

Academia. Архитектура и строительство, № 3, стр. 155–160.
Academia. Architecture and Construction, no. 3, pp. 155–160.

Исследования и теория
Научная статья
УДК 691.328.1, 693.557
DOI: 10.22337/2077-9038-2024-3-155-160

Обобщённые коэффициенты корреляции и их применимость для определения свойств бетона при испытаниях

Арленинов Пётр Дмитриевич (Москва). Кандидат технических наук. Научно-исследовательский, проектно-конструкторский и технологический институт бетона и железобетона НИЦ «Строительство» (Россия, 109428, Москва, 2-я Институтская ул., 6, корп. 5. НИИЖБ им. А.А. Гвоздева); кафедра железобетонных и каменных конструкций Национального исследовательского Московского государственного строительного университета (Россия, 129337, Москва, Ярославское шоссе, 26. НИУ МГСУ). Эл. почта: niizhb_lab8@mail.ru

Крылов Сергей Борисович (Москва). Доктор технических наук, академик РААСН. Лаборатория механики железобетона Научно-исследовательского, проектно-конструкторского и технологического института бетона и железобетона НИЦ «Строительство» (Россия, 109428, Москва, 2-я Институтская ул., 6, корп. 5. НИИЖБ им. А.А. Гвоздева). Эл. почта: niizhb_lab8@mail.ru

Калмакова Полина Сергеевна (Москва). Лаборатория механики железобетона Научно-исследовательского, проектно-конструкторского и технологического института бетона и железобетона НИЦ «Строительство» (Россия, 109428, Москва, 2-я Институтская ул., 6, корп. 5. НИИЖБ им. А.А. Гвоздева). Эл. почта: niizhb_lab8@mail.ru

Корнюшина Марина Петровна (Москва). Лаборатория механики железобетона Научно-исследовательского, проектно-конструкторского и технологического института бетона и железобетона НИЦ «Строительство» (Россия, 109428, Москва, 2-я Институтская ул., 6, корп. 5. НИИЖБ им. А.А. Гвоздева).

Аннотация. Работа посвящена выявлению геометрического смысла коэффициента парной корреляции, а также множественного коэффициента корреляции. Сделаны широкие обобщения этих коэффициентов. Предложен новый вид множественного коэффициента корреляции – между набором независимых переменных и набором зависимых переменных. Приведены примеры использования предложенных обобщённых коэффициентов корреляции для определения свойств бетона при испытаниях. Показаны их преимущества перед обычными коэффициентами корреляции.

Ключевые слова: корреляция, коэффициент корреляции, парная корреляция, множественная корреляция

Для цитирования. Арленинов П.Д., Крылов С.Б., Корнюшина М.П., Калмакова П.С. Обобщённые коэффициенты корреляции и их применимость для определения свойств бетона при испытаниях // Academia. Архитектура и строительство. – 2024. – № 3. – С. 155–160. – DOI: 10.22337/2077-9038-2024-3-155-160.

Generalized Correlation Coefficients and Their Applicability for Determining the Properties of Concrete during Testing

Arleninov Petr D. (Moscow). Candidate of Sciences of Technology. Research, Design and Technological Institute of Concrete and Reinforced Concrete of the "National Research Center 'Construction'" (6, build. 5, 2nd Institutskaia, Moscow, 109428, Russia. NIIZHB named after A.A. Gvozdev); Department of Reinforced Concrete and Masonry Structures of the National Research Moscow State University of Civil Engineering (Russia, 129337, 26, Yaroslavskoye Shosse, Moscow, Russia. NRU MGSU). E-mail: niizhb_lab8@mail.ru

Krylov Sergey B. (Moscow). Doctor of Sciences of Technology, Academician of RAACS. Laboratory of reinforced concrete mechanics of the Research, Design and Technological Institute of Concrete and Reinforced Concrete of the "National Research Center 'Construction'" (6, build. 5, 2nd Institutskaya, Moscow, 109428, Russia. NIIZHB named after A.A. Gvozdev). E-mail: niizhb_lab8@mail.ru

