

Корреляційныя формулы какъ орудіе статистическаго анализа.

Корреляційныя вычисленія—едва ли не тотъ изъ предлагаемыхъ математическою статистикою приѣмовъ статистическаго анализа, который можетъ притязать на наиболѣе актуальное значеніе. Прежде всего, методъ корреляціи не вызываетъ такихъ—можетъ быть основательныхъ, можетъ быть неосновательныхъ, но во всякомъ случаѣ пока неразрѣшенныхъ—принципальныхъ сомнѣній, какія встрѣчаетъ примѣненіе къ явленіямъ соціальной жизни тѣхъ приѣмовъ статистико-математическаго анализа, въ основѣ которыхъ лежатъ формулы, основанныя на гипотезѣ случайныхъ погрѣшностей: „употребляя методъ корреляціи, мы не дѣлаемъ никакихъ допущеній относительно характера распредѣленія отдѣльныхъ значеній нашихъ величинъ, такъ что всѣ формулы теоріи корреляціи сохраняютъ свое значеніе при всякомъ законѣ распредѣленія“¹⁾; говоря иначе, методъ корреляціонныхъ вычисленій принципиально приложимъ къ какимъ угодно, безъ всякаго ограниченія, численнымъ рядамъ. Дальнѣйшимъ преимуществомъ корреляціонныхъ вычисленій является ихъ сравнительная доступность для тѣхъ широкихъ круговъ статистиковъ, которые лишены спеціальной математической подготовки,—доступность какъ въ смыслѣ пониманія, такъ и въ смыслѣ примѣненія. Правда, прослѣдить ходъ вывода корреляціонныхъ формулъ удастся, думается мнѣ, далеко не всякому, обладающему математическими знаніями въ уровнѣ, положимъ, средняго учебнаго заведенія; но такой или даже, можетъ быть, и еще болѣе скромный уровень математической подготовки достаточенъ, чтобы усвоить и осознать смыслъ этихъ формулъ во всѣхъ ихъ составныхъ элементахъ, а вмѣстѣ съ тѣмъ—чтобы разобраться во всѣхъ отѣнкахъ ихъ значенія для характеристики взаимоотношенія статистическихъ рядовъ; достаточенъ, слѣдовательно, чтобы обезпечить вполнѣ сознательное примѣненіе этихъ формулъ. Самый же процессъ примѣненія корреляціонныхъ формулъ, весьма обременительный, при сколько-нибудь

¹⁾ *Е. Е. Смуркій*. Теорія корреляціи и элементы ученія о природѣ распредѣленія. Кіевъ. 1912, стр. 71.

большихъ рядахъ, по количеству вычисленій, слагается однако изъ весьма простыхъ вычислительныхъ операцій, доступныхъ и для статистика съ болѣе, чѣмъ скромнымъ уровнемъ математическихъ познаній и навыковъ. Наконецъ, и самые результаты примѣненія корреляціонныхъ формулъ отличаются такою степенью конкретности и осязательности, какая свойственна далеко не всѣмъ другимъ приемамъ изображенія и анализа статистическихъ рядовъ, предлагаемымъ математической статистикой.

Такова совокупность соображеній, въ силу которыхъ вопросъ о роли корреляціонныхъ формулъ, какъ орудія статистическаго анализа, кажется мнѣ заслуживающимъ особаго вниманія,—соображеній, въ силу которыхъ я рѣшаюсь подѣлиться съ читателями „Статистическаго Вѣстника“ своими размышленіями по этому вопросу.

Мнѣ едва ли есть надобность входить здѣсь въ подробности относительно существа корреляціонныхъ формулъ и тѣхъ цѣлей, какія преслѣдуетъ методъ корреляціи. Цѣль эта—говоря, сначала, въ общихъ чертахъ—уловленіе *корреляціонной зависимости*, понятіе же корреляціонной зависимости тѣсно связывается съ идеями „совокупности“, „собирательнаго понятія“, слѣдовательно съ самымъ существомъ статистическаго метода. „Зависимость между двумя явленіями—говоритъ проф. Орженцкій ¹⁾, можетъ быть изучаема либо общими приемами изслѣдованія, либо статистически. Первый способъ примѣнимъ въ тѣхъ случаяхъ, когда зависимость проявляется тождественно въ каждомъ случаѣ наблюденія, или когда изъ наблюденія путемъ эксперимента могутъ быть устранены всѣ обстоятельства, видоизмѣняющія и нарушающія однообразное проявленіе зависимости. При отсутствіи указанныхъ условій, зависимость двухъ явленій скрывается или извращается въ отдѣльныхъ случаяхъ приводящими обстоятельствами и можетъ быть подмѣчена лишь сведеніемъ многихъ случаевъ въ одну общую совокупность. При количественномъ характерѣ изслѣдованія явленій такого рода способъ изслѣдованія составляетъ одну изъ задачъ статистическаго метода“. Въ первомъ случаѣ мы устанавливаемъ *однозначную функциональную зависимость*, при которой каждому значенію одной изъ связанныхъ зависимостью величинъ соотвѣтствуетъ одно значеніе другой—каждому X соотвѣтствуетъ одинъ опредѣленный Y . Во второмъ передъ нами *корреляціонная зависимость*: каждому значенію

¹⁾ Учебникъ математич. статистики, 1914, стр. 116. Также по существу изображается характеръ статистической зависимости и въ другихъ современныхъ руководствахъ,—но терминъ „корреляціонная“ зависимость не является общепринятымъ.

одной величины, фактора, или каждой совокупности значений всѣхъ факторовъ, существенныхъ для даннаго явленія, соответствуетъ цѣлый комплексъ значений зависимой величины, при чемъ общая мѣра по слѣдней—средняя арифметическая, измѣняется въ зависимости отъ значений фактора или факторовъ, и частота каждаго значенія или каждой комбинаціи значений есть функція этихъ значений ¹⁾. Обычные приемы опредѣленія зависимостей этого рода—разбивка рядовъ на части съ выводомъ среднихъ для этихъ частей, разгруппировка ряда предполагаемыхъ зависимыхъ величинъ по значенію того признака, которому приписывается значеніе вліяющаго фактора, съ выводомъ, опять-таки, частныхъ среднихъ, сравненіе расположенія членовъ сопоставляемыхъ рядовъ по величинѣ выраженныхъ въ нихъ величинъ или признаковъ. Всѣ эти приемы—говоритъ Р. М. Орженцкій ²⁾, „непримѣнимы въ томъ случаѣ, когда число строкъ ряда невелико (повидимому, слѣдуетъ читать „велико“); но и тогда, когда они примѣнимы, они не даютъ *мѣры* соотношенія рядовъ. Такую цѣлесообразно-выбранную мѣру соотношенія рядовъ даетъ приемъ исследования, называемый способомъ корреляціи“. Существенный смыслъ теоріи корреляціи, слѣдовательно, „установить не только *наличность* корреляціи двухъ или нѣсколькихъ признаковъ, но и *степень* этой корреляціи“ ³⁾.

И вотъ спрашивается, *въ какой мѣрѣ вычисленіе коэффициента корреляціи* и другихъ, существенно связанныхъ съ нимъ, корреляціонныхъ формулъ *можетъ замѣнить анализъ статистическихъ рядовъ элементарными приемами* и обратно—*въ какой мѣрѣ корреляціонныя формулы могутъ быть замѣнены элементарными приемами анализа?* иначе сказать: что именно прибавляетъ примѣненіе способа корреляціи къ той характеристикѣ корреляціонной зависимости, какую способенъ дать анализъ статистическихъ рядовъ элементарными приемами,—и что прибавляютъ элементарные приемы къ той характеристикѣ корреляціонной связи, какая получается при примѣненіи метода корреляціи? Насколько я могу судить, среди тѣхъ изъ представителей математическаго направленія въ статистикѣ, которые занимались разработкой метода корреляціонныхъ вычисленій, еще не выработалось вполне опредѣленнаго и однообразнаго отвѣта на этотъ вопросъ. Съ одной стороны мы встрѣчаемся съ такимъ осторожнымъ отношеніемъ къ вопросу, какое проявляетъ, напримѣръ, Udny Yule въ своемъ извѣстномъ руководствѣ по статистической теоріи: онъ настоятельно рекомендуетъ

¹⁾ Случкій, назв. соч., 59.

²⁾ Назв. соч., 117.

³⁾ Forcher, Die statist. Methode als selbstständige Wissenschaft. Leipz. 1913, стр. 345—346.

„быть осторожнымъ въ истолкованіи“ получаемыхъ коэффициентовъ корреляціи, — настоятельно совѣтуетъ имѣть въ виду, что какъ коэффициентъ корреляціи, такъ и родственныи ему по существу коэффициентъ соотвѣтствія (качественныхъ признаковъ!) „даётъ лишь часть тѣхъ свѣдѣній о явленіи (information), какія содержатся въ первоначальныхъ данныхъ или въ корреляціонной таблицѣ“¹⁾, и въ соотвѣтствіи съ этимъ рекомендуетъ, при малѣйшей къ тому возможности, не ограничиваться сообщеніемъ полученныхъ коэффициентовъ, а публиковать корреляціонную таблицу или первоначальныя цифры, — очевидно для того, чтобы дать интересующемуся возможность проанализировать ихъ и другими приёмами. „Вычисленъ ли коэффициентъ или нѣтъ — говоритъ Yule въ другомъ мѣстѣ²⁾, всякая таблица должна быть подвергнута тщательному разсмотрѣнію съ цѣлью выяснитъ, представляетъ ли она какія-либо видимо существенныя особенности въ распределеніи частотъ“, — „подобное детальное разсмотрѣніе часто даётъ возможность обнаружить такія существенныя обстоятельства (points), которыя иначе остались бы незамѣченными“. Въ соотвѣтствіи съ этимъ Yule и поступаетъ на слѣдующихъ страницахъ: онъ детально разбираетъ на одномъ примѣрѣ, помощью уложенныхъ имъ въ своеобразную, нѣсколько схоластическую схему, но по существу вполне элементарныхъ приёмовъ, построенную по двумъ признакамъ „корреляціонную“ или комбинаціонную таблицу; сопоставляя результатъ анализа данной таблицы съ результатами анализа другихъ аналогичныхъ таблицъ, онъ подмѣчаетъ нѣкоторую особенность, свойственную значительному большинству таблицъ „и потому требующую объясненія своего происхожденія“. „Будь данная таблица — говоритъ онъ по этому поводу — обработана однимъ лишь методомъ коэффициента соотвѣтствія или другимъ аналогичнымъ суммарнымъ методомъ, эта особенность могла бы остаться незамѣченной“³⁾. Гораздо радикальнѣе рѣшаетъ вопросъ авторъ извѣстнаго русскаго руководства по теоріи корреляціи, Е. Е. Слудкій: съ его точки зрѣнія изученіе „отдѣльныхъ вигзаговъ“ въ статистическихъ рядахъ не представляетъ интереса — „корреляціонная связь выражается только въ среднихъ величинахъ“; „при изслѣдованіи всякой зависимости всегда необходимо отвѣчаться

¹⁾ Yule, *Introduct. to the theory of stat.*, 2 изд. 1912 г., стр. 187—188.

