже этот новый ряд разностей, и, вычислив для него $M$ и $m$, применить критерий оценки достоверности

$$\frac{M}{m} > 3$$

каковой, при наличии прямой корреляции между ростом братьев и сестер, даст всегда лучшие в отношении достоверности результаты.

В заключение рассмотрим еще следующий случай вычисления среднего арифметического $M$ и среднего квадратического отклонения $\sigma$, иногда встречающийся на практике. Пусть совокупность 46 вариант была предварительно разбита на три части, и для каждой из этих частей порознь были вычислены средние и сигмы, числовые значения которых сохранились и находятся в нашем распоряжении. По этим данным требуется определить теперь $M_0$ и $\sigma_0$ для всех 46 вариант, не прибегая к обычным промежуточным вычислениям. Пусть в числах:

| | | | |
|---|---|---|---|
| для первой части ряда | $n_1 = 20$ | $M_1 = 11$ | $\sigma_1 = \pm 2{,}18$ |
| „ второй „ „ | $n_2 = 16$ | $M_2 = 10$ | $\sigma_2 = \pm 2{,}16$ |
| „ третьей „ „ | $n_3 = 10$ | $M_3 = 8$ | $\sigma_3 = \pm 1{,}82.$ |

Среднее арифметическое $M_0$ для объединенного ряда состоящего из всех трех частей, может быть вычислено по формуле:

$$M_0 = \frac{n_1M_1 + n_2M_2 + n_3M_3}{n_1 + n_2 + n_3}$$

или короче

$$M_0 = \frac{\Sigma nM}{\Sigma n}.$$

Эти формулы показывают, что для вычисления $M_0$ следует среднее арифметическое каждой отдельной части ряда помножить на число вариант в данном ряду, все эти произведения сложить и сумму их разделить на сумму вариант во всех трех частях:

$$\begin{aligned} n_1M_1 &= 20 \cdot 11 = 220 \\ n_2M_2 &= 16 \cdot 10 = 160 \\ n_3M_3 &= 10 \cdot \ 8 = \ 80 \\ \hline \Sigma nM &= 460 \end{aligned}$$

Так как сумма вариант во всех трех частях ряда

$$\Sigma n = n_1 + n_2 + n_3 = 20 + 16 + 10 = 46,$$

то

$$M_0 = \frac{\Sigma nM}{\Sigma n} = \frac{460}{46} = 10.$$

Итак, общее для всего ряда среднее арифметическое [1])

[1]) Такое среднее арифметическое называется „взвешанным“.



$$M = 10.$$

Для вычисления общего среднего квадратичечкого отклонения ($\sigma_0$) служит следующая формула:

$$\sigma_0 = \pm \sqrt{\frac{\Sigma (n-1)\,\sigma^2 + \Sigma n\,(M - M_0)^2}{\Sigma n - 1}}.$$

Разъясним порядок действий, необходимых для вычисления $\sigma_0$ по этой формуле на числовых данных нашего примера. Здесь слева в числителе под корнем стоит выражение:

$$\Sigma (n-1)\,\sigma^2.$$

Более подробно его можно представить в следующем виде:

$$\Sigma (n-1)\,\sigma^2 = (n_1 - 1)\,\sigma_1^2 + (n_2 - 1)\,\sigma_2^2 + (n_3 - 1)\,\sigma_3^2.$$

Следовательно для вычисления величины, обозначенной символом $\Sigma (n-1)\,\sigma^2$, необходимо для каждой отдельной части ряда квадрат соответствующей сигмы помножить на число вариант этой части без единицы и все эти произведения сложить:

$$\begin{aligned} (n_1 - 1)\,\sigma_1^2 &= (20 - 1)\,2{,}18^2 = 19 \cdot 4{,}7524 = \ 90{,}2956 \\ (n_2 - 1)\,\sigma_2^2 &= (16 - 1)\,2{,}16^2 = 15 \cdot 4{,}6656 = \ 69{,}9840 \\ (n_3 - 1)\,\sigma_3^2 &= (10 - 1)\,1{,}82^2 = \ 9 \cdot 3{,}3124 = \ 29{,}8116 \\ \hline \Sigma (n-1)\,\sigma^2 &= 190{,}0912. \end{aligned}$$

Рядом с этим выражением в числителе нашей формулы под корнем стоит другое:

$$\Sigma n\,(M - M_0)^2.$$

Более подробно его можно представить в следующем виде:

$$\Sigma n\,(M - M_0)^2 = n_1\,(M_1 - M_0)^2 + n_2\,(M_2 - M_0)^2 + n_3\,(M_3 - M_0)^2.$$

Следовательно для вычисления этого выражения необходимо для каждой отдельной части ряда вычислить разность между данным частным средним и общим для всего ряда средним $M_0$, эту разность возвысить в квадрат и помножить на число вариант в данной части ряда, после чего все эти произведения сложить:

$$n_1(M_1-M_0)^2=20(11-10)^2=20(+1)^2=20\cdot 1=20$$
$$n_2(M_2-M_0)^2=16(10-10)^2=16\cdot 0^2\quad=16\cdot 0=\ 0$$
$$n_3(M_3-M_0)^2=10(\ 8-10)^2=10\cdot(-2)^2=10\cdot 4=40$$
$$\Sigma n(M-M_0)^2=60.$$

Подставляя эти данные в формулу для вычисления $\sigma_0$ получим:

$$\sigma_0=\pm\sqrt{\frac{\Sigma(n-1)\sigma^2+\Sigma n(M-M_0)^2}{\Sigma n-1}}=\pm\sqrt{\frac{190{,}0912+60}{46-1}}=$$
$$=\pm\sqrt{\frac{250{,}0912}{45}}=\pm\sqrt{5{,}5575}=\pm 2{,}36.$$

Итак, среднее квадратическое отклонение для всех 46 вариант

$$\sigma_0=\pm 2{,}36.$$

Средняя ошибка $m_0$ для всего объединенного ряда, очевидно, может быть вычислена по обычной формуле:

$$m_0=\frac{\sigma_0}{\sqrt{\Sigma n}}=\frac{2{,}36}{\sqrt{46}}=\pm 0{,}35.$$

Запись всех этих вычислений удобно вести в следующей расчетной решетке (см. табл. 86).

ТАБЛИЦА 86.

| Обозначения действий и записей | Частные данные | | | Общие результаты |
|---|---|---|---|---|
| | I ряд | II ряд | III ряд | |
| $n$ | 20 | 16 | 10 | $\Sigma n=46$ |
| $M$ | 11 | 10 | 8 | $M_0=10$ |
| $\sigma$ | 2,18 | 2,16 | 1,82 | $\sigma_0=2{,}36$ |
| $nM$ | 220 | 160 | 80 | $\Sigma nM=460$ |
| $\sigma^2$ | 4,7524 | 4,6656 | 3,3124 | — |
| $(n-1)\sigma^2$ | 90,2956 | 69,9840 | 29,8116 | $\Sigma(n-1)\sigma^2=190{,}0912$ |
| $M-M_0$ | 1 | 0 | −2 | — |
| $(M-M_0)^2$ | 1 | 0 | 4 | — |
| $n(M-M_0)^2$ | 20 | 0 | 40 | $\Sigma n(M-M_0)^2=60$ |



Порядок последовательного заполнения отдельных клеток этой решетки понятен из приведенного выше описания способа соответствующих вычислений.

В тех случаях, когда все сигмы вычисляются по формуле без поправки на малое число измерений (т.-е. когда $\Sigma x^2$ делится под корнем не на $n-1$, а просто на $n$), для определения общего среднего квадратического отклонения $\sigma_0$ следует пользоваться формулой:

$$\sigma_0=\pm\sqrt{\frac{\Sigma n\sigma^2+\Sigma n(M-M_0)^2}{\Sigma n}}.$$

## XVIII. Статистический анализ качественных признаков.

В тех случаях, когда варьирующие признаки изменяются качественно, а не количественно, удобно подразделить все объекты исследования на соответствующие группы (классы) и частоты этих классов выразить в процентах от общего количества измерений (наблюдений). Пусть, например, из 76 исследованных детей 19 оказались туберкулезными, а остальные 57 — здоровыми*). Тогда данный коллектив исследованных объектов распадется по признаку наличия туберкулеза на два класса: больных и здоровых. Процентные частоты ($q$) каждого из этих классов мы найдем по формуле:

$$q=\frac{100.n_1}{n}\ \%,$$

где через $n_1$ обозначено количество объектов, обладающих данным качественным признаком, а через $n$ — общее количество всех исследованных объектов. В нашем примере для туберкулезных детей $n_1=19$, а $n=76$. Следовательно процент туберкулезных детей в данном их коллективе

$$q=\frac{100.19}{76}=25\%.$$

*) Пример вымышленный.

Для возможности сопоставления таких процентных частот друг с другом в целях оценки достоверности обнаруженной между ними разницы необходимо научиться вычислять также и средние ошибки этих величин. С поправкой на ограниченность числа измерений (наблюдений) соответствующая формула для этого будет иметь следующий вид:

$$m_q = \pm \sqrt{\frac{q\,(100 - q)}{n - 1}}\,\%,$$

где через $m_q$ обозначена искомая средняя ошибка, через $q$ — заданный процент, а через $n$ — общее количество наблюдений, из которых этот процент выведен. В нашем последнем примере для туберкулезных детей $q = 25$, а $n = 76$. Следовательно:

$$m_q = \pm \sqrt{\frac{25\,(100 - 25)}{76 - 1}} = \pm \sqrt{\frac{25 . 75}{75}} = \pm \sqrt{25} = \pm 5\%.$$

Итак, в условиях нашего примера:

$$q = 25 \pm 5.$$

Допустим, что в каком-либо другом детском коллективе туберкулезных детей оказалось 13 из 26. Тогда процент больных детей найдем по формуле:

$$q = \frac{100 \cdot 13}{26} = 50\%,$$

а среднюю ошибку этой величины — по формуле:

$$m_q = \pm \sqrt{\frac{50(100 - 50)}{26 - 1}} = \pm \sqrt{\frac{50 . 50}{25}} = \pm \frac{50}{5} = \pm 10\%.$$

Итак, в первом случае $q_1 = 25 \pm 5$

а во втором . . . . . $q_2 = 50 \pm 10.$

Спрашивается, достоверной или случайной является обнаруженная здесь разница в заболеваемости туберкулезом среди детей двух этих коллективов?

Для решения этого вопроса подставим наши числовые данные в формулу для оценки достоверности различия между двумя величинами, вычисленными с их средними ошибками (см. гл. V):

$$\frac{(50 - 25)^2}{5^2 + 10^2} = \frac{(25)^2}{25 + 100} = \frac{625}{125} = 5.$$



Так как в результате этой подстановки мы получили здесь число 5 (а не 9, как это требуется), то обнаруженное различие в обоих процентах (25 и 50) в данном случае является лишь **вероятным**, но еще не вполне доказанным, что объясняется, очевидно, недостаточным количеством произведенных наблюдений (всего лишь 26 и 76).

Этот же способ вычисления средних процентных ошибок применим и в том случае, когда по данному качественному признаку весь материал распадается на несколько групп (например, при подразделении детей на конституциональные типы, при подсчете растений с листьями различной формы и т. д.).

ТАБЛИЦА 87.

| потомки / предки | − | + |
|---|---|---|
| + | 6 | 30 |
| − | 44 | 20 |

ТАБЛИЦА 88.

| | 50 | 50 | |
|---|---|---|---|
| 36 | 6 | 30 | |
| 64 | 44 | 20 | 30 |
| | | 44 | 26 |

Между двумя качественными признаками, так же как и между количественными, может существовать более или менее тесная связь (прямая или обратная), характеризующаяся определенной величиной коэффициента корреляции $r$. Пусть, например, требуется узнать, наследуется ли какой-либо качественный признак (цвет волос или глаз, какой-либо физический недостаток, специальная одаренность, темперамент и т. д.). Предположим, что из ста обследованных объектов 36 человек являлись потомками лиц, **обладающих** данным признаком, причем 30 человек его унаследовали, а 6 нет; предки же остальных 64 человек данным признаком **не обладали**, и из числа их потомков 44 человека этого признака также не имели, а 20 имели. Эти числовые данные непосредственного подсчета можно разместить в обычной корреляционной решетке (см. табл 87). Здесь в заголовках наличие признака условно обозначено знаком +, а отсутствие помечено знаком —.

Имея эти частоты, мы могли бы вычислить коэффициент корреляции $r$ обычным способом (см. гл. XV). Для этого необходимо просуммировать частоты по горизонтальным, вертикальным и диагональным направлениям (см. табл. 88) и результат этого подсчета обработать в обычной расчетной решетке по способу сумм (см. табл. 89).

ТАБЛИЦА 89.

| Ряд $x$ | | | Ряд $y$ | | | Ряд $z$ | | | Контр. сводка (ряд $u$) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0 | 0 | $p$ | 36 | 0 | $p$ | 30 | 0 | $p$ | 66 | 0 |
| | | — | 36 | 36 | — | 30 | 30 | — | 66 | 66 | — |
| 50 | — | — | 64 | — | — | 26 | — | — | 140 | — | — |
| 50 | 50 | — | | | — | 44 | 44 | — | 94 | 94 | — |
| 100 | 50 | 0 | 100 | 0 | 0 | 100 | 44 | — | 300 | 94 | — |
| $S_1 = -50$ | | | $S_1 = +36$ | | | $S_1 = -14$ | | | $S_1 = -28$ | | |
| $S_2 = 50$ | | | $S_2 = 36$ | | | $S_2 = 74$ | | | $S_2 = 160$ | | |
| $\Sigma x^2 = 25$ | | | $\Sigma y^2 = 23{,}04$ | | | $\Sigma z^2 = 72{,}04$ | | | $\Sigma xy = +12$ | | |

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \,.\, \Sigma y^2}} = \frac{+12}{\sqrt{25 \,.\, 23{,}04}} = \frac{+12}{\sqrt{576}} = \frac{+12}{24} = +0{,}5.$$

Но то же самое значение коэффициента корреляции ($+0{,}5$) в данном случае мы можем получить и проще. Обозначим четыре частоты, оказавшиеся в клетках нашей корреляционной решетки, через $p_1$, $p_2$, $p_3$ и $p_4$ (см. табл. 90).

ТАБЛИЦА 90.

| $y$ \\ $x$ | — | + |
|---|---|---|
| + | $p_1$ | $p_2$ |
| — | $p_3$ | $p_4$ |

В нашем примере, следовательно:

$$p_1 = 6 \qquad p_2 = 30$$
$$p_3 = 44 \qquad p_4 = 20.$$

При таком обозначении коэффициент корреляции $r$ качественных признаков может быть вычислен по формуле:

$$r = \frac{p_2 p_3 - p_1 p_4}{\sqrt{(p_1 + p_2)(p_3 + p_4)(p_1 + p_3)(p_2 + p_4)}}.$$



Подставив сюда наши числовые данные, получим:

$$r = \frac{30 \,.\, 44 - 6 \,.\, 20}{\sqrt{(6+30)(44+20)(6+44)(30+20)}} = \frac{1320 - 120}{\sqrt{36 \,.\, 64 \,.\, 50 \,.\, 50}} =$$
$$= \frac{+1200}{2400} = +0{,}5.$$

Средняя ошибка этого коэффициента корреляции найдется по обычной формуле:

$$m_r = \pm \frac{1 - r^2}{\sqrt{n}} = \pm \frac{1 - (0{,}5)^2}{\sqrt{100}} = \pm \frac{1 - 0{,}25}{10} = \pm \frac{0{,}75}{10} = \pm 0{,}08.$$

Следовательно окончательно будем иметь

$$r = +0{,}50 \pm 0{,}08.$$

Здесь отношение $\frac{r}{m_r} = \frac{0{,}50}{0{,}08} = 6{,}25$. Так как отношение это оказывается больше 3, то достоверность наследственной передачи данного признака в условиях последнего примера можно считать доказанной.

При вычислении $r$ с помощью логарифмических таблиц следует найти сперва логарифм выражения:

$$\sqrt{(p_1 + p_2)(p_3 + p_4)(p_1 + p_3)(p_2 + p_4)}$$

Для этого логарифмы всех четырех множителей складываются, и сумма их делится на 2. У нас:

$$+\left\{\begin{array}{l} lg\ (p_1 + p_2) = lg\ 36 = 1{,}55630 \\ lg\ (p_3 + p_4) = lg\ 64 = 1{,}80618 \\ lg\ (p_1 + p_3) = lg\ 50 = 1{,}69897 \\ lg\ (p_2 + p_4) = lg\ 50 = 1{,}69897 \end{array}\right.$$

сумма этих логарифмов $= 6{,}76042$.

Следовательно искомый логарифм этого выражения равен

$$\frac{6{,}76042}{2} = 3{,}38021.$$

Для вычисления коэффициента корреляции $r$ следует, далее, из логарифма выражения $(p_2 p_3 - p_1 p_4)$ вычесть только-что найденный логарифм 3,38021. Числовое значение первого выражения найдем без логарифмирования; оно, очевидно, равно $+1200$. Итак:

$$-\left\{\begin{array}{ll} lg\ 1200 & =3{,}07918 \\ \text{найденный } lg & =3{,}38021 \end{array}\right.$$

$$\text{разность этих логарифмов} = \overline{1}{,}69897.$$

Теперь остается лишь по этому логарифму отыскать соответствующее ему число. По таблицам Пржевальского находим, что логарифму $\overline{1}{,}69897$ соответствует число 0,5.

Следовательно искомый коэффициент корреляции

$$r = +0{,}5.$$

Выше мы разобрали случай корреляции двух качественных признаков, распадающихся всего только на две группы (или класса). В действительности, однако, нередко **приходится** иметь дело и с подразделением каждого такого признака на несколько отличных друг от друга категорий, например, по цвету: красный, синий, белый, желтый и т. д. Такие же подразделения возможны, очевидно, и по форме (например, листьев), по конституциональным особенностям (детей), по профессиям (их родителей) и т. д.

ТАБЛИЦА 91.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| $B_1$ | 1 | 1 | 2 | 16 |
| $B_2$ | 7 | 0 | 2 | 1 |
| $B_3$ | 0 | 33 | 16 | 1 |
| $B_4$ | 2 | 16 | 0 | 2 |

В этом случае распределение всех объектов исследования по нескольким группам двух качественных признаков (например, по форме и по цвету) может быть представлено в виде следующей коррелляционной решетки (см. таблицу 91).

Здесь в заголовках вертикальных столбцов (сверху) символами: $A_1$, $A_2$, $A_3$, и $A_4$ условно обозначена классификация объектов исследования по группам одного из двух качественных признаков (например, по цвету), а в заголовках горизонтальных строчек (слева) символами $B_1$, $B_2$, $B_3$, и $B_4$ указаны группы тех же объектов по другому признаку (например, по форме). Числа, поставленные в отдельных клетках внутри этой решетки, обозначают частоты $p$.

Так, например, частота 16, проставленная в этой таблице в крайней справа верхней клетке, указывает на то,



что среди всех 100 объектов исследования оказалось 16 таких, которые по своему цвету относились к группе $A_4$ и в то же время по форме принадлежали к группе $B_1$ и т. д.

По отношению к этим данным нас может заинтересовать вопрос о том, в какой мере и как именно каждая пара категорий этих двух признаков ($A$ и $B$) являются связанными друг с другом, например, коррелирует ли цвет $A_4$ с формой $B_1$ или цвет $A_2$ с формой $B_3$ и т. д.

В каждом частном случае мы можем разрешить этот вопрос путем вычисления соответствующего коэффициента корреляции $r$ по указанной выше формуле (для двух качественных признаков)

$$r = \frac{p_2 p_3 - p_1 p_4}{\sqrt{(p_1+p_2)(p_3+p_4)(p_1+p_3)(p_2+p_4)}}.$$

Пусть, например, требуется вычислить коэффициент корреляции между цветом $A_4$ и формой $B_1$. Чтобы свести

ТАБЛИЦА 92.

| | | не $A_4$ | | | |
|---|---|---|---|---|---|
| | | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
| | $B_1$ | 1 | 1 | 2 | 16 |
| не $B_1$ | $B_2$ | 7 | 0 | 2 | 1 |
| | $B_3$ | 0 | 33 | 16 | 1 |
| | $B_4$ | 2 | 16 | 0 | 2 |

ТАБЛИЦА 93.

| | (—) не $A_4$ | (+) $A_4$ |
|---|---|---|
| (+) $B_1$ | 4 | 16 |
| (—) не $B_1$ | 76 | 4 |

этот вопрос к корреляции качественных признаков, распадающихся только на две группы, мы должны были бы составить новую корреляционную решетку из 4-х клеток, объединив группы по цвету: $A_1$, $A_2$ и $A_3$ в одну общую группу объектов исследования, не имеющих цвета $A_4$ (т.-е. в группу „не $A_4$“) и, совершенно также, сгруппировать формы $B_2$, $B_3$ и $B_4$ в одну общую категорию („не $B_1$“). Это объединение групп и соответствующий подсчет частот в таких составных группах наглядно представлен на таблицах 92 и 93. В первой из них все частоты перераспределены между 4-мя новыми клетками, а во-второй произведен подсчет частот, оказавшихся после этой перегруппировки в каждой такой клетке. В этом новом виде наша корреляционная решетка указывает распределение всех 100 объектов исследования по наличию (+) и отсутствию (—) признаков $A_4$ и $B_1$, что и дает право применить для вычисления коэффициента корреляции $r$ ранее указанную формулу:

$$r = \frac{p_2 p_3 - p_1 p_4}{\sqrt{(p_1 + p_2)(p_3 + p_4)(p_1 + p_3)(p_2 + p_4)}}.$$

Здесь:

$$p_1 = 4 \qquad p_2 = 16$$
$$p_3 = 76 \qquad p_4 = 4.$$

Подставляя эти данные в нашу формулу, получим:

$$r = \frac{16 . 76 - 4 . 4}{\sqrt{(4+16)(76+4)(4+76)(16+4)}} = \frac{1216 - 16}{\sqrt{20 . 80 . 80 . 20}} = \frac{+1200}{1600} = +0{,}75.$$

Итак, коэффициент корреляции между цветом $A_4$ и формой $B_1$ в условиях нашего примера оказался равным $+0{,}75$, что свидетельствует о наличии довольно значительной прямой связи этих признаков. Этот результат показывает, что объекты исследования, имеющие цвет $A_4$, в большинстве случаев обладают в то же время и формой $B_1$. Если бы этот коэффициент оказался отрицательным, то мы имели бы обратную зависимость, при которой данный цвет $A_4$, как общее правило, чаще всего не наблюдался бы в сочетании с формой $B_1$.