Kalmakova Polina S. (Moscow). Laboratory of reinforced concrete mechanics of the Research, Design and Technological Institute of Concrete and Reinforced Concrete of the "National Research Center 'Construction'" (6, build. 5, 2nd Institutskaya, Moscow, 109428, Russia. NIIZHB named after A.A. Gvozdev). E-mail: niizhb_lab8@mail.ru

Kornyushina Marina P. (Moscow). Laboratory of reinforced concrete mechanics of the Research, Design and Technological Institute of Concrete and Reinforced Concrete of the "National Research Center 'Construction'" (6, build. 5, 2nd Institutskaya, Moscow, 109428, Russia. NIIZHB named after A.A. Gvozdev).

Abstract. The work is devoted to identifying the geometric meaning of the paired correlation coefficient, as well as the multiple correlation coefficient. Broad generalizations of these coefficients are made. A new type of multiple correlation coefficient is proposed – between a set of independent variables and a set of dependent variables. Examples of the use of the proposed generalized correlation coefficients are given. Their advantages over conventional correlation coefficients are shown.

Keywords: correlation, correlation coefficient, pair correlation, multiple correlation

For citation. Arleninov P.D., Krylov S.B., Kalmakova P.S., Korniyushina M.P. Generalized Correlation Coefficients and Their Applicability for Determining the Properties of Concrete during Testing. In: *Academia. Architecture and Construction*, 2024, no. 3, pp. 155–160, doi: 10.22337/2077-9038-2024-3-155-160.

Введение

Понятие «корреляция» в настоящее время рассматривается как статистическая взаимосвязь двух или более случайных величин. Часто коэффициент парной корреляции (для двух величин) рассматривается как показатель близости этой взаимосвязи к функциональной зависимости. Корреляционный анализ также находит своё применение и в строительстве. В частности, коэффициенты корреляции применяются для оценки качества регрессионных зависимостей при обработке результатов экспериментов, для выявления взаимосвязей между различными исследуемыми величинами, для оценки точности расчётных моделей и т.д. Смысл этого понятия рассматривается исключительно с позиций статистики. Соответственно, в терминах статистики формулируются все положения корреляционного анализа, математические выражения коэффициентов корреляции строятся исключительно с использованием статистических параметров. Внимательное изучение этих выражений позволяет предложить иной подход к построению таких зависимостей и придать корреляции более широкий смысл.

Рассмотрение этого понятия исключительно с точки зрения статистики скрывает важные геометрические зависимости. В свою очередь, удобная геометрическая интерпретация позволяет сделать важные обобщения корреляции и применить понятие «корреляция» к решению целого ряда задач, связанных с оценкой состояния строительных конструкций.

Таким образом, основная задача проведённого исследования состоит в выявлении геометрического смысла коэффициента корреляции и обобщении его.

При сравнении основных формул, используемых в дисперсионном анализе, и формул, по которым вычисляются нормы

векторов, можно заметить много общего между ними. Это наводит на мысль о возможности применения элементов теории векторных пространств в дисперсионном анализе, а также в регрессионном и корреляционном анализе. Проведённые исследования позволяют сделать обобщения этого понятия с точки зрения теории векторных пространств. В свою очередь, такие обобщения дают возможность лучше понять смысл корреляции и распространить корреляционный анализ на новые виды задач.

Рассмотрение парного коэффициента линейной корреляции с точки зрения векторных пространств

В области строительства для оценки корреляционных связей наиболее широко применяется коэффициент линейной корреляции Пирсона. При этом для двух случайных переменных X и Y , между которыми предполагается корреляционная зависимость, определяется ковариация (1).

$$cov_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}. \quad (1)$$

Коэффициент корреляции при этом определяется как [1; 2]

$$r_{XY} = \frac{cov_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \quad (2)$$

здесь $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ и $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ – средние значения выборок X и Y .

При этом предполагается, что распределение случайных величин близко к нормальному.