²⁾ Стр. 67. Здѣсь идетъ рѣчь собственно не о коэффициентѣ корреляціи, а о коэффициентѣ соотвѣтствія (contingency) качественныхъ признаковъ, — но какъ по самому существу дѣла, такъ и изъ сопоставленія приведенныхъ словъ съ процитированнымъ выше отрывкомъ на стр. 187—188, ясно, что высказанныя здѣсь соображенія въ полной мѣрѣ относятся и къ коэффициенту корреляціи.

³⁾ Стр. 70—71.

отъ чертъ, присущихъ индивидуальному матеріалу, элиминируя случайныя отклоненія, затемняющія дѣйствіе общихъ тенденцій“¹⁾. Это положеніе проф. Слуцкий формулируетъ, какъ обобщеніе подробно разсмотрѣннаго имъ частнаго случая—извѣстнаго Пирсоновскаго примѣра зависимости роста сыновей отъ роста отцовъ, на которомъ мы еще придется остановиться въ дальнѣйшемъ изложеніи. Нанесенныя на графикъ цифры даютъ ясную линію тѣсной корреляціи съ рѣзкими, однако, отклоненіями въ началѣ и въ концѣ эмпирической ломаной линіи. Проф. Слуцкий констатируетъ, что выраженный данною ломаною линіею законъ передачи роста „на первый взглядъ не отличается простотой: слѣдуя за всѣми зигзагами линіи регрессіи“, приходится отмѣтить, „что въ группѣ наиболѣе низкорослыхъ отцовъ увеличенію роста на 1 дм. соотвѣтствуетъ въ среднемъ увеличеніе роста сыновей также на 1 дм.“; что затѣмъ слѣдуетъ „аномальный интервалъ, въ которомъ увеличенію роста отца соотвѣтствуетъ уменьшеніе средняго роста сыновей“; хотя и менѣе рѣзкія уклоненія отъ строгаго соотвѣтствія наблюдаются и на дальнѣйшихъ интервалахъ. Г. Слуцкий констатируетъ всѣ эти аномаліи, а затѣмъ съ нѣкоторымъ оттенкомъ ироніи продолжаетъ: „изъ анализа линіи регрессіи мы могли бы получить еще много такихъ „законовъ“, которыхъ я не привожу здѣсь вовсе не потому, что имъ мѣсто не въ нашей работѣ, а въ трактатѣ по теоріи наследственности. Последней съ ними тоже нечего дѣлать, ибо—какъ (сейчасъ же замѣчу, ошибочно) полагаетъ Е. В. Слуцкий—они вовсе не существуютъ: если бы мы измѣрили другую тысячу паръ отцовъ-сыновей, то можно утверждать съ полною увѣренностью—въ этомъ новомъ матеріалѣ не осталось-бы и слѣда отъ нашихъ многихъ законовъ. Общее направленіе линіи регрессіи осталось бы прежнимъ, но отдѣльные зигзаги расположились бы, можетъ быть, совершенно иначе. Крайніе зигзаги линіи регрессіи — полагаетъ г. Слуцкий²⁾—„легко находятъ себѣ объясненіе въ малочисленности соотвѣтствующихъ случаевъ“, а потому „не нужно быть большимъ теоретикомъ, чтобы сразу почувствовать весьма малую достовѣрность соотвѣтствующихъ деталей линіи регрессіи“. И отсюда уже процитированное обобщеніе: „коррелятивная связь выражается только въ среднихъ величинахъ“, а потому „при изслѣдованіи зависимости необходимо умѣть отвѣчаться отъ чертъ, присущихъ индивидуальному матеріалу“, —„элиминировать случайныя отклоненія, затемняющія дѣйствіе общихъ тенденцій“. Ниже я надѣюсь показать, что какъ разъ данный примѣръ менѣе всего можетъ обосновать вторично процитированныя только что положенія.

¹⁾ Слуцкий, назв. соч., 68.

²⁾ Тамъ-же.

Сейчасъ я отмѣчу только одно—что и самъ Е. Е. Слуцкій не выдерживаетъ столь категорически формулированной имъ точки зрѣнія: нѣсколькими страницами дальше ¹⁾ онъ и самъ возвращается къ данному вопросу и констатируетъ наличность существенныхъ различій въ соотношеніи роста отцовъ и сыновей въ высокорослыхъ и въ низкорослыхъ группахъ. Не вполне, мнѣ кажется, выдержана точка зрѣнія на интересующій насъ вопросъ и у другого виднаго представителя математической статистики въ Россіи—проф. Орженцкаго. Въ предисловіи къ редактированной имъ весьма интересной работѣ М. Б. Гуревича: „Примѣненіе нѣкоторыхъ приѣмовъ математической статистики“ ²⁾ онъ рѣшительно подчеркиваетъ, что „примѣненіе математическихъ приѣмовъ (въ томъ числѣ, значить, и корреляціонныхъ вычисленій, нашедшихъ себѣ примѣненіе какъ разъ на первыхъ страницахъ работы г. Гуревича); не устраняетъ необходимости тщательнаго качественного изученія матеріала и знакомства съ природой изучаемыхъ явленій“,—едва ли подѣ „качественнымъ изученіемъ“ разумѣется что-либо иное, нежели обработка статистическихъ рядовъ элементарными приѣмами. Между тѣмъ въ „Учебникѣ математической статистики“ проф. Орженцкаго элементарные приѣмы, можно сказать, совершенно игнорируются: характеристикѣ ихъ посвящено лишь нѣсколько строкъ, за которыми слѣдуетъ уже процитированное выше указаніе на непримѣнимость элементарнаго приѣма при небольшомъ числѣ строкъ и на предпочтительность способа корреляціи, какъ приѣма, дающаго цѣлесообразно-выбранную мѣру соотношенія рядовъ. Если же обратиться къ узко-спеціальной по своему предмету, но чрезвычайно цѣнной по своему обще-методологическому значенію работѣ Р. М. Орженцкаго: „Урожай ржи на крестьянскихъ земляхъ Ярославской губерніи“ ³⁾, то здѣсь все изслѣдованіе основывается исключительно на вычисленіи коэффиціентовъ корреляціи и на нѣкоторыхъ другихъ приѣмахъ математической статистики; къ элементарнымъ приѣмамъ проф. Орженцкій абсолютно не прибѣгаетъ; разложеніе рядовъ на части у него хотя и встрѣчается, но исключительно въ цѣляхъ вычисленія, опять-таки, коэффиціентовъ корреляціи между выхитрымъ по медіанамъ интерваловъ факторіальнымъ признакомъ и зависящимъ отъ него функциональнымъ. Приходится, какъ будто, обѣдать виномъ, что въ главахъ проф. Орженцкаго вычисленіе коэффиціента корреляціи и другихъ корреляціонныхъ формулъ факти-

¹⁾ Стр. 86.

²⁾ Полное заглавіе: М. Б. Гуревичъ. Примѣненіе нѣкоторыхъ приѣмовъ математической статистики. Оцѣнка недвижимыхъ имуществъ г. Ростова, Ярославской губ. Выпускъ первый. Изд. стат. отд. Яросл. губ. земства. Ярославль 1912.

³⁾ Стат. отдѣлъ Яросл. губ. земства, вып. 86. 1912 г.

чески совершенно устраняетъ всякую необходимость въ дополнительномъ анализѣ статистическихъ рядовъ элементарными приемами.

Между тѣмъ, уже по чисто-априорнымъ соображеніямъ позволительно усомниться въ томъ, чтобы корреляціонныя вычисленія давали достаточно полное и всестороннее представленіе о характерѣ зависимости между статистическими рядами,—чтобы они могли совершенно устранить необходимость въ анализѣ изучаемыхъ рядовъ при помощи элементарныхъ приемовъ. Мало того: априорныя же соображенія позволяютъ усомниться въ томъ, чтобы коэффициентъ корреляціи и соотносительныя съ нимъ формулы, по самому существу своему, всегда могли давать болѣе точное и опредѣленное, притомъ болѣе объективно-обоснованное и болѣе независимое отъ субъективной оцѣнки представленіе о характерѣ соотношеній между статистическими рядами, нежели элементарные приемы.

Въ самомъ дѣлѣ: вѣдь ни коэффициентъ корреляціи, ни зависящие отъ него коэффициенты регрессіи не имѣютъ, сами по себѣ, рѣшающаго значенія—они должны быть разсматриваемы въ связи съ ихъ средними или, что по существу дѣла безразлично, съ ихъ вѣроятными ошибками. Между тѣмъ, это обстоятельство, какъ легко видѣть, отнимаетъ у коэффициента корреляціи, у коэффициентовъ регрессіи, а слѣдовательно и у тѣхъ соотношеній, которыя находятъ себѣ выраженіе въ уравненіяхъ регрессіи, значительную долю точности и объективной убѣдительности. Прежде всего—предѣлы убѣдительности коэффициента корреляціи и коэффициентовъ регрессіи зависятъ вѣдь даже не непосредственно отъ средняго отклоненія или вѣроятной ошибки—величинъ, какъ никакъ, поддающихся точному вычисленію и не зависящихъ въ своемъ количественномъ выраженіи отъ личнаго взгляда исследователя,—а отъ *кратнаго* вычисленной величины. Но *какого* кратнаго? Здѣсь сейчасъ же открывается широкій просторъ для личнаго взгляда и произвола. По Случайному для достовѣрности вывода необходимо, чтобы найденная величина „по крайней мѣрѣ въ пять разъ“ превышала свою вѣроятную ошибку“ ¹⁾. По *крайней мѣрѣ* въ пять? а можетъ быть и въ шесть? Можетъ быть, въ самомъ дѣлѣ, и въ шесть разъ—потому что въ другомъ мѣстѣ ²⁾ тотъ же проф. Случайскій требуетъ, чтобы разница сравниваемыхъ величинъ превышала „по крайней мѣрѣ унятеренную или ушестеренную ошибку разности“. Но вѣдь будемъ ли мы исходить изъ унятеренной или изъ ушестеренной вѣроятной ошибки, а можетъ быть изъ еще ~~большой~~ ея кратнаго („по крайней мѣрѣ“) —это, въ известномъ случаѣ, ~~с~~

1) Случайскій, назв. соч., 97.

2) То же, 13.

мымъ рѣшительнымъ образомъ отразится на степени того довѣрія, съ какимъ мы отнесемся къ полученнымъ коэффициентамъ. У другихъ авторовъ, напр., у Udny Yul'я или у Р. М. Орженцкаго, мы находимъ, на первый взглядъ, болѣе опредѣленные указанія: критеріемъ у нихъ является, безъ всякихъ оговорокъ, *утроенное* среднее отклоненіе. Читатель, конечно, помнитъ, почему именно *утроенное*—потому, что утроенное отклоненіе даетъ вѣроятность въ 0,999979, значить, вѣроятность чрезвычайно близкую къ достовѣрности, что разница цифръ имѣетъ не случайный, а существенный характеръ. Но вѣдь требовать *непременно такой*, чрезвычайно высокой, степени вѣроятности, *такой* степени „практической достовѣрности“ результата—дѣло чисто субъективнаго усмотрѣнія. Если принять въ соображеніе малую степень точности всѣхъ вообще данныхъ, съ какимъ оперируетъ социальная статистика, и приближенный, суммарный характеръ всѣхъ ея выводовъ, то легко будетъ удовольствоваться вѣроятностью 0,9996, которая соотвѣтствуетъ взятому *два съ половиной* раза среднему отклоненію, или даже вѣроятностью 0,995, соотвѣтствующую *удвоенному* отклоненію. А между тѣмъ, опять-таки, разъ мы будемъ исходить изъ требованія не утроеннаго, а помноженнаго на 2½ или удвоеннаго отклоненія,—это позволитъ намъ придавать рѣшающее значеніе гораздо болѣе низкимъ коэффициентамъ корреляціи, позволить довѣрять данному невысокому коэффициенту при значительно меньшемъ числѣ членовъ изучаемыхъ въ ихъ взаимоотношеніи статистическихъ рядовъ.

