

Пасеков В.П.

*Федеральный исследовательский центр «Информатика и управление»  
Российской академии наук, ул. Вавилова, д. 40, Москва, 119991, Россия*

## О ПРИМЕНЕНИИ БИОМЕТРИИ В АНТРОПОЛОГИИ И ОСНОВНЫХ ЗАДАЧАХ НАУКИ

**Введение.** Задачей биометрии в антропологии является описание и анализ случайной изменчивости, присущей человеку как объекту исследования. В настоящее время для этого широко используются многомерные биометрические модели, в которых предполагается, что наблюдаемые количественные признаки организма выражаются в виде взвешенных сумм скрытых случайных факторов (например, анализ главных компонент, факторный и дискриминантный анализы, канонические переменные и др.). Данное предположение согласуется с существованием общих и специфических биологических факторов-причин, определяющих признаки организма. Поскольку основной задачей науки является изучение причинно-следственных связей в природе, важно проанализировать взаимоотношения между статистическими и причинно-следственными зависимостями.

**Результаты и обсуждение.** Получаемые в результате обработки данных статистические факторы (такие как общий размер, форма и т. д.) в общем случае не только не согласуются с причинными факторами, но могут даже противоречить им. Продемонстрировано, что статистические коэффициенты корреляции, регрессии и ковариации даже в простейшем случае двух признаков, случайно изменяющихся под влиянием одной общей причины и нескольких специфических факторов, могут принимать бесконечно много значений при одной и той же причинной связи, причем при той же самой статистической закономерности изменчивости существует бесконечно много ее причинно-следственных объяснений.

**Заключение.** Интерпретация в отношении причинно-следственных закономерностей детерминации признаков и другие причинно-следственные выводы не должны вытекать из результатов применения одномерных и многомерных биометрических моделей, использующих в качестве входных данных указанные статистические коэффициенты.

**Ключевые слова:** биометрия; многомерные модели; причинные законы; регрессия; корреляции; общие и специфические факторы; антропология

## Введение

Использование биометрических методов в биологии, особенно, в антропологических исследованиях, имеет длительную историю, начинающуюся уже с момента появления биометрических подходов (вспомним коэффициенты корреляции и расового сходства Пирсона, дисперсионный анализ Фишера и что термин регрессия был введен в практику биометрических исследований Гальтоном). Дальнейшее развитие биометрии привело к широкому применению методов многомерного статистического анализа (скажем, таких как расстояние Махаланобиса, анализ главных компонент, факторный анализ, канонические переменные, многомерный дисперсионный и дискриминантный анализы, и др. (см., например, [Дерябин, 1983; Rao, 1952]). При этом основная задача биометрии состоит, как правило, в описании и анализе существующей изменчивости исследуемых объектов. Биометрический анализ позволяет отделить случайные различия от реально существующих. Он неocenим для решения таких задач как оптимальное оценивание параметров популяции, разбиение обнаруженных статистических зависимостей на достоверные и чисто случайно сложившиеся в выборочных наблюдениях, для определения значимости различий между группами, для перехода от наблюдаемых признаков к новым случайным переменным, позволяющим судить о принадлежности наблюдаемого объекта к одной из нескольких групп, и кроме того, в результате подобного перехода легче увидеть закономерности случайной вариабельности и т. д.

Здесь результаты применения формальных биометрических методов оставляют как бы за кадром причинно-следственные закономерности, лежащие в основе обнаруживаемых статистических связей. Их интерпретация предоставляется исследователям. У меня при первоначальном знакомстве с биометрией вызывало сожаление отсутствие в учебниках причинно-следственные выводов из результатов статистической обработки материалов. Возникало впечатление, что авторы старательно обходят, казалось бы, лежащую на поверхности причинную интерпретацию выявляемых статистических связей. Достоинство такой подачи материала в том, что она обеспечивает универсальность чисто статистического взгляда.

Сейчас использование биометрии в антропологии базируется на относительно простом решении возникающих статистических задач на

компьютерах помощью соответствующих пакетов программ. В результате освобождается время для причинного анализа изменчивости вместо затрат на трудоемкие рутинные вычисления. Выяснение соотношений между причинным и статистическим анализами является важной задачей. Стремление проанализировать использование биометрии в данном направлении мотивировано тем, что *основной задачей науки является изучение причинно-следственных законов функционирования и развития изучаемых объектов в целях описания, объяснения и понимания, а также прогнозирования их состояний.*

Причинные механизмы позволяют увидеть вместо нагромождения фактов логичную картину явления в целом, редуцировать огромный объем информации к сравнительно немногим принципам, лаконичные схемы которых дают возможность прогнозировать многочисленные свойства-следствия. Подобно этой картине многомерные методы статистики редуцируют море статистических данных к немногим показателям, которые допускают наглядную интерпретацию и позволяют увидеть основные закономерности случайной изменчивости. Однако эти закономерности, как мы увидим, лежат как бы в другой плоскости, нежели скрытые под ними причинные связи.

Применительно к типичным материалам в антропологии, собираемым путем обследования индивидуумов в популяции, будем понимать под причинными связями законы, действующие на уровне организма (и независимые от популяционного окружения), а статистические связи отражают эти индивидуальные закономерности на популяционном уровне. Мы ограничимся далее случаем *функциональных связей* (законов) для организма, когда результаты применения закона *однозначны*, что характерно для макромира. Например, известный из школьного курса закон Ньютона о движении тела под действием приложенных сил определяет траекторию динамики тела единственным образом.

В антропологии хотелось бы, скажем, найти закон, по которому в процессе развития происходят изменения размеров различных признаков организма под действием концентрации гормона роста в крови в течение созревания, что приводит к соответствующей траектории динамики морфологического облика индивидуума с возрастом. Так как скорости изменения размеров признаков отличаются между собой, то соотношения этих размеров меняются в ходе развития от характерных для ребенка значений до

типичных для взрослого состояния. Если бы такой закон роста был известен, то можно было бы предсказывать пропорции данного человека к взрослому состоянию и пр. в зависимости от пола, длительности и интенсивности процессов роста и концентрации гормона роста.

В частности, тогда можно предсказать качественное и количественное изменение грацильного детского черепа мальчика до матуризованного черепа мужчины с прогнозом результатов вплоть до выяснения теоретической возможности появления на черепе гребней, если бы рост продолжался далее существующих пределов. Аналогично изменение действия фактора роста в эволюционном процессе приводит к закономерному, например, аллометрическому соотношению размеров признаков организма не с возрастом, а в серии ископаемых останков.

Задача выяснения законов функционирования и динамики объектов обычно решается в точных науках на основе экспериментального подхода. Для человека такой путь недопустим по этическим соображениям. В то же время данные по естественной изменчивости в окружающей действительности словно бы дают в руки результаты экспериментов, поставленных самой природой. Это производит впечатление возможности заменить проведение экспериментов на обработку таких данных. Статистические связи в естественной изменчивости, казалось бы, позволяют оценить лежащие в основе согласованной изменчивости скрытые неизвестные природные законы, порождающие наблюдаемое разнообразие, лишь бы были правильно использованы адекватные статистические методы.