Для выявления геометрического смысла коэффициента корреляции, перепишем формулу (2) по-другому:

$$r_{xy} = \sum_i \left[\frac{x_i - \bar{x}}{\sqrt{\sum_j (x_j - \bar{x})^2}} \cdot \frac{y_i - \bar{y}}{\sqrt{\sum_j (y_j - \bar{y})^2}} \right]. \quad (3)$$

Также рассмотрим два вектора a и b :

$$a = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \dots \\ x_n - \bar{x} \end{pmatrix} b = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \dots \\ y_n - \bar{y} \end{pmatrix}. \quad (4)$$

Координаты каждого из этих двух векторов представляют собой отклонение значений величин X и Y от их средних значений – это графически показано на рисунке 1 для трёх точек.

Тогда нормы этих векторов в евклидовом пространстве [3] определяются как

$$\|a\| = \sqrt{\sum_j (x_j - \bar{x})^2} \quad \|b\| = \sqrt{\sum_j (y_j - \bar{y})^2}. \quad (5)$$

Пронормируем векторы a и b . При этом получим единичные векторы. Обозначим их соответственно e_a и e_b

$$e_a = \begin{pmatrix} \frac{x_1 - \bar{x}}{\sqrt{\sum_j (x_j - \bar{x})^2}} \\ \dots \\ \frac{x_n - \bar{x}}{\sqrt{\sum_j (x_j - \bar{x})^2}} \end{pmatrix} e_b = \begin{pmatrix} \frac{y_1 - \bar{y}}{\sqrt{\sum_j (y_j - \bar{y})^2}} \\ \dots \\ \frac{y_n - \bar{y}}{\sqrt{\sum_j (y_j - \bar{y})^2}} \end{pmatrix}. \quad (6)$$

Сравнивая компоненты векторов e_a и e_b (6) с компонентами формулы (3), можно сделать вывод, что правая часть выражения (3) представляет собой скалярное произведение векторов e_a и e_b , то есть

$$r_{xy} = (e_a \cdot e_b). \quad (7)$$

Поскольку векторы e_a и e_b являются нормированными, то их скалярное произведение равно косинусу угла между ними

$$r_{xy} = \|e_a\| \|e_b\| \cos \alpha = 1 \cdot 1 \cdot \cos \alpha. \quad (8)$$

Таким образом, коэффициент линейной корреляции Пирсона является косинусом угла между векторами, компоненты которых состоят из отклонений двух случайных величин от их средних значений. По своей сути он выражает степень близости указанных векторов к линейной зависимости.

Обобщения коэффициента парной корреляции

Выяснив геометрический смысл коэффициента корреляции, можем обобщить его, расширив области возможных значений параметров, входящих в него.

1. Вид распределения случайных величин не имеет значения при предлагаемом подходе.

2. Поскольку коэффициент корреляции представляет собой косинус угла между векторами, то можно использовать любую функцию угла, а в некоторых случаях удобно пользоваться

не функцией угла, а самим углом α . Поскольку обычный коэффициент корреляции изменяется в пределах от -1 до $+1$, то удобнее в качестве данного обобщённого коэффициента рассматривать относительный угол, например: $r_\alpha = \frac{\alpha}{90^\circ}$ или $r_\alpha = \frac{\alpha}{\pi/2}$, где угол α измеряется соответственно в градусах или радианах (в этой записи предполагается, что угол приведён к величине в пределах от 0 до 90 градусов).

3. Как было показано выше, коэффициент корреляции Пирсона соответствует векторам отклонений значений переменных X и Y от их средних значений. Можно обобщить понятие коэффициента корреляции на случай, когда вычисления отклонений выполняются не относительно среднего значения каждой переменной, а относительно произвольной точки на числовой оси. Эта точка может быть связана с областью значений переменных X и Y (например, привязана к их средним значениям, как коэффициент корреляции Пирсона) или привязана к какой-то фиксированной точке на каждой числовой оси (например, к началу координат) (рис. 2).

В названии такого обобщённого коэффициента корреляции целесообразно указать название точки, относительно которой выполняются вычисления.