Но и независимо отъ этого: кажущаяся опредѣленность результатовъ, къ какимъ приводитъ примѣненіе корреляціонныхъ вычисленій, умалется уже въ виду сложности тѣхъ элементовъ, отъ которыхъ зависитъ даже непосредственно-вычисляемая величина средней либо вѣроятной ошибки. Это прежде всего, какъ ясно видно изъ формулы $\Sigma = \delta \sqrt{1 - r^2}$, величина самаго коэффициента корреляціи: если она незначительна, средняя ошибка уравненія регрессіи мало отличается отъ средняго отклоненія даннаго ряда,—„слѣдовательно предсказаніе, которое мы можемъ сдѣлать по данной величинѣ X относительно Y, должно быть весьма несовершеннымъ“¹⁾, а самое значеніе коэффициента регрессіи—весьма ограниченнымъ и условнымъ. Напротивъ, при величинѣ коэффициента корреляціи, приближающейся къ единицѣ, средняя ошибка будетъ весьма незначительна, „при данной величинѣ одного перемѣннаго распредѣленіе величинъ другого будетъ весьма сжатымъ“, а значеніе коэффициента регрессіи сдѣлается весьма опредѣленнымъ и недопускающимъ сомнѣнія. Съ другой стороны—что еще важнѣе—при каждой данной величинѣ коэффициента корреляціи,

1) Слупскій, назв. соч., стр. 82.

величина средней ошибки, а следовательно значение формулы регрессии и самого коэффициента корреляции, будет зависеть и от второго множителя—среднего квадратического отклонения ряда, которое, в свою очередь, зависит не только от непосредственной колеблемости чисел, но также и от числа членов ряда,—при прочих равных условиях среднее квадратическое отклонение ряда, а с ним и средняя квадратическая ошибка формулы регрессии, будет тем больше, чем короче ряд. Поэтому, „чем больше корреляция, тем меньше может быть число случаев“—иначе говоря, число членов ряда, „достаточное для того, чтобы с несомненностью определить наличие корреляционной связи и ее величину“¹⁾; и обратно, чем число членов ряда больше, тем меньшая величина коэффициента корреляции дает право на достоверное заключение относительно наличия корреляционной связи. Отсюда прямой вывод—что каждая данная величина коэффициента корреляции свидетельствует о наличии корреляционной связи и дает ее меру лишь при соответственном ей минимальном числе членов ряда, и что при числе членов, не достигающем некоторого, так сказать, абсолютного минимума, определение коэффициента корреляции вообще не может дать убедительных результатов. *Камия* числа необходимы для твердого обоснования каждой данной величины коэффициента корреляции, и где лежит абсолютный минимум,—для решения этого вопроса, сколько мы известно, не существует объективных оснований. Все дело сводится, употребляя выражение Е. Е. Слупцаго, к „практическим правилам“; вопрос, иначе сказать, решается в зависимости от субъективного взгляда исследователя. Сам проф. Слупский предлагает руководствоваться таким „практическим правилом“, что „при N между 20 и 30 коэффициент корреляции должен быть не менее 0,5, чтобы о самом существовании корреляционной связи можно было говорить с некоторой уверенностью“: как показали делавшиеся для выяснения данного вопроса опыты, при $N=21$ и при фактическом отсутствии корреляции коэффициент корреляции может случайно оказаться большим $+0,5$ и меньшим $-0,5$, и эта случайность может иметь место в двух случаях из ста. В конце концов проф. Слупский с сожалением признает, что „до дальнейшего усовершенствования теории статистики не следует применять метод корреляции к группам меньше, чем из 20 случаев“²⁾. Таким же чисто „практическим“, т.е. более или менее произвольным способом решает вопрос и Удну-Йуле: „Если говорить о нем—мы напишем на карточ-

1) Там же, 97.

2) Там же, 97—98.

кахъ серію строго независимыхъ значеній X и Y и затѣмъ вычислимъ коэффициенты корреляціи для группъ, положимъ, по 40 или 50 карточекъ, взятыхъ наудачу, мы едва ли когда либо найдемъ точно $r=0$, но получимъ рядъ положительныхъ и отрицательныхъ чиселъ, колеблющихся около нуля. Нельзя поэтому, продолжаетъ онъ, — разъ только число наблюдений незначительно, возлагать большого довѣрія на малыя и даже на умѣренно большія значенія r и видѣть въ нихъ указанія на дѣйствительное существованіе корреляціи. Такъ, напримѣръ, если $N=36$, значеніе $r=0,5$ можетъ быть чисто случайнымъ, хотя и рѣдкимъ результатомъ; равнымъ образомъ, при $N=100$, $r=0,3$ можетъ быть результатомъ случайностей отбора, хотя также нечасто встрѣчающимся. Столь же рѣдкою, но возможною случайностью отбора можетъ оказаться $r=0,1$ при $N=900$ ¹⁾, и это главное основаніе, почему исследователь „долженъ быть остороженъ въ истолкованіи получающихся у него коэффициентовъ“¹⁾.

Едва ли, послѣ сказаннаго, подлежитъ сомнѣнію, что коэффициенты корреляціи и прочія соотносительныя съ ними формулы не обладаютъ тою степенью точности и, главное, объективности, независимости отъ субъективнаго взгляда исследователя, какая имъ приписывается: и при пользованіи методомъ корреляціонныхъ вычисленій остается широкій просторъ для „практическихъ“ правилъ оцѣнки получаемыхъ результатовъ, иначе сказать для субъективнаго взгляда и усмотрѣнія, — едва ли менѣе широкій, чѣмъ при анализѣ статистическихъ рядовъ элементарными приемами.

Изъ сказаннаго вытекаетъ или, можетъ быть, лучше сказать, сказаннымъ объясняется еще нѣчто другое: недостаточность корреляціонныхъ формулъ для характеристики тѣхъ безконечно разнообразныхъ степеней и проявленій корреляціонной зависимости, какія могутъ существовать между даваемыми эмпирическою дѣйствительностью парами или вообще серіями статистическихъ рядовъ. Разсуждая абстрактно, степень корреляціонной зависимости можетъ быть выражена всѣми мыслимыми дробями между 0 и ± 1 (въ одну сторону) и 0 и -1 въ другую: методъ корреляціи располагаетъ, следовательно, безконечнымъ разнообразіемъ численныхъ обозначеній, достаточнымъ, чтобы уловить всѣ мыслимыя оцѣнки корреляціонной зависимости. Въ дѣйствительности, однако, дѣло обстоитъ не такъ. Прежде всего, изъ изложеннаго выше ясно, что малыя значенія коэффициента корреляціи — такіа, какъ 0,1 при любой длинѣ сопоставляемыхъ рядовъ, какъ 0,2 и 0,3 при небольшомъ числѣ членовъ въ этихъ рядахъ — не даютъ, вообще, права на какой бы то ни было выводъ относительно наличности или

¹⁾ Назв. соч. 187.

отсутствия корреляционной зависимости; — фактически, значить, въ распоряженіи исследователя остаются только дроби, лежащія: при очень длинныхъ рядахъ въ промежуткѣ примѣрно между 0,2 и 1, въ болѣе обычныхъ случаяхъ — между 0,3, 0,4, даже 0,5 и единицею. Далѣе: разъ при оцѣнкѣ полученнаго изъ вычисленія коэффициента нельзя исходить непосредственно изъ полученнаго численнаго значенія, а приходится сообразоваться съ числомъ наблюдений и съ вѣроятною ошибкой, вообще руководствоваться „практическими правилами“, являющимися въ значительной степени синонимомъ субъективнаго взгляда исследователя, то это не можетъ не отразиться, такъ сказать, на степени чувствительности корреляціоннаго аппарата: если вычисленія въ одномъ случаѣ дали коэффициентъ, положимъ, 0,57, а въ другомъ 0,54, то изъ этого, при охарактеризованномъ только-что положеніи вещей, отнюдь не слѣдуетъ, и никто практически не сдѣлаетъ вывода, чтобы корреляціонная зависимость въ первомъ случаѣ была дѣйствительно сильнѣе, и именно въ такой степени сильнѣе, нежели во второмъ: выраженная въ сотыхъ доляхъ разница легко можетъ быть результатомъ случайныхъ колебаній, дѣйствительная же корреляціонная зависимость въ первомъ случаѣ можетъ быть и такою же, и болѣе слабою, нежели во второмъ. Говоря вообще: разница во второмъ десятичномъ знакѣ практически лишена всякаго значенія. Да и не говоря уже о сотыхъ доляхъ — даже разница, выраженная въ цѣлой десятой доли, — и та, при извѣстной, достаточно обычной комбинаціи условій — единичныхъ рядахъ и большой средней либо вѣроятной ошибкѣ, легко можетъ болѣе или менѣе цѣликомъ свестись къ вліянію случайныхъ колебаній; значить, и такая разница далеко не всегда можетъ быть принята за несомнѣнное доказательство и тѣмъ болѣе — за точную мѣру разницы въ степени дѣйствительной корреляціонной зависимости. Пусть коэффициентъ корреляціи для одной пары рядовъ получился 0,47, а для другой 0,58, или просто для одной 0,5, для другой 0,6 — это не рѣшаетъ безусловно даже вопроса о томъ, дѣйствительно ли корреляціонная зависимость во второй парѣ сильнѣе, нежели въ первой, не говоря уже о томъ, что разница въ дѣйствительной степени этой зависимости можетъ быть и гораздо болѣею, и гораздо меньшею, нежели соответствовало бы полученнымъ изъ вычисленія цифрамъ. Такимъ образомъ, оказывается, что столь точный и чувствительный — по видимости — корреляціонный аппаратъ въ дѣйствительности предоставляет въ распоряженіе исследователя лишь очень грубую, въ смыслѣ числа, могущихъ быть использованными дѣленій, и притомъ весьма расплывчатую, въ смыслѣ точности получаемыхъ доказаній, шкалу: видимыхъ дѣленій на этой шкалѣ, правда, бесконечное множество, — но только тѣ различія получающихся изъ вычисленія

значений коэффициента корреляции, которые отражаются на первом десятизначном знаке, дают действительное, да и то не всегда безспорное право заключать о наличии действительных различий в степени корреляционной зависимости. В концѣ-концовъ 0,2 или 0,3—корреляционная зависимость весьма слабая или даже сомнительная; 0,4 или 0,5—зависимость не слишкомъ ясно выраженная; 0,6, 0,7, 0,8—разныя степени болѣе или менѣе сильной корреляционной зависимости; 0,9 съ болѣе или менѣе значительными сотыми долями—зависимость, приближающаяся къ строгой функциональной; вотъ и все разнообразіе показаній корреляціоннаго аппарата, съ которыми приходится считаться при заключеніяхъ относительно наличия и степени корреляционной зависимости; всѣ остальные, болѣе дробныя различія получающихся изъ вычисления коэффициентовъ имѣютъ лишь численное значеніе и не могутъ обосновывать сколько-нибудь достовѣрныхъ заключеній.