Сейчас хорошо известно, что статистические связи не эквивалентны причинным, *хотя наличие статистических зависимостей обусловлено причинными влияниями. Цель настоящей работы состоит в анализе соотношения статистических выводов и причинных закономерностей на примере использования широко распространенных биометрических методов (множественной линейной регрессии, анализа главных компонент, факторного анализа, канонических переменных и др. им подобных).* Для этого исследуется принципиальная возможность определения природных закономерностей по данным, представляющим собой наблюдения над случайной неконтролируемой изменчивостью организмов. Изложение базируется на выступлениях автора на различных конференциях (например, [Пасеков, 2001]) и, главным образом,

на препринте [Пасеков, 2005], развиваемых и объединяемых на основе исследования возможности использования биометрии для причинного анализа популяционных антропологических данных. Обоснование приводимых далее выражений для коэффициентов регрессии, корреляции и ковариации в рассматриваемых ситуациях приведено в приложении.

Гетерогенность популяции по рассматриваемым признакам обусловлена, главным образом, различиями индивидуумов по значениям присущих им факторов-причин, их разными комбинациями и изменчивостью условий среды. Предположим, что причинные зависимости значений признаков от внешних и внутренних факторов (допустим, от интегрального уровня гормона роста в крови или от уровня питания) описываются некоторыми функциональными законами, выражающими устойчивые общие отношения, свойственные всем явлениям данного рода. Тогда в гетерогенной популяции при наличии случайной изменчивости будут наблюдаться вызванные этими зависимостями статистические связи между признаками.

При получении знаний о началах биометрии появляется желание связать статистическую зависимость с причинной. Однако получаемое на популяционном уровне статистическое отражение индивидуальных закономерностей может быть искаженным. Например, реальные (функциональные) связи между признаками, вообще говоря, отличаются от регрессионных зависимостей [Кендалл, Стюарт, 1973; Пасеков, 2001].

Для иллюстрации основных идей в дальнейшем отвлечемся от важных, но излишних для наших целей подробностей, например, от учета выборочных ошибок, искажающих теоретические закономерности (т.е. будем рассматривать свойства генеральной совокупности).

## Результаты и обсуждение

*Коэффициенты корреляции и регрессии и важность отдельной причины для признака*

Рассмотрим такую простую и популярную характеристику статистической зависимости как коэффициент корреляции (Пирсона). Кажется, что чем больше абсолютная величина коэффициента корреляции двух признаков, тем более один из них влияет на другой или сильнее воздействует на признаки неизвестный или предполагаемый общий фактор. Чем ближе коэффициент

корреляции к  $\pm 1$ , тем больше теснота статистической связи, и считается, что тем важнее она статистически. Можно ли говорить, что тем важнее она и биологически?

**А.** Приведем один из наиболее простых вариантов детерминации признака  $Y$  в виде суммы независимых случайных положительных факторов-причин  $\{X_i\}$ , например, когда  $Y \equiv X_1 + X_2 + X_3$  (см. рис. 1а). Реально ими могут быть, например, составные части размерного признака, скажем, составным признаком будет длина тела, складывающаяся из высоты стопы, длины голени, бедра и т. д. (если отвлечься в иллюстративных целях от возможных зависимостей между ними).

При малости отдельных отклонений  $Y$  от среднего значения (когда для всех индивидуумов относительные вклады  $\{X_i\}$  в  $Y$  сходны) можно определить *важность  $i$ -го фактора  $X_i$  для признака  $Y$  как отношение среднего вклада фактора в среднее значение признака  $E\{X_i\}/E\{Y\}$* . Здесь символ  $E$  означает операцию получения средней величины.

Возьмем какую-либо отдельную составную часть ( $X_i$ ) признака  $Y$ . Известно, что если к каждому значению случайной величины  $X_i$  добавить одно и то же допустимое число, то ее среднее значение  $E\{X_i\}$  сдвинется на ту же величину. Очевидно, на столько же изменится и  $E\{Y\}$ , но отношение  $E\{X_i\}/E\{Y\}$  в зависимости от величины и знака добавка может стать близким к нулю или единице. При этом изменчивость (дисперсия) не изменится. Дело в том, что в определении дисперсии используются только отклонения от среднего, сохраняющиеся при такой модификации. Точно также и коэффициент корреляции

между признаком и его составной частью (фактором) останется фиксированным на прежней величине, если все значения фактора (а значит, и принятый нами показатель важности данной причины – ее вклад в признак) изменить на какую-либо допустимую величину, поскольку и корреляция определяется отклонениями от средних.

Поэтому *какова бы ни была фиксированная величина коэффициента корреляции (теснота статистической связи) между фактором-причиной и признаком-следствием, относительная величина среднего вклада фактора в среднее значение признака (важность данной причины) может быть как угодно близкой к нулю, так и произвольно близкой к единице. Значит, в данной схеме коэффициент корреляции не характеризует важность причины для следствия.*

**Б.** Пусть теперь  $Y(X) \equiv \sum c_i X_i$ , где отдельная фактор-причина, обозначенная как  $X$ , влияет на признак  $Y$  по закону  $Y(X) = cX$ . Коэффициент  $c$  определяет интенсивность причинной зависимости  $Y$  от фактора  $X$ . Чем больше отличается от нуля коэффициент  $c$ , тем сильнее воздействует причина  $X$  на признак  $Y$ , и можно сказать, что тем важнее она для  $Y$ . Если причина  $X$  случайно варьирует, то это не сказывается на зависимости  $Y(X)$ , по которой значение  $Y$  находится для каждого из значений  $X$ . Однако, когда на  $Y$  дополнительно влияют случайные специфические факторы даже с нулевым средним значением, то такого соответствия между заданном  $X$  и  $Y$  уже не будет из-за добавления к значению  $Y(X)$  случайного фактора или нескольких случайных факторов.

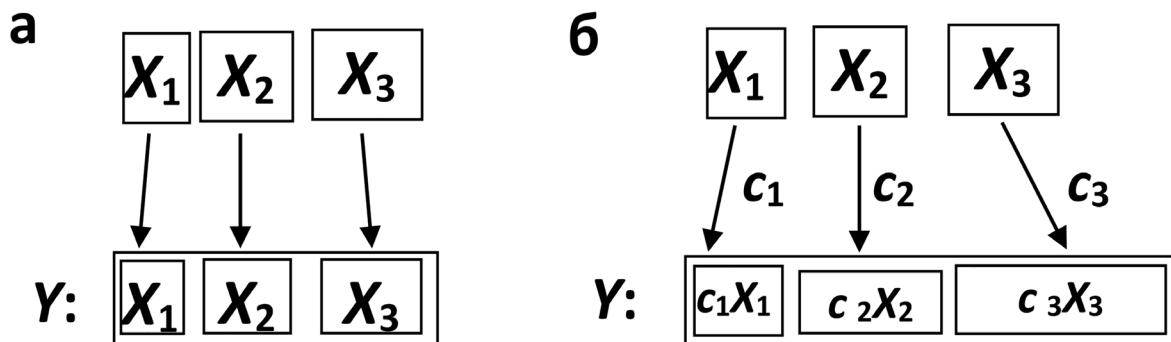


Рисунок 1. Простейшие схемы детерминации признака  $Y$  факторами-причинами  $\{X_i\}$   
Figure 1. The simplest schemes for determining the trait  $Y$  by factors-causes  $\{X_i\}$