Совершенно также мы могли бы вычислить и другие коэффициенты корреляции, например, между цветом $A_2$ и формой $B_3$, между цветом $A_3$ и формой $B_4$ и т. д.

Однако, при вычислении всех коэффициентов по указанной выше формуле, работу эту очень усложняет предварительное перераспределение частот первичной корреляционной решетки по новым 4-м клеткам. Это затруднение можно устранить, если при вычислении коэффициентов $r$ пользоваться другой формулой, способ применения которой мы разъясним здесь на том же самом числовом примере (см. таблицу 94).

Подсчитаем прежде всего общее количество объектов исследования по каждому цвету отдельно и, независимо от этого, количество их по каждой форме. Для этого придется



сложить заданные частоты по всем вертикальным столбцам и по всем горизонтальным строчкам. В результате мы получим итоги ($p_x$) суммирования частот по столбцам:

$$10, \quad 50, \quad 20 \text{ и } 20$$

и такие же итоги ($p_y$) суммирования по строчкам:

$$20, \quad 10, \quad 50 \text{ и } 20.$$

Эти итоги ($p_x$ и $p_y$) мы запишем снизу и справа нашей расчетной решетки, как показано на таблице 94.

ТАБЛИЦА 94.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $p_y$ |
|---|---|---|---|---|---|
| $B_1$ | 1 | 1 | 2 | 16 | 20 |
| $B_2$ | 7 | 0 | 2 | 1 | 10 |
| $B_3$ | 0 | 33 | 16 | 1 | 50 |
| $B_4$ | 2 | 16 | 0 | 2 | 20 |
| $p_x$ | 10 | 50 | 20 | 20 | |

Очевидно, что общая сумма всех итогов $p_x$ равна сумме всех итогов $p_y$, представляя собой ничто иное, как общее количество всех объектов исследования ($n$).

$$\Sigma p_x = \Sigma p_y = n$$

При этих условиях любой коэффициент корреляции $r$ между двумя признаками (например $A_4$ и $B_1$) мы можем вычислить по формуле:

$$r = \frac{np - p_x p_y}{\sqrt{p_x(n - p_x)}\ \sqrt{p_y(n - p_y)}}$$

где через $p$—обозначена частота той клетки корреляционной решетки (в данном случае 16), которая находится на пересечении нужного столбца ($A_4$) с соответствующей строчкой ($B_1$), через $p_x$—общий итог (сумма частот) данного столбца ($A_4$), а через $p_y$ — итог (сумма частот) соответствующей строчки ($B_1$).

В нашем примере при вычислении коэффициента корреляции между цветом $A_4$ и формой $B_1$

$$n = 100 \qquad p_x = 20$$
$$p = 16 \qquad p_y = 20$$

Следовательно искомый коэффициент корреляции

$$r = \frac{100 \cdot 16 - 20 \cdot 20}{\sqrt{20(100-20)}\ \sqrt{20(100-20)}} = \frac{1600-400}{\sqrt{20 \cdot 80}\ \sqrt{20 \cdot 80}} =$$
$$= \frac{+1200}{40 \cdot 40} = \frac{+1200}{1600} = +0{,}75$$

что мы имели уже и раньше.

Для вычисления коэффициента корреляции между цветом $A_3$ и формой $B_2$ мы должны будем подставить в эту формулу следующие числовые значения:

$$n = 100 \qquad p_x = 20$$
$$p = 2 \qquad p_y = 10$$

Тогда искомый коэффициент корреляции

$$r = \frac{100 \cdot 2 - 20 \cdot 10}{\sqrt{20(100-20)}\ \sqrt{10(100-10)}} = \frac{200-200}{\sqrt{20 \cdot 80}\ \sqrt{10 \cdot 90}} =$$
$$= \frac{0}{40 \cdot 30} = 0.$$

Заметим, что два выражения, стоящие в знаменателе этой формулы, т.-е.

$$\sqrt{p_x(n-p_x)} \quad \text{и} \quad \sqrt{p_y(n-p_y)}$$

по величине своей зависят только лишь от $p_x$ и $p_y$ (при постоянном $n$). Следовательно первое из них $\sqrt{p_x(n-p_x)}$ будет иметь постоянную величину для всех клеток одного и того же вертикального столбца (сумма частот которого равна данному итогу $p_x$), а второе: $\sqrt{p_y(n-p_y)}$ сохранит неизменную величину для всех клеток каждой горизонтальной строчки (с данным итогом $p_y$). Поэтому нам выгоднее раз навсегда для каждого $p_x$ вычислить соответствующую ему величину $\sqrt{p_x(n-p_x)}$ и точно также для каждого $p_y$ — величину $\sqrt{p_y(n-p_y)}$. Обозначив первую из них через $t_x$, а вторую через $t_y$, будем иметь:



$$t_x = \sqrt{p_x(n-p_x)} \qquad t_y = \sqrt{p_y(n-p_y)}$$

Тогда формула для вычисления коэффициента корреляции $r$ примет следующий окончательный вид:

$$r = \frac{np - p_x\ p_y}{t_x\ t_y}$$

Вычислим же для каждого итога $p_x$ соответствующую ему величину $t_x$, а для каждого итога $p_y$ — величину $t_y$:

для $p_x = 20$

$$t_x = \sqrt{20(100-20)} = \sqrt{20 \cdot 80} = \sqrt{1600} = 40$$

для $p_x = 20$

$$t_x = \sqrt{20(100-20)} = \sqrt{20 \cdot 80} = \sqrt{1600} = 40$$

для $p_x = 50$

$$t_x = \sqrt{50(100-50)} = \sqrt{50 \cdot 50} = \sqrt{2500} = 50$$

для $p_x = 10$

$$t_x = \sqrt{10(100-10)} = \sqrt{10 \cdot 90} = \sqrt{900} = 30$$

Точно также: для $p_y = 20$

$$t_y = \sqrt{20(100-20)} = \sqrt{20 \cdot 80} = \sqrt{1600} = 40$$

для $p_y = 10$

$$t_y = \sqrt{10(100-10)} = \sqrt{10 \cdot 90} = \sqrt{900} = 30$$

для $p_y = 50$

$$t_y = \sqrt{50(100-50)} = \sqrt{50 \cdot 50} = \sqrt{2500} = 50$$

для $p_y = 20$

$$t_y = \sqrt{20(100-20)} = \sqrt{20 \cdot 80} = \sqrt{1600} = 40$$

Эти величины $t_x$ и $t_y$ мы проставим для ясности около соответствующих итогов $p_x$ и $p_y$ в нашей первоначально заданной корреляционной решетке, после чего она примет следующий вид (см. табл. 95).

ТАБЛИЦА 95.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $p_y$ | $t_y$ |
|---|---|---|---|---|---|---|
| $B_1$ | 1 | 1 | 2 | 16 | 20 | 40 |
| $B_2$ | 7 | 0 | 2 | 1 | 10 | 30 |
| $B_3$ | 0 | 33 | 16 | 1 | 50 | 50 |
| $B_4$ | 2 | 16 | 0 | 2 | 20 | 40 |
| $p_x =$ | 10 | 50 | 20 | 20 | | $n = 100$ |
| $t_x =$ | 30 | 50 | 40 | 40 | | |

Если теперь нам потребовалось бы вычислить тот же коэффициент корреляции между цветом $A_4$ и формой $B_1$, то мы могли бы воспользоваться формулой

$$r = \frac{np - p_x\, p_y}{t_x\, t_y}$$

В нашем случае:

$$n = 100 \qquad p_x = 20 \qquad t_x = 40$$
$$p = 16 \qquad p_y = 20 \qquad t_y = 40$$

Следовательно:

$$r = \frac{100 \cdot 16 - 20 \cdot 20}{40 \cdot 40} = \frac{1600 - 400}{1600} = \frac{+1200}{1600} = +0{,}75$$

Для вычисления коэффициента корреляции между цветом $A_1$ и формой $B_3$ мы должны будем подставить в эту формулу:

$$n = 100 \qquad p_x = 10 \qquad t_x = 30$$
$$p = 0 \qquad p_y = 50 \qquad t_y = 50$$

Откуда:

$$r = \frac{100 \cdot 0 - 10 \cdot 50}{30 \cdot 50} = \frac{0 - 500}{1500} = \frac{-500}{1500} = -0{,}33$$

Вычислив этим путем все 16 коэффициентов корреляции, получим следующие их числовые значения, занесенные для ясности в соответствующие клетки нашей решетки (см. таблицу 96).



ТАБЛИЦА 96.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| $B_1$ | −0,08 | −0,45 | −0,13 | +0,75 |
| $B_2$ | +0,67 | −0,33 | 0 | −0,08 |
| $B_3$ | −0,33 | +0,32 | +0,30 | −0,45 |
| $B_4$ | 0 | +0,30 | −0,25 | −0,13 |

Теперь следовало бы оценить достоверность каждого коэффициента корреляции $r$, вычислив его среднюю ошибку $m_r$ по обычной формуле

$$m_r = \frac{1 - r^2}{\sqrt{n}}$$

и убедиться в том, что абсолютная величина самого коэффициента $r$ превышает его среднюю ошибку $m_r$ по крайней мере в 3 раза, т.-е. воспользоваться неравенством:

$$\frac{r}{m_r} > 3$$

Однако при постоянном числе измерений $n$ (у нас $n = 100$ для всех коэффициентов корреляции $r$) можно избежать этих добавочных вычислений, определив то минимальное значение $r$, которое при данном $n$ будет удовлетворять указанному выше условию достоверности.

Эта наименьшая величина достоверных коэффициентов корреляции (минимум $r$) может быть вычислена по формуле:

$$\min\, r = \frac{\sqrt{n + 36} - \sqrt{n}}{6}$$

В нашем примере $n = 100$, следовательно:

$$\min\, r = \frac{\sqrt{100 + 36} - \sqrt{100}}{6} = \frac{\sqrt{136} - \sqrt{100}}{6}$$

Так как $\sqrt{136} = 11{,}6619$, а $\sqrt{100} = 10$, то

$$\min\, r = \frac{11{,}6619 - 10}{6} = \frac{1{,}6619}{6} = 0{,}28$$

Итак, $\min\, r = 0{,}28$.

Этот результат показывает, что из всех 16 вычисленных нами коэффициентов корреляции $r$ только те будут превышать свою среднюю ошибку в 3 раза и более, которые по абсолютной своей величине (т.-е. не принимая во внимание знака $+$ или $-$) окажутся больше этого min. $r = 0{,}28$. Остальные же коэффициенты этому условию удовлетворять не будут и, как недостоверные, не должны быть принимаемы во внимание.

Нашему условию ($r > 0{,}28$) удовлетворяет здесь только 9 коэффициентов корреляции $r$ из общего их числа 16. Оставив их в решетке на соответствующих местах и выкинув 7 недостоверных коэффициентов (меньших 0,28), получим следующие окончательные результаты корреляционного анализа наших первоначальных данных (см. табл. 97).

ТАБЛИЦА 97.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| $B_1$ | | $-0{,}45$ | | $+0{,}75$ |
| $B_2$ | $+0{,}67$ | $-0{,}33$ | | |
| $B_3$ | $-0{,}33$ | $+0{,}32$ | $+0{,}30$ | $-0{,}45$ |
| $B_4$ | | $+0{,}30$ | | |

Эти результаты показывают, что прямая корреляция наблюдается здесь между:

цветом $A_4$ и формой $B_1$
цветом $A_1$ и формой $B_2$
цветом $A_2$ и формой $B_3$
цветом $A_3$ и формой $B_3$
цветом $A_2$ и формой $B_4$

а обратная зависимость констатирована между:

цветом $A_2$ и формой $B_1$
цветом $A_2$ и формой $B_2$
цветом $A_1$ и формой $B_3$
цветом $A_4$ и формой $B_3$

Связь же между остальными парами цветов и форм при данных условиях осталась невыясненной (не является достоверной).



Итак, при вычислении комплекта коэффициентов корреляции для выяснения связи между двумя рядами качественных признаков, необходимо выполнить следующие операции:

1. Составить корреляционную решетку и в соответствующие клетки ее занести частоты $p$.

2. Просуммировать эти частоты по всем вертикальным столбцам и по всем горизонтальным строчкам. Получим итоги: $p_x$ и $p_y$.

3. Для каждого итога $p_x$ вычислить величину $t_x$ по формуле:

$$t_x = \sqrt{p_x(n - p_x)},$$

а для каждого итога $p_y$ величину $t_y$ по формуле:

$$t_y = \sqrt{p_y(n - p_y)}$$

4. Для каждой пары признаков вычислить коэффициент корреляции $r$ по формуле:

$$r = \frac{np - p_x p_y}{t_x\, t_y}$$

5. Определить наименьшую величину достоверных коэффициентов корреляции $r$ по формуле:

$$\min\, r = \frac{\sqrt{n+36} - \sqrt{n}}{6}$$

6. Сделать соответствующие выводы на основании рассмотрения только тех коэффициентов корреляции $r$, которые по своей абсолютной величине окажутся больше этого минимального их значения.

В разобранном выше числовом примере оба качественных признака (цвет и форма) распадались на одинаковое число категорий (по 4 в каждом). Однако это не является обязательным, и мы можем представить себе случай, когда учитываются, например, 3 цвета и 5 форм. Порядок вычисления, очевидно, при этом останется прежним.

В некоторых случаях нас могут интересовать не все категории какого-либо признака, а лишь некоторые из них, например, только некоторые болезни, при наличии еще целого ряда других, для данного исследования не представляющих особого значения. Тогда все эти лишние для нас подразделения данного качественного признака следует объединить в одну общую рубрику, и непременно принять их в расчет при подсчете итогов и общего числа наблюдений $n$. Не выполнив этого условия, мы рискуем получить ошибочные результаты, т. к. в этом случае все коэффициенты корреляций будут указывать на наличие прямой или обратной зависимости признаков лишь внутри той искусственно выделенной группы объектов исследования, которая у нас осталась за вычетом некоторого числа непринятых во внимание объектов.

ТАБЛИЦА 98.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $p_y$ |
|---|---|---|---|---|---|
| $B_1$ | 1 | 1 | 2 | 16 | 20 |
| $B_2$ | 7 | 0 | 2 | 1 | 10 |
| $B_3$ | 0 | 33 | 16 | 1 | 50 |
| $B_4$ | 2 | 16 | 0 | 2 | 20 |
| $p_x$ | 10 | 50 | 20 | 20 | 100 |

Представим, теперь, случай, когда нас не интересует корреляция между отдельными категориями признаков (например, связь между определенным цветом $A_4$ и определенной же формой $B_1$), но зато требуется выяснить, существует ли вообще какая либо зависимость между двумя этими признаками в целом (например, между цветом и формой вообще).

Вычертив заново нашу первичную корреляционную решетку, просуммируем в ней все вертикальные столбцы и горизонтальные строчки и получим, как и раньше, итоги $p_x$ и $p_y$ (см. таблицу 98).

Здесь мы имеем фактическое распределение частот $p$ по 16 клеткам этой решетки, именно то их распределение, которое может быть как раз и вызвано наличием искомой корреляции между цветом и формой. Зададимся же для сравнения вопросом о том, каково было бы это распределение частот по клеткам, если бы корреляции между данными признаками не наблюдалось вовсе, т.-е. в том случае, когда все 100 наших объектов исследования рассеялись бы по этим



клеткам в чисто случайном порядке (вернее в полном беспорядке). Теория вероятностей указывает путь к разрешению этого вопроса. Оказывается, что для каждой клетки нашей решетки мы можем определить то количество объектов исследования ($p_0$), которое попало бы в нее при вполне случайном распределении вариант, пользуясь для этого следующей простой формулой:

$$p_0 = \frac{p_x p_y}{n}.$$

Так, например, в крайней слева клетке, при фактическом распределении объектов исследования по нашим категориям двух качественных признаков, оказался зарегистрированным лишь 1 такой объект. Это значит, что сочетание цвета $A_1$, с формой $B_1$ встретилось среди наших 100 случаев всего только один раз. Посмотрим, сколько же раз это сочетание ($A_1$ с $B_1$), должно было бы встретиться при вполне случайном распределении объектов исследования по этим категориям цвета и формы, т.-е. при полном отсутствии корреляции между данными признаками. Так как общее число объектов исследования цвета $A_1$, т.-е.

$$p_x = 10,$$

а общее число всех объектов, обладающих формой $B_1$, т.-е.

$$p_y = 20,$$

то при $n = 100$ искомое число случайных сочетаний $A_1$ с $B_1$ т.-е.

$$p_0 = \frac{p_x\, p_y}{n} = \frac{10\,.\,20}{100} = \frac{200}{100} = 2.$$

Итак, при случайном распределении (при отсутствии корреляции) в этой клетке было бы зарегистрировано 2 объекта исследования, а не 1, как у нас получилось.

Для соседней справа клетки (в которой фактически оказался также 1 объект исследования) $p_x = 50$, а $p_y = 20$. Следовательно

$$p_0 = \frac{p_x\, p_y}{n} = \frac{50\,.\,20}{100} = \frac{1000}{100} = 10.$$

Эти теоретически вычисленные значения $p_0$ мы назовем **теоретическими** частотами, в отличие от занесенных в нашу первичную решетку **эмпирических** частот $p$.

Вычислив для всех 16 клеток соответствующие им теоретические частоты $p_0$, получим следующее их распределение, отвечающее случаю полного отсутствия какой-либо корреляции между цветом и формой (см. табл. 99).

Нетрудно заметить, что эти теоретические частоты $p_0$ по всем вертикальным столбцам и горизонтальным строчкам дают в сумме те же самые итоги $p_x$ и $p_y$, которые ранее мы получили и для эмпирических частот $p$.

ТАБЛИЦА 99.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $p_y$ |
|---|---|---|---|---|---|
| $B_1$ | 2 | 10 | 4 | 4 | 20 |
| $B_2$ | 1 | 5 | 2 | 2 | 10 |
| $B_3$ | 5 | 25 | 10 | 10 | 50 |
| $B_4$ | 2 | 10 | 4 | 4 | 20 |
| $p_x$ | 10 | 50 | 20 | 20 | |

Имея в своем распоряжении эту картину теоретического распределения, мы можем теперь сравнить ее с нашим фактическим распределением частот, представленным на таблице 98. Если бы оба эти распределения оказались вполне совпадающими друг с другом, мы вправе были бы утверждать, что корреляция между формой и цветом не была обнаружена и в наших первичных данных. Наоборот, всякое различие в этих двух распределениях мы должны будем приписать уже существованию определенной связи между этими признаками и при том связи тем большей, чем заметнее окажутся различия между распределением эмпирических и теоретических частот $p$ и $p_0$ в каждой клетке.

Таким образом определение интересующей нас связи между цветом и формой свелось здесь к выяснению различий между $p$ и $p_0$ во всех 16 клетках этой таблицы. Вычислим же разности между $p$ и $p_0$ и результаты этих вычитаний запишем для ясности (с тем или иным знаком) в соответствующих клетках нашей корреляционной решетки (см.табл. 100).



ТАБЛИЦА 100.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| $B_1$ | $1-2=-1$ | $1-10=-9$ | $2-4=-2$ | $16-4=+12$ |
| $B_2$ | $7-1=+6$ | $0-5=-5$ | $2-2=0$ | $1-2=-1$ |
| $B_3$ | $0-5=-5$ | $33-25=+8$ | $16-10=+6$ | $1-10=-9$ |
| $B_4$ | $2-2=0$ | $16-10=+6$ | $0-4=-4$ | $2-4=-2$ |

Так как мы поставили себе целью выяснение вопроса о корреляции между цветом и формой вообще, и нас не интересует здесь зависимость между отдельными категориями этих признаков, то для суждения о большем или меньшем различии между эмпирическими и теоретическими частотами $p$ и $p_0$ мы должны будем учесть величину полученных здесь разностей $p-p_0$ для всех 16 клеток вместе взятых (суммарно, а не порознь). Но как это сделать? Просто сложив все 16 разностей, мы всегда получим в сумме 0, так как сумма положительных разностей (у нас $+38$) всегда окажется равной по своей абсолютной величине сумме отрицательных разностей (у нас $-38$). Для устранения этого затруднения можно прибегнуть к обычному приему возвышения всех отдельных разностей в квадрат *). Выполнив эту операцию со всеми 16 разностями, получим следующие результаты, занесенные для ясности в соответствующие клетки нашей решетки (см. табл. 101).

ТАБЛИЦА 101.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| $B_1$ | $(-1)^2=1$ | $(-9)^2=81$ | $(-2)^2=4$ | $(+12)^2=144$ |
| $B_2$ | $(+6)^2=36$ | $(-5)^2=25$ | $(0)^2=0$ | $(-1)^2=1$ |
| $B_3$ | $(-5)^2=25$ | $(+8)^2=64$ | $(+6)^2=36$ | $(-9)^2=81$ |
| $B_4$ | $(0)^2=0$ | $(+6)^2=36$ | $(-4)^2=16$ | $(-2)^2=4$ |

Здесь мы получили лишь абсолютную величину всех 16 разностей между $p$ и $p_0$ (вернее квадратов этих разностей) и не приняли еще во внимание неодинаковой величины самих

*) В конце вычисления это придется учесть и исправить результат путем извлечения квадратного корня.

теоретических частот $p_0$. В самом деле, различие, скажем, на 6 единиц (36) в той клетке, для которой $p_0 = 1$, несомненно, должно считаться очень значительным, в то время как то же различие при $p_0 = 10$ уже менее существенно. Чтобы уравнять и в этом отношении все наши разности (вернее их квадраты), разделим их на теоретические частоты $p_0$ (см. табл. 102).

Сложив теперь эти результаты, получим в сумме:

$$0{,}5 + 36 + 5 + 0 + 8{,}1 + 5 + 2{,}56 + 3{,}6 + 1 + 0 + 3{,}6 + 4 + 36 + 0{,}5 + 8{,}1 + 1 = 114{,}96.$$

ТАБЛИЦА 102.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| $B_1$ | $\frac{1}{2} = 0{,}5$ | $\frac{81}{10} = 8{,}1$ | $\frac{4}{4} = 1$ | $\frac{144}{4} = 36$ |
| $B_2$ | $\frac{36}{1} = 36$ | $\frac{25}{5} = 5$ | $\frac{0}{2} = 0$ | $\frac{1}{2} = 0{,}5$ |
| $B_3$ | $\frac{25}{5} = 5$ | $\frac{64}{25} = 2{,}56$ | $\frac{36}{10} = 3{,}6$ | $\frac{81}{10} = 8{,}1$ |
| $B_4$ | $\frac{0}{2} = 0$ | $\frac{36}{10} = 3{,}6$ | $\frac{16}{4} = 4$ | $\frac{4}{4} = 1$ |

Найденная величина представляет собой сумму квадратов разностей между эмпирическими и теоретическими частотами, отнесенных к величине соответствующих теоретических частот. Эта сумма квадратов накопилась у нас для всех 100 учтенных объектов исследования, стало быть, на долю каждого из них приходится по $\frac{1}{100}$ этой суммы, т.-е.

$$\frac{114{,}96}{100} = 1{,}1496.$$

Это число (1,1496) и дает представление о средней относительной величине квадрата разностей между $p$ и $p_0$. Так как нас интересует не средняя величина ее квадрата, а средняя величина самой разности, то из полученного выше числа (1,1496) следует еще извлечь квадратный корень

$$\sqrt{1{,}1496} = 1{,}07.$$



Итак, среднее относительное различие между эмпирическими и теоретическими частотами во всех 16 клетках нашей корреляционной решетки выражается величиной 1,07. Эта условная числовая характеристика различий между $p$ и $p_0$ носит название коэффициента „взаимной сопряженности“ (двух качественных признаков: у нас цвета и формы) и обозначается греческой буквой $\varphi$ (фи).