Можно ли, при такихъ условіяхъ, ожидать, чтобы та „общая мѣра“, какою является коэффициентъ корреляции и остальные соотносительныя съ нимъ формулы, была достаточна для характеристики степеней и формъ существующей между явленіями, во всей ихъ безконечной жизненной сложности и разнообразіи, корреляционной зависимости? чтобы эта на первый взглядъ столь точная и чувствительная, въ действительности же столь грубая „общая мѣра“ могла дать болѣе въ данномъ направленіи, нежели могутъ дать все-же болѣе тонкіе, все-же болѣе разнообразные элементарные приемы? мало того—чтобы она была способна хотя бы съ тою же степенью ясности и отчетливости охарактеризовать взаимоотношеніе между статистическими рядами, а слѣдовательно и между изображаемыми ими явленіями, съ какою это достигается „глазомѣрными“ приемами „общей статистики“? Отвѣтъ на этотъ вопросъ едва ли можетъ быть благоприятнымъ для метода корреляціонныхъ вычисленій.

Попробуемъ, однако, проверить это апіорное заключеніе. Возьмемъ сначала простѣйшіе случаи—именно, тѣ схематическіе примѣры, на которыхъ проф. Оржанцкій показываетъ приемъ вычисленія коэффициента корреляции¹⁾. Первый примѣръ—ряды 1, 2, 3, 4, 5 и 2, 4, 6, 8, 10—коэффициентъ корреляции 1 (у проф. Оржанцкаго 0,994); второй—ряды 1, 2, 3, 4, 5 и 1, 3, 3,5, 5,5, 6,—коэффициентъ корреляции 0,55. Нужны ли, спрашивается, какія бы то ни было вычисленія, вообще какіе бы то ни было искусственныя приемы для того, чтобы увидѣть, что въ первомъ случаѣ мы имѣемъ дѣло со строгою пропорциональностью; что и во второмъ имѣется налицо совершенно

¹⁾ Учебникъ 123.

отчетливо выраженная, но въ то же время гораздо менѣе строгая прямая зависимость? и не достаточно ли бѣлаго взгляда на сопоставляемые ряды, чтобы убѣдиться въ томъ, что во второмъ случаѣ коэффициентъ корреляціи не даетъ никакого представленія о прерывистомъ или волнообразномъ характерѣ зависимости между заданными рядами?

Перейдемъ отъ схематическихъ примѣровъ къ реальнымъ. Беремъ наудачу одну изъ табличекъ изъ уже цитированной работы проф. Орженцкаго „Урожай ржи“ ¹⁾—именно таблицу 4-ю. Даны ряды: одинъ (групповыя среднія) 3,2 — 3,6 — 4,1 — 4,6 — 5,0 — 5,3 — 5,8 — 6,9, другой (медианы групповыхъ привнаковъ) 2,7 — 3,2 — 3,7 — 4,2 — 4,7 — 5,2 — 5,7 — 6,2. Вычисляя коэффициентъ корреляціи, проф. Орженцкій получаетъ = 0,987; не довольствуясь этимъ, онъ выводитъ формулу уравненія регрессіи и по этой формулѣ вычисляетъ теоретическія значенія для второго ряда: 2,8 — 3,2 — 3,7 — 4,2 — 4,7 — 5,0 — 5,5 — 6,6. Какъ и слѣдовало ожидать, эти вычисленные значенія „весьма близко совпадаютъ съ действительными“. Но проф. Орженцкій совершенно справедливо замѣчаетъ, что „уже на глазъ, безъ вычисленій, видно, что между данными рядами существуетъ полная корреляція“. А разъ она сама собой видна—и она *действительно* видна!—то зачѣмъ, спрашивается, нужно вычислять теоретическій рядъ, и что прибавляетъ коэффициентъ корреляціи 0,987 къ нашему уже „на глазъ“ получающемуся убѣжденію въ наличности полной корреляціи? Или возьмемъ болѣе сложные случаи—серію „корреляціонныхъ“ таблицъ, помѣщенныхъ у Uduy Yul'я на стр. 158—163. Всѣхъ таблицъ шесть—для первыхъ четырехъ изъ нихъ я вычислилъ $r = 0,94, 0,79, 0,49$ и $0,18$,—для остальныхъ двухъ отсутствіе корреляціи слишкомъ очевидно. Я прошу теперь читателя раскрыть книгу Yul'я. Беремъ первую таблицу,—и здѣсь сразу-же, „на глазъ, безъ всякихъ вычисленій“, обнаруживается вполне соответствующая высокому коэффициенту корреляціи, чрезвычайно тѣсная зависимость между сопоставляемыми биологическими признаками: цифры располагаются строго по діагонали, при чемъ числа въ каждомъ интервалѣ, взятомъ по значенію X'a, располагаются не болѣе, чѣмъ въ трехъ интервалахъ по значенію Y-a, главною своею массою сгучиваясь въ одномъ изъ интерваловъ; и чего коэффициентъ корреляціи уже не показываетъ,—въ низшихъ по значенію X'a и Y'a группахъ господствующій интервалъ—это тотъ, гдѣ встрѣчаются одинаковыя значенія того и другого, въ высшихъ—тотъ, гдѣ данному значенію X'a соответствуетъ слѣдующее по порядку высшее значеніе Y'a. Ясаую прямую корреляціонную зависимость, но, въ соответствіи

¹⁾ Стр. 13.

съ существенно меньшимъ коэффициентомъ корреляціи, значительно меньше тѣсную, обнаруживаетъ и таблица 2-я и здѣсь всѣ цифры располагаются по діагонали въ направленіи, соответствующемъ положительной зависимости, и по каждому значенію X'a соответствуетъ уже не два или три, а отъ восьми до десяти значеній Y'a, съ явнымъ въ каждой группѣ преобладаніемъ двухъ или трехъ ближайшихъ къ общему направленію діагонали значеній. Еще значительно слабѣе, опять-таки въ полномъ соответствіи съ значительно болѣе низкимъ коэффициентомъ корреляціи, проявляется уловимая на глазъ зависимость въ таблицѣ 3-й: цифры въ ней разбросаны почти по всей таблицѣ и тяготѣніе ихъ къ „положительной“—если такъ можно выразиться—діагонали выражено очень слабо; если, однако, отбросить крайнія, очень малочисленные группы; именно четыре низшихъ и двѣ высшихъ по росту отца, то въ остальныхъ нетрудно подмѣтить тяготѣніе главной массы цифръ къ небольшому числу, по большей части къ четыремъ или къ пяти смежно лежащимъ интерваламъ. Легко замѣтить притомъ, что эти преобладающіе интервалы составляютъ въ совокупности сплошную полосу съ наклономъ, нѣсколько отстающимъ отъ наклона основной діагонали; иначе сказать—что нарастаніе помѣщенного въ боковомъ заголовкѣ признака, роста сына, нѣсколько отстаетъ отъ нарастанія, помѣщенного въ верхнемъ заголовкѣ—роста отца. Еще слабѣе—пока мы исходимъ, напоминая, изъ простого „глазomѣра“—уловляется на глазъ зависимость въ 4-й таблицѣ. Однако, два лежащія въ сторонѣ отъ „положительной“ діагонали угла таблицы и здѣсь остаются свободными отъ цифръ, которыя замѣтно скучиваются къ серединѣ и отъ нея—къ лѣвому верхнему углу таблицы, что, очевидно, соответствуетъ, хотя и очень слабо выраженному, положительному діагональному направленію преобладающихъ интерваловъ. Наклонъ полосы преобладающихъ интерваловъ и здѣсь видимо отстаетъ отъ наклона общей діагонали таблицы;—иначе сказать, возрастаніе помѣщенного въ боковомъ заголовкѣ признака, числа дѣтей у дочерей, отстаетъ отъ роста признака, помѣщенного въ верхнемъ заголовкѣ—числа дѣтей у матерей. Въ пятой таблицѣ цифры располагаются весьма своеобразно: въ совокупности своей онѣ образуютъ полосу, приблизительно параллельную „отрицательной“ діагонали, слѣдовательно, обнаруживаютъ—какъ и слѣдовало ожидать, обратную зависимость между золотыми запасами банка и уровнемъ учетнаго процента. Но полоса эта значительно приподнята надъ общею діагональю таблицы, такъ что уже не много повышенному уровню золотого запаса соответствуетъ замѣтное пониженіе учетнаго процента, относительно небольшому пониженію перваго—весьма рѣзкое повышеніе второго. Наконецъ, одного взгляда на таблицу № 6 достаточно, чтобы убѣдиться въ совершенномъ от-

существовании корреляционной связи между сопоставляемыми адъёс признаками—распределением округов по общему числу рождений и долей мальчиковъ въ общемъ числѣ рождений: какъ среди самыхъ большихъ, такъ и среди самыхъ мелкихъ, одинаково преобладаютъ двѣ группы округовъ—съ 507—509 и съ 510—512 мальчиками на каждую тысячу рождений, и за ними непосредственно слѣдуютъ прилегающія къ нимъ сверху и снизу группы съ 504—506 и съ 513—515 мужскими на каждую тысячу общаго числа рождений,—величина же округовъ отражается только на *разстоянн*, дисперсии цифръ: въ болѣе крупныхъ по общему числу рождений группахъ округовъ господствующая тенденція получаетъ полное проявленіе, и въ соответствии съ этимъ доля мужскихъ рождений нигдѣ не падаетъ ниже 504 и нигдѣ не поднимается выше 512 на тысячу. Начиная отъ группы общинъ съ общимъ числомъ рождений въ 44—48 тысячъ, доля мужскихъ рождений начинаетъ выступать то въ ту, то въ другую сторону за указанные нормальные предѣлы, въ группахъ же мальчайшихъ округовъ, съ числомъ рождений не превышающимъ четырехъ тысячъ, доля мужскихъ рождений колеблется въ такихъ широкихъ предѣлахъ, какъ 465 на тысячу съ одной и 545 на тысячу съ другой стороны.

Мы видимъ, такимъ образомъ, что во многихъ случаяхъ и простого сопоставленія рядовъ, и болѣе сложной корреляціонной группировки какъ наличность вообще, такъ и степень, а въ значительной мѣрѣ и самый характеръ корреляціонной зависимости съ достаточною ясностью видны уже „на глазъ, безъ вычисленій“, и для выясненія этой зависимости нѣтъ надобности не только въ вычисленіи коэффициента корреляціи, но даже въ примѣненіи какихъ-либо элементарныхъ приемовъ обработки рядовъ: съ нѣкоторымъ вниманіемъ всматриваясь въ таблицу, мы получаемъ отнюдь не менѣе ясное представленіе о силѣ и о характерѣ корреляціонной зависимости, нежели какое даютъ намъ корреляціонныя формулы. Само собою ясно, однако, что такое просто „глазomѣрное“ выясненіе вопроса о наличности и характерѣ корреляціонной зависимости далеко не всегда можетъ привести къ удовлетворительному результату: простой глазомѣръ окажется недостаточнымъ во всѣхъ тѣхъ случаяхъ, гдѣ приходится имѣть дѣло съ болѣе длинными и потому трудно обозрѣваемыми рядами, равно какъ и въ тѣхъ, когда основная тенденція одного или обоихъ сопоставляемыхъ рядовъ маскируется неправильными колебаніями отдельныхъ чиселъ или группъ чиселъ. И въ этихъ случаяхъ, однако, элементарные приемы обработки рядовъ могутъ дать не менѣе, а иногда и болѣе опредѣленный и отчетливый результатъ, нежели корреляціонныя вычисленія. Само собою разумѣется, съ другой стороны, что и въ случаяхъ, подобныхъ только что рассмотрѣннымъ, они способствуютъ

значительному уточненію полученныхъ изъ „глазомера“ представлений.