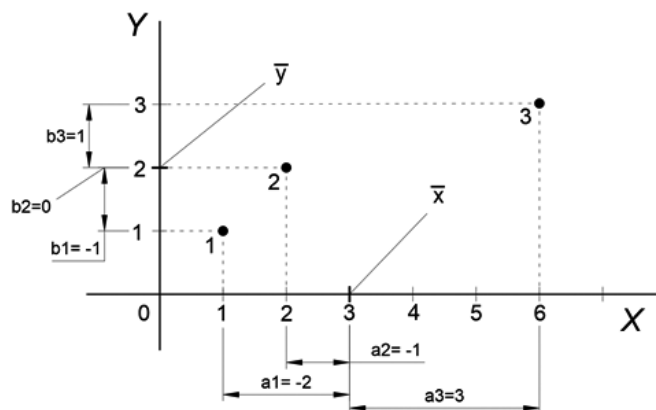


Рис. 1. Отклонение значений случайных величин X и Y от средних значений (координаты векторов a и b)

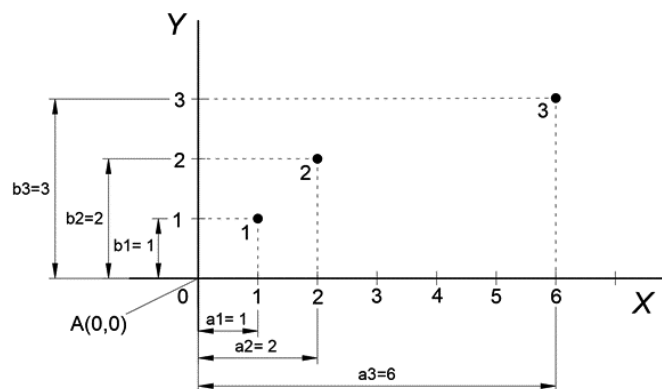


Рис. 2. Координаты векторов a и b , необходимые для вычисления обобщённого коэффициента корреляции, соответствующего точке A

¹ Все приведённые в статье рисунки выполнены её авторами.

Информативность коэффициента корреляции

Введённые обобщения позволяют оценить информативность коэффициента парной корреляции. Рассмотрим это на типичном примере оценки точности регрессионной зависимости по методике ГОСТ 22690-2015 «Бетоны. Определение прочности механическими методами неразрушающего контроля»² (также такую обработку можно провести для любых других результатов испытаний бетонных и железобетонных образцов, включая длительные испытания). Рассмотрим случай, когда у нас есть бетон, для которого известны несколько фактических значений прочности $R_{\text{фн}}$. Пусть прочность измерялась прибором неразрушающего контроля и были получены показания прибора (косвенные значения) H_i . Для определённости примем значения {1, 2, 3}. После этого показания прибора были пересчитаны в значения прочности по данным неразрушающего контроля $R_{\text{нн}}$ по градуировочной зависимости вида

$$R_{\text{нн}} = H.$$

Обычно для оценки точности построенной градуировочной зависимости вычисляется коэффициент парной корреляции между измеренными и фактическими значениями прочности. При высокой корреляции градуировочная зависимость признаётся хорошей. Сразу заметим, что нам требуется не корреляция, а, по возможности, точное совпадение фактических и измеренных значений. Для большей наглядности будем рассматривать случай, когда обе переменные $R_{\text{ф}}$ и $R_{\text{н}}$ принимают по три значения и все соответствующие точки расположены на одной прямой в координатах XOY (рис. 3).

В данном примере коэффициент корреляции между фактическими значениями прочности и значениями, измеренными и пересчитанными по градуировочной зависимости, равен единице.

Теперь рассмотрим случай, когда использована ошибочная градуировочная зависимость, тоже линейная. В качестве такой зависимости выберем произвольную линию, например, описываемую уравнением

² <https://docs.cntd.ru/document/1200124396>

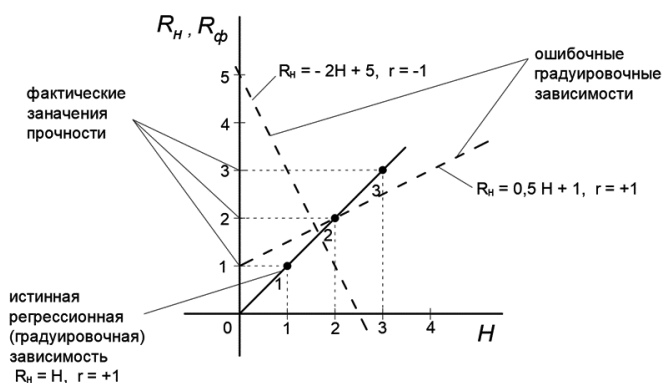


Рис. 3. Истинная и ошибочные градуировочные зависимости, имеющие одинаковое абсолютное значение коэффициента парной корреляции Пирсона