В данном случае следовательно:

$$\varphi = 1{,}07.$$

Если бы все эмпирические частоты $p$ в точности совпали с соответствующими им теоретическими частотами $p_0$ (при отсутствии корреляции), то каждая разность $p - p_0$, а, стало быть, и коэффициент $\varphi$ оказались бы $= 0$. Итак, при $\varphi = 0$ корреляции между данными признаками не существует. Чем больше $\varphi$, тем и теснее связь между двумя изучаемыми признаками.

Между коэффициентом взаимной сопряженности $\varphi$ и обычным коэффициентом корреляции $r$ существует приближенная зависимость, определяемая формулой:

$$r = \sqrt{\frac{\varphi^2}{\varphi^2 + 1}}.$$

В нашем примере $\varphi^2 = 1{,}1496$. Следовательно:

$$r = \sqrt{\frac{1{,}1496}{1{,}1496 + 1}} = \sqrt{\frac{1{,}1496}{2{,}1496}} = \sqrt{0{,}53} = 0{,}73.$$

Ошибка коэффициента корреляции $r$, вычисленного при посредстве коэффициента взаимной сопряженности $\varphi$, определяется по формуле:

$$m_r = \frac{4}{3}\left(\frac{1 - r^2}{\sqrt{n}}\right).$$

В нашем примере $r^2 = 0{,}53$ $n = 100$, следовательно:

$$m_r = \frac{4}{3}\left(\frac{1 - 0{,}53}{\sqrt{100}}\right) = \frac{4}{3}\left(\frac{0{,}47}{10}\right) = \pm 0{,}06.$$

Так как сам коэффициент корреляции $r = 0{,}73$ превышает свою среднюю ошибку $m_r = 0{,}06$ более, чем в три раза:

$$\frac{r}{m_r} = \frac{0{,}73}{0{,}06} = 12{,}17,$$

то наличие связи между цветом и формой в данном случае можно считать доказанным.

Заметим, кстати, что ошибку $m_r$ можно было бы вычислять и по обычной формуле:

$$m_r = \frac{1 - r^2}{\sqrt{n}},$$

т.-е. не умножая величины $\left(\frac{1-r^2}{\sqrt{n}}\right)$ на $\frac{4}{3}$, но тогда условием достоверности коэффициента $r$ служило бы неравенство:

$$\frac{r}{m_r} > 4,$$

показывающее, что сам коэффициент корреляции $r$ должен превышать свою ошибку $m_r$ по крайней мере в 4 (а не в 3) раза.

При вычислении $r$ с помощью $\varphi$ коэффициент корреляции $r$ пишется без знака ($+$ или $-$), так как здесь не может быть речи о прямой или обратной зависимости, и определяется лишь наличие или отсутствие зависимости вообще. В самом деле, одни пары цвета $A$ и формы $B$ могут быть связаны прямой корреляцией, другие обратной, но и в том и в другом случае существует все же какая то зависимость между цветом и формой, зависимость, количественную сторону которой суммарно и определяет этот коэффициент взаимной сопряженности $\varphi$, а, стало быть, и вычисленный с его помощью коэффициент корреляции $r$.

Указанный выше обычный способ вычисления $\varphi$ представляет многие технические неудобства и является слишком сложным.

В самом деле, для того чтобы произвести нужные вычисления, в каждой клетке корреляционной решетки приходится выполнить пять последовательных действий:



1) Умножить $p_x$ на $p_y$.
2) Разделить это произведение на $n$ (получим $p_0$).
3) Вычесть $p_0$ из $p$.
4) Возвести полученную разность в квадрат и
5) Разделить этот квадрат разности на $p_0$.

Кроме того, так как $p_0$ в общем случае всегда получается в виде целого числа с дробью (или просто в виде десятичной дроби), то все эти действия приходится производить в сущности с многозначными числами (не менее, чем с 2 цифрами после запятой) *).

ТАБЛИЦА 103.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $p_y$ |
|---|---|---|---|---|---|
| $B_1$ | 1 | 1 | 2 | 16 | 20 |
| $B_2$ | 7 | 0 | 2 | 1 | 10 |
| $B_3$ | 0 | 33 | 16 | 1 | 50 |
| $B_4$ | 2 | 16 | 0 | 2 | 20 |
| $p_x$ | 10 | 50 | 20 | 20 | 100 |

Для устранения этих неудобств можно воспользоваться другим приемом вычисления $\varphi$ (вернее $\varphi^2$), при котором в каждой клетке приходится производить всего лишь по 2 действия и притом исключительно с целыми числами.

Поясним этот способ на том же самом примере (см. таблицу 103).

ТАБЛИЦА 104.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $p_y$ |
|---|---|---|---|---|---|
| $B_1$ | $1^2 = 1$ | $1^2 = 1$ | $2^2 = 4$ | $16^2 = 256$ | 20 |
| $B_2$ | $7^2 = 49$ | $0^2 = 0$ | $2^2 = 4$ | $1^2 = 1$ | 10 |
| $B_3$ | $0^2 = 0$ | $33^2 = 1089$ | $16^2 = 256$ | $1^2 = 1$ | 50 |
| $B_4$ | $2^2 = 4$ | $16^2 = 256$ | $0^2 = 0$ | $2^2 = 4$ | 20 |
| $p_x$ | 10 | 50 | 20 | 20 | |

Здесь мы имеем частоты $p$ и итоги $p_x$ и $p_y$, представляющие собой, очевидно, всегда целые числа.

*) У нас все $p_0$ оказались целыми числами лишь благодаря искусственному подбору данных при составлении этого примера.

Первое из двух действий, которое нам нужно выполнить в каждой из 16 клеток этой решетки, заключается в возвышении всех этих частот $p$ в квадрат (см. табл. 104).

ТАБЛИЦА 105.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $p_y$ |
|---|---|---|---|---|---|
| $B_1$ | $\frac{1}{10}=0{,}10$ | $\frac{1}{50}=0{,}02$ | $\frac{4}{20}=0{,}20$ | $\frac{256}{20}=12{,}80$ | 20 |
| $B_2$ | $\frac{49}{10}=4{,}90$ | $\frac{0}{50}=0$ | $\frac{4}{20}=0{,}20$ | $\frac{1}{20}=0{,}05$ | 10 |
| $B_3$ | $\frac{0}{10}=0$ | $\frac{1089}{50}=21{,}78$ | $\frac{256}{20}=12{,}80$ | $\frac{1}{20}=0{,}05$ | 50 |
| $B_4$ | $\frac{4}{10}=0{,}40$ | $\frac{256}{50}=5{,}12$ | $\frac{0}{20}=0$ | $\frac{4}{20}=0{,}20$ | 20 |
| $p_x$ | 10 | 50 | 20 | 20 | |

Второе действие, выполняемое тоже в каждой из 16 клеток этой решетки, заключается в делении полученных чисел ($p^2$) на соответствующие нижние итоги $p_x$ (для всех клеток каждого столбца этот итог $p_x$ будет, очевидно, один и тот же). Результат этого деления представлен на таблице 105.

На этом и кончаются операции, выполняемые во всех 16 клетках корреляционной решетки. Далее следует сложить полученные результаты в каждой строчке отдельно и полученные суммы разделить на соответствующие итоги $p_y$ (см. табл. 106).

ТАБЛИЦА 106.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | |
|---|---|---|---|---|---|
| $B_1$ | 0,10 | 0,02 | 0,20 | 12,80 | $\frac{13{,}12}{20}=0{,}656$ |
| $B_2$ | 4,90 | 0 | 0,20 | 0,05 | $\frac{5{,}15}{10}=0{,}515$ |
| $B_3$ | 0 | 21,78 | 12,80 | 0,05 | $\frac{34{,}63}{50}=0{,}6926$ |
| $B_4$ | 0,40 | 5,12 | 0 | 0,20 | $\frac{5{,}72}{20}=0{,}286$ |



Складывая, например, все числа в верхней горизонтальной строчке, получим в сумме:

$$13{,}12.$$

Эту сумму нам нужно теперь разделить на соответствующий итог данной строчки, т.-е. на $p_y=20$

$$\frac{13{,}12}{20}=0{,}656.$$

Аналогичные действия выполним и для всех остальных строчек нашей решетки. Результаты этих действий выписаны на полях таблицы 106. Теперь надо сложить все эти 4 числа:

$$0{,}656+0{,}515+0{,}6926+0{,}286=2{,}1496.$$

Полученную этим путем сумму (2,1496) мы обозначим буквой $S$. Итак у нас:

$$S=2{,}1496.$$

Между суммой $S$ и квадратом показателя взаимной сопряженности $\varphi^2$ существует вполне точная зависимость определяемая формулой:

$$\varphi^2=S-1.$$

Так как у нас $S=2{,}1496$, то искомая величина

$$\varphi^2=2{,}1496-1=1{,}1496,$$

что мы имели и раньше.

При посредстве суммы $S$ искомый коэффициент корреляции $r$ может быть вычислен по формуле:

$$r=\sqrt{\frac{S-1}{S}}.$$

В нашем примере, при $S=2{,}1496$, будем иметь, как и раньше:

$$r=\sqrt{\frac{2{,}1496-1}{2{,}1496}}=\sqrt{\frac{1{,}1496}{2{,}1496}}=\sqrt{0{,}53}=0{,}73.$$

Далее следовало бы вычислить среднюю ошибку $m_r$ этого коэффициента корреляции $r$ по формуле:

$$m_r=\frac{4}{3}\left(\frac{1-r^2}{\sqrt{n}}\right)$$

и оценить достоверность вывода, убедившись в том, что

$$\frac{r}{m_r} > 3.$$

Однако это можно сделать и проще, вовсе не прибегая к вычислению $r$ и $m_r$.

Так как коэффициент $r$ зависит от суммы $S$, а ошибка $m_r$ от коэффициента $r$ и стало быть также от $S$, то мы можем заранее вычислить то наименьшее значение $S$ (min $S$), при котором $r$ будет по крайней мере в 3 раза больше своей ошибки $m_r$. Вычисление это производится по формуле:

$$\min S = 0{,}5 + \sqrt{0{,}25 + \frac{16}{n}},$$

где $n$, как всегда, обозначает общее число измерений (наблюдений). В нашем примере $n = 100$, следовательно:

$$\min S = 0{,}5 + \sqrt{0{,}25 + \frac{16}{100}} = 0{,}5 + \sqrt{0{,}25 + 0{,}16} = $$
$$= 0{,}5 + \sqrt{0{,}41}.$$

Так как

$\sqrt{0{,}41} = 0{,}64$, то окончательно будем иметь:

$$\min S = 0{,}5 + 0{,}64 = 1{,}14.$$

Итак, если сумма $S$ оказалась бы меньше, чем 1,14, то коэффициент корреляции $r$ не удовлетворял бы условию тройного превышения своей средней ошибки $m_r$. Так как в данном случае $S = 2{,}1496$, т.-е. оказалась больше чем min $S$, то условие это будет выполнено, и наш коэффициент корреляции $r$ (фактически не вычисленный) будет отвечать обычному критерию достоверности.

При определении коэффициента $r$ с помощью суммы $S$ (или только $S$ и min $S$) мы можем воспользоваться также и другим приемом вычисления $S$, при помощи которого удобно проверять выкладки по их конечным результатам. Этот видоизмененный прием вычисления $S$ заключается в делении квадратов частот ($p^2$) не на итоги $p_x$, а на итоги $p_y$, с последующим сложением этих результатов уже



по вертикальным столбцам и делением полученных сумм на соответствующие итоги $p_x$.

Поясним этот прием на числовом примере, разобранном выше. На таблице 107 квадраты частот ($p^2$) в каждой клетке разделены на итоги $p_y$, т.-е., например, в первой строчке на 20, во второй на 10 и т. д.

ТАБЛИЦА 107.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $p_y$ |
|---|---|---|---|---|---|
| $B_1$ | $\frac{1}{20} = 0{,}05$ | $\frac{1}{20} = 0{,}05$ | $\frac{4}{20} = 0{,}20$ | $\frac{256}{20} = 12{,}80$ | 20 |
| $B_2$ | $\frac{49}{10} = 4{,}90$ | $\frac{0}{10} = 0$ | $\frac{4}{10} = 0{,}40$ | $\frac{1}{10} = 0{,}10$ | 10 |
| $B_3$ | $\frac{0}{50} = 0$ | $\frac{1089}{50} = 21{,}78$ | $\frac{256}{50} = 5{,}12$ | $\frac{1}{50} = 0{,}02$ | 50 |
| $B_4$ | $\frac{4}{20} = 0{,}20$ | $\frac{256}{20} = 12{,}8$ | $\frac{0}{20} = 0$ | $\frac{4}{20} = 0{,}20$ | 20 |
| $p_x$ | 10 | 50 | 20 | 20 | |

Далее все эти результаты следует сложить по столбцам и полученные суммы разделить на итоги этих столбцов $p_x$ (см. табл. 108).

ТАБЛИЦА 108.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| $B_1$ | 0,05 | 0,05 | 0,20 | 12,80 |
| $B_2$ | 4,90 | 0 | 0,40 | 0,10 |
| $B_3$ | 0 | 21,78 | 5,12 | 0.02 |
| $B_4$ | 0,20 | 12,80 | 0 | 0.20 |
| | $\frac{5{,}15}{10} = 0{,}515$ | $\frac{34{,}63}{50} = 0{,}6926$ | $\frac{5{,}72}{20} = 0{,}286$ | $\frac{13{,}12}{20} = 0{,}656$ |

Сложив найденные этим путем результаты, мы и получим искомую сумму $S$:

$$0{,}515 + 0{,}6926 + 0{,}286 + 0{,}656 = 2{,}1496.$$

Здесь, как и раньше, $S=2{,}1496$.

Итак, для вычисления коэффициента корреляции $r$, определяющего суммарную зависимость между двумя качественными признаками, распадающимися на несколько категорий, следует:

1) Записать частоты $p$ в соответствующие клетки корреляционной решетки.

2) Просуммировать эти частоты по всем вертикальным столбцам и по всем горизонтальным строчкам. Получим итоги $p_x$ и $p_y$.

3) Во всех клетках решетки возвысить частоты $p$ в квадрат. Получим $p^2$.

4) Квадрат частоты $p^2$ в каждой клетке разделить на итог всего данного столбца $p_x$. Получим величины вида $\frac{p^2}{p_x}$.

5) Все эти величины сложить по строчкам.

6) Каждую такую сумму разделить на итог данной строчки $p_y$.

7) Все эти результаты сложить. Получим сумму $S$.

8) Для проверки вычислений проделать их еще раз в другом порядке, т. е. в каждой клетке разделить $p^2$ на $p_y$, результаты просуммировать по столбцам и разделить их на $p_x$. При сложении этих чисел должна получиться та же сумма $S$.

9) Вычислить коэффициент корреляции $r$ по формуле:

$$r=\sqrt{\frac{S-1}{S}}.$$

10) Вычислить среднюю ошибку $m_r$ этого коэффициента по формуле:

$$m_r=\frac{4}{3}\left(\frac{1-r^2}{\sqrt{n}}\right).$$

11) Оценить достоверность коэффициента корреляции $r$ по критерию:

$$\frac{r}{m_r}>3$$



или, если ошибка $m_r$ была вычислена по обычной формуле:

$$m_r=\frac{1-r^2}{\sqrt{n}},$$

то выполнить эту оценку по измененному критерию:

$$\frac{r}{m_r}>4.$$

12) Взамен вычисления $r$ и $m_r$ с последующим применением указанного выше критерия достоверности можно вычислить наименьшее значение суммы $S$ по формуле:

$$\min S=0{,}5+\sqrt{0{,}25+\frac{16}{n}}$$

и, сравнив полученное ранее $S$ с этим значением $\min S$ прямо судить о достоверности коэффициента корреляции $r$ на основании условия

$$S>\min S.$$

Все записи при выполнении этих вычислений удобно производить в расчетной решетке следующей формы: (см. таблицу 109).

Здесь первое из чисел, записанных в каждой клетке, представляет собой частоту $p$, второе — $p^2$, третье — $\frac{p^2}{p_x}$, четвертое — $\frac{p^2}{p_y}$. Дроби, проставленные справа у каждой горизонтальной строчки, в числителе имеют сумму выражений вида $\frac{p^2}{p_x}$, полученную для данной строчки, а в знаменателе — соответствующий итог этой строчки $p_y$. Дроби, проставленные снизу под каждым вертикальным столбцом, имеют в числителе сумму выражений вида $\frac{p^2}{p_y}$, полученную для данного столбца, а в знаменателе — соответствующий итог этого столбца $p_x$.

Полезно запомнить, что при вычислении $S$ удобно итти путем выполнения однотипных операций. Так, при заполнении всех клеток расчетной решетки нецелесообразно производить сперва все действия в одной клетке, переходя затем к следующим. Практика показала, что гораздо скорее и легче выполнить однородные действия во всех клетках, затем следующую операцию опять во всех клетках решетки и т. д.

Выше мы рассматривали случаи корреляции между такими признаками, которые совершенно не подаются никакому количественному измерению. Однако, иногда представляется

ТАБЛИЦА 109.

| | $A_1$ | $A_2$ | $A_3$ | $A_4$ | |
|---|---|---|---|---|---|
| $B_1$ | 1<br>1<br>0,10<br>0,05 | 1<br>1<br>0,02<br>0,05 | 2<br>4<br>0,20<br>0,20 | 16<br>256<br>12,80<br>12,80 | $\frac{13,12}{20} = 0,656$ |
| $B_2$ | 7<br>49<br>4,90<br>4,90 | 0<br>0<br>0<br>0 | 2<br>4<br>0,20<br>0,40 | 1<br>1<br>0,05<br>0,10 | $\frac{5,15}{10} = 0,515$ |
| $B_3$ | 0<br>0<br>0<br>0 | 33<br>1089<br>21,78<br>21,78 | 16<br>256<br>12,80<br>5,12 | 1<br>1<br>0,05<br>0,02 | $\frac{34,63}{50} = 0,6926$ |
| $B_4$ | 2<br>4<br>0,40<br>0,20 | 16<br>256<br>5,12<br>12,80 | 0<br>0<br>0<br>0 | 2<br>4<br>0,20<br>0,20 | $\frac{5,72}{20} = 0,286$ |
| | $\frac{5,15}{10} =$<br>$= 0,515$ | $\frac{34,63}{50} =$<br>$= 0,6926$ | $\frac{5,72}{20} =$<br>$= 0,286$ | $\frac{13,12}{20} =$<br>$= 0,656$ | $S = 2,1496$ |

возможность, если и не измерить величину признака в строгом смысле слова, то хотя бы приближенно охарактеризовать степень его интенсивности по сравнению с какой-либо скалой условных градаций этой интенсивности. Примером такой приближенной количественной оценки интенсивности качественных признаков могут служить различные оттенки яркости одного какого-либо цвета, или произвольно взятые ступени постепенного перехода одного цвета в другой, расположение



учащихся в ряд по возрастанию (или убыванию) их успешности в занятиях и т. д. Во всех этих случаях мы устанавливаем некоторую последовательность нарастания данного признака (цвета, успешности и т. д.) и каждого отдельного индивидуума характеризуем соответствующим № (или рангом) его в этом ряду.

Пусть, например, имеем группу объектов исследования

$$A, B, C, D, E, F, G, H \text{ и } K,$$

расположенных в этом ряду в порядке постепенного нарастания какого-либо признака. Это значит, что объект исследования $A$ имеет самую малую степень данного признака, объект $B$ — уже несколько большую, следующий объект $C$ — еще большую и т. д. Таким образом, самая большая интенсивность данного признака будет наблюдаться у крайнего справа объекта исследования $K$.

Пронумеруем этих объектов слева направо последовательными числами от № 1 до № 9 включительно:

| №№ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ | $G$ | $H$ | $K$ |

При невозможности точно измерить величину интересующего нас признака у каждого из 9 индивидуумов мы вынуждены условно принять данные №№ за количественную оценку этого признака и считать, что величина (интенсивность) признака у объекта исследования $A$ равна 1 у $B = 2$ и т. д.

При этом необходимо помнить, что такой способ количественной оценки признака в сущности очень неточен. В самом деле разница в интенсивности признака у $A$ и у $B$ может и не равняться в действительности разнице между $B$ и $C$. Наши же нумера в обоих случаях дают одинаковое увеличение на 1 (от 1 к 2 и от 2 к 3). Далее, объект исследования $B$ отмечен у нас № 2, а объект исследования $D$ — № 4. Можно ли отсюда заключить, что данный признак у $D$ вдвое интенсивнее, чем у $B$? Очевидно, такое заключение было бы опрометчивым. Наша нумерация имеет здесь чисто условный характер и применена лишь за невозможностью произвести точные измерения.

Пусть, на ряду с первым признаком ($x$) те же 9 объектов исследования были расположены по рангам какого-либо другого признака ($y$), причем порядок их нумеров в данном случае определился так:

| №№ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| | *D* | *A* | *C* | *G* | *B* | *F* | *K* | *H* | *E* |

Если бы между признаками $x$ и $y$ существовала полная прямая зависимость (корреляция), то, как в первом, так и во втором случае все 9 объектов исследования расположились бы в совершенно одинаковые ряды, сохранив те же самые №№ и во втором случае. Однако мы этого здесь не наблюдаем. Измененный порядок их расположения во втором ряду (по признаку $y$) свидетельствует о нарушении прямой корреляции, которая, таким образом, здесь уже не может считаться полной.

Зададимся целью вычислить соответствующий коэффициент корреляции $r$ по имеющимся у нас данным.

Приняв №№ первого ряда за варианты ряда $x$ (т.-е. за $V_x$), а №№ этих же 9 объектов исследования по второму признаку за варианты ряда $y$ (т.-е. за $V_y$) и применив способ непосредственного вычисления, получим следующий результат (см. таблицу 110).