Разсмотримъ опять простѣйшій схематическій примѣръ, который заимствуемъ, на этотъ разъ, у Yul'я. Даны два ряда: 1, 2, 3, 4, 5 и 2, 5, 3, 8, 7. Коэффициентъ корреляціи, вычисленный съ однимъ десятичнымъ знакомъ, равенъ 0,8—свидѣтельствуетъ, слѣдовательно, о достаточно тѣсной корреляціонной зависимости. Тотъ же результатъ получается, однако, и путемъ элементарнаго приѣма: беремъ группы по три (при чемъ третьи члены входятъ и въ первую, и во вторую группы), получаемъ групповыя среднія 2—4 и 3,3—6; или беремъ пары цифръ (при чемъ цифры отъ второй до четвертой входятъ въ двѣ послѣдовательныя пары),—получаемъ среднія для паръ 1,5—2,5—3,5—4,3 и 3,5—4—5,5—7,5; тѣсная прямая зависимость цифръ и въ томъ и въ другомъ случаѣ выявляется съ совершенною отчетливостью, а группировка по парамъ показываетъ, кромѣ того, что характеръ этой зависимости въ нижней части сопоставляемыхъ рядовъ нѣсколько иной, нежели въ верхней.

Переходимъ теперь къ примѣрамъ конкретнаго характера. Беремъ одинъ изъ нихъ на стр. 10 многократно цитированной работы проф. Орженцкаго „Урожай ржи“: два ряда цифръ, по 16 членовъ въ каждомъ, изображающихъ поволостныя цифры урожайности по двумъ различнымъ источникамъ. Вычисленіе коэффициента корреляціи даетъ $r=0,37$. Величина и знакъ этого коэффициента—говоритъ проф. Орженцкій—„указываютъ на то, что между рядами существуетъ зависимость, но въ малой степени“, а слѣдовательно „совпаденіе или несовпаденіе характеристикъ отдѣльныхъ волостей является въ данномъ случаѣ болѣе или менѣе случайнымъ“¹⁾. Что случай дѣйствительно играетъ здѣсь видную роль,—въ этомъ легко убѣдиться и путемъ элементарнаго приѣма: располагаемъ волости въ порядкѣ возрастающей величины членовъ перваго изъ сопоставляемыхъ рядовъ и опредѣляемъ мѣста ихъ по величинѣ членовъ второго ряда,—получаемъ такіе два ряда нумеровъ, разбитыхъ каждый на четыре соответственно равныя части:

1—2—3—4 5—6—7—8 9—10—11—12 13—14—15—16

1—4—2—14 11—10—3—8 12—6—16—7 9—12—5—15

Только первыя, низшія по величинѣ урожая группы совпадаютъ въ большинствѣ членовъ,—значитъ, волости наименѣе урожайныя по одному источнику являются таковыми же по другому; въ остальныхъ группахъ нельзя уловить никакого видимаго соотвѣтствія между цифрами, и, такимъ образомъ, колебанія поволостныхъ цифръ дѣйствительно

¹⁾ „Урожай ржи“, 11.

являются болѣе или менѣе случайными. Не устраняется вліяніе случая и при выводѣ групповыхъ среднихъ—для перваго ряда получаютъ среднія 4,4—4,75—5,25—5,75, для втораго 4,77—5,05—6,02—5,5; въ первыхъ трехъ группахъ повышение средней въ первомъ рядѣ сопровождается повышеніемъ ея и во второмъ, и, слѣдовательно, сквозъ случайныя колебанія отдѣльныхъ поволостныхъ цифръ выступаетъ общая тенденція приблизительнаго соотвѣтствія показаній того и другаго источника,—въ четвертой группѣ повышенная средняя въ первомъ рядѣ сопровождается пониженною во второмъ,—слѣдовательно, здѣсь даже и въ средней общая тенденція соотвѣтствія цифръ оказывается заслоненною случайными колебаніями.

Беремъ еще одинъ примѣръ изъ той же работы—поволостныя же данныя объ урожаяхъ по показаніямъ старостъ и по мѣстному изслѣдованію, приведенныя и разобранныя на стр. 25—26. Вычисленіе коэффициента корреляціи даетъ $r=0,82$, слѣдовательно обнаруживаетъ „весьма значительное совпаденіе“. Значительная степень соотвѣтствія обнаруживается и при примѣненіи элементарныхъ приѣмовъ, при чемъ, однако, выясняются нѣкоторыя не лишеныя значенія частности. Располагая волости по величинѣ членовъ перваго изъ сопоставляемыхъ рядовъ и подписывая порядковые нумера по величинѣ соотвѣтственныхъ членовъ втораго ряда, получаемъ такіе два ряда нумеровъ, разбитыхъ каждый на три части:

1—2—3—4—5—6 7—8—9—10—11—12 13—14—15—16—17—18—19
8—10—3—2—5—1 12—9—4—6—13—15 11—17—7—14—16—18—19

Мы видимъ, такимъ образомъ, что волости съ высшею урожайностью по одному источнику всѣ, кромѣ одной, оказываются таковыми же и по второму; нѣсколько меньшее, но все же достаточное совпаденіе оказывается въ составѣ низшей по размѣру урожайности группы; въ третьей, средней группѣ всякое соотвѣтствіе исчезаетъ—среднему уровню урожайности по показаніямъ одного источника соотвѣтствуетъ весьма разнообразная, начиная отъ низшихъ до самыхъ высшихъ категорій, урожайность по показаніямъ другаго источника. Если затѣмъ разбить оба сравниваемыхъ ряда на пять группъ и вывести групповыя среднія, то получатся такіа цифры: для перваго ряда 3,0—3,6—3,8—4,05—4,8, для втораго 3,5—3,6—3,9—4,2—5,5—темпъ измѣненій уровня урожайности, очень близкій въ среднихъ группахъ обихъ рядовъ, оказывается очень различнымъ въ крайнихъ: въ нижней парѣ группъ довольно рѣзкому повышенію урожайности по первому источнику соотвѣтствуетъ едва замѣтное повышение по второму, въ высшей парѣ группъ рѣзкое повышение урожайности по первому источнику сопровождается еще болѣе рѣзкимъ повышеніемъ по второму.

Беру затѣмъ два другихъ примѣра изъ моей еще неопубликованной работы по писцовымъ книгамъ Шелонской пятины, относимымъ издателями къ 1498 году. Сначала—два ряда, изображающихъ: среднее число дворовъ на деревню и среднее число „коробей“ посѣва на дворъ—сопоставленіе, имѣющее цѣлью выяснитъ зависимость между величиной деревни и размѣромъ запашки. Вычисленіе коэффициента корреляціи даетъ $r = -0,17$, обнаруживаетъ, слѣдовательно, обратную, но крайне слабую зависимость; тѣмъ болѣе слабую, что сопоставляемые ряды состоятъ всего изъ 43 членовъ каждый, слѣдовательно, данная малая величина коэффициента корреляціи является еще въ значительной мѣрѣ случайною. Затѣмъ—другая пара рядовъ, изображающихъ: одинъ—также среднее число дворовъ въ деревнѣ, другой—среднее на дворъ количество копенъ сѣнокоса. Вычисленіе коэффициента корреляціи даетъ $r = -0,53$, значитъ, обнаруживаетъ гораздо болѣе сильную обратную зависимость, нежели въ предыдущемъ случаѣ. Прибѣгаемъ затѣмъ къ простѣйшему элементарному приему—разбиваемъ наши 43 погоста на четыре группы по среднему числу дворовъ въ деревнѣ, а именно: погосты со среднимъ числомъ дворовъ 1,5 и менѣе, отъ 1,5 до 2,5, отъ 2,5 до 3, свыше трехъ дворовъ,—и выводимъ для каждой такой группы среднее на дворъ число коробей посѣва и число копенъ сѣнокоса. Получаемъ для посѣва рядъ среднихъ 4,9—5,0—4,5—4,5; для сѣнокоса—54—30—25—30. Разница въ *степени* зависимости оказывается далеко не столь рѣзкою, какъ можно было бы судить по коэффициентамъ корреляціи ¹⁾, и въ то же самое время обнаруживается существенное различіе *въ самомъ характерѣ* зависимости между величиной селеній и размѣромъ запашки, съ одной стороны, величиной селенія и количествомъ покоса, съ другой: по обеспеченности покосами рѣзко выдѣляется только *одна* группа самыхъ мелкихъ селеній, приблизительно вдвое лучше обеспеченныхъ покосомъ, нежели остальные селенія; дальнѣйшія различія въ величинѣ селеній на обеспеченности покосами не отражаются. По обеспеченности пашней выдѣляются въ выгодную сторону не одна, а *два* группы болѣе мелкихъ селеній; эти двѣ группы обнаруживаютъ почти тождественную (4,9 и 5,0) степень, двѣ группы болѣе

¹⁾ Особенно, если принять во вниманіе, съ одной стороны, выясненный мною въ другомъ мѣстѣ, въ значительной мѣрѣ условный, характеръ писцовыхъ записей о пашнѣ, а съ другой—гораздо меньшую, если можно такъ выразиться, растяжимость обеспечения пашней по сравненію съ обеспеченностью сѣнокосами: первая, въ условіяхъ данной мѣстности и данной эпохи, ограничивается рабочею силою семьи, послѣдняя зависитъ болѣе или менѣе исключительно отъ природнаго богатства данной мѣстности сѣнокосными угодьями.

крупных—замѣтно меньшую, но въ свою очередь совершенно тождественную степень обезпеченности пашней.

Беремъ затѣмъ рядъ примѣровъ болѣе сложнаго характера, которые заимствуемъ опять-таки у Yul'я. Первый примѣръ или „задача“ на вычисленіе корреляціи, предлагаемый Yul'емъ на стр. 178 и 189—два ряда цифръ за 38-лѣтній періодъ, изображающихъ: одинъ—общій процентъ лицъ, пользовавшихся общественнымъ призрѣніемъ (пауперизмъ!), другой—процентъ пользовавшихся общественною помощью внѣ рабочихъ домовъ (outdoor). Вычисленіе коэффиціента корреляціи даетъ $r=0,57$,—обнаруживаетъ, значить, прямую, умѣренно-ясно выраженную зависимость. Разбиваемъ 38-лѣтній періодъ на четыре восьмилѣтія и одно (последнее) шестилѣтіе и вычисляемъ частныя среднія,—получаемъ: для общаго пауперизма 1,8, 3,0, 3,95, 4,6, 5,5, для помощи внѣ рабочихъ домовъ 3,65, 4,4, 6,2, 5,7, 9,9. Общая тенденція, конечно, та же, на какую указываетъ коэффиціентъ корреляціи: ростъ общаго пауперизма сопровождается ростомъ доли получающихъ помощь внѣ рабочихъ домовъ; тенденція эта, однако, рѣзко нарушается въ четвертомъ періодѣ, когда довольно сильный ростъ общаго пауперизма сопровождается, напротивъ, почти столь же замѣтнымъ паденіемъ доли пользующихся помощью на дому,—значить, и элементарный анализъ обнаруживаетъ ту же прямую, но не вполне отчетливо выраженную зависимость.