При этом степень нарушения соответствия можно охарактеризовать коэффициентом корреляции Пирсона  $\rho_{YX}$ . Если  $\rho_{YX} = \pm 1$ , то нарушения соответствия не будет. Когда  $\rho_{YX} = 0$ , соответствие возможно только случайно. Напомним, что коэффициент корреляции  $\rho_{YX}$  находится через случайные отклонения  $X$  и  $Y$  от своих средних значений как отношение ковариации  $\text{COV}(Y, X)$  признака  $Y$  с  $X$  к квадратному корню из произведения их дисперсий  $V_Y$  и  $V_X$ :

$$\rho_{YX} \equiv \text{COV}(Y, X) / \sqrt{V_Y V_X};$$

$$\text{COV}(Y, X) \equiv E\{(X - E\{X\})(Y - E\{Y\})\},$$

$$V_X \equiv \text{COV}(X, X).$$

Как известно, дисперсия  $V_Y$  равна сумме дисперсий, суммируемых в  $Y$  независимых случайных величин, вклад  $W_X$  фактора  $X$  в дисперсию  $V_Y$  равняется  $c^2 V_X$ , а ковариация в данном случае будет равна произведению  $c V_X$ . Следовательно, при  $c > 0$

$$\rho_{YX} = c V_X / \sqrt{V_Y V_X} = c \sqrt{V_X / V_Y} = \sqrt{c^2 V_X / V_Y} = \sqrt{W_X / V_Y}.$$

Значит, при одной и той же причинной зависимости  $Y(X) = cX$  между признаком  $Y$  и причиной  $X$  (при ее фиксированной важности  $c$ ) коэффициент корреляции  $\rho_{YX}$  может принимать бесконечно много разных значений в зависимости от отношения  $W_X / V_Y$  компонента  $W_X$  дисперсии признака  $Y$  из-за фактора  $X$  ко всей дисперсии  $V_Y$  признака. Отсюда вытекает, что величина коэффициента корреляции, будь она близкой к нулю или к  $\pm 1$ , ничего не говорит о важности причины  $X$  для  $Y$  (об интенсивности причинной зависимости).

**В.** Перейдем к сравнению характера (формы) статистической и причинной зависимостей. Когда  $Y = c_1 X_1 + c_2 X_2 + c_3 X_3$  (см. рис. 1б), степень влияния причины  $X_i$  на признак  $Y$  определяется коэффициентом  $c_i$ . Графически причинная зависимость  $Y$  от  $X_i$  представляет собой прямую линию с угловым коэффициентом  $c_i$ . Если факторы-причины случайно варьируют, то в качестве показателя связи между  $Y$  и  $X_i$  кажется привлекательным использовать соответствующие коэффициенты регрессии, определяющие углы наклона регрессионных прямых.

Возможность предсказания по линии регрессии среднего значения «зависимого» признака, соответствующего заданному значению «ведущего» (независимого), ассоциируется со следующими экспериментами. Пусть нужно исследовать влияние некоторого фактора на анали-

зируемый признак. Для этого в равных условиях, отличающихся лишь величиной рассматриваемого фактора, производится серия опытов (в которых изменяется фактор-причина) и регистрируются значения анализируемого признака. Изменения последнего от опыта к опыту показывают влияние фактора на признак. Поскольку линия регрессии представляет собой график изменений среднего значения зависимого признака при изменениях значений ведущего, то, казалось бы, она дает полную аналогию с указанной серией опытов.

При статистической связи двух признаков  $Y_1$  и  $Y_2$  каждому фиксированному значению первой случайной переменной  $Y_1$  соответствует распределение значений второй  $Y_2$  и наоборот. Если взять совокупность индивидуумов с фиксированным значением случайно варьирующего признака  $Y_1$ , то при этом среднее значение у  $Y_2$  для данного распределения находится по формуле линейной регрессии как  $Y_2 = a + b_{21} Y_1$ . Здесь  $b_{21}$  обозначает коэффициент регрессии второго признака  $Y_2$  от первого  $Y_1$ . Вычисление коэффициента регрессии следствия ( $Y$ ) от причины (скажем, от  $X_1$ ) для схемы на рис. 1б дают коэффициент, равный  $c_1$ . В данном случае коэффициент регрессии совпадает с влиянием  $c_1$  причины  $X_1$  на признак-следствие  $Y$ .

Если поменять местами роли случайных величин (формально рассматривать  $X_1$  как зависимую величину), то коэффициент  $b_{X_1 Y}$  регрессии  $X_1$  от  $Y$  не будет прямо связан с законом влияния причины на следствие. Он не равен  $1/c_1$ , как можно было наивно ожидать, а будет равняться

$$\frac{1}{c_1} \frac{2}{\rho_{YX_1}},$$

что обычно меньше абсолютной величины у коэффициента  $1/c_1$  в функциональной зависимости  $X_1$  от  $Y$ , так как квадрат любого коэффициента корреляции не больше единицы.

Таким образом, при одной и той же причинной зависимости  $Y(X) = cX$  между признаком  $Y$  и причиной  $X$  коэффициент регрессии  $b_{X_1 Y}$  причины  $X$  от признака  $Y$  может принимать бесконечно много разных значений

$$\frac{1}{c} \frac{2}{\rho_{YX}},$$

в зависимости от (бесконечного согласно предыдущему п. Б) числа значений у  $\rho_{YX}$ . Отсюда вытекает, что величина коэффициента регрес-

сии ничего не говорит о важности причины  $X$  для  $Y$  (об интенсивности причинной зависимости).

Г. Проанализируем отражение зависимости двух признаков из-за общей причины в статистической связи между ними. В биологии широко распространено предположение о детерминации комплексов признаков общими и специфическими (действующих лишь на отдельные признаки) факторами. Такая детерминация обеспечивает гибкость как при индивидуальном росте и развитии, так и в эволюционных преобразованиях. Пусть, например, зависимость между признаками  $Y_1$  и  $Y_2$  обусловлена единственной общей причиной  $P$  при наличии случайных специфических факторов  $A, B, C$  и  $D, E$  (см. рис. 2) для первого и второго признаков соответственно.

В простейшей формализации данной картины признаки выражаются как соответствующие суммы общих и специфических факторов-причин (как линейные комбинации факторов).

Рассмотрим в данной ситуации соотношения между регрессиями и причинными связями при независимости случайных факторов-причин, суммирующихся в признаке (эта ситуация допускает простые точные вычисления коэффициен-

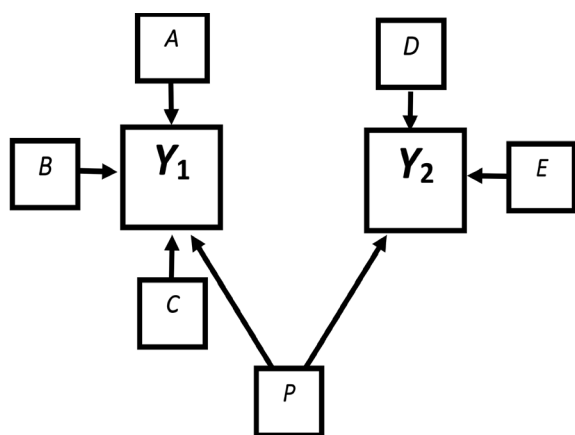


Рисунок 2. Концептуальная схема статистической зависимости признаков  $Y_1$  и  $Y_2$ . Зависимость обусловлена общей причиной  $P$  при наличии специфических факторов  $A, B, C$  и  $D, E$ , стрелки указывают на направление причинных влияний

Figure 1. Conceptual scheme of the statistical dependence of traits  $Y_1$  and  $Y_2$ . The dependence is due to a common cause  $P$  in the presence of specific factors  $A, B, C$  and  $D, E$ , arrows indicate the direction of causal influences

тов регрессии и корреляции). Если выбрать из всей совокупности часть индивидуумов со значением признака  $Y_1$ , отклоняющимся от среднего значения, то в ней ожидаем соответствующих средних отклонений каждого из влияющих на  $Y_1$  факторов (специфических  $A, B, C$  и общего фактора  $P$ ). При этом среднее значение признака  $Y_2$  в отобранной части будет отклоняться от общего среднего для всей совокупности (регрессия), но только за счет индуцированных изменений общего фактора  $P$  (поскольку другие изменившиеся специфические факторы  $A, B, C$  не влияют на  $Y_2$ ).