ТАБЛИЦА 110.

| Объекты исследов. | $V_x$ | $V_y$ | $x$ | $x^2$ | $y$ | $y^2$ | $xy$ |
|---|---|---|---|---|---|---|---|
| *A* | 1 | 2 | — 4 | 16 | — 3 | 9 | +12 |
| *B* | 2 | 5 | — 3 | 9 | 0 | 0 | 0 |
| *C* | 3 | 3 | — 2 | 4 | — 2 | 4 | + 4 |
| *D* | 4 | 1 | — 1 | 1 | — 4 | 16 | + 4 |
| *E* | 5 | 9 | 0 | 0 | + 4 | 16 | 0 |
| *F* | 6 | 6 | + 1 | 1 | + 1 | 1 | + 1 |
| *G* | 7 | 4 | + 2 | 4 | — 1 | 1 | — 2 |
| *H* | 8 | 8 | + 3 | 9 | + 3 | 9 | + 9 |
| *K* | 9 | 7 | + 4 | 16 | + 2 | 4 | + 8 |
| Σ | 45 | 45 | 0 | 60 | 0 | 60 | +36 |



Здесь среднее арифметическое ряда $x$, т.-е.

$$M_x = \frac{\Sigma V_x}{n} = \frac{45}{9} = 5$$

и среднее арифметическое ряда $y$, т.-е.

$$M_y = \frac{\Sigma V_y}{n} = \frac{45}{9} = 5$$

Вычислим все отклонения $x$ и $y$, перемножим их попарно, затем возвысим каждый $x$ и каждый $y$ в квадрат и просуммируем соответствующие столбцы в расчетной решетке. В результате получим:

$$\Sigma x^2 = 60 \qquad \Sigma y^2 = 60 \qquad \Sigma xy = +36$$

Откуда искомый коэффициент корреляции:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2\, \Sigma y^2}} = \frac{+36}{\sqrt{60\,.\,60}} = \frac{+36}{60} = +0{,}6$$

Ошибка этого коэффициента:

$$m_r = \frac{1 - r^2}{\sqrt{n}} = \frac{1 - 0{,}6^2}{\sqrt{9}} = \frac{1 - 0{,}36}{3} = \frac{0{,}64}{3} = \pm 0{,}21$$

Отношение:

$$\frac{r}{m_r} = \frac{0{,}60}{0{,}21} = 2{,}86,$$

откуда заключаем, что условие достоверности $r$ в данном случае не выполнено.

Однако этот общий способ вычисления $r$ в данном случае может быть в значительной степени упрощен с помощью применения особого приема.

Выпишем в первом столбце расчетной решетки (см. таблицу 111) порядковые №№ всех 9-ти наших объектов исследования по первому признаку ($N_x$), а в соседнем столбце их порядковые №№ по второму признаку ($N_y$). Далее из каждого числа столбца $N_x$ вычтем соответствующее ему число столбца $N_y$ и полученные разности $N_x - N_y = \delta$ (дельта) с соответствующими знаками (+ или —) занесем в клетки третьего столбца, озаглавленного буквою $\delta$.

Сумма этих разностей (с учетом знаков) должна равняться 0, что мы и имеем здесь в действительности.

Далее, каждую разность $\delta$ следует возвысить в квадрат и записать в крайнем справа столбце, озаглавленном символом $\delta^2$.

ТАБЛИЦА 111.

| $N_x$ | $N_y$ | $\delta$ | $\delta^2$ |
|---|---|---|---|
| 1 | 2 | −1 | 1 |
| 2 | 5 | −3 | 9 |
| 3 | 3 | 0 | 0 |
| 4 | 1 | +3 | 9 |
| 5 | 9 | −4 | 16 |
| 6 | 6 | 0 | 0 |
| 7 | 4 | +3 | 9 |
| 8 | 8 | 0 | 0 |
| 9 | 7 | +2 | 4 |
| — | — | 0 | 48 |

Сумма чисел этого столбца, т. е. сумма квадратов всех 9-ти разностей между нумерами $N_x$ и $N_y$ в данном случае равна 48. Обозначив ее через $\Sigma\delta^2$, получим:

$$\Sigma\delta^2 = 48.$$

Искомый коэффициент корреляции $r$ связан с этой величиной ($\Sigma\delta^2$) следующей формулой:

$$r = 1 - \frac{6\,\Sigma\delta^2}{n\,(n^2 - 1)}$$

В нашем примере $\Sigma\delta^2 = 48$, а $n = 9$. Следовательно:

$$r = 1 - \frac{6\,.\,48}{9\,.\,(9^2 - 1)} = 1 - \frac{2\,.\,48}{3\,(81 - 1)} = 1 - \frac{96}{3\,.\,80} = 1 - \frac{2}{5} = \frac{3}{5} = +0{,}6$$

Итак, при этом способе вычислений, так же как и раньше, коэффициент корреляции

$$r = +0{,}6$$

В тех случаях, когда несколько объектов исследования оказываются по интенсивности какого-либо признака совершенно одинаковыми, и у нас нет никаких оснований снабдить их различными №№, всем им дается один и тот же нумер, соответствующий среднему арифметическому этих №№. Так, например, если 4 совершенно равноценных объекта исследования занимают в ряду отдельные места, соответствующие №№ 5, 6, 7 и 8, то всем им следует присвоить один и тот же №, вычисленный по формуле:

$$\text{№} = \frac{5 + 6 + 7 + 8}{4} = 6{,}5$$



или, проще, как полусумму крайних нумеров 5 и 8

$$\text{№} = \frac{5 + 8}{2} = \frac{13}{2} = 6{,}5$$

При этих условиях, следовательно, некоторые из ранговых №№ могут оказаться и дробными числами.

## XIX. Графический способ оценки достоверности окончательных выводов и приближенного вычисления средних ошибок.

В главе V был указан и разъяснен на нескольких числовых примерах общий способ оценки достоверности тех или иных заключений и выводов, получаемых при сопоставлении друг с другом различных числовых характеристик варьирующих признаков. Способ этот заключается в применении особого количественного критерия, на основании которого подлежащая оценке величина должна превышать свою среднюю ошибку по крайней мере в 3 раза.

Так, для суждения о достоверности наличия обнаруженной связи между двумя признаками, мы должны убедиться в том, что абсолютная величина коэффициента корреляции $r$ превышает абсолютную величину его средней ошибки $m_r$ не менее, чем в 3 раза. Это условие можно выразить в форме следующего неравенства:

$$\frac{r}{m_r} > 3.$$

Знак неравенства $>$, поставленный острием вправо, указывает здесь на то, что отношение $\frac{r}{m}$ должно оказаться больше 3. Только при соблюдении этого требования мы можем считать данную зависимость доказанной; в противном случае, как бы ни был сам по себе велик коэффициент корреляции $r$, мы все же должны сомневаться в его достоверности.

При оценке достоверности какой-либо разницы между двумя величинами (напр., между двумя средними арифметическими $M_1$ и $M_2$, двумя средними квадратическими отклонениями $\sigma_1$ и $\sigma_2$ или двумя коэффициентами корреляции $r_1$ и $r_2$) следует также абсолютную величину этой разницы разделить на абсолютную величину ее средней ошибки, которая в данном случае будет равна квадратному корню из суммы квадратов двух ошибок этих величин (см. главу V). Если сравниваемые величины обозначить через $T_1$ и $T_2$, а их средние ошибки — через $m_1$ и $m_2$, то критерий достоверности данной разницы выразится следующим неравенством:

$$\frac{T_1 - T_2}{\sqrt{m_1^2 + m_2^2}} > 3.$$

Чтобы избежать извлечения квадратного корня в знаменателе, можно обе части этого неравенства возвысить в квадрат и пользоваться им уже в таком виде:

$$\frac{(T_1 - T_2)^2}{m_1^2 + m_2^2} > 9.$$

Некоторые частные случаи применения этих формул были указаны в числовых примерах главы V (полезно просмотреть эту главу еще раз).

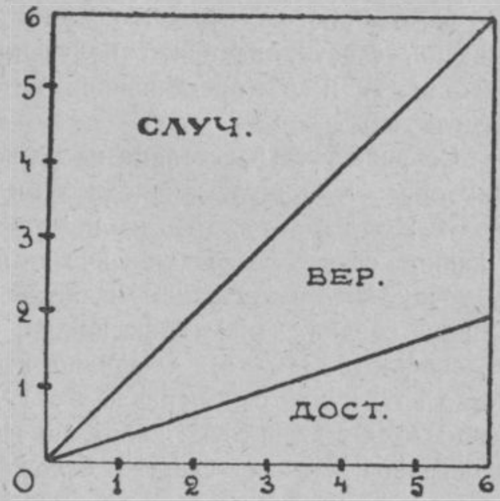
Покажем теперь более легкий графический способ оценки достоверности окончательных выводов, основанный на применении тех же формул, но уже без всяких вычислений, с помощью лишь простого отмера циркулем расстояния на особом „графике".

График этот представляет собой квадрат произвольного размера, вычерченный на листе миллиметровой бумаги или, в крайнем случае, на простой писчей бумаге, разграфленной в клетку (см. черт. 1). От нижнего левого угла этого квадрата вверх и вправо по его сторонам следует наметить деления какого-либо произвольного масштаба и против этих делений проставить соответствующие цифры (на нашем чертеже левая и нижняя стороны квадрата для простоты разделены всего лишь на 6 частей; в действительности же удобнее делить эти стороны на значительно большее число мелких частей). Далее, на правой стороне квадрата, на расстоянии одной трети полной его высоты, наметим точку и соединим ее прямой линией с левым нижним углом квадрата. Кроме этой линии проведем еще диагональ, соединяющую левый нижний угол квадрата с его правым верхним углом. Этими



двумя линиями вся площадь квадрата разделится на 3 части, обозначенные на нашем чертеже следующими сокращенными надписями: „случ." (случайная разница), „вер". (вероятная разница) и „дост." (достоверная разница). В этом окончательном виде наш график готов к употреблению.

ЧЕРТ. 1.

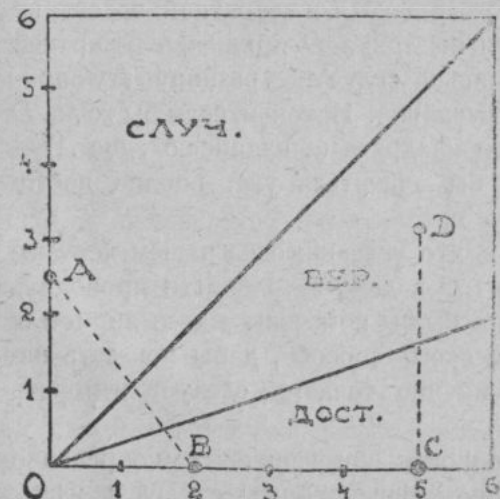



Покажем способ пользования этим графиком на числовом примере. Пусть требуется сравнить следующие два средних арифметических $M_1$ и $M_2$ с их средними ошибками $m_1$ и $m_2$.

$$M_1 = 15 \qquad m_1 = \pm 2{,}5$$
$$M_2 = 10 \qquad m_2 = \pm 2.$$

Прежде всего на левой и на правой стороне квадрата (см. черт. 2) наметим две точки $A$ и $B$, соответствующие в данном масштабе двум нашим ошибкам $m_1$ и $m_2$. Так как $m_1 = 2{,}5$, то первая точка $A$ ляжет посредине между 2-м и 3-м делением левой стороны квадрата. Вторая точка $B$, соответствующая второй ошибке $m_2=2$, придется, очевидно, как раз на 2-м де-

ЧЕРТ. 2.



лении нижней стороны квадрата (заметим кстати, что можно было бы поступить и наоборот, т.-е. на левой стороне квадрата наметить точку, соответствующую ошибке $m_2$, а на нижней — $m_1$). Далее, с помощью циркуля с двумя остриями или даже просто с помощью бумажки, сложенной в виде линейки, следует измерить полученное расстояние между точками $A$ и $B$ (обозначенное на чертеже пунктирной линией $AB$).

Сохранив это расстояние на разведенном соответствующим образом циркуле или пометками на бумажной линейке, следует, далее, отметить на нижней стороне квадрата (и в данном случае уже непременно на нижней, а не на левой) ту точку, котор я в нашем масштабе соответствует данной разнице средних арифметических $M_1 — M_2$ (в этом примере $M_1 — M_2 = 15 — 10 = 5$). Очевидно, эта точка (обозначим ее буквою $C$) придется как раз на 5-м делении нижней стороны квадрата. От этой точки вверх, по направлению линий миллиметровой бумаги, следует отложить полученный ранее отрезок $AB$ (зафиксированный циркулем или на бумажке). На верхнем конце этого вертикального отрезка мы получим точку $D$, по положению которой и можно судить о достоверности оцениваемой разницы между двумя средними арифметическими $M_1$ и $M_2$.

Если, как в данном случае, точка эта окажется в средней части квадрата (с надписью: „вер.“), данная разница средних должна быть признана хотя и вероятной, но еще не вполне доказанной. Если бы точка $D$ пришлась в верхней части квадрата (с надписью „случ)“ — разницу эту пришлось бы считать лишь случайной. Наконец, если бы точка $D$ оказалась в нижней части квадрата (с надписью „дост.“) — разница средних могла бы считаться уже вполне достоверной.

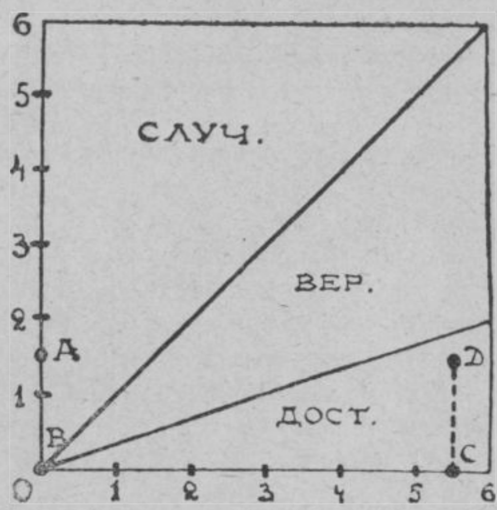
Необходимо заметить, что указанные на нашем чертеже пунктирные линии $AB$ и $CD$ в действительности проводить на графике не надо. Они были помечены здесь лишь для ясности описания графического способа, дабы показать наглядно, между какими именно точками следует измерять расстояние циркулем.

При оценке достоверности коэффициента корреляции $r$ операции с этим графиком еще упрощаются. Пусть, напр.,



требуется выяснить достоверность наличия обратной связи двух признаков, характеризующейся коэффициентом корреляции $r = -0,55$ со средней ошибкой $m_r = \pm 0,15$. Чтобы подогнать эти данные к масштабу нашего графика, помножим абсолютную величину коэффициента корреляции $r$ (0,55) и его средней ошибки $m_r$ (0,15) на одно и то же число, в нашем случае на 10 (такое умножение или деление мы всегда имеем право сделать, так как нас интересуют не самые числа 0,55 и 0,15, а их отношение 0,55 : 0,15, которое, очевидно, при таком умножении или делении не изменится). В результате взамен $r$ и $m_r$ получим два новых числа: 5,5 и 1,5. Эти числа уже подходят к принятому на графике масштабу, и потому их можно откладывать на сторонах заготовленного нами квадрата (см. черт. 3).

ЧЕРТ. 3.



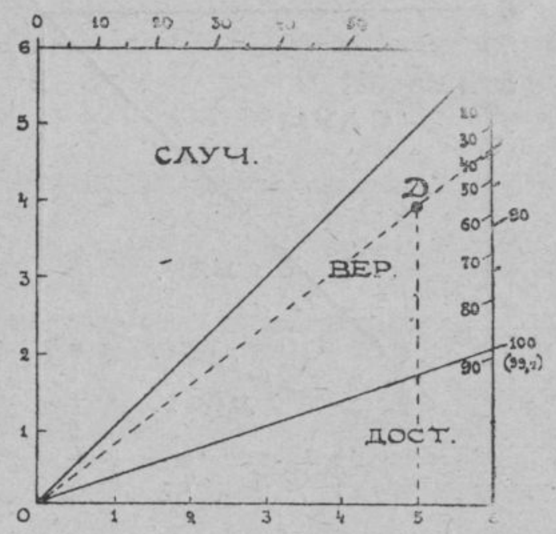

Наметив на левой стороне квадрата точку $A$, соответствующую в данном масштабе средней ошибке коэффициента корреляции $m_r$ (т.-е. теперь уже числу 1,5), измерим циркулем расстояние от точки $B$ до точки $A$. Далее, на нижней стороне квадрата наметим точку $C$, соответствующую в том же масштабе самому коэффициенту корреляции $r$ (т.-е. числу 5,5), и отложим от этой точки $C$ вверх по направлению линий миллиметровой бумаги только-что измеренный отрезок $AB = CD$. На верхнем конце этого отрезка получим точку $D$, по положению которой и можно судить о достоверности заданного коэффициента корреляции $r$.

В нашем примере точка $D$ оказалась в той части квадрата, которая озаглавлена сокращенным названием „дост.“ Следовательно наличие обратной корреляции при этих условиях можно считать вполне достоверным.

При большом количестве таких оценок применение описанного графического способа может дать существенную экономию времени, так как вся эта операция при известном навыке занимает здесь всего лишь несколько секунд, в то время, как соответствующие вычисления по формуле требуют значительно большей затраты времени и внимания.

Описанная здесь упрощенная форма этого графика позволяет делать оценку достоверности того или иного вывода лишь в самой грубой форме, пользуясь понятиями „случайных“, „вероятных“ и „достоверных“ различий. Однако в некоторых случаях может представить интерес и более детальная количественная формулировка такой оценки, выраженная хотя бы в числе шансов (из 100) за то, что данная обнаруженная частным опытом разница существует в действительности, а не является простым результатом случайного подбора объектов исследования в сравниваемых группах.

ЧЕРТ. 4.




В целях установления таких промежуточных градаций в оценке достоверности отдельных выводов и заключений можно на правой и верхней стороне квадрата нанести особые (не равные друг другу) деления, соответствующие значениям так называемого „нормального интеграла вероятностей“ (подробности об этих значениях будут изложены во 2-ой части данного руководства). В этом несколько усложненном виде наш график представлен на черт. 4.

Ранее при пользовании этим графиком мы получали где либо в его плоскости конечную точку *D* (см. описание чертежа 2-го) и по положению ее в одной из трех зон



(„случ.“, „вер.“ или „дост.“) делали окончательное заключение о достоверности оцениваемого вывода. При наличии нанесенных на правую и верхнюю сторону квадрата новых делений мы имеем возможность уточнить это суждение о той или иной степени достоверности вывода, соединив точку *D* с угловой точкой графика *O* и заметив на продолжении этой линии (она обозначена на чертеже 4 пунктиром) то деление масштаба правой стороны квадрата, которое эта линия отметила. Чтобы не портить графика проведением этой линии в действительности, можно просто приложить край линейки к точке *D* и к точке *O* и, поставив острие циркуля к соответствующей стороне квадрата у края линейки, посмотреть, на каком делении пришлась эта точка. В данном случае пунктирная линия *OD* пересекла правую сторону квадрата на уровне деления, соответствующего цифре 80. Это указывает на наличие 80 шансов из 100 за то, что оцениваемая разница средних (или оцениваемый коэффициент корреляции *r*) не является случайным. Стало быть, если бы мы повторили наш опыт еще 99 раз (а всего, следовательно, 100 раз), то приблизительно в 80 случаях (из этих 100) мы могли бы ожидать подтверждения того же самого заключения, а в остальных 20 случаях (100—80) мог бы получиться и обратный (противоречащий нашему) вывод.

Если бы пунктирная линия (соединяющая точку *O* и *D*) пересекла верхнюю сторону квадрата в точке, соответствующей, например, 50, то мы имели бы 50 шансов из 100 за то, что данный вывод не является случайным, т.-е. при повторении опыта могли бы ожидать подтверждения его только в 50 случаях из 100 (и, стало быть, неподтверждения тоже в 50 случаях), что лишило бы этот вывод всякой практической ценности.

В конце книги приложен этот график *) с намеченными уже делениями на всех четырех сторонах квадрата. Слева и снизу помечены деления равномерные (более мелкие, чем на описанных здесь чертежах), а справа и сверху оценочные деления, указывающие число шансов из 100 за то, что данный вывод не является случайным. На внутренней части

*) Имеется в отдельной продаже в книжном складе кооп. т-ва „Современник“ (Ленинград, пр. Володарского 27).

правой стороны квадрата, там, где снаружи идут деления от 69 до 100 (сверху — вниз), помечены особые мелкие неравномерные деления от 0 (сверху) до 100 (снизу). Эти деления обозначают то количество шансов, за которое можно ручаться во всяком случае. Например, там где у нас по наружному масштабу получилось число шансов = 80, отсчет по внутреннему масштабу даст приблизительно только 40 шансов (см. черт. 4). Это значит, что мы ожидаем приблизительно 80 шансов за правильность данного вывода и во всяком случае не менее 40 шансов.

На том же графике вычерчены внутри плоскости квадрата еще несколько линий. С их помощью можно графически определять приближенную величину средних ошибок.

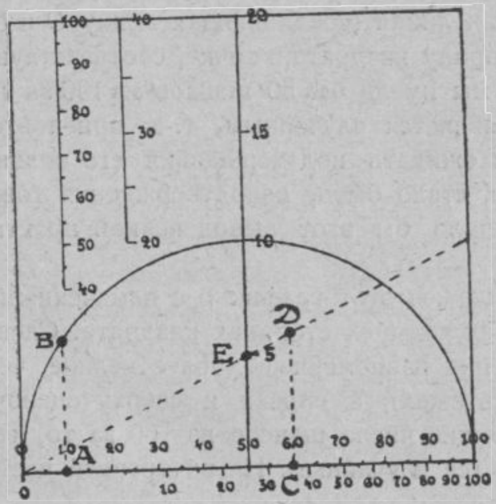
Так, для определения средней ошибки процентного отношения ($m_q$), которая, как известно, вычисляется по формуле (см. стр. 234):

$$m_q = \pm \sqrt{\frac{q\,(100-q)}{n-1}}\ \%,$$

мы поступаем следующим образом (см. черт. 5).

На нижней стороне квадрата по верхнему (равномерному) масштабу отыскиваем точку $A$, соответствующую данному проценту $q$ (здесь $q = 10\%$). В эту точку мы ставим одну ножку циркуля, а другую ножку поднимаем (раздвигая циркуль) до линии полуокружности и ставим ее в точку $B$. Далее осторожно снимаем циркуль с чертежа и по нижнему (уже неравномерному) масштабу той же (нижней) стороны квадрата отыскиваем деление, соответствующее числу измерений без 1, т. е. $n-1$

ЧЕРТ. 5.