Въ данномъ случаѣ примѣненіе метода корреляціи приводитъ, приблизительно, къ тому-же результату, что и элементарный приемъ анализа. Въ рядѣ другихъ случаевъ, частью уже затронутыхъ въ предыдущемъ изложеніи, элементарные приемы анализа даютъ зависимости между цифрами, въ значительной мѣрѣ, иное освѣщеніе, нежели какое получается при одномъ только вычисленіи коэффиціента корреляціи. Беремъ таблицу 1-ю на стр. 158 книги Yul'я—отношеніе двухъ измѣреній діаметра у зоологическаго вида *Pecten opercularis*. Коэффиціентъ корреляціи $r=0,94$,—значить рѣзкая прямая зависимость. Но, какъ мы видѣли, уже внимательнаго взгляда на корреляціонную таблицу достаточно, чтобы убѣдиться въ томъ, что сказать „рѣзкая прямая зависимость“—еще далеко не значитъ дать понятіе о ея характерѣ. И въ самомъ дѣлѣ: обозначая порядковый номеръ группы или „строю“ по „дорсовентральному діаметру“ черезъ n , мы видимъ, что въ первыхъ (снизу) четырехъ „строяхъ“ преобладающею группой является n -ая группа по „антеропостеріорному“ діаметру, въ слѣдующихъ двухъ очень сильно представлены и n -ая и $n-1$ -ая группы, въ остальныхъ пяти (двѣ послѣднія отбрасываю въ виду ихъ малочисленности) рѣшительное преобладаніе переходитъ къ $n-1$ -ой группѣ. Это ясно видно изъ слѣдующей таблички:

Дорсовентр. діам.	Экземп. въ группѣ.	Въ т. ч. съ ант.-пост. діам. п.	% ихъ.	Дорсовентр. діам.	Экземп. въ группѣ.	Въ т. ч. съ з.-п. діам. п и п-1.	% ихъ.	Дорсовентр. діам.	Экземп. въ группѣ.	Изъ нихъ съ з.-п. діам. п-1.	% ихъ.
37—39	5	4	80	49—51	44	44	100	55—57	103	68	66
40—42	13	12	92	52—54	47	46	98	58—60	112	90	80
43—45	42	35	83					61—63	88	59	67
46—48	49	35	71					64—66	20	13	65
								67—69	8	5	62

„Антеропостеріорный діаметръ“, такимъ образомъ, неуклонно растеть вмѣстѣ съ „дорсовентральнымъ“,—но ростъ перваго хотя и медленно, но столь же неуклонно отстаетъ отъ роста второго: при маломъ „дорсовентральномъ“ діаметрѣ „антеропостеріорный“ діаметръ приблизительно равенъ ему, при среднемъ размѣрѣ перваго второй равенъ ему или на одинъ интервалъ меньше, среди же экземпляровъ съ наибольшимъ „дорсовентральнымъ“ діаметромъ замѣтное преобладаніе принадлежитъ меньшему на одинъ интервалъ антеропостеріорному діаметру.

Переходимъ къ 2-й таблицѣ, приводимой Yul'емъ на стр. 159,—таблицѣ, изображающей корреляцію возраста мужей съ возрастомъ женъ по даннымъ англійской переписи 1901 года. Коэффициентъ корреляціи, мы видѣли, 0,79,—прямая и достаточно ясно выраженная зависимость. Наличие именно такой прямой, достаточно ясной, но въ то же время гораздо менѣе рѣзкой, чѣмъ въ предыдущемъ примѣрѣ, зависимости легко можно, какъ отмѣчалось выше, уловить даже просто „на глазъ, безъ вычисленій“. Но элементарный анализъ бросаетъ такой свѣтъ на характеръ этой зависимости, на который корреляціонныя вычисленія не дадутъ ни малѣйшаго намека. Въ самомъ дѣлѣ: беремъ „строи“ по возрасту женъ, по каждому такому „строю“ подсчитываемъ число (тысячъ) паръ съ мужьями въ возрастъ, соответствующемъ п-му и п+1-му,—соответственному и слѣдующему высшему интерваламъ, и затѣмъ число паръ съ мужьями въ возрастъ, соответствующемъ п-1-му и всѣмъ остальнымъ по направленію книзу интерваламъ,—иначе сказать, съ болѣе молодыми мужьями. Соответственные цифры съ выведенными изъ нихъ процентными отношеніями даютъ нижеслѣдующую таблицу: (см. стр. 21).

Преобладаніе все время остается, такимъ образомъ, затѣми двумя группами, гдѣ мужья приблизительно въ одномъ возрастѣ съ женами или нѣсколько старше ихъ. Но степень преобладанія супружествъ этого типа уменьшается съ поразительною правильностью по мѣрѣ перехода отъ младшихъ къ старшимъ возрастамъ, и въ томъ же направленіи, съ не менѣе поразительною правильностью, растеть доля

Группы по возр. женъ.	Всего въ группѣ (тыс. паръ).	Изъ нихъ по возр. мужей, въ группѣ.		% паръ въ группѣ.	
		n и $n+1$	$n-1$ и ниже	n и $n+1$	$n-1$ и ниже
15—	23	18	0	78	0
20—	414	358	2	86	0,5
25—	808	667	46	82	6
30—	854	662	88	77	10
35—	781	588	95	75	12
40—	669	487	94	73	14
45—	550	398	81	72	15
50—	437	305	74	70	17
55—	317	222	57	70	18
60—	226	154	48	68	21
65—	134	89	32	66	24
70—	68	45	18	66	26
75—	27	17	9	63	33
80—	8	4	4	50	50
85—	1	—	1	—	(100)

такихъ супружествъ, гдѣ мужья, напротивъ, моложе своихъ женъ: процентъ, такъ сказать, нормальныхъ супружествъ, въ младшихъ „строяхъ“ достигающій и даже превышающій четыре пятыхъ, въ среднихъ возрастахъ понижается до трехъ четвертей, въ старшихъ постепенно падаетъ до двухъ третей и даже до половины общаго числа брачныхъ паръ,—въ самой же старшей группѣ всѣ мужья естественно моложе женъ. Процентъ супружествъ съ болѣе молодыми мужьями отъ нуля очень быстро поднимается до десяти, затѣмъ медленно, но безъ малѣйшаго уклоненія отъ правильности, продолжаетъ возрастать, составляя въ среднихъ возрастныхъ группѣ около одной шестой и поднимаясь въ старшихъ до четверти, трети и даже половины общаго числа супружествъ. Обо всемъ этомъ коэффициентъ корреляціи, очевидно, не даетъ ни малѣйшаго понятія.

Слѣдующій примѣръ,—таблица 3-я на стр. 161 книги Yul'я, изображающая корреляцію между ростомъ отцовъ и ростомъ ихъ сыновей. Коэффициентъ корреляціи получился 0,49—значить, прямая, но довольно слабо выраженная зависимость. „На глазъ“, безъ вычисленія, такая зависимость усматривается, въ данномъ случаѣ, лишь при извѣстномъ усилии вниманія; но элементарные приемы обработки выявляютъ такого рода зависимость—притомъ гораздо болѣе тѣсную, нежели можно было бы судить на основаніи коэффициента корреляціи—съ совершенною отчетливостію. Беремъ, въ самомъ дѣлѣ, группы или „строи“ по росту отца и вычисляемъ для каждой такой группы средній ростъ сына, который и сопоставляемъ со среднимъ (взятымъ по групповой медианѣ) ростомъ отцовъ даннаго „строга“. Параллельно

примѣняемъ еще другой приемъ: обозначая номеръ каждаго даннаго интервала по росту отцовъ черезъ n , обозначаемъ соответственными формулами совокупность тѣхъ группъ по росту сыновей, которымъ, взятымъ вмѣстѣ, принадлежитъ преобладаніе въ данномъ „строѣ“; сосчитываемъ число случаевъ въ этихъ преобладающихъ группахъ и вычисляемъ процентъ ихъ къ общему числу случаевъ даннаго „строа“. Получаемъ слѣдующую таблицу:

№ строевъ въ таблицѣ.	Ростъ отцовъ.	Средній ростъ сыновей.	Болѣе (+) или менѣе (—) средняго интервала.	Преобладающія группы.	Численность всего строя.	Преобл. группъ.	% преоблад. группъ.
5	62,5—63,5	66,04	+3,04	$n+3$ до $n+6$	33,5	21,75	65
6	—64,5	66,7	+2,7	$n+1$ до $n+4$	61,5	42,75	69
7	—65,5	67,2	+2,2	n до $n+4$	95,5	67	70
8	—66,5	67,5	+1,5	n до $n+4$	142	164,5	74
9	—67,5	68,2	+1,2	$n-1$ до $n+2$	137,5	90,75	66
10	—68,5	69,15	+1,15	$n-1$ до $n+3$	154	105,5	69
11	—69,5	69,4	+0,4	$n-1$ до $n+1$	141,5	81,0	57
12	—70,5	69,7	—0,3	$n-1$ до $n+1$	116	61,75	52
13	—71,5	70,5	—0,5	$n-2$ до $n+1$	78	45,25	58
14	—72,5	70,9	—1,1	$n-2$ до $n+1$	49	29	59
15	—73,5	72	—1,0	$n-3$ до $n+1$	28,5	17,75	62

Совершенно правильный ростъ цифръ въ столбцѣ средняго роста сыновей свидѣтельствуешь о наличности вполне отчетливой—повторяю, гораздо болѣе отчетливой, нежели можно было бы предположить по величинѣ коэффиціента корреляціи—прямой зависимости между ростомъ отцовъ и ростомъ сыновей. Средній ростъ сыновей поднимается, однако, значительно медленнѣе, нежели средній ростъ отцовъ, и благодаря этому отношеніе роста тѣхъ и другихъ въ разныхъ „строяхъ“ существенно различно: въ строяхъ съ ростомъ 68,5—69,5 и 69,5—70,5 средній ростъ сыновей почти не отличается отъ средняго роста отцовъ; у болѣе низкорослыхъ отцовъ сыновья выше ихъ, и разница въ ростѣ тѣмъ больше, чѣмъ ниже ростъ отцовъ; напротивъ, у наиболѣе рослыхъ отцовъ сыновья меньше ихъ ростомъ, при чемъ разница, въ общемъ незначительная, однако, замѣтно увеличивается по мѣрѣ перехода къ наивысшимъ по росту „строямъ“. То же по существу даетъ и подсчетъ преобладающихъ группъ по росту сыновей: въ двухъ „строяхъ“ наиболѣе низкорослыхъ отцовъ преобладаютъ сыновья съ значительно болѣе высокимъ ростомъ, въ двухъ слѣдующихъ къ числу преобладающихъ присоединяется группа съ сыновьями одинаковаго роста, а затѣмъ, чѣмъ дальше, тѣмъ въ большей степени преобладаніе переходитъ къ группамъ съ сыновьями болѣе низкаго роста, нежели ихъ отцы.