Из этих качественных соображений [Пасек, 2001] следует, что регрессия признака  $Y_2$  по признаку  $Y_1$  показывает изменения  $Y_2$  только за счет общих с  $Y_1$  причин (у нас из-за единственной причины  $P$ ) при изменении признака  $Y_1$  за счет всех влияющих на него факторов (в том числе специфических  $A, B, C$ ).

Понятно, что если рассматривать другую регрессию (признака  $Y_1$  от признака  $Y_2$ ), то получим иную линию, поскольку теперь изменения  $Y_1$  будут только за счет общих с  $Y_2$  причин при изменении признака  $Y_2$  за счет всех влияющих на  $Y_2$  факторов, включая свои специфические  $D, E$  не участвующие в предыдущей регрессии. Получается, что ни одна из двух линий регрессии не совпадает с зависимостью обоих признаков от общего фактора. Искомая функциональная связь двух признаков получается, если рассматривать одновременные изменения  $Y_1$  и  $Y_2$  в результате изменений общей причины  $P$  [Пасек, 2001; Пасек, 2012].

Регрессионная зависимость не дает причинный закон, а использование коэффициентов регрессии самих по себе для характеристики причинных закономерностей малопродуктивно. Таким образом, форма статистической связи (расположение линии регрессии) для зависящих от общей причины признаков, вообще говоря, не соответствует форме причинной зависимости, хотя их направление (знак) совпадают. Количественно различия между регрессионной и причинной зависимостями рассматриваются в приложении.

Итак, при одних и тех же причинных закономерностях возможны разные коэффициенты регрессии и корреляции, а при одном и том же наборе таких коэффициентов возможны разные причинные зависимости.

Простота приведенных выше схем влияния причинных факторов (законов природы) не позволяет экстраполировать полученные для них результаты на гораздо более сложные биологические феномены. Однако качественное понимание того, как могут отражаться причинные закономерности в статистических показателях, во многом сохраняется. Далее рассмотрим соотношение более сложных биометрических моделей с причинными зависимостями.

#### *Множественная линейная регрессия*

Одним из наиболее распространенных методов многомерного анализа является множественная линейная регрессия (см., например, [Кендалл, Стюарт, 1973; Дерябин 1983]). Пусть задача состоит в причинной интерпретации статистических связей некоторого множества признаков. Здесь можно попытаться объяснить каждую из статистических зависимостей, а можно для более глубокого понимания пойти дальше и рассматривать связи при исключении одного, двух и т. д. признаков. В биометрии статистические зависимости между количественными признаками при исключении одного или ряда других характеризуются частными коэффициентами регрессии или корреляции.

Например, под *множественной линейной регрессией* понимается статистическая зависимость среднего значения *зависимого признака*, скажем,  $Y_0$  от значений ряда других  $Y_1, \dots, Y_n$ , рассматриваемых как *признаки-предикторы*:

$$Y_0 = a + b_1 Y_1 + b_2 Y_2 + \dots + b_n Y_n.$$

Коэффициент  $b_i$  называется *частным коэффициентом регрессии*  $Y_0$  от  $Y_i$  (регрессии при условии исключения прочих признаков-предикторов). *Под исключением признака подразумевается ситуация, когда при изучении статистической связи между другими двумя (и более) признаками исключаемый принимает постоянное значение* и, тем самым, зависимость между остальными признаками не может быть вызваны им. Более подробно данная формула интерпретируется следующим образом. Если выбрать совокупность индивидуумов с фиксированными значениями случайно варьирующих признаков  $Y_1, Y_2, \dots, Y_n$ , то в данной совокупности у  $Y_0$  сохраняется какая-то изменчи-

вость с некоторым распределением. Среднее значение  $Y_0$  в совокупности в зависимости от значений фиксированных признаков находится по приведенной формуле линейной регрессии.

Нередко исключение признаков трактуется как прием для выявления эффекта отдельного из предикторов (отражаемого соответствующим частным коэффициентом регрессии  $b_i$  или частным коэффициентом корреляции) на зависимый признак в чистом виде при удалении косвенных влияний через другие предикторы. Однако частным показателям связи в общем случае трудно придать реальный причинный смысл.

Обсудим, в какой степени соображения о мешающих и косвенных влияниях исключаемого признака соотносятся с линейной причинной детерминацией рассматриваемого комплекса признаков скрытыми общими и специфическими причинами. Можно ли при использовании частных зависимостей ожидать более яркого проявления и разделения влияний отдельных причин, так как исключение признаков способно ослабить или даже устранить одну из них и подчеркнуть другие. Может ли величина частных показателей связи и их сравнение с безусловными показателями прояснить характер элементарных причинных влияний, скрытый под статистическими зависимостями в многофакторной структуре?

Далее ограничимся для простоты анализом эффектов исключения только одного признака, поскольку оно приводит к изменениям причинной структуры качественно такого же типа, как при исключении нескольких. Поэтому получаемые при исключении одного признака выводы, допустим, в отношении регрессии и корреляции в качественном отношении имеют общий характер, но при этом достигается выигрыш в простоте.

В рамках предположения о детерминации признака  $Y_0$  рядом причин устранение изменчивости  $Y_0$  дает эффекты двоякого рода.

1. Исключение признака приводит не к исключению изменчивости влияющих на него причин, а, как известно, лишь к уменьшению их варибельности (согласно теории статистики, дисперсия отдельной причины падает в  $1 - \rho^2$  раз, где  $\rho$  есть коэффициент корреляции данной причины с признаком). Так как при статистической связи с исключаемым признаком изменчивость причин все же остается, то и *остается вклад причин в согласованную изменчивость, но в (по-разному) ослабленном виде*.

2. При сохраняющейся изменчивости причин всякое изменение одной из них должно ком-

пенсироваться соответствующими изменениями других, чтобы значение исключаемого признака оставалось постоянным. Тем самым между всеми влияющими на исключаемый признак причинами *возникает ложная статистическая связь с ложным вкладом в связь между признаками*, под которой нет никакого реально существующего (а не искусственно вызванного исключением) основания. Появление ложных корреляций *искажает существующие зависимости вплоть до проявления связи признаков там, где ее не существует*.