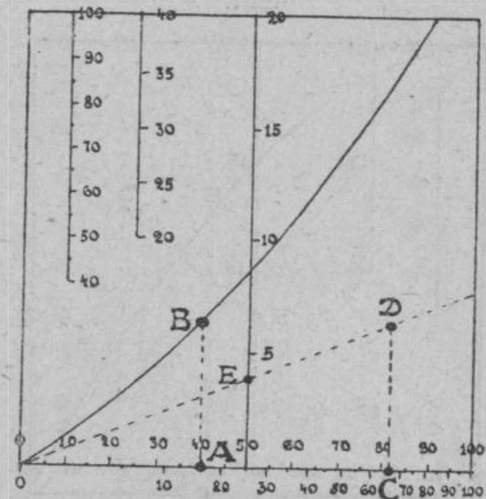

(у нас $n = 37$, а $n - 1 = 36$). В эту точку ($C$) мы ставим одну из ножек циркуля, а другую ножку поднимаем вверх (не раздвигая циркуля) и ставим ее в точку, в которой она окажется (в точку $D$) *). Таким образом отрезок $CD$ в точности равен измеренному перед тем отрезку $AB$. Соединив затем краем линейки точку $D$ с левым нижним углом квадрата ($O$), мы циркуль снимаем и одну из его ножек ставим, прижимая к краю линейки, на вертикальную масштабную прямую в точке $E$. Место этой точки и определяет приблизительную величину искомой ошибки $m_q$.

В нашем примере точка $E$ пришлась на 5-м делении этой линии. Следовательно

$$m_q = \pm 5\%.$$

Если бы край линейки, приложенной к точкам $D$ и $O$, вовсе не пересек бы этой вертикальной прямой, то мы сделали бы отметку на параллельной ей масштабной прямой (с делениями от 20 до 40), расположенной несколько левее, и на ней прочли бы желаемый результат. Если бы и эта линия не попала в пересечение с краем линейки, то мы воспользовались бы самой крайней слева масштабной прямой (с делениями от 40 до 100).

ЧЕРТ. 6.



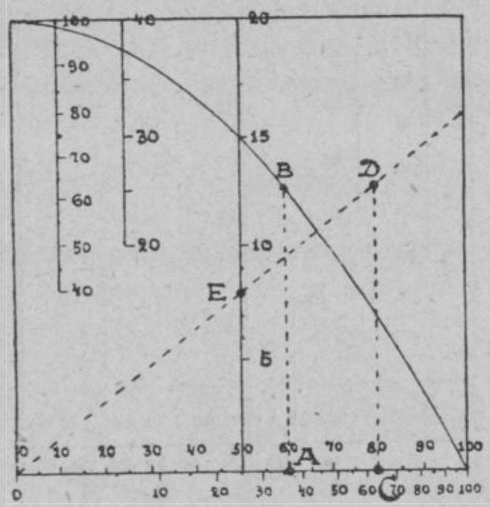

Аналогичным построением можно определить и ошибку $m_v$ вариационного коэффициента $v$, вычисляемую по формуле (см. стр. 192):

$$m_v = \frac{v\sqrt{0{,}5 + \left(\frac{v}{100}\right)^2}}{\sqrt{n}}$$

*) Чтобы облегчить отложение отрезка $CD$ вверх по направлению строго перпендикулярному нижней стороне квадрата, на графике, приложенном в конце этой книги, проведена сеть параллельных друг другу вертикальных прямых.

Для этого (см. черт. 6) по равномерному масштабу нижней стороны квадрата отыскиваем точку $A$, соответствующую значению вариационного коэффициента $v$ (пусть $v = 40\%$), и измеряем циркулем расстояние кверху от этой точки до вычерченной на графике специальной кривой в точке $B$. Затем по неравномерному нижнему масштабу отыскиваем точку $C$, соответствующую заданному числу измерений $n$ (например $= 66$), и от этой точки $C$ откладываем вверх только что измеренное перед тем расстояние ($AB = CD$). Полученную точку $D$ соединяем с $O$ линейкой и на пересечении ее края с вертикальной масштабной прямой находим точку $E$, положение которой и дает искомую величину ошибки вариационного коэффициента. В данном случае точка $E$ пришлась приблизительно на 4-м делении соответствующего масштаба. Следовательно:

ЧЕРТ. 7.




$$m_v = \pm 4\%.$$

Если бы край линейки не дал пересечения со средней масштабной прямой, то точку $E$ пришлось бы искать на одной из двух слева расположенных вертикальных линий.

С помощью того же графика можно определить и ошибку коэффициента корреляции $r$, вычисляемую по формуле (см. стр. 153):

$$m_r = \frac{1 - r^2}{\sqrt{n}}$$

Для этого следует воспользоваться кривой (параболой), соединяющей левый верхний угол квадрата с его правым нижним углом (см. черт. 7). Схема построения остается прежней, с той лишь разницей, что значение коэффициента корреляции следует сперва увеличить в 100 раз, а конечный



отсчет по вертикальной масштабной прямой, наоборот, уменьшить в 100 раз.

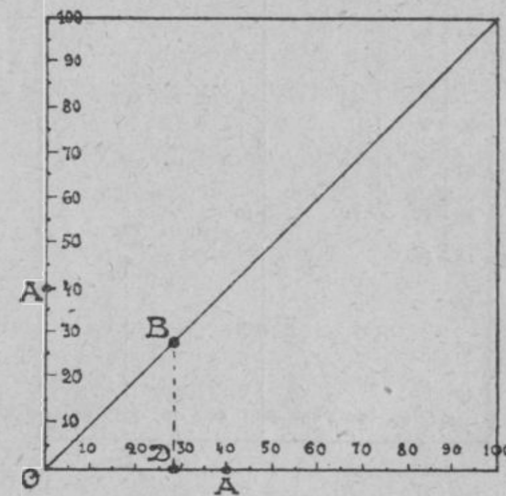
Пусть, например, $r = 0{,}60$. Увеличив его в 100 раз, получим 60. Отыскав по горизонтальному равномерному масштабу точку $A$, соответствующую 60, измеряем расстояние от $A$ до $B$ (до параболы) и откладываем его от точки $C$ (соответствующей числу измерений $n = 64$) кверху от $C$ до $D$ ($CD = AB$). Соединив $D$ с $O$ (по линейке), найдем на вертикальной масштабной прямой точку $E$, которая в нашем примере пришлась на 8 делении. Этот отсчет следует уменьшить в 100 раз. Следовательно искомая ошибка коэффициента корреляции

$$m_r = \pm 0{,}08.$$

Для определения приближенной величины ошибки ($m_\sigma$) среднего квадратического отклонения ($\sigma$), вычисляемой по формуле (см. стр. 187):

$$m_\sigma = \frac{m}{\sqrt{2}},$$

ЧЕРТ. 8.




необходимо выполнить на графике следующее построение (см. черт. 8).

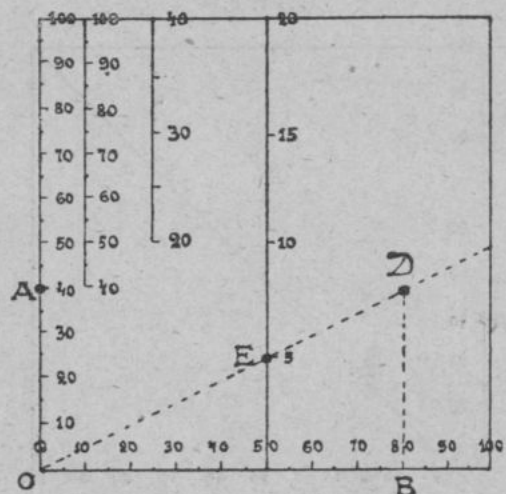
Отмерив от точки $O$ по равномерному масштабу нижней (или левой) стороны квадрата отрезок $OA$, соответствующий величине $m$, откладываем его на диагонали ($OB = OA$). Перпендикуляр, опущенный из точки $B$ до нижней стороны квадрата (до точки $D$) в том же масштабе, и даст искомую величину ошибки сигмы.

Наконец, с помощью этого графика можно находить и приближенную величину процентного отношения одного числа к другому и, в частности, определять величину вариационного коэффициента $v$, вычисляемого по формуле (см. стр. 190):

$$v = \frac{100\,\sigma}{M}\,\%$$

Для этого (см. черт. 9) на левой стороне квадрата от точки $O$ отложим отрезок $OA$, соответствующий в данном масштабе величине $\sigma$ (пусть $\sigma = 40$). Затем, отыскав по равномерному масштабу нижней стороны точку $B$, соответствующую величине $M$ $(= 80)$, восставим из $B$ перпендикуляр $BD = OA$. Соединив точку $D$ с $O$, определим деление $E$ на вертикальной масштабной прямой, по которой и произведем нужный нам отсчет. Этот отсчет (5) следует всегда увеличить в 10 раз (получим $\upsilon = 50\%$).

ЧЕРТ. 9.




Если бы $M = 8$, а $\sigma = 4$, то, чтобы избежать слишком мелких построений, удобнее предварительно увеличить оба эти числа в 10 раз; если же, наоборот, заданная величина $M$ и $\sigma$ выходит за пределы масштаба (например, $M = 800$, а $\sigma = 400$), то оба эти числа следует еще до построения уменьшить в одинаковое число раз (например, в 10).

Эти увеличения и уменьшения $M$ и $\sigma$ в одинаковое число раз при данном построении совершенно не отражаются на конечном расчете, который, таким образом, во всех случаях необходимо в свою очередь увеличивать в 10 раз.

Но несколько иначе должно быть выполняемо приспособление к масштабу графика при построении средних ошибок (см. черт. 5, 6 и 7). Там изменять можно лишь величину отрезка, откладываемого по нижнему неравномерному масштабу нижней стороны квадрата (деления этого масштаба нанесены пропорционально квадратному корню из числа измерений $n$, и соответствующие цифры от 0 до 100 проставлены снизу масштабных делений).



Если это число измерений $n$ выходит за пределы данного масштаба (т.-е. оказывается больше 100), то при отыскании точки $C$ его следует уменьшить в 100 раз, а при окончательном отсчете по вертикальному масштабу, т.-е. при определении на нем точки $E$, соответствующий отсчет увеличить в 10 раз.

Полезно запомнить, что при пользовании этим графиком для определения числа шансов из 100 за тот или иной вывод (черт. 1, 2, 3 и 4) можно увеличить или уменьшить в одинаковое число раз все величины, откладываемые по равномерным масштабам левой и нижней стороны квадрата. Так, при оценке достоверности различия между двумя средними арифметическими $(M_1 - M_2)$, вычисленными с их средними ошибками $m_1$ и $m_2$, можно все эти три величины $(M_1 - M_2,\ m_1$ и $m_2)$ предварительно умножить или разделить на постоянное число, а при оценке достоверности коэффициента корреляции $r$, вычисленного со средней ошибкой $m_r$, можно $r$ и $m_r$ изменять в любое одинаковое число раз.

# КРАТКИЙ БИБЛИОГРАФИЧЕСКИЙ УКАЗАТЕЛЬ.

Бетц, В. Проблема корреляции в психологии (о соотношении психических способностей). М. 1923.

Беляев-Башкиров, Б. В. Статистический метод в психологии и педологии М. 1927.

Вихляев, П. А. Очерки теоретической статистики. 2-ое изд. М. 1926.

Иванов, А. А. Теория ошибок и способ наименьших квадратов. Пгр. 1921.

Кауфман, А. А. Теория и методы статистики. Л. 1928.

Лахтин, Л. К. Кривые распределения и построение для них интерполяционных формул по способам Пирсона и Брунса. М. 1922.

Леонтович, А. В. Элементарное пособие к применению методов Gauss'a и Pearson'a. Часть I, II, III. Киев. 1909—1911.

Митропольский, А. К. Основы статистики (статистическое исчисление) Т. I. Л. 1925.

Орженцкий, Р. М. Учебник математической статистики. 1914.

Принцинг, Ф. Методы санитарной статистики. М. 1925.

Ритц, Г. Л. Математические методы в статистике. М. 1927.

Романовский. В. И. Элементарный курс математической статистики. Л. 1924.

Сапегин, А. А. Вариационная статистика. 3-ье изд. Л. 1928.

Слуцкий, Е. Е. Теория корреляции и элементы учения о кривых распределения. Киев. 1912.

Статистический метод в научном исследовании. Под редакцией М. Смит и А. Тимирязева. М. 1925.

Уиппль, Дж. Ч. и Новосельский, С. А. Основы демографической и санитарной статистики. М. 1929.

Филипченко, Ю. А. Изменчивость и методы ее изучения (основы биологической вариационной статистики). Л. 1927.

Хотимский, В. И. Выравнивание статистических рядов по методу наименьших квадратов (способ Чебышева). Л. 1925.

Эгиз С. А. Корреляционный метод в полевом опыте. М. 1925.

Эльдертон, В. П. Кривые распределения численностей и корреляция. М. 1924.

Наиболее доступными для читателей не математиков являются книги проф. Ю. А. Филипченко (Изменчивость и методы ее изучения) и проф. А. А. Сапегина (Вариационная статистика). Для лиц с достаточной математической подготовкой можно рекомендовать книгу проф. А. К. Митропольского (Основы статистики). В книге того же автора „Предмет статистики", Л. 1927, помещен подробный систематический указатель литературы по „Статистическому Исчислению".

# ТАБЛИЦА

**квадратных корней и квадратов чисел от 1 до 1000.**

| $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1,0000 | 1 | 36 | 6,0000 | 1296 | 71 | 8,4261 | 5041 |
| 2 | 1,4142 | 4 | 37 | 6,0828 | 1369 | 72 | 8,4853 | 5184 |
| 3 | 1,7321 | 9 | 38 | 6,1644 | 1444 | 73 | 8,5440 | 5329 |
| 4 | 2,0000 | 16 | 39 | 6,2450 | 1521 | 74 | 8,6023 | 5476 |
| 5 | 2,2361 | 25 | **40** | 6,3246 | 1600 | 75 | 8,6603 | 5625 |
| 6 | 2,4495 | 36 | 41 | 6,4031 | 1681 | 76 | 8,7178 | 5776 |
| 7 | 2,6458 | 49 | 42 | 6,4807 | 1764 | 77 | 8,7750 | 5929 |
| 8 | 2,8284 | 64 | 43 | 6,5574 | 1849 | 78 | 8,8318 | 6084 |
| 9 | 3,0000 | 81 | 44 | 6,6332 | 1936 | 79 | 8,8882 | 6241 |
| **10** | 3,1623 | 100 | 45 | 6,7082 | 2025 | **80** | 8,9443 | 6400 |
| 11 | 3,3166 | 121 | 46 | 6,7823 | 2116 | 81 | 9,0000 | 6561 |
| 12 | 3,4641 | 144 | 47 | 6,8557 | 2209 | 82 | 9,0554 | 6724 |
| 13 | 3,6056 | 169 | 48 | 6,9282 | 2304 | 83 | 9,1104 | 6889 |
| 14 | 3,7417 | 196 | 49 | 7,0000 | 2401 | 84 | 9,1652 | 7056 |
| 15 | 3,8730 | 225 | **50** | 7,0711 | 2500 | 85 | 9,2195 | 7225 |
| 16 | 4,0000 | 256 | 51 | 7,1414 | 2601 | 86 | 9,2736 | 7396 |
| 17 | 4,1231 | 289 | 52 | 7,2111 | 2704 | 87 | 9,3274 | 7569 |
| 18 | 4,2426 | 324 | 53 | 7,2801 | 2809 | 88 | 9,3808 | 7744 |
| 19 | 4,3589 | 361 | 54 | 7,3485 | 2916 | 89 | 9,4340 | 7921 |
| **20** | 4,4721 | 400 | 55 | 7,4162 | 3025 | **90** | 9,4868 | 8100 |
| 21 | 4,5826 | 441 | 56 | 7,4833 | 3136 | 91 | 9,5394 | 8281 |
| 22 | 4,6904 | 484 | 57 | 7,5498 | 3249 | 92 | 9,5917 | 8464 |
| 23 | 4,7958 | 529 | 58 | 7,6158 | 3364 | 93 | 9,6437 | 8649 |
| 24 | 4,8990 | 576 | 59 | 7,6811 | 3481 | 94 | 9,6954 | 8836 |
| 25 | 5,0000 | 625 | **60** | 7,7460 | 3600 | 95 | 9,7468 | 9025 |
| 26 | 5,0990 | 676 | 61 | 7,8102 | 3721 | 96 | 9,7980 | 9216 |
| 27 | 5,1962 | 729 | 62 | 7,8740 | 3844 | 97 | 9,8489 | 9409 |
| 28 | 5,2915 | 784 | 63 | 7,9373 | 3969 | 98 | 9,8995 | 9604 |
| 29 | 5,3852 | 841 | 64 | 8,0000 | 4096 | 99 | 9.9499 | 9801 |
| **30** | 5,4772 | 900 | 65 | 8,0623 | 4225 | **100** | 10,0000 | 10000 |
| 31 | 5,5678 | 961 | 66 | 8,1240 | 4356 | 101 | 10,0499 | 10201 |
| 32 | 5,6569 | 1024 | 67 | 8,1854 | 4489 | 102 | 10,0995 | 10404 |
| 33 | 5,7446 | 1089 | 68 | 8,2462 | 4624 | 103 | 10,1489 | 10609 |
| 34 | 5,8310 | 1156 | 69 | 8,3066 | 4761 | 104 | 10,1980 | 10816 |
| 35 | 5,9161 | 1225 | **70** | 8,3666 | 4900 | 105 | 10,2470 | 11025 |

| $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ |
|---|---|---|---|---|---|---|---|---|
| 101 | 10,0499 | 10201 | 151 | 12,2882 | 22801 | 201 | 14,1774 | 40401 |
| 102 | 10,0995 | 10404 | 152 | 12,3288 | 23104 | 202 | 14.2127 | 40804 |
| 103 | 10,1489 | 10609 | 153 | 12,3693 | 23409 | 203 | 14,2478 | 41209 |
| 104 | 10,1980 | 10816 | 154 | 12,4097 | 23716 | 204 | 14,2829 | 41616 |
| 105 | 10,2470 | 11025 | 155 | 12,4499 | 24025 | 205 | 14,3178 | 42025 |
| 106 | 10,2956 | 11236 | 156 | 12,4900 | 24336 | 206 | 14,3527 | 42436 |
| 107 | 10,3441 | 11449 | 157 | 12,5300 | 24649 | 207 | 14,3875 | 42849 |
| 108 | 10,3923 | 11664 | 158 | 12,5698 | 24964 | 208 | 14,4222 | 43264 |
| 109 | 10,4403 | 11881 | 159 | 12,6095 | 25281 | 209 | 14,4568 | 43681 |
| **110** | 10,4881 | 12100 | **160** | 12,6491 | 25600 | **210** | 14,4914 | 44100 |
| 111 | 10,5357 | 12321 | 161 | 12,6886 | 25921 | 211 | 14,5258 | 44521 |
| 112 | 10,5830 | 12544 | 162 | 12,7279 | 26244 | 212 | 14,5602 | 44944 |
| 113 | 10,6301 | 12769 | 163 | 12,7671 | 26569 | 213 | 14,5945 | 45369 |
| 114 | 10,6771 | 12996 | 164 | 12.8062 | 26896 | 214 | 14,6287 | 45796 |
| 115 | 10,7238 | 13225 | 165 | 12,8452 | 27225 | 215 | 14,6629 | 46225 |
| 116 | 10,7703 | 13456 | 166 | 12,8841 | 27556 | 216 | 14,6969 | 46656 |
| 117 | 10,8167 | 13689 | 167 | 12,9228 | 27889 | 217 | 14,7309 | 47089 |
| 118 | 10,8628 | 13924 | 168 | 12,9615 | 28224 | 218 | 14,7648 | 47524 |
| 119 | 10,9087 | 14161 | 169 | 13,0000 | 28561 | 219 | 14,7986 | 47961 |
| **120** | 10,9545 | 14400 | **170** | 13,0384 | 28900 | **220** | 14,8324 | 48400 |
| 121 | 11,0000 | 14641 | 171 | 13,0767 | 29241 | 221 | 14,8661 | 48841 |
| 122 | 11,0454 | 14884 | 172 | 13,1149 | 29584 | 222 | 14,8997 | 49284 |
| 123 | 11,0905 | 15129 | 173 | 13,1529 | 29929 | 223 | 14,9332 | 49729 |
| 124 | 11,1355 | 15376 | 174 | 13,1909 | 30276 | 224 | 14,9666 | 50176 |
| 125 | 11,1803 | 15625 | 175 | 13,2288 | 30625 | 225 | 15,0000 | 50625 |
| 126 | 11,2250 | 15876 | 176 | 13,2665 | 30976 | 226 | 15,0333 | 51076 |
| 127 | 11,2694 | 16129 | 177 | 13,3041 | 31329 | 227 | 15,0665 | 51529 |
| 128 | 11,3137 | 16384 | 178 | 13,3417 | 31684 | 228 | 15,0997 | 51984 |
| 129 | 11,3578 | 16641 | 179 | 13,3791 | 32041 | 229 | 15,1327 | 52441 |
| **130** | 11,4018 | 16900 | **180** | 13,4164 | 32400 | **230** | 15,1658 | 52900 |
| 131 | 11,4455 | 17161 | 181 | 13,4536 | 32761 | 231 | 15,1987 | 53361 |
| 132 | 11,4891 | 17424 | 182 | 13,4907 | 33124 | 232 | 15,2315 | 53824 |
| 133 | 11,5326 | 17689 | 183 | 13,5277 | 33489 | 233 | 15,2643 | 54289 |
| 134 | 11,5758 | 17956 | 184 | 13,5647 | 33856 | 234 | 15,2971 | 54756 |
| 135 | 11,6190 | 18225 | 185 | 13,6015 | 34225 | 235 | 15,3297 | 55225 |
| 136 | 11,6619 | 18496 | 186 | 13,6382 | 34596 | 236 | 15,3623 | 55696 |
| 137 | 11,7047 | 18769 | 187 | 13,6748 | 34969 | 237 | 15,3948 | 56169 |
| 138 | 11,7473 | 19044 | 188 | 13,7113 | 35344 | 238 | 15,4272 | 56644 |
| 139 | 11,7898 | 19321 | 189 | 13,7477 | 35721 | 239 | 15,4596 | 57121 |
| **140** | 11,8322 | 19600 | **190** | 13,7840 | 36100 | **240** | 15,4919 | 57600 |
| 141 | 11,8743 | 19881 | 191 | 13,8203 | 36481 | 241 | 15,5242 | 58081 |
| 142 | 11,9164 | 20164 | 192 | 13,8564 | 36864 | 242 | 15,5563 | 58564 |
| 143 | 11,9583 | 20449 | 193 | 13,8924 | 37249 | 243 | 15,5885 | 59049 |
| 144 | 12,0000 | 20736 | 194 | 13,9284 | 37636 | 244 | 15,6205 | 59536 |
| 145 | 12,0416 | 21025 | 195 | 13,9642 | 38025 | 245 | 15,6525 | 60025 |
| 146 | 12,0830 | 21316 | 196 | 14,0000 | 38416 | 246 | 15,6844 | 60516 |
| 147 | 12,1244 | 21609 | 197 | 14,0357 | 38809 | 247 | 15,7162 | 61009 |
| 148 | 12,1655 | 21904 | 198 | 14,0712 | 39204 | 248 | 15,7480 | 61504 |
| 149 | 12,2066 | 22201 | 199 | 14,1067 | 39601 | 249 | 15,7797 | 62001 |
| **150** | 12,2474 | 22500 | **200** | 14,1421 | 40000 | **250** | 15,8114 | 62500 |