или

$$R_{\text{н}} = -2 \cdot H + 5 \quad (9)$$

$$R_{\text{н}} = 0,5 \cdot H + 1 \quad (10)$$

Если произвести непосредственные вычисления, можно убедиться, что абсолютное значение коэффициента корреляции Пирсона при ошибочных зависимостях по-прежнему получается равным единице. И если не проконтролировать точность этой зависимости по графику (зрительно) или по средней ошибке аппроксимации, то ошибка не будет обнаружена. Это происходит потому, что при смещении значений по оси измеренной прочности на пять единиц в уравнении (9), координаты вектора отклонений от нового значения $R_{\text{н,ср}}$ остаются прежними и длина вектора не меняется. При умножении на 2 соотношения между этими координатами остаются прежними и направление упомянутого вектора отклонений также не изменяется. Из приведённого примера следует, что коэффициент корреляции Пирсона в общем случае не позволяет оценить точность совпадения фактических и измеренных результатов. Высокий коэффициент корреляции лишь означает, что между измеренными и фактическими результатами существует линейная зависимость, параметры которой могут быть определены только из некоторых дополнительных условий (минимум погрешности, например). Данный результат может показаться парадоксальным с точки зрения инженера, но в действительности он известен в теории корреляции. Здесь этот пример приведён для подтверждения изложенных выше векторных построений.

Применение обобщённого коэффициента парной корреляции в данной ситуации может сразу выявить ошибку. Выберем случайную точку, относительно которой будем вычислять векторы отклонений. Для генерации случайных координат этой точки воспользуемся функцией RandomReal из пакета компьютерной математики Wolfram Mathematica. Для воспроизводимости расчётов генератор инициализировался числом 1234 (SeedRandom [1234]). Генерировалось три случайных числа, имеющих равномерное распределение в интервале от 0 до 5. Получены значения {4,38304; 2,60982; 0,431117}. Эти числа использовались в качестве координат случайной точки (координат случайного вектора A). В соответствии с изложенным выше алгоритмом получим координаты вектора отклонений истинных значений от выбранного случайного вектора:

$$A - R_{\text{ф}} = \{3,38304; 0,609821; -2,56888\},$$

и координаты вектора отклонений измеренных значений по уравнению (9) от выбранного случайного вектора:

$$A - R_{\text{н}} = \{1,38304; 1,60982; 1,43112\}$$

Косинус угла между этими векторами равен 0,180631, то есть векторы почти ортогональны. Тем самым ошибка выявлена. При использовании коэффициента Пирсона, как было отмечено, ошибка не выявляется. Справедливости ради следует отметить, что такой результат получился из-за удачного расположения случайной точки. Тем не менее пример является показательным.

Основываясь на изложенном выше представлении о переменных как векторах, можно сделать вывод, что для проверки точности регрессионной зависимости следует рассмотреть, как угол между векторами фактических и измеренных прочностей, так и нормы этих векторов. Это справедливо как для обобщённых парных коэффициентов корреляции, так и для коэффициентов корреляции Пирсона.

В соответствии со сделанными обобщениями можем выбрать произвольную фиксированную точку на числовой оси, относительно которой будем вычислять координаты векторов. Выберем в качестве такой точки начало координат. Тогда следует рассмотреть не векторы отклонений, а векторы непосредственно измеренных и фактических значений. То есть необходимо рассмотреть следующие соотношения.