Этот примѣръ—замѣчу между прочимъ—представляетъ нѣкоторый особый интересъ. Элементарный анализъ, какъ мы видимъ, обнаруживаетъ совершенно полную и строгую закономерность. А между тѣмъ данный примѣръ—какъ разъ тотъ самый, гдѣ Е. Е. Слуцкій столь рѣшительно отказывается входить въ разсмотрѣніе тѣхъ „законовъ“ (кавычки Е. Е. Слуцкаго), которые выявляются при детальномъ разсмотрѣніи цифръ,—тотъ самый, на которомъ онъ иллюстрируетъ то общее положеніе, что задачей изслѣдованія является исключительно выявленіе „общихъ тенденцій“, и что „единственнымъ для этого средствомъ въ статистикѣ является нахожденіе числового выраженія для всякой характеристики (т.-е. вычисленіе коэффициента корреляціи и сравненіе его съ его вѣроятною ошибкою“¹⁾). Нѣсколькими страницами дальше г. Слуцкій—уже не на данныхъ конкретныхъ числахъ, а въ схематическомъ видѣ,—исходя изъ $r=0,5$ и приблизительно одинаковой колеблемости обоихъ уравненій регрессіи, выясняетъ существующее между ростомъ отцовъ и ростомъ сыновей „простое соотношеніе“, выражающееся въ томъ, что „средній размѣръ отклоненія величины признака у сыновей окажется (при коэффициентѣ корреляціи равномъ 0,5) равнымъ половинѣ отклоненія величины этого признака у отца. Такимъ образомъ, продолжаетъ г. Слуцкій, группа отцовъ ростомъ выше средняго на 20 см. будетъ имѣть сыновей, средній ростъ которыхъ отличается отъ средняго только на 10 см.; группа отцовъ ростомъ ниже средняго будетъ имѣть сыновей, средній ростъ которыхъ окажется занимающимъ середину между ростомъ отцовъ и среднимъ ростомъ всего населенія. Вообще, отбирая группу отцовъ, мы получимъ сыновей, отклоняющихся отъ средняго уровня въ томъ же направленіи, но только болѣе близкихъ къ нему, чѣмъ отцы“²⁾. Это выводимый изъ формулъ корреляціи частный случай того, что англійскіе статистики-біологи называютъ наслѣдственной „регрессіей“. Между тѣмъ мы видѣли, что путемъ элементарнаго анализа это явленіе выясняется если не съ большею, то, во всякомъ случаѣ, не съ меньшею отчетливостію.

Наконецъ, послѣдній примѣръ—таблица № 4 на стр. 161 книги Юл'я, изображающая корреляцію между числомъ дѣтей у матерей и числомъ дѣтей у ихъ дочерей. Вычисленіе коэффициента корреляціи даетъ $r=0,18$, т.-е. коэффициентъ, почти не дающій права заключать о наличности какой-либо зависимости. Прибѣгаемъ къ простѣйшему элементарному приему—вычисляемъ для cadaго „строга“, образованнаго

¹⁾ Стр. 63.

²⁾ Стр. 86.

по числу дѣтей у матери, среднее число дѣтей у дочерей. Результатъ вычисленія выразится въ слѣдующей табличкѣ:

Число дѣтей у ма- тери.	Число дочерей.	У нихъ дѣтей.		Число дѣтей у ма- тери.	Число дочерей.	У нихъ дѣтей.	
		Всего.	Въ сред- немъ.			Всего.	Въ сред- немъ.
1	53	177	3,8	8	92	446	4,9
2	67	201	3,5	9	76	389	5,1
3	100	364	3,6	10	52	267	5,1
4	132	531	4,0	11	25	138	5,5
5	140	572	4,1	12	22	93	4,3
6	124	531	4,3	13	10	81	8,1
7	113	530	4,7	14—15	4	31	7,7

Въ противоположность тому, чего можно было ожидать въ виду ничтожной величины коэффициента корреляціи, мы видимъ, такимъ образомъ, вполне определенную и отчетливо выраженную въ одиннадцати первыхъ по порядку „строяхъ“ зависимость: число дѣтей у дочерей съ совершенною правильностью растетъ вмѣстѣ съ числомъ дѣтей у матерей ихъ—плодовитость тѣхъ и другихъ находится, значитъ, въ отчетливо выраженномъ прямомъ отношеніи; при этомъ, однако, среднее число дѣтей у дочерей возрастаетъ гораздо медленнѣе, чѣмъ растетъ принятое за основаніе группировки число дѣтей у матерей,—благодаря чему получается такая же, въ существѣ дѣла, „регрессія“, какъ и въ предыдущемъ случаѣ: въ низшихъ по плодовитости матерей группахъ среднее число дѣтей у дочерей больше, чѣмъ было дѣтей у ихъ матерей; въ группѣ четырехдѣтныхъ среднее число дѣтей у дочерей сравнивается съ числомъ дѣтей у матерей, въ слѣдующихъ, высшихъ по плодовитости матерей группахъ среднее число дѣтей у дочерей все болѣе отстаетъ отъ взятаго за групповой признакъ числа дѣтей у матерей. Какъ упомянуто, такого рода весьма рѣзко выраженная зависимость обнаруживается только въ первыхъ по порядку одиннадцати группахъ или „строяхъ“, — въ послѣднихъ трехъ она исчезаетъ, и колебанія средняго числа дѣтей у дочерей не обнаруживаютъ какой-либо правильности. Это, скорѣе всего, результатъ малочисленности данныхъ трехъ „строевъ“. Если, въ самомъ дѣлѣ, мы объединимъ ихъ въ одну группу—матерей, имѣвшихъ 12 и болѣе человекъ дѣтей, то получимъ групповую среднюю 5,7,—значитъ, установленная нами зависимость останется въ силѣ и для этой сводной группы. Едва ли будетъ ошибкою сказать, что ничтожная величина коэффициента даетъ, въ данномъ случаѣ, совершенно неправильное представленіе о степени зависимости между двумя сопоставленными явленіями.

Число примѣровъ, очевидно, могло бы быть увеличено до безконечности. Но и приведенныхъ примѣровъ, мнѣ кажется, достаточно, чтобы рѣшиться на нѣкоторые обобщенія. Я едва ли ошибусь именно, если скажу, что ни въ одномъ изъ разнообразныхъ примѣровъ вычисленіе коэффициента корреляціи не раскрыло чего-либо такого, что не могло бы быть, съ далеко меньшею затратой вычислительнаго труда, раскрыто помощью элементарныхъ приѣмовъ статистическаго анализа—въ послѣднемъ же изъ разобранныхъ случаевъ результатъ корреляціонныхъ вычисленій даже заслонилъ совершенно ясную и сразу обнаруживающуюся при помощи простѣйшаго изъ элементарныхъ приѣмовъ корреляціонную зависимость. Коэффициентъ корреляціи даетъ, правда, *численную мѣру* этой зависимости. Но даже оставляя въ сторонѣ тѣ изложенныя выше соображенія теоретическаго свойства, въ силу которыхъ численныя значенія коэффициента корреляціи утрачиваютъ значительную долю присущей имъ, на первый взглядъ, точности и объективной убѣдительности, позволительно спросить себя: даетъ ли получаемая въ результатѣ длинныхъ вычисленій цифра 0,3, 0,7 или 0,9 болѣе отчетливое представленіе о степени зависимости между сравниваемыми рядами, нежели какое получается въ однихъ случаяхъ непосредственно „на глазъ“, безъ примѣненія какихъ бы то ни было вычислительныхъ приѣмовъ¹⁾, въ другихъ—при помощи требующихъ несравненно меньшаго количества вычисленій элементарныхъ приѣмовъ? Думаю, что отвѣтъ на этотъ вопросъ будетъ зависеть, если не цѣликомъ, то въ значительной мѣрѣ, отъ индивидуальнаго склада каждаго даннаго статистика-ислѣдователя или каждаго даннаго читателя. Инымъ,—допускаю, что довольно многимъ,—цифра скажетъ больше, нежели какой-либо иной способъ характеристики корреляціонной связи. Я, напротивъ, принадлежу къ числу тѣхъ, для кого цифровая характеристика—по крайней мѣрѣ въ данныхъ условіяхъ—не обладаетъ такою степенью наглядности и убѣдительности,—кто гораздо больше выноситъ изъ непосредственнаго изученія строенія изученныхъ рядовъ, въ первоначально данномъ или въ соответственномъ концентрированномъ элементарномъ обработкою видѣ.

¹⁾ Въ дополненіе къ разобраннымъ на предыдущихъ страницахъ примѣрамъ приведу еще одинъ, гдѣ ненужность вычисленія корреляціи бросается въ глаза особенно рѣзко: это анализъ данныхъ о зависимости степени прочности жилыхъ и хозяйственныхъ построекъ, приводимый на стр. 134 „учебника“ проф. Орженцаго, какъ примѣръ вычисленій корреляціи качественныхъ признаковъ: нужны ли, спрашивается, сложныя вычисленія, составляющія прибѣгать къ таблицамъ логарифмовъ, если 197 и 129 случаямъ совпаденія „прочности“ либо „ветхости“ жилыхъ и хозяйственныхъ построекъ противостоятъ 3 и 5 случаевъ расхожденія, и что прибавляютъ $r = 0,95$ къ простому „общему впечатлѣнію“, какое производитъ составленная изъ приведенныхъ только что цифръ табличка?

Болѣ категорически рѣшается, какъ мнѣ кажется, дальнѣйшій вопросъ: способны ли корреляціонныя формулы дать *исчерпывающую* характеристику зависимости между статистическими рядами? можетъ ли, другими словами, вычисленіе коэффициента корреляціи и другихъ соотносительныхъ съ нимъ формулъ избавить отъ необходимости анализа статистическихъ рядовъ при помощи элементарныхъ приѣмовъ? Уже по чисто апріорнымъ соображеніямъ можно было ожидать отрицательнаго отвѣта на поставленный такимъ образомъ вопросъ,—можно было предполагать, что шаблонъ корреляціонныхъ формулъ окажется недостаточнымъ, чтобы уловить безконечное разнообразіе тѣхъ формъ, въ какія можетъ облекаться корреляціонная зависимость. И дѣйствительно—на разобранныхъ нами примѣрахъ мы видѣли, что приспособленные къ особенностямъ каждаго даннаго матеріала и къ характеру данной познавательной задачи элементарные приѣмы въ большинствѣ случаевъ обнаруживали наличность такихъ болѣ сложныхъ и тонкихъ зависимостей, какія не находятъ себѣ никакого выраженія въ коэффициентѣ корреляціи и которыя только въ нѣкоторыхъ случаяхъ и лишь до извѣстной степени выявляются при помощи уравненій регрессіи. Мы видимъ, что коэффициентъ корреляціи дѣйствительно „даетъ *лишь часть тѣхъ свѣдѣній* о явленіи, какія содержатся въ первоначальныхъ данныхъ или въ корреляціонной таблицѣ“, что поэтому, „вычисленъ ли коэффициентъ или нѣтъ,—всякая таблица должна быть подвергнута тщательному разсмотрѣнію съ цѣлью выяснитъ, представляетъ ли она какія-либо видимо существенныя особенности въ распредѣленіи частотъ“; что подобное детальное разсмотрѣніе — иначе сказать, примѣненіе элементарныхъ приѣмовъ анализа, дѣйствительно, „часто даетъ возможность обнаружить такія существенныя обстоятельства, которыя иначе остались бы незамѣченными“ (Yule); что, словомъ, статистика не можетъ ограничиваться тѣмъ констатированіемъ и, въ лучшемъ случаѣ, тою „мѣрою“ корреляціонной зависимости, какую даютъ корреляціонныя формулы, а должна стремиться дать ту болѣ тонкую, въ значительной мѣрѣ качественную характеристику этой зависимости, во всѣхъ ея конкретныхъ жизненныхъ особенностяхъ и отѣнкахъ, какая, въ видѣ правила, можетъ получиться лишь при помощи элементарныхъ приѣмовъ статистическаго анализа.