| $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ |
|---|---|---|---|---|---|---|---|---|
| 251 | 15,8430 | 63001 | 301 | 17,3494 | 90601 | 351 | 18 7350 | 123201 |
| 252 | 15,8745 | 63504 | 302 | 17,3781 | 91204 | 352 | 18,7617 | 123904 |
| 253 | 15,9060 | 64009 | 303 | 17,4069 | 91809 | 353 | 18,7883 | 124609 |
| 254 | 15,9374 | 64516 | 304 | 17,4356 | 92416 | 354 | 18,8149 | 125316 |
| 255 | 15,9687 | 65025 | 305 | 17,4642 | 93025 | 355 | 18,8414 | 126025 |
| 256 | 16,0000 | 65536 | 306 | 17,4929 | 93636 | 356 | 18,8680 | 126736 |
| 257 | 16,0312 | 66049 | 307 | 17,5214 | 94249 | 357 | 18,8944 | 127449 |
| 258 | 16,0624 | 66564 | 308 | 17,5499 | 94864 | 358 | 18,9209 | 128164 |
| 259 | 16,0935 | 67081 | 309 | 17,5784 | 95481 | 359 | 18,9473 | 128881 |
| **260** | 16,1245 | 67600 | **310** | 17,6068 | 96100 | **360** | 18,9737 | 129600 |
| 261 | 16,1555 | 68121 | 311 | 17,6352 | 96721 | 361 | 19,0000 | 130321 |
| 262 | 16,1864 | 68644 | 312 | 17,6635 | 97344 | 362 | 19,0263 | 131044 |
| 263 | 16,2173 | 69169 | 313 | 17,6918 | 97969 | 363 | 19,0526 | 131769 |
| 264 | 16,2481 | 69696 | 314 | 17,7200 | 98596 | 364 | 19,0788 | 132496 |
| 265 | 16,2788 | 70225 | 315 | 17,7482 | 99225 | 365 | 19,1050 | 133225 |
| 266 | 16,3095 | 70756 | 316 | 17,7764 | 99856 | 366 | 19,1311 | 133956 |
| 267 | 16,3401 | 71289 | 317 | 17,8045 | 100489 | 367 | 19,1572 | 134689 |
| 268 | 16,3707 | 71824 | 318 | 17,8326 | 101124 | 368 | 19,1833 | 135424 |
| 269 | 16,4012 | 72361 | 319 | 17,8606 | 101761 | 369 | 19,2094 | 136161 |
| **270** | 16,4317 | 72900 | **320** | 17,8885 | 102400 | **370** | 19,2354 | 136900 |
| 271 | 16,4621 | 73441 | 321 | 17,9165 | 103041 | 371 | 19,2614 | 137641 |
| 272 | 16,4924 | 73984 | 322 | 17,9444 | 103684 | 372 | 19,2873 | 138384 |
| 273 | 16,5227 | 74529 | 323 | 17,9722 | 104329 | 373 | 19,3132 | 139129 |
| 274 | 16,5529 | 75076 | 324 | 18,0000 | 104976 | 374 | 19,3391 | 139876 |
| 275 | 16,5831 | 75625 | 325 | 18,0278 | 105625 | 375 | 19,3649 | 140625 |
| 276 | 16,6132 | 76176 | 326 | 18,0555 | 106276 | 376 | 19,3907 | 141376 |
| 277 | 16,6433 | 76729 | 327 | 18,0831 | 106929 | 377 | 19,4165 | 142129 |
| 278 | 16,6733 | 77284 | 328 | 18,1108 | 107584 | 378 | 19,4422 | 142884 |
| 279 | 16,7033 | 77841 | 329 | 18,1384 | 108241 | 379 | 19,4679 | 143641 |
| **280** | 16,7332 | 78400 | **330** | 18,1659 | 108900 | **380** | 19,4936 | 144400 |
| 281 | 16,7631 | 78961 | 331 | 18,1934 | 109561 | 381 | 19,5192 | 145161 |
| 282 | 16,7929 | 79524 | 332 | 18,2209 | 110224 | 382 | 19,5448 | 145924 |
| 283 | 16,8226 | 80089 | 333 | 18,2483 | 110889 | 383 | 19,5704 | 146689 |
| 284 | 16,8523 | 80656 | 334 | 18,2757 | 111556 | 384 | 19,5959 | 147456 |
| 285 | 16,8819 | 81225 | 335 | 18,3030 | 112225 | 385 | 19,6214 | 148225 |
| 286 | 16,9115 | 81796 | 336 | 18,3303 | 112896 | 386 | 19,6469 | 148996 |
| 287 | 16,9411 | 82369 | 337 | 18,3576 | 113569 | 387 | 19,6723 | 149769 |
| 288 | 16,9706 | 82944 | 338 | 18,3848 | 114244 | 388 | 19,6977 | 150544 |
| 289 | 17,0000 | 83521 | 339 | 18,4120 | 114921 | 389 | 19,7231 | 151321 |
| **290** | 17,0294 | 84100 | **340** | 18,4391 | 115600 | **390** | 19,7484 | 152100 |
| 291 | 17,0587 | 84681 | 341 | 18,4662 | 116281 | 391 | 19,7737 | 152881 |
| 292 | 17,0880 | 85264 | 342 | 18,4932 | 116964 | 392 | 19,7990 | 153664 |
| 293 | 17,1172 | 85849 | 343 | 18,5203 | 117649 | 393 | 19,8242 | 154449 |
| 294 | 17,1464 | 86436 | 344 | 18,5472 | 118336 | 394 | 19,8494 | 155236 |
| 295 | 17,1756 | 87025 | 345 | 18,5742 | 119025 | 395 | 19,8746 | 156025 |
| 296 | 17,2047 | 87616 | 346 | 18,6011 | 119716 | 396 | 19,8997 | 156816 |
| 297 | 17,2337 | 88209 | 347 | 18,6279 | 120409 | 397 | 19,9249 | 157609 |
| 298 | 17,2627 | 88804 | 348 | 18,6548 | 121104 | 398 | 19,9499 | 158404 |
| 299 | 17,2916 | 89401 | 349 | 18,6815 | 121801 | 399 | 19,9750 | 159201 |
| **300** | 17,3205 | 90000 | **350** | 18,7083 | 122500 | **400** | 20,0000 | 160000 |

| $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ |
|---|---|---|---|---|---|---|---|---|
| 401 | 20,0250 | 160801 | 451 | 21,2368 | 203401 | 501 | 22,3830 | 251001 |
| 402 | 20,0499 | 161604 | 452 | 21,2603 | 204304 | 502 | 22,4054 | 252004 |
| 403 | 20,0749 | 162409 | 453 | 21,2838 | 205209 | 503 | 22,4277 | 253009 |
| 404 | 20,0998 | 163216 | 454 | 21,3073 | 206116 | 504 | 22,4499 | 254016 |
| 405 | 20,1246 | 164025 | 455 | 21,3307 | 207025 | 505 | 22,4722 | 255025 |
| 406 | 20,1494 | 164836 | 456 | 21,3542 | 207936 | 506 | 22,4944 | 256036 |
| 407 | 20,1742 | 165649 | 457 | 21,3776 | 208849 | 507 | 22,5167 | 257049 |
| 408 | 20,1990 | 166464 | 458 | 21,4009 | 209764 | 508 | 22,5389 | 258064 |
| 409 | 20,2237 | 167281 | 459 | 21,4243 | 210681 | 509 | 22,5610 | 259081 |
| **410** | 20,2485 | 168100 | **460** | 21,4476 | 211600 | **510** | 22,5832 | 260100 |
| 411 | 20,2731 | 168921 | 461 | 21,4709 | 212521 | 511 | 22,6053 | 261121 |
| 412 | 20,2978 | 169744 | 462 | 21,4942 | 213444 | 512 | 22,6274 | 262144 |
| 413 | 20,3224 | 170569 | 463 | 21,5174 | 214369 | 513 | 22,6495 | 263169 |
| 414 | 20,3470 | 171396 | 464 | 21,5407 | 215296 | 514 | 22,6716 | 264196 |
| 415 | 20,3715 | 172225 | 465 | 21,5639 | 216225 | 515 | 22,6936 | 265225 |
| 416 | 20,3961 | 173056 | 466 | 21,5870 | 217156 | 516 | 22,7156 | 266256 |
| 417 | 20,4206 | 173889 | 467 | 21,6102 | 218089 | 517 | 22,7376 | 267289 |
| 418 | 20,4450 | 174724 | 468 | 21,6333 | 219024 | 518 | 22,7596 | 268324 |
| 419 | 20,4695 | 175561 | 469 | 21,6564 | 219961 | 519 | 22,7816 | 269341 |
| **420** | 20,4939 | 176400 | **470** | 21,6795 | 220900 | **520** | 22,8035 | 270400 |
| 421 | 20,5183 | 177241 | 471 | 21,7025 | 221841 | 521 | 22,8254 | 271441 |
| 422 | 20,5426 | 178084 | 472 | 21,7256 | 222784 | 522 | 22,8473 | 272484 |
| 423 | 20,5670 | 178929 | 473 | 21,7486 | 223729 | 523 | 22,8692 | 273529 |
| 424 | 20,5913 | 179776 | 474 | 21,7715 | 224676 | 524 | 22,8910 | 274576 |
| 425 | 20,6155 | 180625 | 475 | 21,7945 | 225625 | 525 | 22,9129 | 275625 |
| 426 | 20,6398 | 181476 | 476 | 21,8174 | 226576 | 526 | 22,9347 | 276676 |
| 427 | 20,6640 | 182329 | 477 | 21,8403 | 227529 | 527 | 22,9565 | 277729 |
| 428 | 20,6882 | 183184 | 478 | 21,8632 | 228484 | 528 | 22,9783 | 278784 |
| 429 | 20,7123 | 184041 | 479 | 21,8861 | 229441 | 529 | 23,0000 | 279841 |
| **430** | 20,7364 | 184900 | **480** | 21,9089 | 230400 | **530** | 23,0217 | 280900 |
| 431 | 20,7605 | 185761 | 481 | 21,9317 | 231361 | 531 | 23,0434 | 281961 |
| 432 | 20,7846 | 186624 | 482 | 21,9545 | 232324 | 532 | 23,0651 | 283024 |
| 433 | 20,8087 | 187489 | 483 | 21,9773 | 233289 | 533 | 23,0868 | 284089 |
| 434 | 20,8327 | 188356 | 484 | 22,0000 | 234256 | 534 | 23,1084 | 285156 |
| 435 | 20,8567 | 189225 | 485 | 22,0227 | 235225 | 535 | 23,1301 | 286225 |
| 436 | 20,8806 | 190096 | 486 | 22,0454 | 236196 | 536 | 23,1517 | 287296 |
| 437 | 20,9045 | 190969 | 487 | 22,0681 | 237169 | 537 | 23,1733 | 288369 |
| 438 | 20,9284 | 191844 | 488 | 22,0907 | 238144 | 538 | 23,1948 | 289444 |
| 439 | 20,9523 | 192721 | 489 | 22,1133 | 239121 | 539 | 23,2164 | 290521 |
| **440** | 20,9762 | 193600 | **490** | 22,1359 | 240100 | **540** | 23,2379 | 291600 |
| 441 | 21,0000 | 194481 | 491 | 22,1585 | 241081 | 541 | 23,2594 | 292681 |
| 442 | 21,0238 | 195364 | 492 | 22,1811 | 242064 | 542 | 23,2809 | 293764 |
| 443 | 21.0476 | 196249 | 493 | 22,2036 | 243049 | 543 | 23,3024 | 294849 |
| 444 | 21,0713 | 197136 | 494 | 22,2261 | 244036 | 544 | 23,3238 | 295936 |
| 445 | 21,0950 | 198025 | 495 | 22,2486 | 245025 | 545 | 23,3452 | 297025 |
| 446 | 21,1187 | 198916 | 496 | 22,2711 | 246016 | 546 | 23,3666 | 298116 |
| 447 | 21,1424 | 199809 | 497 | 22,2935 | 247009 | 547 | 23,3880 | 299209 |
| 448 | 21,1660 | 200704 | 498 | 22,3159 | 248004 | 548 | 24,4094 | 300304 |
| 449 | 21,1896 | 201601 | 499 | 22,3383 | 249001 | 549 | 23,4307 | 301401 |
| **450** | 21,2132 | 202500 | **500** | 22,3607 | 250000 | **550** | 23,4521 | 302500 |



| $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ |
|---|---|---|---|---|---|---|---|---|
| 551 | 23,4734 | 303601 | 601 | 24,5153 | 361201 | 651 | 25,5147 | 423801 |
| 552 | 23,4947 | 304704 | 602 | 24,5357 | 362404 | 652 | 25,5343 | 425104 |
| 553 | 23,5160 | 305809 | 603 | 24,5561 | 363609 | 653 | 25,5539 | 426409 |
| 554 | 23,5372 | 306916 | 604 | 24,5764 | 364816 | 654 | 25,5734 | 427716 |
| 555 | 23,5584 | 308025 | 605 | 24,5967 | 366025 | 655 | 25,5930 | 429025 |
| 556 | 23,5797 | 309136 | 606 | 24,6171 | 367236 | 656 | 25,6125 | 430336 |
| 557 | 23,6008 | 310249 | 607 | 24,6374 | 368449 | 657 | 25,6320 | 431649 |
| 558 | 23,6220 | 311364 | 608 | 24,6577 | 369664 | 658 | 25,6515 | 432964 |
| 559 | 23,6432 | 312481 | 609 | 24,6779 | 370881 | 659 | 25,6710 | 434281 |
| **560** | 23,6643 | 313600 | **610** | 24,6982 | 372100 | **660** | 25,6905 | 435600 |
| 561 | 23,6854 | 314721 | 611 | 24,7184 | 373321 | 661 | 25,7099 | 436921 |
| 562 | 23,7065 | 315844 | 612 | 24,7386 | 374544 | 662 | 25,7294 | 438244 |
| 563 | 23,7276 | 316969 | 613 | 24,7588 | 375769 | 663 | 25,7488 | 439569 |
| 564 | 23,7487 | 318096 | 614 | 24,7790 | 376996 | 664 | 25,7682 | 440896 |
| 565 | 23,7697 | 319225 | 615 | 24,7992 | 378225 | 665 | 25,7876 | 442225 |
| 566 | 23,7908 | 320356 | 616 | 24,8193 | 379456 | 666 | 25,8070 | 443556 |
| 567 | 23,8118 | 321489 | 617 | 24,8395 | 380689 | 667 | 25,8263 | 444889 |
| 568 | 23,8328 | 322624 | 618 | 24,8596 | 381924 | 668 | 25,8457 | 446224 |
| 569 | 23,8537 | 323761 | 619 | 24,8797 | 383161 | 669 | 25,8650 | 447561 |
| **570** | 23,8747 | 324900 | **620** | 24,8998 | 384400 | **670** | 25,8844 | 448900 |
| 571 | 23,8956 | 326041 | 621 | 24,9199 | 385641 | 671 | 25,9037 | 450241 |
| 572 | 23,9165 | 327184 | 622 | 24,9399 | 386884 | 672 | 25,9230 | 451584 |
| 573 | 23,9374 | 328329 | 623 | 24,9600 | 388129 | 673 | 25,9422 | 452929 |
| 574 | 23,9583 | 329476 | 624 | 24,9800 | 389376 | 674 | 25,9615 | 454276 |
| 575 | 23,9792 | 330625 | 625 | 25,0000 | 390625 | 675 | 25,9808 | 455625 |
| 576 | 24,0000 | 331776 | 626 | 25,0200 | 391876 | 676 | 26,0000 | 456976 |
| 577 | 24,0208 | 332929 | 627 | 25,0400 | 393129 | 677 | 26,0192 | 458329 |
| 578 | 24,0416 | 334084 | 628 | 25,0599 | 394384 | 678 | 26,0384 | 459684 |
| 579 | 24,0624 | 335241 | 629 | 25,0799 | 395641 | 679 | 26,0576 | 461041 |
| **580** | 24,0832 | 336400 | **630** | 25,0998 | 396900 | **680** | 26,0768 | 462400 |
| 581 | 24,1039 | 337561 | 631 | 25,1197 | 398161 | 681 | 26,0960 | 463761 |
| 582 | 24,1247 | 338724 | 632 | 25,1396 | 399424 | 682 | 26,1151 | 465124 |
| 583 | 24,1454 | 339889 | 633 | 25,1595 | 400689 | 683 | 26,1343 | 466489 |
| 584 | 24,1661 | 341056 | 634 | 25,1794 | 401956 | 684 | 26,1534 | 467856 |
| 585 | 24,1868 | 342225 | 635 | 25,1992 | 403225 | 685 | 26,1725 | 469225 |
| 586 | 24,2074 | 343396 | 636 | 25,2190 | 404496 | 686 | 26,1916 | 470596 |
| 587 | 24,2281 | 344569 | 637 | 25,2389 | 405769 | 687 | 26,2107 | 471969 |
| 588 | 24,2487 | 345744 | 638 | 25,2587 | 407044 | 688 | 26,2298 | 473344 |
| 589 | 24,2693 | 346921 | 639 | 25,2784 | 408321 | 689 | 26,2488 | 474721 |
| **590** | 24,2899 | 348100 | **640** | 25,2982 | 409600 | **690** | 26,2679 | 476100 |
| 591 | 24,3105 | 349281 | 641 | 25,3180 | 410881 | 691 | 26,2869 | 477481 |
| 592 | 24,3311 | 350464 | 642 | 25,3377 | 412164 | 692 | 26,3059 | 478864 |
| 593 | 24,3516 | 351649 | 643 | 25,3574 | 413449 | 693 | 26,3249 | 480249 |
| 594 | 24,3721 | 352836 | 644 | 25,3772 | 414736 | 694 | 26,3439 | 481636 |
| 595 | 24,3926 | 354025 | 645 | 25,3969 | 416025 | 695 | 26,3629 | 483025 |
| 596 | 24,4131 | 355216 | 646 | 25,4165 | 417316 | 696 | 26,3818 | 484416 |
| 597 | 24,4336 | 356409 | 647 | 25,4362 | 418609 | 697 | 26,4008 | 485809 |
| 598 | 24,4540 | 357604 | 648 | 25,4558 | 419904 | 698 | 26,4197 | 487204 |
| 599 | 24,4745 | 358801 | 649 | 25,4755 | 421201 | 699 | 26,4386 | 488601 |
| **600** | 24,4949 | 360000 | **650** | 25,4951 | 422500 | **700** | 26,4575 | 490000 |