Уменьшение изменчивости какой-либо причины оказывает неоднозначное влияние на такие статистические показатели, как коэффициенты регрессии и корреляции. Оно изменяет не только вклад в ковариацию признаков (числитель указанных показателей), но и в соответствующие средние квадратические отклонения (в знаменателе). Итог для частных коэффициентов зависит от соотношения изменений и величин числителя и знаменателя, определяемых качественной причинной структурой и количественными характеристиками связей с причиной. К одним и тем же значениям безусловных и частных показателей связи могут привести разные (и *неизвестные заранее*) схемы и разные значения причинных влияний. В точности эти же эффекты присущи исключению не одного, а ряда признаков. Соответственно нельзя придавать однозначный причинный смысл коэффициентам уравнения множественной регрессии (частным коэффициентам регрессии) и результатам использования этого уравнения.

*В целом, коэффициенты множественной регрессии отражают не только наложенные друг на друга скрытые причинные влияния, но и зависят от дисперсий причин при неизвестной степени их уменьшения из-за исключения признаков, вызывающего также количественно неизвестные ложные зависимости между причинами.* Использование статистических зависимостей (как безусловных, так и получаемых при исключении ряда признаков) для выявления отдельных элементарных причин и их эффектов невозможно без дополнительной информации. В конце концов, *частные коэффициенты регрессии или корреляции выражаются через привычные безусловные коэффициенты связей, которые могут принимать разные значения при одних и тех же причинных зависимостях*, как показано выше. Следовательно, при этом обычно получают разные частные коэффициенты.

В итоге для общего случая использование множественной регрессии в целях причинного анализа практически бесполезно.

#### *Метод главных компонент*

Метод главных компонент является одним из популярных инструментов многомерного статистического анализа (см., например, [Кендалл, Стюарт, 1976; Дерябин, 1983]). Обычно, чтобы избежать затруднений из-за измерений признаков в разных шкалах (скажем, при измерениях длины и веса), используются данные о корреляциях между признаками. *Главные компоненты определяются в виде сумм признаков с соответствующими весами и обратно – признаки являются некоторыми суммами главных компонент. Сами главные компоненты интерпретируются как скрытые и недоступные напрямую измерению фундаментальные свойства, лежащие в основе наблюдаемой изменчивости.*

Так как значения главных компонент равны соответствующим суммам значений признаков, а признаки, в свою очередь, выражаются как взвешенные суммы скрытых причин, то и сами главные компоненты представляют собой суммы причин, определяющих признаки. Поэтому главные компоненты можно рассматривать как некоторые новые признаки, линейно детерминированные теми же самыми причинами, что и исходные, и интерпретация взаимоотношений главных компонент несколько не проще интерпретации рассмотренных ранее наблюдаемых признаков.

Противоречия результатов анализа главных компонент и причинной структуры видны из следующего примера. Если общие причины влияют в одном направлении на рассматриваемые признаки, то матрица корреляций между ними будет состоять из положительных коэффициентов. Из теории положительных матриц следует, что при этом все нагрузки (влияния статистического фактора) первой главной компоненты положительны. Тогда у следующих компонент *обязательно* имеются нагрузки разных знаков, так как по определению все главные компоненты независимы (ортогональны). Значит, для каждой следующей главной компоненты найдутся признаки с разными знаками зависимости от нее. Однако предполагалось, что причин, противоположно



влияющих на признаки, нет. Значит, здесь главные компоненты не могут рассматриваться как причинные факторы, так как дают ложные результаты, противоречащие реальным влияниям причин на признаки.

Прямое изучение соотношений между результатами многомерного анализа и причинной детерминацией количественных признаков возможно с помощью имитационных экспериментов (статистических испытаний). В них с помощью компьютера генерируются случайные "выборки наблюдений" при заданных причинных связях. Полученные выборки можно анализировать многомерными методами и проверять, насколько статистические результаты отражают лежащую в основе генерированных данных заранее заданную причинную структуру.

Подобный подход использовался для изучения возможности судить о причинной структуре с помощью анализа общих главных компонент [Houle, 2002] для нескольких популяций с обсуждением эволюционно-генетических приложений. Оказалось, что результаты такого статистического анализа часто противоречат биологической интуиции, существует тенденция обнаруживать сходную структуру при разной причинной и находить различия, когда причинная структура одинакова.

Суммируя результаты, мы можем сказать, что метод главных компонент дает лишь удобный способ *описания* совместной изменчивости признаков.

#### Факторный анализ

Предположения модели факторного анализа (см., например, [Иберла, 1980; Дерябин, 1983]) находятся в тесном согласии с используемыми в настоящей работе предпосылками о линейной детерминации признаков общими и специфическими факторами. При этом общие факторы могут, в свою очередь, определяться факторами следующего порядка и т. д. По первому впечатлению кажется, что факторный анализ наконец-то дает возможность статистическими методами решать задачи о влиянии факторов-причин на признаки и оценивать значения факторов у отдельных индивидуумов.

Напомним, что существует бесконечное множество значений корреляций признаков с ненаблюдаемыми причинами, порождающих наблюдаемые корреляции между признаками, в том числе *не только* класс тех множеств, которые переходят одно в другое при соответствующем

повороте осей координат. В факторном анализе внимание концентрируется на таком классе. Попытки выбрать в нем единственное множество основываются на каких-либо постулируемых критериях. Например, согласно принципу *простой структуры* поворотом осей достигается положение, когда каждый фактор оказывает максимальное влияние на группу признаков, предположительно обладающими свойствами определенного рода.

Можно сказать, что при простой структуре каждый фактор влияет по возможности на наибольшее число признаков, а каждый признак определяется небольшим числом факторов. Получаемая при использовании простой структуры картина огрубленно соответствует разбиению признаков на группы, корреляции внутри которых больше, чем межгрупповые. В биологии данную картину для признаков организма нельзя признать универсальной и теоретически обоснованной.

При использовании канонических переменных мы опять представляем их как новые признаки, сконструированные в виде взвешенных сумм исходных. В рамках детерминации последних общими и специфическими причинами новые признаки по характеру детерминации не отличаются от исходных, и поэтому их нельзя отождествлять с причинными факторами по прежде изложенным основаниям. Аналогично, как бы ни была привлекательна попытка причинной интерпретации канонических переменных, по ним также нельзя судить о реальных причинах. Достаточно напомнить, что канонические переменные определяются по данным о корреляциях признаков, которые могут принимать различные значения и, соответственно, определять разные канонические переменные при одних и тех же причинных влияниях.

#### Заключение

Резюмируя, мы можем сказать, что такой простой показатель зависимости как величина коэффициента корреляции ничего не говорит о биологической важности соответствующей статистической связи, и утверждения типа «... *большая корреляция элемента кисти с другим признаком человеческого тела является указанием на большую морфологическую зависимость между обоими признаками*» [Ярхо, 1926, с. 53] не являются корректными. Подкупающие предположения более сложных моделей главных компонент, факторного анализа и канонических переменных находятся в тесном согласии с используемыми в

настоящей работе предпосылками о линейной детерминации признаков общими и специфическими факторами. Однако даже в тех случаях, когда удается правильно оценить качественный характер причинной детерминации признаков (направление влияния причинного фактора), связь статистических факторов с признаками не совпадает с причинной количественно. Она даже может отличаться качественно, что особенно наглядно проявляется в соотношениях между главными компонентами и причинными зависимостями.