И въ конечномъ результатѣ: примѣненіе корреляціонныхъ формулъ—съ этого я началъ и этимъ закончу—не встрѣчаетъ принципиальныхъ возраженій. Пользованіе этими формулами, можетъ быть, отвѣчаетъ потребности изслѣдователей и читателей довольно распространеннаго мыслительнаго склада,—тѣхъ, кому мѣра и число говорить больше, нежели впечатлѣніе и описаніе. Но я не вижу основаній считать, чтобы способъ корреляціи долженъ былъ быть отнесенъ

къ числу *необходимыхъ* орудій статистическаго анализа; сомнѣваюсь чтобы то обогащеніе нашего знанія, какое достигается путемъ примѣненія корреляціонныхъ формулъ, окупало связанныя съ этимъ затраты вычислительнаго труда. И на чемъ я самымъ рѣшительнымъ образомъ настаиваю — вычисленіе корреляціонныхъ формулъ представляетъ собою слишкомъ суммарный и шаблонный способъ характеристики корреляціонной зависимости, а потому ни въ какомъ случаѣ не можетъ устранить необходимости въ примѣненіи приспособленныхъ къ условіямъ и требованіямъ каждаго даннаго случая элементарныхъ приѣмовъ статистическаго анализа.

А. Кауфманъ.

Post scriptum. Эта статья была сдана въ редакцію, когда появилась 1—2 книга „Статистическаго Вѣстника“ со статьею С. Н. Прокоповича „Объ основаніяхъ выбора признаковъ для построенія комбинаціонныхъ таблицъ“, въ которой дѣлается попытка путемъ вычисленія коэффициентовъ корреляціи рѣшить вопросъ о признакахъ, долженствующихъ полагаться въ основу комбинаціонныхъ группировокъ. Статья эта не заставляетъ меня внести какихъ-либо измѣненій или поправокъ въ сдѣланные мною выше выводы. Впослѣдствіи, можетъ быть, я буду имѣть случай войти въ болѣе подробное разсмотрѣніе предлагаемаго г. Прокоповичемъ приѣма строго объективнаго рѣшенія вопроса о наиболѣе цѣлесообразномъ основаніи для комбинаціонныхъ группировокъ. Сейчасъ скажу лишь кратко, что этотъ приѣмъ, по моему убѣжденію, не разрѣшаетъ и не можетъ — по самому существу дѣла не можетъ разрѣшить вопроса.

Г. Прокоповичъ — какъ читатель помнитъ — вычисляетъ коэффициенты корреляціи для посѣвной площади и для другого весьма обычнаго группировочнаго признака — надѣла; убѣждается въ томъ, что „корреляціонная связь между площадью посѣва и другими элементами крестьянскаго двора... тѣснѣе, чѣмъ между тѣми же элементами и величиной надѣла, или надѣльной и купчей земли вмѣстѣ“, и отсюда приходитъ къ заключенію, что „статистики были совершенно правы, отдавъ предпочтеніе для построенія комбинаціонныхъ таблицъ посѣвной площади передъ площадью надѣла“ (стр. 156).

Что корреляціонная связь посѣвной площади сильнѣе, — иначе сказать, что посѣвная площадь стоитъ въ болѣе непосредственномъ соотношеніи съ другими элементами крестьянскаго двора, нежели на-

дѣльная или надѣльная плюсъ купчая земля, — это изъ сопоставленія коэффициентовъ корреляціи вытекаетъ съ очевидностью. Но это было ясно и безъ всякихъ вычисленій: посѣвная площадь — это именно тотъ наиболѣе симптоматическій признакъ, въ которомъ въ мѣстности съ преобладаніемъ земледѣлія находить себѣ выраженіе опредѣленная степень состоятельности крестьянскаго двора; естественно, что именно этотъ признакъ стоитъ въ наиболѣе непосредственномъ соотношеніи со всѣми тѣми другими признаками, которые болѣе или менѣе непосредственно зависятъ отъ благосостоянія (грамотность, наемъ и отпускъ рабочихъ рукъ) или принадлежать къ числу обусловливающихъ его факторовъ (рабочая сила семьи). Съ посѣвною площадью, въ роли симптоматическаго признака, можетъ конкурировать живой инвентарь земледѣльческаго хозяйства — рабочій скотъ; неудивительно, что именно этотъ признакъ даетъ съ посѣвною площадью наиболѣе высокіе коэффициенты корреляціи.

Поскольку, такимъ образомъ, корреляціонныя вычисленія даютъ положительный результатъ, иначе сказать — высокіе коэффициенты корреляціи, они доказываютъ лишь то, что не требовало доказательства, — что давно было установлено не „подсчетомъ голосовъ“, а коллективною мыслью русскихъ земскихъ статистиковъ: что посѣвъ и, на ряду съ нимъ, рабочій скотъ являются наиболѣе цѣлесообразнымъ группировочнымъ признакомъ, — если, конечно, мы имѣемъ дѣло съ земледѣльческимъ по преимуществу хозяйствомъ и если притомъ наша цѣль — выяснить вліяніе наличной степени крестьянскаго благосостоянія на другіе признаки, или, говоря общѣе, связь его съ этими послѣдними. Но что, для даннаго вопроса, доказываютъ корреляціонныя вычисленія, если они даютъ отрицательный результатъ — иначе сказать, низкіе коэффициенты корреляціи? Ничего или очень мало. Коэффициентъ корреляціи — не такая мѣрка, которую достаточно механически приложить, чтобы получить готовое рѣшеніе столь сложнаго вопроса, какъ вопросъ о цѣлесообразномъ выборѣ группировочныхъ признаковъ. „Коэффициентъ корреляціи — говоритъ Udny Yule въ своемъ многократно цитированномъ въ текстѣ этой статьи руководствѣ (стр. 191) — даетъ только облеченное въ суммарную и легко понятную форму, но одностороннее представленіе (only one particular aspect) о тѣхъ фактахъ, на которыхъ онъ основанъ. Данная величина коэффициента можетъ быть совмѣстима съ нѣкоторою данною гипотезою, но можетъ быть совмѣстима и съ другими; и при обсужденіи такихъ возможныхъ гипотезъ необходимы не только осторожность и вдумчивость, но также и основательное знаніе фактовъ во всестороннемъ ихъ освѣщеніи (in all other possible aspects)“. Въ частности, сколь угодно низкій коэффициентъ корреляціи, даже равный нулю, еще не доказываетъ отсутствія кор-

реляціонной зависимости. Вѣдь дѣло отнюдь не можетъ быть сведено къ тому простѣйшему типу корреляціи, съ которымъ только и оперируетъ г. Прокоповичъ. Анализъ формулъ „многократной корреляціи“ показываетъ, что конечный результатъ въ видѣ полевого коэффициента корреляціи можетъ получиться при наличности сколь угодно тѣсной корреляціонной связи между отдѣльными факторами, участвующими въ многократной корреляціи; при такомъ положеніи вещей, какъ подчеркиваетъ Yule (стр. 251), „мы легко можемъ неправильно истолковать коэффициентъ корреляціи, равный нулю“, — легко можемъ принять за отсутствіе корреляціонной зависимости то, что есть лишь болѣе сложная и потому труднѣе уловимая корреляціонная связь.

Или въ примѣненіи къ занимающему насъ вопросу: если надѣльная земля или любой иной группировочный признакъ слабо реагируетъ на вычисленіе коэффициента корреляціи, то это доказываетъ лишь одно: что между этимъ признакомъ и другими элементами крестьянскаго двора нѣтъ *такой простой, такой непосредственной* связи, какая есть у этихъ элементовъ съ посѣвною площадью. Но, съ одной стороны, въ этомъ вѣдь нѣтъ ничего новаго и неожиданнаго: тѣмъ же коллективнымъ разумомъ земскихъ статистиковъ давно установлено, что вліяніе какъ „надѣльной земли“, такъ и „надѣльной плюсь купчая“, можетъ быть, доминировавшее въ первое время послѣ надѣленія, перекрещивается и въ значительной мѣрѣ парализуется вліяніемъ разнообразныхъ другихъ факторовъ, а потому, сплошь и рядомъ, не даетъ достаточно ощутительной реакціи на цифрахъ. А съ другой—и это главное: отсутствіе *простой и непосредственной* связи, уловимой помощью простѣйшей корреляціонной формулы, отнюдь не доказываетъ отсутствія *болѣе сложной*, труднѣе уловимой зависимости, и тѣмъ болѣе не устраняетъ и не умаляетъ интереса ея выясненія. Я—напомню во избѣжаніе всякихъ недоразумѣній—отнюдь не принадлежу къ числу рѣшительныхъ сторонниковъ группировки по надѣльной землѣ, да и вообще какой бы то ни было другой однообразно-проводимой группировки. Я считаю несомнѣннымъ, что крестьянское хозяйство въ каждой мѣстности, дѣйствительно, „имѣетъ свое особое строеніе, свой особый организаціонный планъ“, и вмѣстѣ съ В. В. думаю, что комбинаціонная таблица „должна сочетать въ разныхъ случаяхъ не одни и тѣ же признаки, отвѣчая каждый разъ на особый вопросъ“¹⁾, при чемъ вопросы въ каждомъ данномъ случаѣ должны ставиться и группировочные признаки должны выбираться въ зависимости, съ одной стороны, отъ конкретныхъ, частью заранѣе извѣстныхъ, частью выясненныхъ

¹⁾ См. мое сообщеніе въ „Трудахъ комиссіи по вопросамъ земской статистики 20—23 февраля 1913 г.“, стр. 98—99.

даннымъ изслѣдованіемъ условій данной мѣстности и даннаго времени, съ другой—отъ индивидуальнаго интереса изслѣдователя. Но если совокупность условій мѣста и времени, вмѣстѣ съ индивидуальными воззрѣніями статистика, вызываютъ въ немъ особый интересъ къ выясненію вліянія надѣльнаго землевладѣнія или иного опредѣленнаго признака; низкій коэффициентъ корреляціи, иначе сказать—отсутствіе простой, непосредственной связи этого признака съ другими, недостаточно для того, чтобы побудить его доискиваться болѣе сложныхъ и тонкихъ, а потому труднѣе уловимыхъ связей. Наоборотъ, оно послужитъ для него только лишнимъ толчкомъ для того, чтобы примѣнить къ рѣшенію интересующаго его вопроса и болѣе сложные и тонкіе приемы, въ частности—и болѣе сложную комбинаціонную группировку, способную выявить вліяніе даннаго признака, изолировать его отъ другихъ, съ нимъ перекрещивающихся и въ значительной мѣрѣ его парализующихъ.

А. К.