| $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ |
|---|---|---|---|---|---|---|---|---|
| 701 | 26,4764 | 491401 | 751 | 27,4044 | 564001 | 801 | 28,3019 | 641601 |
| 702 | 26,4953 | 492804 | 752 | 27,4226 | 565504 | 802 | 28,3196 | 643204 |
| 703 | 26,5141 | 494209 | 753 | 27,4408 | 567009 | 803 | 28,3373 | 644809 |
| 704 | 26,5330 | 495616 | 754 | 27,4591 | 568516 | 804 | 28,3549 | 646416 |
| 705 | 26,5518 | 497025 | 755 | 27,4773 | 570025 | 805 | 28,3725 | 648025 |
| 706 | 26,5707 | 498436 | 756 | 27,4955 | 571536 | 806 | 28,3901 | 649636 |
| 707 | 26,5895 | 499849 | 757 | 27,5136 | 573049 | 807 | 28,4077 | 651249 |
| 708 | 26,6083 | 501264 | 758 | 27,5318 | 574564 | 808 | 28,4253 | 652864 |
| 709 | 26,6271 | 502681 | 759 | 27,5500 | 576081 | 809 | 28,4429 | 654481 |
| **710** | 26,6458 | 504100 | **760** | 27,5681 | 577600 | **810** | 28,4605 | 656100 |
| 711 | 26,6646 | 505521 | 761 | 27,5862 | 579121 | 811 | 28,4781 | 657721 |
| 712 | 26,6833 | 506944 | 762 | 27,6043 | 580644 | 812 | 28,4956 | 659344 |
| 713 | 26,7021 | 508369 | 763 | 27,6225 | 582169 | 813 | 28,5132 | 660969 |
| 714 | 26,7208 | 509796 | 764 | 27,6405 | 583696 | 814 | 28,5307 | 662596 |
| 715 | 26,7395 | 511225 | 765 | 27,6586 | 585225 | 815 | 28,5482 | 664225 |
| 716 | 26,7582 | 512656 | 766 | 27,6767 | 586756 | 816 | 28,5657 | 665856 |
| 717 | 26,7769 | 514089 | 767 | 27,6948 | 588289 | 817 | 28,5832 | 667489 |
| 718 | 26,7955 | 515524 | 768 | 27,7128 | 589824 | 818 | 28,6007 | 669124 |
| 719 | 26,8142 | 516961 | 769 | 27,7308 | 591361 | 819 | 28,6182 | 670761 |
| **720** | 26,8328 | 518400 | **770** | 27,7489 | 592900 | **820** | 28,6356 | 672400 |
| 721 | 26,8514 | 519841 | 771 | 27,7669 | 594441 | 821 | 28,6531 | 674041 |
| 722 | 26,8701 | 521284 | 772 | 27,7849 | 595984 | 822 | 28,6705 | 675684 |
| 723 | 26,8887 | 522729 | 773 | 27,8029 | 597529 | 823 | 28,6880 | 677329 |
| 724 | 26,9072 | 524176 | 774 | 27,8209 | 599076 | 824 | 28,7054 | 678976 |
| 725 | 26,9258 | 525625 | 775 | 27,8388 | 600625 | 825 | 28,7228 | 680625 |
| 726 | 26,9444 | 527076 | 776 | 27,8568 | 602176 | 826 | 28,7402 | 682276 |
| 727 | 26,9629 | 528529 | 777 | 27,8747 | 603729 | 827 | 28,7576 | 683929 |
| 728 | 26,9815 | 529984 | 778 | 27,8927 | 605284 | 828 | 28,7750 | 685584 |
| 729 | 27,0000 | 531441 | 779 | 27,9106 | 606841 | 829 | 28,7924 | 687241 |
| **730** | 27,0185 | 532900 | **780** | 27,9285 | 608400 | **830** | 28,8097 | 688900 |
| 731 | 27,0370 | 534361 | 781 | 27,9464 | 609961 | 831 | 28,8271 | 690561 |
| 732 | 27,0555 | 535824 | 782 | 27,9643 | 611524 | 832 | 28,8444 | 692224 |
| 733 | 27,0740 | 537289 | 783 | 27,9821 | 613089 | 833 | 28,8617 | 693889 |
| 734 | 27,0924 | 538756 | 784 | 28,0000 | 614656 | 834 | 28,8791 | 695556 |
| 735 | 27,1109 | 540225 | 785 | 28,0179 | 616225 | 835 | 28,8964 | 697225 |
| 736 | 27,1293 | 541696 | 786 | 28,0357 | 617796 | 836 | 28,9137 | 698896 |
| 737 | 27,1477 | 543169 | 787 | 28,0535 | 619369 | 837 | 28,9310 | 700569 |
| 738 | 27,1662 | 544644 | 788 | 28,0713 | 620944 | 838 | 28,9482 | 702244 |
| 739 | 27,1846 | 546121 | 789 | 28,0891 | 622521 | 839 | 28,9655 | 703921 |
| **740** | 27,2029 | 547600 | **790** | 28,1069 | 624100 | **840** | 28,9828 | 705600 |
| 741 | 27,2213 | 549081 | 791 | 28,1247 | 625681 | 841 | 29,0000 | 707281 |
| 742 | 27,2397 | 550564 | 792 | 28,1425 | 627264 | 842 | 29,0172 | 708964 |
| 743 | 27,2580 | 552049 | 793 | 28,1603 | 628849 | 843 | 29,0345 | 710649 |
| 744 | 27,2764 | 553536 | 794 | 28,1780 | 630436 | 844 | 29,0517 | 712336 |
| 745 | 27,2947 | 555025 | 795 | 28,1957 | 632025 | 845 | 29,0689 | 714025 |
| 746 | 27,3130 | 556516 | 796 | 28,2135 | 633616 | 846 | 29,0861 | 715716 |
| 747 | 27,3313 | 558009 | 797 | 28,2312 | 635209 | 847 | 29,1033 | 717409 |
| 748 | 27,3496 | 559504 | 798 | 28,2489 | 636804 | 848 | 29,1204 | 719104 |
| 749 | 27,3679 | 561001 | 799 | 28,2666 | 638401 | 849 | 29.1376 | 720801 |
| **750** | 27,3861 | 562500 | **800** | 28,2843 | 640000 | **850** | 29,1548 | 722500 |



| $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ | $N$ | $\sqrt{N}$ | $N^2$ |
|---|---|---|---|---|---|---|---|---|
| 851 | 29,1719 | 724201 | 901 | 30,0167 | 811801 | 951 | 30,8383 | 904401 |
| 852 | 29,1890 | 725904 | 902 | 30,0333 | 813604 | 952 | 30,8545 | 906304 |
| 853 | 29,2062 | 727609 | 903 | 30,0500 | 815409 | 953 | 30,8707 | 908209 |
| 854 | 29,2233 | 729316 | 904 | 30,0666 | 817216 | 954 | 30,8869 | 910116 |
| 855 | 29,2404 | 731025 | 905 | 30,0832 | 819025 | 955 | 30,9031 | 912025 |
| 856 | 29,2575 | 732736 | 906 | 30,0998 | 820836 | 956 | 30,9192 | 913936 |
| 857 | 29,2746 | 734449 | 907 | 30,1164 | 822649 | 957 | 30,9354 | 915849 |
| 858 | 29,2916 | 736164 | 908 | 30,1330 | 824464 | 958 | 30,9516 | 917764 |
| 859 | 29,3087 | 737881 | 909 | 30,1496 | 826281 | 959 | 30,9677 | 919681 |
| **860** | 29,3258 | 739600 | **910** | 30,1662 | 828100 | **960** | 30,9839 | 921600 |
| 861 | 29,3428 | 741321 | 911 | 30,1828 | 829921 | 961 | 31,0000 | 923521 |
| 862 | 29,3598 | 743044 | 912 | 30,1993 | 831744 | 962 | 31,0161 | 925444 |
| 863 | 29,3769 | 744769 | 913 | 30,2159 | 833569 | 963 | 31,0322 | 927369 |
| 864 | 29,3939 | 746496 | 914 | 30,2324 | 835396 | 964 | 31,0483 | 929296 |
| 865 | 29,4109 | 748225 | 915 | 30,2490 | 837225 | 965 | 31,0644 | 931225 |
| 866 | 29,4279 | 749956 | 916 | 30,2655 | 839056 | 966 | 31,0805 | 933156 |
| 867 | 29,4449 | 751689 | 917 | 30,2820 | 840889 | 967 | 31,0966 | 935089 |
| 868 | 29,4618 | 753424 | 918 | 30,2985 | 842724 | 968 | 31,1127 | 937024 |
| 869 | 29,4788 | 755161 | 919 | 30,3150 | 844561 | 969 | 31,1288 | 938961 |
| **870** | 29,4958 | 756900 | **920** | 30,3315 | 846400 | **970** | 31,1448 | 940900 |
| 871 | 29,5127 | 758641 | 921 | 30,3480 | 848241 | 971 | 31,1609 | 942841 |
| 872 | 29,5296 | 760384 | 922 | 30,3645 | 850084 | 972 | 31,1769 | 944784 |
| 873 | 29,5466 | 762129 | 923 | 30,3809 | 851929 | 973 | 31,1929 | 946729 |
| 874 | 29,5635 | 763876 | 924 | 30,3974 | 853776 | 974 | 31,2090 | 948676 |
| 875 | 29,5804 | 765625 | 925 | 30,4138 | 855625 | 975 | 31,2250 | 950625 |
| 876 | 29,5973 | 767376 | 926 | 30,4302 | 857476 | 976 | 31,2410 | 952576 |
| 877 | 29,6142 | 769129 | 927 | 30,4467 | 859329 | 977 | 31,2570 | 954529 |
| 878 | 29,6311 | 770884 | 928 | 30,4631 | 861184 | 978 | 31,2730 | 956484 |
| 879 | 29,6479 | 772641 | 929 | 30,4795 | 863041 | 979 | 31,2890 | 958441 |
| **880** | 29,6648 | 774400 | **930** | 30,4959 | 864900 | **980** | 31,3050 | 960400 |
| 881 | 29,6816 | 776161 | 931 | 30,5123 | 866761 | 981 | 31,3209 | 962361 |
| 882 | 29,6985 | 777924 | 932 | 30,5287 | 868624 | 982 | 31,3369 | 964324 |
| 883 | 29,7153 | 779689 | 933 | 30,5450 | 870489 | 983 | 31,3528 | 966289 |
| 884 | 29,7321 | 781456 | 934 | 30,5614 | 872356 | 984 | 31,3688 | 968256 |
| 885 | 29,7489 | 783225 | 935 | 30,5778 | 874225 | 985 | 31,3847 | 970225 |
| 886 | 29,7658 | 784996 | 936 | 30,5941 | 876096 | 986 | 31,4006 | 972196 |
| 887 | 29,7825 | 786769 | 937 | 30,6105 | 877969 | 987 | 31,4166 | 974169 |
| 888 | 29,7993 | 788544 | 938 | 30,6268 | 879844 | 988 | 31,4325 | 976144 |
| 889 | 29,8161 | 790321 | 939 | 30,6431 | 881721 | 989 | 31,4484 | 978121 |
| **890** | 29,8329 | 792100 | **940** | 30,6594 | 883600 | **990** | 31,4643 | 980100 |
| 891 | 29,8496 | 793881 | 941 | 30,6757 | 885481 | 991 | 31,4802 | 982081 |
| 892 | 29,8664 | 795664 | 942 | 30,6920 | 887364 | 992 | 31,4960 | 984064 |
| 893 | 29,8831 | 797449 | 943 | 30,7083 | 889249 | 993 | 31,5119 | 986049 |
| 894 | 29,8998 | 799236 | 944 | 30,7246 | 891136 | 994 | 31,5278 | 988036 |
| 895 | 29,9166 | 801025 | 945 | 30,7409 | 893025 | 995 | 31,5436 | 990025 |
| 896 | 29,9333 | 802816 | 946 | 30,7571 | 894916 | 996 | 31,5595 | 992016 |
| 897 | 29,9500 | 804609 | 947 | 30,7734 | 896809 | 997 | 31,5753 | 994009 |
| 898 | 29,9666 | 806404 | 948 | 30,7896 | 898704 | 998 | 31,5911 | 996004 |
| 899 | 29,9833 | 808201 | 949 | 30,8058 | 900601 | 999 | 31,6070 | 998001 |
| **900** | 30,0000 | 810000 | **950** | 30,8221 | 902500 | **1000** | 31,6228 | 1000000 |

Таблица значений величины $\gamma = v\sqrt{0,5+\left(\frac{v}{100}\right)^2}$, входящей в числитель упрощенной формулы для определения средней ошибки вариационного коэффициента $m_v = \frac{\gamma}{\sqrt{n}}$, для всех значений этого коэффициента в пределах от 1 до 125 включительно.

(См. описание на стр. 195).

| $v$ | $\gamma$ | $\Delta$ | $v$ | $\gamma$ | $\Delta$ | $v$ | $\gamma$ | $\Delta$ | $v$ | $\gamma$ | $\Delta$ | $v$ | $\gamma$ | $\Delta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,71 | 0,070 | 26 | 19,59 | 0,085 | 51 | 44,46 | 0,118 | 76 | 78,89 | 0,161 | 101 | 124,58 | 0,207 |
| 2 | 1,41 | 0,071 | 27 | 20,44 | 0,085 | 52 | 45,64 | 0,120 | 77 | 80,50 | 0,162 | 102 | 126,60 | 0,208 |
| 3 | 2,12 | 0,071 | 28 | 21,29 | 0,087 | 53 | 46,84 | 0,120 | 78 | 82,12 | 0,164 | 103 | 128,68 | 0,211 |
| 4 | 2,83 | 0,071 | 29 | 22,16 | 0,088 | 54 | 48,04 | 0,123 | 79 | 83,76 | 0,166 | 104 | 130,79 | 0,213 |
| 5 | 3,54 | 0,071 | 30 | 23,04 | 0,089 | 55 | 49,27 | 0,124 | 80 | 85,42 | 0,168 | 105 | 132,92 | 0,215 |
| 6 | 4,25 | 0,072 | 31 | 23,93 | 0,091 | 56 | 50,51 | 0,126 | 81 | 87,09 | 0,170 | 106 | 135,07 | 0,216 |
| 7 | 4,97 | 0,072 | 32 | 24,84 | 0,091 | 57 | 51,77 | 0,127 | 82 | 88,79 | 0,171 | 107 | 137,23 | 0,219 |
| 8 | 5,69 | 0,072 | 33 | 25,75 | 0,093 | 58 | 53,04 | 0,129 | 83 | 90,50 | 0,173 | 108 | 139,42 | 0,220 |
| 9 | 6,41 | 0,073 | 34 | 26,68 | 0,093 | 59 | 54,33 | 0,131 | 84 | 92,23 | 0,175 | 109 | 141,62 | 0,222 |
| 10 | 7,14 | 0,073 | 35 | 27,61 | 0,096 | 60 | 55,64 | 0,133 | 85 | 93,98 | 0,177 | 110 | 143,84 | 0,225 |
| 11 | 7,87 | 0,074 | 36 | 28,57 | 0,096 | 61 | 56,97 | 0,134 | 86 | 95,75 | 0,179 | 111 | 146,09 | 0,226 |
| 12 | 8,61 | 0,074 | 37 | 29,53 | 0,097 | 62 | 58,31 | 0,135 | 87 | 97,54 | 0,180 | 112 | 148,35 | 0,228 |
| 13 | 9,35 | 0,074 | 38 | 30,50 | 0,099 | 63 | 59,66 | 0,138 | 88 | 99,34 | 0,183 | 113 | 150.63 | 0,230 |
| 14 | 10,09 | 0,075 | 39 | 31,49 | 0,101 | 64 | 61,04 | 0,139 | 89 | 101,17 | 0,184 | 114 | 152,93 | 0,232 |
| 15 | 10,84 | 0,076 | 40 | 32,50 | 0,101 | 65 | 62,43 | 0,141 | 90 | 103,01 | 0,186 | 115 | 155,25 | 0,234 |
| 16 | 11,60 | 0,076 | 41 | 33,51 | 0,103 | 66 | 63,84 | 0,143 | 91 | 104,87 | 0,188 | 116 | 157,59 | 0,236 |
| 17 | 12,36 | 0,077 | 42 | 34,54 | 0,105 | 67 | 65,27 | 0,144 | 92 | 106,75 | 0,190 | 117 | 159,95 | 0,238 |
| 18 | 13,13 | 0,078 | 43 | 35,59 | 0,105 | 68 | 66,71 | 0,146 | 93 | 108,65 | 0,192 | 118 | 162,33 | 0,239 |
| 19 | 13,91 | 0,079 | 44 | 36,64 | 0,108 | 69 | 68,17 | 0,148 | 94 | 110,57 | 0,194 | 119 | 164,72 | 0,242 |
| 20 | 14,70 | 0,079 | 45 | 37,72 | 0,108 | 70 | 69,65 | 0,150 | 95 | 112,51 | 0,195 | 120 | 167,14 | 0,244 |
| 21 | 15,49 | 0,080 | 46 | 38,80 | 0,111 | 71 | 71,15 | 0,151 | 96 | 114,46 | 0,198 | 121 | 169,58 | 0,245 |
| 22 | 16,29 | 0,081 | 47 | 39,91 | 0,111 | 72 | 72,66 | 0,153 | 97 | 116,44 | 0,199 | 122 | 172,03 | 0,248 |
| 23 | 17,10 | 0,082 | 48 | 41,02 | 0,113 | 73 | 74,19 | 0,155 | 98 | 118,43 | 0,201 | 123 | 174,51 | 0,249 |
| 24 | 17,92 | 0,083 | 49 | 42,15 | 0,115 | 74 | 75,74 | 0,157 | 99 | 120,44 | 0,203 | 124 | 177,00 | 0,252 |
| 25 | 18,75 | 0,084 | 50 | 43,30 | 0,116 | 75 | 77,31 | 0,158 | 100 | 122,47 | 0,206 | 125 | 179,52 | 0,253 |



Таблица значений величин $1-r^2$, соответствующих значениям коэффициентов корреляции $r$ в пределах от 0,01 до 1,00 включительно и наименьшего числа измерений $n$, при наличии которого данный коэффициент $r$ будет превышать свою среднюю ошибку $m_r$ по крайней мере в 3 раза.

(См. описание на стр. 154).

| $r$ | $1-r^2$ | $n$ | $r$ | $1-r^2$ | $n$ | $r$ | $1-r^2$ | $n$ | $r$ | $1-r^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,01 | 0,9999 | 89983 | 0,26 | 0,9324 | 116 | 0,51 | 0,7399 | 19 | 0,76 | 0,4224 |
| 0,02 | 0,9996 | 22483 | 0,27 | 0,9271 | 107 | 0,52 | 0,7296 | 18 | 0,77 | 0,4071 |
| 0,03 | 0,9991 | 9982 | 0,28 | 0,9216 | 98 | 0,53 | 0,7191 | 17 | 0,78 | 0,3916 |
| 0,04 | 0,9984 | 5608 | 0,29 | 0,9159 | 90 | 0,54 | 0,7084 | 16 | 0,79 | 0,3759 |
| 0,05 | 0,9975 | 3583 | 0,30 | 0,9100 | 83 | 0,55 | 0,6975 | 15 | 0,80 | 0,3600 |
| 0,06 | 0,9964 | 2483 | 0,31 | 0,9039 | 77 | 0,56 | 0,6864 | 14 | 0,81 | 0,3439 |
| 0,07 | 0,9951 | 1819 | 0,32 | 0,8976 | 71 | 0,57 | 0,6751 | 13 | 0,82 | 0,3276 |
| 0,08 | 0,9936 | 1389 | 0,33 | 0,8911 | 66 | 0,58 | 0,6636 | 12 | 0,83 | 0,3111 |
| 0,09 | 0,9919 | 1094 | 0,34 | 0,8844 | 61 | 0,59 | 0,6519 | 11 | 0,84 | 0,2944 |
| 0,10 | 0,9900 | 883 | 0,35 | 0,8775 | 57 | 0,60 | 0,6400 | 11 | 0,85 | 0,2775 |
| 0,11 | 0,9879 | 726 | 0,36 | 0,8704 | 53 | 0,61 | 0,6279 | 10 | 0,86 | 0,2604 |
| 0,12 | 0,9856 | 608 | 0,37 | 0,8631 | 49 | 0,62 | 0,6156 | 9 | 0,87 | 0,2431 |
| 0,13 | 0,9831 | 515 | 0,38 | 0,8556 | 46 | 0,63 | 0,6031 | 9 | 0,88 | 0,2256 |
| 0,14 | 0,9804 | 442 | 0,39 | 0,8479 | 43 | 0,64 | 0,5904 | 8 | 0,89 | 0,2079 |
| 0,15 | 0,9775 | 383 | 0,40 | 0,8400 | 40 | 0,65 | 0,5775 | 8 | 0,90 | 0,1900 |
| 0,16 | 0,9744 | 334 | 0,41 | 0,8319 | 38 | 0,66 | 0,5644 | 7 | 0,91 | 0,1719 |
| 0,17 | 0,9711 | 294 | 0,42 | 0,8236 | 35 | 0,67 | 0,5511 | 7 | 0,92 | 0,1536 |
| 0,18 | 0,9676 | 261 | 0,43 | 0,8151 | 33 | 0,68 | 0,5376 | 6 | 0,93 | 0,1351 |
| 0,19 | 0,9639 | 232 | 0,44 | 0,8064 | 31 | 0,69 | 0,5239 | 6 | 0,94 | 0,1164 |
| 0,20 | 0,9600 | 208 | 0,45 | 0,7975 | 29 | 0,70 | 0,5100 | 5 | 0,95 | 0,0975 |
| 0,21 | 0,9559 | 187 | 0,46 | 0,7884 | 27 | 0,71 | 0,4959 | 5 | 0,96 | 0,0784 |
| 0,22 | 0,9516 | 169 | 0,47 | 0,7791 | 25 | 0,72 | 0,4816 | 5 | 0,97 | 0,0591 |
| 0,23 | 0,9471 | 153 | 0,48 | 0,7696 | 24 | 0,73 | 0,4671 | 4 | 0,98 | 0,0396 |
| 0,24 | 0,9424 | 139 | 0,49 | 0,7599 | 22 | 0,74 | 0,4524 | 4 | 0,99 | 0,0199 |
| 0,25 | 0,9375 | 127 | 0,50 | 0,7500 | 21 | 0,75 | 0,4375 | 4 | 1,00 | 0,0000 |

# ПЕРЕЧЕНЬ ОБОЗНАЧЕНИЙ.

| | СТРАН. |
|---|---|
| $V$ — варианта (числовой результат отдельного измерения) | 50 |
| $p$ — частота́ (количество вариант данного значения $V$) | 65 |
| $n$ — общее число измерений (наблюдений) | 50 |
| $\Sigma$ — знак суммирования (напр. $\Sigma p = n$) | 50 |
| $M$ — среднее арифметическое | 10 |
| $x$ (или $y$) — центральное отклонение (от $M$) | 50 |
| $C$ (или $C_x$ или $C_y$) — начало отсчетов (произвольное) | 57 |
| $x_c$ (или $y_c$) — вспомогательное отклонение (от $C$) | 57 |
| $a_1$, $a_2$, $b_1$, $b_2$ — итоги сложения частот по вертикальным столбцам при вычислениях, производимых по способу сумм | 77—78 |
| $S_1$ — сумма вспомогательных отклонений | 58 |
| $S_2$ — сумма квадратов вспомогательных отклонений | 58 |
| $\Sigma x$ — сумма центральных отклонений ($\Sigma x = 0$) | 52 |
| $\Sigma x^2$ — сумма квадратов центральных отклонений | 53 |
| $k$ (или $k_x$ или $k_y$) — классовый промежуток | 96 |
| $\sigma$ — среднее квадратическое отклонение | 14 |
| $m$ — средняя ошибка (среднего арифметического $M$) | 19 |
| $m_\sigma$ — ошибка сигмы ($\sigma$) | 47 |
| $v$ — вариационный коэффициент | 190 |
| $P$ — показатель точности | 197 |
| $m_v$ — ошибка вариационного коэффициента ($v$) | 192 |
| $\gamma$ — вспомогательная величина для вычисления ошибки $m_v$, определяемая по таблице | 195 |
| $\Delta$ — табличная разность при определении величины $\gamma$ | 196 |
| $m_P$ — ошибка показателя точности ($P$) | 199 |
| $q$ — процентная частота́ (для качественных признаков) | 233 |
| $m_q$ — ошибка процентной частоты ($q$) | 234 |
| $\Sigma xy$ — сумма произведений центральных отклонений | 148 |
| $r$ — коэффициент корреляции | 27 |
| $m_r$ — ошибка коэффициента корреляции ($r$) | 37 |
| $z$ — сумма центральных отклонений ($z = x + y$) | 158 |
| $p_1$, $p_2$, $p_3$, $p_4$ — часто́ты в четырех клетках корреляционной решетки при вычислении коэффициента $r$ для двух качественных признаков | 236 |



| | |
|---|---|
| $p_x$, $p_y$ — итоги подсчета часто́т по столбцам и по строчкам при вычислении коэффициентов $r$ для качественных признаков | 241 |
| $t_x$, $t_y$ — вспомогательные величины (зависящие от $p_x$ и $p_y$) при вычислении коэффициентов $r$ для качественн. признаков | 243 |
| $\delta$ — разность между ранговыми (порядковыми) номерами | 265 |
| $R_y$ — коэффициент регрессии $y$ на $x$ | 201 |
| $R_x$ — коэффициент регрессии $x$ на $y$ | 201 |
| $m_y$ — ошибка коэффициента регрессии $y$ на $x$ ($R_y$) | 206 |
| $m_x$ — ошибка коэффициента регрессии $x$ на $y$ ($R_x$) | 206 |
| $\varphi$ — показатель взаимной сопряженности | 253 |
| $S$ — вспомогательная сумма, вычисляемая при определении показателя $\varphi$ или коэффициента $r$ (для многих качественных признаков) | 257 |
| $T$ — общее обозначение любых величин, подлежащих сравнению друг с другом | 268 |

# СВОДКА ФОРМУЛ

## I. Общие числовые характеристики.

1) Среднее арифметическое: страм.