Кроме того, как уже мы неоднократно подчеркивали, при одних и те же причинных влияниях на признаки коэффициенты корреляции и ковариации между ними могут принимать бесконечно много значений. Попытки выбрать для них единственную структуру причинных влияний статистическими методами основываются на каких-либо постулируемых критериях (например, на максимуме дисперсии, простой структуре, максимальной корреляции), не имеющих биологического обоснования, и априори их нельзя считать выполненными в конкретных исследованиях. Поэтому не обоснованы утверждения типа «... (статистические) факторы имеют реальный смысл, т. е. являются математическим выражением конкретной деятельности внутренних систем организма: гуморальной, нервной» [Русалов, 1979, с. 61]. Более чем полувекковой промежуток времени, разделяющий цитируемые высказывания, говорит о живучести подобных представлений. Согласно Дж. Тэннеру [Тэннер, 1979, с. 451] «Факторы представляют собой, по существу, статистические величины, такие же, как средние квадратические отклонения. А *rigor* их нельзя связывать с генетическими и физиологическими механизмами». Как пишут известные специалисты по теории статистики [Кендалл, Стюарт, 1973, с. 375], «... причинная зависимость не может быть выведена ни из какого наблюдаемого совместного изменения, даже самого тесного».

В связи с приведенной аргументацией возникает вопрос о конечной ценности биометрического описания данных в антропологии. Есть впечатление о некотором культе такого подхода, понимаемого во многом чисто статистически. Биометрический анализ дает описательную, а не причинную картину изменчивости. Описательный этап является только начальным при изучении любых феноменов, а конечной целью исследования служит построение и проверка гипотез о причинных механизмах, лежащих в их основе. Хотя помимо статистического описа-

ния в антропологических работах всегда привлекаются соображения о том, чем могут быть обусловлены выявляемые связи, анализ чаще всего базируется на статистической аргументации. Однако *ни форма статистической связи между признаками (расположение линий регрессии), ни ее теснота (величина коэффициента корреляции) не связаны жестко с характером причинной детерминации признаков.*

**Замечание.** Размер составного признака индивидуума складывается из значения его части, рассматриваемой как общий фактор, плюс значения специфических факторов (других частей). Результат измерения в технике состоит из точного значения признака неживого объекта плюс независимая случайная ошибка (из аналогов значений общего и специфического факторов). Поэтому результаты анализа роли ошибок измерения [Кендалл, Стюарт, 1973, Гл. 29], казалось бы, полностью приложимы и к биометрическим исследованиям популяций. Однако между контролируемыми с помощью повторных измерений случайными ошибками в технике и неконтролируемыми специфическими биологическими факторами есть *принципиальные различия.*

Скорректировать искажения функциональной зависимости, вносимые ошибками измерения, можно с помощью учета их дисперсии (дисперсии специфического фактора), которая оценивается по повторным измерениям. Размер неживого измеряемого объекта не меняется, но при каждом измерении к нему добавляется независимая случайная ошибка, играющая роль специфического фактора. В результате с помощью повторных измерений тех же самых объектов можно оценить дисперсию подобных «специфических» факторов (ошибок измерения).

В рассматриваемой биологической ситуации при повторных измерениях можно выделить два типа специфических факторов – биологические факторы и ошибки измерения. Каждое повторное измерение признака индивидуума включает все его составные части, поэтому наблюдаемое значение признака представляет собой сумму реализованных и *фиксированных* у индивидуума значений как общих, так и биологических специфических факторов (случайных в популяции, но постоянных у каждого отдельного индивидуума) плюс *варьирующая* реальная ошибка измерения. Повторные измерения отдельного индивидуума не позволят оценить необходимую для коррекции дисперсию специфического биологического фактора, обусловленную изменчивостью

между разными индивидуумами, а дают возможность найти только дисперсию реальной ошибки измерения. Поэтому избежать искажений при оценивании причинной биологической зависимости от общего фактора с помощью повторных измерений невозможно.

Анализируемые выше биометрические методы изложены, например, в выгодно отличающейся доступностью подачи материалов книге В.Е. Дерябина [Дерябин, 1983]. Интерпретация в ней статистических факторов типа один из них является показателем общей длины тела, другой показателем формы [Дерябин, 1983, с. 47–48] может натолкнуть читателя на мысль связать факторы с реальными причинами (факторами общего и специфического роста с соответствующими количественными зависимостями признаков от факторов). В свете вышеизложенного данная интерпретация, в сущности, просто фигура речи, удобная для ориентации в результатах статистической обработки.

Еще раз отметим, что сделанные выводы не зависят от вида распределения признаков, а также от размера выборки (относятся к генеральной совокупности). Они не являются биологической спецификой и справедливы для феноменов любой природы, детерминированных общими и специфическими факторами.

## Библиография

- Дерябин В.Е. Многомерная биометрия для антропологов. М.: Изд-во МГУ. 1983. 227с.
- Иберла К. Факторный анализ. М.: Статистика. 1980. 398 с.
- Кендалл М., Стюарт А. Статистические выводы и связи. М.: Наука. 1973. 899 с.
- Кендалл М., Стюарт А. Многомерный статистический анализ и временные ряды. М.: Наука. 1976. 736 с.
- Пасеков В.П. Концептуальная схема соотношений причинного и статистического анализов на примере антропологических исследований // Бунаковские чтения. Тезисы докладов. Ч. I. М.: Старый сад, 2001. С. 96–99.
- Пасеков В.П. О теоретических проблемах биометрического и причинного подходов в популяционных исследованиях. I. Конспективное изложение. Москва: ВЦ РАН. 2005. 64 с.
- Пасеков В.П. О принципах решения проблемы размера и формы // Вестник Московского университета. Серия 23. Антропология, 2012. № 2. С. 17–28.
- Русалов В.М. Биологические основы индивидуально-психологических различий. М.: Наука. 1979. 352 с.
- Тэннер Дж. Рост и конституция человека / Биология человека. М.: Мир, 1979. С. 366–471.
- Ярхо А.И. О некоторых морфологических свойствах человеческой кисти // Русский антропологический журнал, 1926. Т. 15. Вып. 1–2. С. 50–74.

### Сведения об авторах

Пасеков Владимир Петрович, к.б.н., [pass40@mail.ru](mailto:pass40@mail.ru).

Поступила в редакцию 20.09.2022,  
принята к публикации 27.10.2022.

Passekov V.P.

*Federal Research Center “Computer Science and Control”  
of Russian Academy of Sciences, Vavilova str., 40, Moscow, Russia, 119991*

## ON THE APPLICATION OF BIOMETRICS IN ANTHROPOLOGY AND THE MAIN TASKS OF SCIENCE

**Introduction.** *The task of biometrics in anthropology is to describe and analyze the random variability inherent in a person as an object of study. At the modern time, multivariate biometric models are widely used for this, in which it is assumed that the observed quantitative organism traits are expressed as weighted sums of latent random factors (for example, principal component analysis, factor and discriminant analyses, canonical variables, etc.). The given assumption is consistent with the existence of general and specific biological factors-causes that determine the organism traits. Since the main task of science is to study the laws of cause-and-effect in nature, it is important to analyze the relationship between statistical and causal dependences.*

**Results and discussion.** *Statistical factors obtained as a result of data processing (such as the general size, shape, etc.) in the general case not only disagree with the causal factors, but may even contradict them. It is shown that the statistical coefficients of correlation, regression, and covariance, even in the simplest case of two traits randomly varying under the influence of one common cause and several specific factors, can take on infinitely many values under the same causal relationship, and with the same statistical pattern of variability, there are infinitely many causal explanations it.*

**Conclusions.** *Interpretation in relation to the cause-and-effect patterns of trait determination and other causal conclusions should not follow from the results of the application of one-dimensional and multidimensional biometric models that use the indicated statistical coefficients as input data.*

**Keywords:** biometrics; multivariate models; causal laws; regression; correlations; general and specific factors; anthropology

## References

Deryabin V.E. *Mnogomernaya biometriya dlya antropologov* [Multivariate biometrics for anthropologists]. Moscow, MSU Publ., 1983. 227 p. (In Russ.).