$M = \frac{\Sigma V}{n}$ . . . . . . . . . . . . 50

$M = C + \frac{kS_1}{n}$ . . . . . . . . . . . . 96

2) Среднее квадратическое отклонение:

$\sigma = \pm k \sqrt{\frac{\Sigma x^2}{n-1}}$ . . . . . . . . . . . . 96

$\sigma = \pm k \sqrt{\frac{\Sigma x^2}{n}}$ . . . . . . . . . . . . 187

3) Вариационный коэффициент $v = \frac{100\,\sigma}{M}\,\%$ . . . . . . . . . . 190

4) Показатель точности: $P = \frac{100\,m}{M}\,\%$ . . . . . . . . . . 197

$P = \frac{v}{\sqrt{n}}\,\%$ . . . . . . . . . . 198

5) Коэффициент корреляции (для количественных признаков):

$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$ . . . . . . . . . . 148

$r = \frac{\Sigma xy}{n\sigma_x\sigma_y}$ . . . . . . . . . . 188

6) Коэффициент корреляции (для двух качественных признаков):

$r = \frac{p_2 p_3 - p_1 p_4}{\sqrt{(p_1+p_2)(p_3+p_4)(p_1+p_3)(p_2+p_4)}}$ . . . . 236

$r = \frac{np - p_x\, p_y}{t_x\, t_y}$ . . . . . . . . . . 243

7) Коэффициент корреляции (для многих качественных признаков):

$r = \sqrt{\frac{\varphi^2}{1+\varphi^2}}$ . . . . . . . . . . 253

$r = \sqrt{\frac{S-1}{S}}$ . . . . . . . . . . 257



8) Коэффициент корреляции (для рангов) $r = 1 - \frac{6\,\Sigma\delta^2}{n(n^2-1)}$ . . . 266

9) Коэффициенты регрессии:

$R_y = \frac{k_y}{k_x} \cdot \frac{\Sigma xy}{\Sigma x^2}$ . . . . . . . . . . 202

$R_x = \frac{k_x}{k_y} \cdot \frac{\Sigma xy}{\Sigma y^2}$ . . . . . . . . . . 202

$R_y = \frac{\sigma_y}{\sigma_x}\,r$ . . . . . . . . . . . . 205

$R_x = \frac{\sigma_x}{\sigma_y}\,r$ . . . . . . . . . . . . 205

## II. Средние ошибки.

10) Средняя ошибка среднего арифметического ($M$)

$m = \frac{\sigma}{\sqrt{n}}$ . . . . . . . . . . 50

$m = \pm k \sqrt{\frac{\Sigma x^2}{n(n-1)}}$ . . . . . . . 96

11) Ошибка среднего квадратического отклонения ($\sigma$)

$m_\sigma = \frac{\sigma}{\sqrt{2n}}$ . . . . . . . . . . 47

$m_\sigma = \frac{m}{\sqrt{2}}$ . . . . . . . . . . 187

12) Ошибка вариационного коэффициента ($v$)

$m_v = \pm \frac{v\sqrt{0{,}5 + \left(\frac{v}{100}\right)^2}}{\sqrt{n}}\,\%$ . . . . . . . 192

$m_v = \frac{\gamma}{\sqrt{n}}$ . . . . . . . . . . 195

13) Ошибка показателя точности ($P$)

$m_P = \pm P \sqrt{\frac{1}{2n} + \left(\frac{P}{100}\right)^2}\,\%$ . . . . . . . 199

14) Ошибка процентной частоты ($q$)

$m_q = \pm \sqrt{\frac{q\,(100-q)}{n-1}}$ . . . . . . . . . 234

15) Ошибка коэффициента корреляции ($r$)

$m_r = \pm \frac{1-r^2}{\sqrt{n}}$ . . . . . . . . . . 153

16) Ошибка коэффициента корреляции, вычисленного при посредстве $\varphi$ (или $S$)

$$m_r = \pm\frac{4}{3}\left(\frac{1-r^2}{\sqrt{n}}\right) \quad \ldots\ldots\ldots \quad 253$$

17) Ошибки коэффициентов регрессии ($R_y$ и $R_x$):

$$m_y = \pm\frac{k_y}{k_x}\sqrt{\frac{\Sigma y^2}{\Sigma x^2}\left(\frac{1-r^2}{n}\right)} \quad \ldots\ldots \quad 206$$

$$m_x = \pm\frac{k_x}{k_y}\sqrt{\frac{\Sigma x^2}{\Sigma y^2}\left(\frac{1-r^2}{n}\right)} \quad \ldots\ldots \quad 206$$

$$m_y = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r^2}{n}} \quad \ldots\ldots\ldots \quad 207$$

$$m_x = \frac{\sigma_x}{\sigma_y}\sqrt{\frac{1-r^2}{n}} \quad \ldots\ldots\ldots \quad 207$$

## III. Технические и контрольные формулы.

18) Центральное отклонение

$$x = V - M \quad \ldots\ldots\ldots \quad 50$$

19) Вспомогательное отклонение

$$x_c = V - C \quad \ldots\ldots\ldots \quad 57$$

20) Сумма центральных отклонений

$$\Sigma x = 0 \quad \ldots\ldots\ldots \quad 52$$

21) Сумма вспомогательных отклонений

$$S_1 = \Sigma V - nC \quad \ldots\ldots\ldots \quad 58$$

$$S_1 = a_1 - b_1 \quad \ldots\ldots\ldots \quad 78$$

22) Сумма квадратов вспомогательных отклонений

$$S_2 = a_1 + b_1 + 2a_2 + 2b_2 \quad \ldots\ldots\ldots \quad 78$$

23) Проверка величин $a_1$, $b_1$, $a_2$ и $b_2$

$$\left.\begin{array}{l} \Sigma a_1 = a_1 \text{ в сводке} \\ \Sigma b_1 = b_1 \text{ „ „} \\ \Sigma a_2 = a_2 \text{ „ „} \\ \Sigma b_2 = b_2 \text{ „ „} \end{array}\right\} \quad \ldots\ldots\ldots \quad \left\{\begin{array}{l} 123 \\ 124 \end{array}\right.$$

24) Проверка величин $S_1$ и $S_2$

$$\left.\begin{array}{l} \Sigma S_1 = S_1 \text{ в сводке} \\ \Sigma S_2 = S_2 \text{ „ „} \end{array}\right\} \quad \ldots\ldots\ldots \quad \left\{\begin{array}{l} 124 \\ 125 \end{array}\right.$$

25) Сумма квадратов центральных отклонений

$$\Sigma x^2 \text{ (или } \Sigma y^2\text{, или } \Sigma z^2) = S_2 - \frac{S_1^2}{n} \quad \ldots\ldots \quad 160$$

26) Сумма произведений центральных отклонений

$$\Sigma xy = \frac{\Sigma z^2 - \Sigma y^2 - \Sigma x^2}{2} \quad \ldots\ldots\ldots \quad 161$$



27) Проверка величин $\Sigma x^2$, $\Sigma y^2$ и $\Sigma z^2$

$$\Sigma x^2 + \Sigma y^2 + \Sigma z^2 = S_{2(u)} - \frac{\Sigma S_1^2}{n} \quad \ldots\ldots \quad 173$$

28) Проверка коэффициентов регрессии

$$R_y \cdot R_x = r^2 \quad \ldots\ldots\ldots \quad 203$$

29) Проверка ошибок коэффициентов регрессии

$$m \cdot m_x = \frac{1-r^2}{n} \quad \ldots\ldots\ldots \quad 207$$

## IV. Формулы для оценки достоверности выводов.

30) Условие достоверности различия между двумя величинами $T_1$ и $T_2$, вычисленными со средними ошибками $m_1$ и $m_2$

$$\frac{T_1 - T_2}{\sqrt{m_1^2 + m_2^2}} > 3 \quad \ldots\ldots\ldots \quad 45$$

или

$$\frac{(T_1 - T_2)^2}{m_1^2 + m_2^2} > 9 \quad \ldots\ldots\ldots \quad 46$$

31) Условие достоверности коэффициента корреляции

$$\frac{r}{m_r} > 3 \quad \ldots\ldots\ldots \quad 43$$

32) Условие достоверности коэффициентов регрессии

$$\frac{R_y}{m_y} > 3 \quad \ldots\ldots\ldots \quad 209$$

$$\frac{R_x}{m_x} > 3 \quad \ldots\ldots\ldots \quad 209$$

33) Наименьшее значение коэффициента корреляции $r$, при котором этот коэффициент будет превышать свою среднюю ошибку ($m_r$) по крайней мере в 3 раза

$$\min r = \frac{\sqrt{n+36} - \sqrt{n}}{6} \quad \ldots\ldots\ldots \quad 245$$

34) Наименьшее значение вспомогательной суммы $S$, при котором зависящий от нее коэффициент корреляции $r$ (для качественных признаков) будет превышать свою среднюю ошибку $m_r$ по крайней мере в 3 раза

$$\min S = 0{,}5 + \sqrt{0{,}25 + \frac{16}{n}} \quad \ldots\ldots\ldots \quad 258$$

## V. Формулы, упрощающие вычисления в некоторых частных случаях.

35) Если ко всем вариантам $V$ прибавить постоянную величину $a$, то общие числовые характеристики нового ряда ($M_0$, $\sigma_0$ и $m_0$) будут находиться в следующем взаимоотношении к старым ($M$, $\sigma$ и $m$):

$M_0 = M + a$ . . . . . . . . . . . 211
$\sigma_0 = \sigma$ . . . . . . . . . . . . . 211
$m_0 = m$ . . . . . . . . . . . . . 214

36) Если из всех вариант $V$ вычесть постоянную величину $a$, то

$M_0 = M - a$ . . . . . . . . . . . 211
$\sigma_0 = \sigma$ . . . . . . . . . . . . . 21
$m_0 = m$ . . . . . . . . . . . . . 214

37) Если все варианты $V$ умножить на постоянную величину $a$, то

$M_0 = aM$ . . . . . . . . . . . . 211
$\sigma_0 = a\sigma$ . . . . . . . . . . . . .
$m_0 = am$ . . . . . . . . . . . . 214

38) Если все варианты $V$ разделить на постоянную величину $a$, то

$M_0 = \frac{M}{a}$ . . . . . . . . . . . . 212

$\sigma_0 = \frac{\sigma}{a}$ . . . . . . . . . . . . 212

$m_0 = \frac{m}{a}$ . . . . . . . . . . . . 214

39) Если все варианты ряда $y$ (т.-е. $V_y$) и, независимо от этого, все варианты ряда $x$ (т.-е. $V_x$) изменить на какую-либо постоянную величину, то новый коэффициент корреляции $r_0$ и оба новых коэффициента регрессии $R_{0y}$ и $R_{0x}$ окажутся равными прежним их значениям $r$, $R_y$ и $R_x$

$r_0 = r$ . . . . . . . . . . . . . 216
$R_{0y} = R_y$ . . . . . . . . . . . . 217
$R_{0x} = R_x$ . . . . . . . . . . . . 217

40) Если все варианты ряда $y$ (т.-е. $V_y$) умножить на постоянное число $a$, то

$r_0 = r$ . . . . . . . . . . . . . 216
$R_{0y} = aR_y$ . . . . . . . . . . . 218
$R_{0x} = \frac{R_x}{a}$ . . . . . . . . . . . . 218



41) Если все варианты ряда $y$ (т.-е. $V_y$) разделить на постоянное число $a$, то

$r_0 = r$ . . . . . . . . . . . . . 216
$R_{0y} = \frac{R_y}{a}$ . . . . . . . . . . . 218
$R_{0x} = aR_x$ . . . . . . . . . . . 218

42) Если все варианты ряда $x$ (т.-е. $V_x$) умножить на постоянное число $a$, то

$r_0 = r$ . . . . . . . . . . . . . 216
$R_{0y} = \frac{R_y}{a}$ . . . . . . . . . . . 217
$R_{0x} = aR_x$ . . . . . . . . . . . 217

43) Если все варианты ряда $x$ (т.-е. $V_x$) разделить на постоянное число $a$, то

$r_0 = r$ . . . . . . . . . . . . . 216
$R_{0y} = aR_y$ . . . . . . . . . . . 218
$R_{0x} = \frac{R_x}{a}$ . . . . . . . . . . . 218

44) Если новые варианты $V_3$ образуются путем сложения попарно двух соответствующих вариант $V_1$ и $V_2$ первоначальных рядов, то общие числовые характеристики нового ряда ($M_3$, $\sigma_3$ и $m_3$) будут находиться в следующем взаимоотношении к старым ($M_1$, $M_2$, $\sigma_1$, $\sigma_2$, $m_1$ и $m_2$) и к коэффициенту корреляции $r$ между двумя суммируемыми рядами

$M_3 = M_1 + M_2$ . . . . . . . . . . . . 219
$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2 + 2r\sigma_1\sigma_2}$ . . . . . . . 220
$m_3 = \sqrt{m_1^2 + m_2^2 + 2r\,m_1 m_2}$ . . . . . . 225

45) В частности, при полной прямой корреляции суммируемых рядов (т.-е. при $r = +1$)

$\sigma_3 = \sigma_1 + \sigma_2$ . . . . . . . . . . . . . 220
$m_3 = m_1 + m_2$ . . . . . . . . . . . . 225

46) При полной обратной корреляции этих рядов (т.-е. при $r = -1$)

$\sigma_3 = \sigma_1 - \sigma_2$ . . . . . . . . . . . . . 220
$m_3 = m_1 - m_2$ . . . . . . . . . . . . 225

47) При отсутствии корреляции между ними (т.-е. при $r = 0$)

$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2}$ . . . . . . . . . . 220
$m_3 = \sqrt{m_1^2 + m_2^2}$ . . . . . . . . . . 225

48) Коэффициент корреляции

$r = \frac{\sigma_3^2 - \sigma_2^2 - \sigma_1^2}{2\sigma_1\sigma_2}$ . . . . . . . . . 223

49) Если новые варианты $V_3$ образуются путем вычитания вариант $V_2$ из $V_1$ (т.-е. если $V_3 = V_1 - V_2$), то

$M_3 = M_1 - M_2$ . . . . . . . . . . . . . . . . 223

$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\,\sigma_1\,\sigma_2}$ . . . . . . . . . . 223

$m_3 = \sqrt{m_1^2 + m_2^2 - 2r\,m_1\,m_2}$ . . . . . . . . . 225

50) В частности, при полной прямой корреляции этих рядов (т.-е. при $r = +1$)

$\sigma_3 = \sigma_1 - \sigma_2$ . . . . . . . . . . . . . 223

$m_3 = m_1 - m_2$ . . . . . . . . . . . . . 226

51) При полной обратной их корреляции (т.-е. при $r = -1$)

$\sigma_3 = \sigma_1 + \sigma_2$ . . . . . . . . . . . . . 223

$m_3 = m_1 + m_2$ . . . . . . . . . . . . . 226

52) При отсутствии корреляции между этими рядами (т.-е. при $r = 0$)

$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2}$ . . . . . . . . . . 223

$m_3 = \sqrt{m_1^2 + m_2^2}$ . . . . . . . . . . 226

53) Коэффициент корреляции

$r = \frac{\sigma_1^2 + \sigma_2^2 - \sigma_3^2}{2\,\sigma_1\,\sigma_2}$ . . . . . . . . . . 224

54) Если новые варианты $V$ образуются путем сложения трех соответствующих вариант $V_1$, $V_2$ и $V_3$, то общие числовые характеристики нового ряда ($M$, $\sigma$ и $m$) будут находиться в следующем взаимоотношении к старым ($M_1$, $M_2$, $M_3$, $\sigma_1$, $\sigma_2$, $\sigma_3$, $m_1$, $m_2$ и $m_3$) и к коэффициентам корреляции ($r_{12}$, $r_{13}$ и $r_{23}$) между этими рядами

$M = M_1 + M_2 + M_3$ . . . . . . . . . . . . . . . . . . . . . . 225

$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2r_{12}\,\sigma_1\,\sigma_2 + 2r_{13}\,\sigma_1\,\sigma_3 + 2r_{23}\,\sigma_2\,\sigma_3}$ . . . . . 225

$m = \sqrt{m_1^2 + m_2^2 + m_3^2 + 2r_{12}\,m_1\,m_2 + 2r_{13}\,m_1\,m_3 + 2r_{23}\,m_2\,m_3}$

55) Если вариационный ряд был разбит на части и для каждой из этих частей были определены $n$, $M$, $\sigma$ и $m$, то общие для всего ряда числовые его характеристики $n_0$, $M_0$, $\sigma_0$ и $m_0$ могут быть вычислены по формулам

$n_0 = \Sigma n$ . . . . . . . . . . . . . . . . . . . . 230

$M_0 = \frac{\Sigma nM}{\Sigma n}$ . . . . . . . . . . . . . . . . . . 230

$\sigma_0 = \pm\sqrt{\frac{\Sigma(n-1)\,\sigma^2 + \Sigma n\,(M - M_0)^2}{\Sigma n - 1}}$ . . . . . . . 231

$m_0 = \frac{\sigma_0}{\sqrt{\Sigma n}}$ . . . . . . . . . . . . . . . . . . 232

56) В частности, при достаточно большом числе измерений $n_0$

$\sigma_0 = \pm\sqrt{\frac{\Sigma\,n\,\sigma^2 + \Sigma\,n\,(M - M_0)^2}{\Sigma n}}$ . . . . . . . . 233



## ДЛЯ ЗАМЕТОК:

# ОГЛАВЛЕНИЕ.

СТРАН.

От автора . . . . . . . . . . 3
I. Среднее арифметическое ($M$) . . . . . . . . 7
II. Среднее квадратическое отклонение ($\sigma$) . . . . . . 10
III. Средняя ошибка ($m$) . . . . . . . . . 16
IV. Коэффициент корреляции ($r$) . . . . . . . 23
V. Оценка достоверности окончательных выводов . . . . . . 39
VI. Способ непосредственного вычисления . . . . . . . 50
VII. Способ округления . . . . . . . . . 55
VIII. Способ произведений . . . . . . . . 63
IX. Способ сумм . . . . . . . . . 70
X. Пользование таблицей квадратных корней . . . . . . . 84
XI. Разбивка вариационного ряда на классы . . . . . . 95
XII. Одновременная обработка нескольких вариационных рядов со сводкой . . . . . . . . 116
XIII. Дополнительный учет отдельных вариант с большими отклонениями . . . . . . . . 135
XIV. Вычисление коэффициента корреляции при малом числе измерений . . . . . . . . 147
XV. Вычисление коэффициента корреляции при большом числе измерений . . . . . . . . 163
XVI. Вариационный коэффициент ($v$), показатель точности ($P$) и коэффициенты регрессии ($R$) . . . . . . . . 186
XVII. Некоторые частные случаи статистической обработки вариационных рядов . . . . . . . . 210
XVIII. Статистический анализ качественных признаков . . . . . . 233
XIX. Графический способ оценки достоверности окончательных выводов и приближенного вычисления средних ошибок . . . 267
Краткий библиографический указатель . . . . . . . 280
Таблица квадратов и квадратных корней . . . . . . . 281
Таблица значений величины $\gamma = v\sqrt{0{,}5 + \left(\frac{v}{100}\right)^2}$ для вычисления средней ошибки вариационного коэффициента . . . . . 288
Таблица значений $1 - r^2$ и наименьшего количества измерений ($n$), для вычисления ошибки коэффициента корреляции . . . . . 289
Перечень обозначений . . . . . . . . . 290
Сводка формул . . . . . . . . . 292
Приложения: а) расчетные решетки, формы №№ 1 и 2,
б) график для приближенного вычисления вероятностей и средних ошибок и
в) статистическая карточка.

**Цена 3 руб.**
Переплет 50 коп.

*Того же автора.* **Вариационная статистика. Часть II.**

Содержание. Графическое изображение вариационных рядов. Понятие о нормальном распределении. Построение кривой нормального распределения. Таблица процентного распределения вариант. Ассиметрия и эксцесс. Упрощенные способы вычисления показателей ассиметрии и эксцесса. Мода, медиана и квартиль. Методы оценки индивидуальных вариант с помощью стандартов, выведенных на массовом материале. Оценочные зоны. Децисигмальная оценка и перцентиляж. Составление оценочных таблиц и изготовление оценочных графиков. Комплексная оценка индивидуума по совокупности нескольких признаков. Индивидуальный профиль. Изготовление оценочного трафарета для вычерчивания индивидуальных профилей. Корреляция между отдельными компанентами единого комплекса суммарно оцениваемых признаков (интеркомпанентная корреляция). Парциальный коэффициент корреляции. Эмпирическая линия регрессии. Криволинейная корреляция. Вычисление корреляционных отношений по способу сумм. Построение теоретической линии регрессии 1-го, 2-го и 3-го порядка. Упрощенные способы выравнивания ломаных линий.

*Д-р мед. Н. Шеповальников и проф. Ю. Поморский.* **Возрастно-половые эволютивные особенности школьника и методика разработки стандартов антропометрических величин.** Л. Ц. 1 р. 25 к.

---

**Для вариационно-статистических вычислений имеются:**

| | |
|---|---|
| расчетные решетки с рабочими и контрольными формулами для вычисления констант (форма № 1) и для корреляций (форма № 2), 100 экз. | 4 р. — к. |
| график ошибок и вероятностей, на картоне, 1 экз. | — „ 75 „ |
| статистические карточки, 100 экз. | 2 „ — „ |
| масштабные сетки для вычерчивания сигмальных профилей (антропометрических, психометрич. и пр.), 100 экз. | 2 „ — „ |
| картонные трафареты с прорезями и масштабными делениями для накладывания на сетки при вычерчивании сигмальных профилей (по специальным заданиям), 1 экз. | 2 „ — „ |
| раскладочный ящик для статистических карточек с 32 отделениями, 1 экз. | 10 „ — „ |

УПАКОВКА И ПЕРЕСЫЛКА ПО ДЕЙСТВИТ. СТОИМОСТИ.

ИЗДАНИЕ НА СКЛАДЕ
Кооперативного Товарищества „СОВРЕМЕННИК"
Ленинград 28. Просп. Володарского 27. Тел. 160-10.