Iberla K. *Faktorenanalyse*. Berlin, Springer-Verlag, 1977.

Kendall M., Stuart A. *Statisticheskie vyvody i svyazi* [The Advanced Theory of Statistics. V. 2: Inference and Relationship]. London, Griffin & Co., 1961. 676 p.

Kendall M.G., Stuart A. *Mnogomernyj statisticheskij analiz i vremennye ryady* [The Advanced Theory of Statistics. V. 3: Design and Analysis, and Time-Series]. London & High Wycombe, Charles Griffin & Co. Ltd. 1976.

Passekov V.P. *Konceptual'naya skhema sootnoshenij prichinnogo i statisticheskogo analizov na primere antropologicheskikh issledovaniy* [Conceptual scheme of relationships between causal and statistical analyzes on the example of anthropological research]. In *Bunakovskie chteniya. Tezisy dokladov. Chast' I* [Bunak Reading. Abstracts. Prt 1]. Moscow. Staryj sad Publ., 2001, pp. 96–99. (In Russ.).

Passekov V.P. *O teoreticheskikh problemah biometricheskogo i prichinnogo podhodov v populyacionnykh issledovaniyah. I. Konspektivnoe izlozhenie*. [On theoretical problems of biometric and causal approaches in population studies. I. Concize treating]. Moscow, VC RAN Publ., 2005. 64 p. (In Russ.).

Passekov V.P. *O principah resheniya problemy razmera i formy* [On the principles of solving the problem of size and shape]. *Moscow University Anthropology Bulletin* [Vestnik Moskovskogo universiteta Seriya 23. Antropologiya], 2012, 2, pp. 17–28. (In Russ.).

Rusalov V.M. *Biologicheskie osnovy individual'no-psihologicheskikh razlichij* [Biological basis of individual psychological differences]. Moscow, Nauka Publ., 1979. 352 p. (In Russ.).

Tanner J.M. *Growth and constitution*. In *Human biology*. Oxford, Oxford Univ. Press, 1977, pp. C. 366–471.

Yarkho A.I. *O nekotorykh morfologicheskikh svoystvakh chelovecheskoy kisti* [On some morphological properties of the human wrist]. *Russkij antropologicheskij zhurnal* [Russian Journal of Anthropology], 1926, 15 (1–2), pp. 50–74. (In Russ.).

Houle D., Mezey J., Galperin P. *Interpretation of the results of common principal analysis*. *Evolution*, 2002, 56 (3), pp. 433–440.

Rao C.R. *Advanced Statistical Methods in Biometric Research*. New York, Wiley. 1952. 390 p.

## Information about Authors

*Passekov Vladimir Petrovich, PhD, pass40@mail.ru.*

## Приложение Appendix

В данном приложении приводится вывод формул для характеристик статистических связей признаков в использованных выше простейших схемах их детерминации и суммируются полученные результаты. В рамках существования общих и специфических факторов рассматриваются элементарные зависимости между причинами и следствием, где определение важности отдельных причин достаточно наглядно и возможно легко сравнить статистическую и причинную зависимости количественно. Повторим, что получаемые далее соотношения относятся к свойствам популяции как генеральной совокупности (выборочные ошибки отсутствуют), а на распределение признаков не наложено особых ограничений.

**А.** Ранее мы показали, что *какова бы ни была фиксированная величина коэффициента корреляции (теснота статистической связи)*

*между фактором-причиной  $X_i$  и признаком-следствием  $Y$ , равным  $\sum X_i$ , относительная величина среднего вклада фактора в среднее значение признака (важность данной причины) может быть как угодно близкой к нулю, так и произвольно близкой к единице. Значит, здесь коэффициент корреляции не характеризует важность причины для следствия.*

**Б.** Теперь обратимся к случаю, когда признак  $Y$  является суммой независимых причин  $\{X_i\}$  с весами  $\{c_i\}$  и дисперсиями  $\{V_i\}$ , т. е.  $Y = \sum c_i X_i$ . Найдем ковариацию признака  $Y$  с отдельной причиной, скажем, с  $X_1$ . По свойствам ковариации, когда один из аргументов ( $Y$ ) у ковариации  $\text{COV}(Y, X_1)$  является суммой случайных величин, то  $\text{COV}(Y, X_1)$  равна сумме ковариаций каждого из слагаемых с другим аргументом ( $X_1$ ). В случае трех причин (см. рис. 16)



$$\begin{aligned}
 Y &= c_1X_1 + c_2X_2 + c_3X_3; \\
 \text{COV}(Y, X_1) &= \text{COV}(c_1X_1 + c_2X_2 + c_3X_3, X_1) \\
 &= \text{COV}(c_1X_1, X_1) + \text{COV}(c_2X_2, X_1) \\
 &\quad + \text{COV}(c_3X_3, X_1) \\
 &= c_1\text{COV}(X_1, X_1) = c_1V_1. \quad (1)
 \end{aligned}$$

Здесь  $V_1$  обозначает дисперсию фактора-причины  $X_1$ , и мы учли, что ковариации независимых факторов равны нулю, а постоянный множитель ( $c_1$ ) у  $\text{COV}(c_1X_1, X_1)$  можно вынести из-под знака ковариации. Кроме того, напомним, что по определению

$$\text{COV}(X_1, X_1) = E\{(X_1 - E\{X_1\})^2\} = V_1.$$

В итоге для любого количества факторов в сумме  $Y = \sum X_i$  ковариация  $\text{COV}(Y, X_1)$  признака  $Y$  с отдельной причиной ( $X_1$ ) при фиксированном причинном влиянии ( $c_1$ ) может принимать бесконечное множество значений  $c_1V_1$ , в зависимости от величины дисперсии  $V_1$  фактора-причины  $X_1$ .

Напомним также, что дисперсия  $V_Y$  признака  $Y$ , равного сумме независимых случайных величин  $c_iX_i$ , будет суммой дисперсий слагаемых, постоянный множитель  $c_i$  можно выносить из-под символа дисперсии, и при этом его следует возвести в квадрат ( $V(cX) = c^2V(X)$ ). Значит,

$$V_Y = c_1^2V_1 + c_2^2V_2 + c_3^2V_3 = W_1 + W_2 + W_3, \quad W_i \equiv c_i^2V_i,$$

где  $W_i$  интерпретируется как компонент дисперсии признака  $Y$ , индуцированный  $i$ -й причиной.

Отсюда коэффициент корреляции  $\rho$  признака  $Y$  с одной из причин ( $X_1$ ) находится как

$$\rho_{YX_1} \equiv \text{COV}(Y, X_1) / \sqrt{V_Y V_1} = c_1V_1 / \sqrt{V_Y V_1} = c_1\sqrt{V_1/V_Y},$$

$$\rho_{YX_1}^2 = c_1^2V_1/V_Y \equiv W_1/V_Y. \quad (2)$$

Таким образом, квадрат коэффициента корреляции признака с одной из влияющих на него независимых причин равен отношению компонента дисперсии признака из-за этой причины ко всей дисперсии признака. *Значит, коэффициент корреляции признака  $Y$  с причиной  $X_1$ , влияющей на него по одному и тому же неизменному закону  $Y = c_1X_1$ , может принимать разные значения в зависимости от*

*дисперсии  $V_1$  причины  $X_1$  (и дисперсий других причин, определяющих величины компонентов  $\{W_i\}$ , суммирующихся в общей дисперсии  $V_Y$  признака  $Y$ ). Следовательно, указанный коэффициент корреляции  $\rho_{YX_1} = c_1\sqrt{V_1/V_Y}$  не характеризует закон влияния причины  $X_1$  на признак  $Y$ .*

**В.** Заметим, что можно попробовать взять в качестве показателя важности фактора не коэффициент корреляции, а такую естественную *характеристику важности* ( $c_i$ )  $i$ -го фактора, которая показывает как сильно меняется признак (скорость его изменений) при варьировании данной причины. Такая функциональная (неслучайная) зависимость как  $Y = cX$  графически отображается проходящей через начало координат прямой с угловым коэффициентом  $c$ . Отсюда можно выразить  $X$  как

$$X = \frac{1}{c}Y.$$

Если данная функциональная связь нарушается присутствием случайных факторов  $\{X_k\}$ , то  $Y = cX + \sum c_kX_k$ . Здесь кажется, что в качестве аналога коэффициента  $c$  можно рассматривать коэффициент регрессии  $b_{YX}$  признака  $Y$  от причины  $X$  (угловой коэффициент соответствующей линии регрессии). Согласно (1)

$$b_{YX} \equiv \text{COV}(Y, X)/V_X = cV_X/V_X = c.$$

Итак, в данном случае *причинная функциональная связь признака  $Y$  с причиной  $X$  и регрессионная связь условного среднего значения  $Y$  от  $X$  совпадают независимо от дисперсий факторов:*

$$Y = cX, \quad Y = b_{YX}X = cX.$$

Если рассматривать коэффициент другой регрессии  $b_{XY}$  причины  $X$  от признака  $Y$ , то согласно (2)

$$b_{XY} \equiv \text{COV}(Y, X)/V_Y = cV_X/V_Y = \frac{1}{c}c^2V_X/V_Y$$

$$= \frac{1}{c}W_X/V_Y = \frac{1}{c}\rho_{YX}^2.$$

Так как величина коэффициента корреляции не выше единицы, то абсолютная величина углового коэффициента  $b_{YX}$  линии регрессии причины  $X$  от зависимого от нее признака  $Y$  обычно меньше, чем у коэффициента  $1/c$  линии функциональной связи.

Мы получили, что *причинная функциональная связь и регрессионная зависимость причины от зависимого признака не совпадают*:

$$X = \frac{1}{c}Y, X = b_{YX}Y = \left(\frac{1}{c} \rho_{YX}^2\right)Y.$$

Г. Пусть теперь функциональная зависимость между двумя признаками  $Y_1$  и  $Y_2$  вызвана общей причиной  $P$ , скажем,

$$Y_1 = cP, Y_2 = dP.$$

Если из первого уравнения выразить  $P$  как функцию  $Y_1$  и подставить ее во второе уравнение, то найдем зависимость  $Y_2$  от  $Y_1$  (аналогично находится зависимость  $Y_1$  от  $Y_2$ ):

$$P = \frac{1}{c}Y_1; Y_2 = dP = \frac{d}{c}Y_1, Y_1 = \frac{c}{d}Y_2.$$

Далее обратимся к ситуации, когда данная функциональная связь  $Y_1$  с  $Y_2$  нарушается присутствием независимых случайных факторов  $\{X_k\}$  и  $\{Z_j\}$ , скажем,

$$Y_1 = cP + \sum c_k X_k, Y_2 = dP + \sum d_j Z_j,$$

$$\text{COV}(Y_1, Y_2) \equiv \text{COV}_{12} = \text{COV}(cP, dP) = cdV_P.$$

$$\rho(Y_1, Y_2) \equiv \rho_{12} = \text{COV}_{12} / \sqrt{V_1 V_2}$$

$$= cdV_P / \sqrt{V_1 V_2} = c \sqrt{V_P/V_1} d \sqrt{V_P/V_2}.$$

Вспомним только что полученное выражение (2) для коэффициента корреляции причины с признаком. Для  $P$  и  $Y_1$  оно принимает вид

$$\rho_{Y_1 P} = c \sqrt{V_P/V_1}, \text{ а } \rho_{Y_2 P} = d \sqrt{V_P/V_2}.$$

Отсюда  $\rho_{12} = \rho_{Y_1 P} \rho_{Y_2 P}$ .

Перейдем к вычислению коэффициентов регрессии. Начнем с коэффициента  $b_{21}$  регрессии признака  $Y_2$  от  $Y_1$ .

$$b_{21} \equiv \text{COV}_{12}/V_1 = cdV_P/V_1 = \frac{d}{c} c^2 V_P/V_1 = \frac{d}{c} \rho_{Y_1 P}^2$$

согласно (2). Таким образом, абсолютная величина углового коэффициента  $b_{21}$  линии регрессии принятого за зависимый признака  $Y_2$  от признака  $Y_1$  обычно меньше, чем у углового коэффициента

$\frac{d}{c}$  соответствующей функциональной связи  $Y_2$  от признака  $Y_1$ , поскольку  $\rho^2 \leq 1$ . Геометрически это означает, что рассматриваемая линия регрессии получается поворотом линии функциональной зависимости признаков  $Y_2 = \frac{d}{c}Y_1$  из-за общей причины к оси координат зависимого признака (параллельно ей).

Аналогично коэффициент  $b_{12}$  регрессии зависимого признака  $Y_1$  от  $Y_2$ , меньше по абсолютной величине углового коэффициента  $\frac{c}{d}$  соответствующей функциональной связи  $Y_1$  от  $Y_2$ :

$$b_{12} \equiv \text{COV}_{12}/V_2 = cdV_P/V_2 = \frac{c}{d} d^2 V_P/V_2 = \frac{c}{d} \rho_{Y_2 P}^2.$$

Мы получили, что *функциональная связь признаков, вызванная общей причиной, и регрессионные зависимости при наложении на эту связь случайных специфических факторов не совпадают*:

$$Y_2 = \frac{d}{c}Y_1, Y_2 = b_{21}Y_1 = \frac{d}{c} \rho_{Y_1 P}^2 Y_1;$$

$$Y_1 = \frac{c}{d}Y_2, Y_1 = b_{12}Y_2 = \frac{c}{d} \rho_{Y_2 P}^2 Y_2.$$

На плоскости с осями координат  $OY_1$  и  $OY_2$  одна из линий регрессии получается поворотом линии функциональной зависимости признаков из-за общей причины к первой из осей координат, а другая к второй. Величины поворотов вокруг центра с координатами  $(E\{Y_1\}, E\{Y_2\})$  обычно различны и изменяются при изменениях дисперсий общего и специфического факторов, а линия зависимости признаков из-за общей причины лежит между обеими регрессиями, и ее положение постоянно.

Таким образом, *ни теснота статистической связи (квадрат коэффициента корреляции), ни регрессионные зависимости не характеризуют ни скрытые причинные закономерности, ни важность причины согласованной совместной изменчивости признаков*.