Для оценки совпадения направлений:

$$r = \left(\frac{R_H}{\|R_H\|} \cdot \frac{R_\Phi}{\|R_\Phi\|} \right) = \frac{\sum_j R_{H,j} \cdot R_{\Phi,j}}{\sqrt{\sum_j R_{H,j}^2} \sqrt{\sum_j R_{\Phi,j}^2}},$$

и для оценки совпадения длин векторов:

$$d = \frac{\|R_H\|}{\|R_\Phi\|} = \frac{\sqrt{\sum_j R_{H,j}^2}}{\sqrt{\sum_j R_{\Phi,j}^2}}.$$

Вернёмся к данным из приведённого примера с истинными значениями переменных {1, 2, 3}. В качестве ошибочной градуировочной зависимости рассмотрим уравнение:

$$R_H = 0,5 \cdot H + 5. \quad (11)$$

Выполнив вычисления относительно начала координат, получим для истинной градуировочной зависимости $r = 1, d = 1$. Для той же ошибочной зависимости (11) получим $r = 0,95, d = 2,7$. Таким образом, направления векторов почти совпадают, а длины отличаются почти в три раза, что позволяет выявить ошибку. То же для ошибочной зависимости и коэффициента корреляции Пирсона (вычисления относительно среднего значения): $r = 1, d = 0,5$. Таким образом, коэффициент корреляции Пирсона остался равен 1, а длины векторов истинных и измеренных данных, сформированных по Пирсону, отличаются в два раза.

Всё же основным способом проверки качества регрессионной зависимости следует считать оценку её погрешности. Обычно для этого вычисляется средняя ошибка аппроксимации, но, учитывая изложенное, методически более правильно было бы использовать отношение нормы вектора погрешности к норме вектора независимой переменной. То есть необходимо рассмотреть относительную погрешность

$$\delta = \frac{\|R_H - R_\Phi\|}{\|R_\Phi\|} = \frac{\sqrt{\sum_j (R_{H,j} - R_{\Phi,j})^2}}{\sqrt{\sum_j R_{\Phi,j}^2}}. \quad (12)$$

Выполнив такую проверку, получим в данном примере для истинной градуировочной зависимости $\delta = 0$ (погрешность равна нулю), для ошибочной зависимости (11) и вычисления

ях относительно начала координат получим относительную погрешность $\delta = 1,86$, а для векторов, сформированных по Пирсону $\delta = 0,5$. Таким образом, по одной оценке получаем ошибку почти двукратную, по другой оценке – ошибка строго двукратная. Коэффициенты корреляции, вычисленные разными методами, при этом почти одинаково высоки.

В качестве ещё одного примера было выполнено исследование корреляции между нагрузками и прогибами конструкции при нелинейной работе материала при ступенчатом нагружении. Проведённые исследования показали, что коэффициент парной корреляции Пирсона при оценке нелинейной работы конструкции (на примере балки) является мало информативным. Он позволяет явно определить развитие пластических деформаций в материале, только когда эти деформации охватят значительную часть поперечного сечения. На начальных этапах пластической работы наблюдается практически полная корреляция между нагрузкой и, например, прогибами конструкции. В этих условиях предложенный обобщённый коэффициент парной корреляции в виде угла между векторами нагрузок и перемещений оказывается значительно более чувствительным и практически сразу реагирует на появление нелинейности в работе материала в сечении.

Множественный коэффициент корреляции

Аналогичный смысл имеет множественный коэффициент корреляции. Так, проведя непосредственные вычисления, можно убедиться, что если зависимую переменную рассмотреть

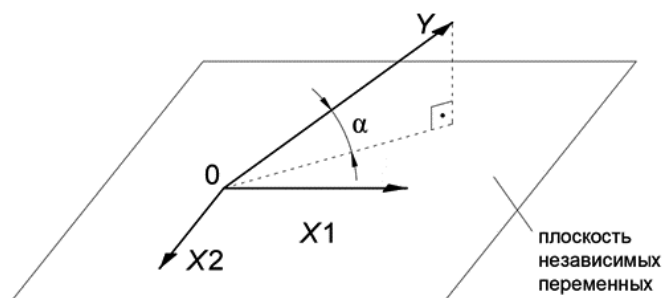


Рис. 4. Геометрический смысл классического коэффициента множественной корреляции.

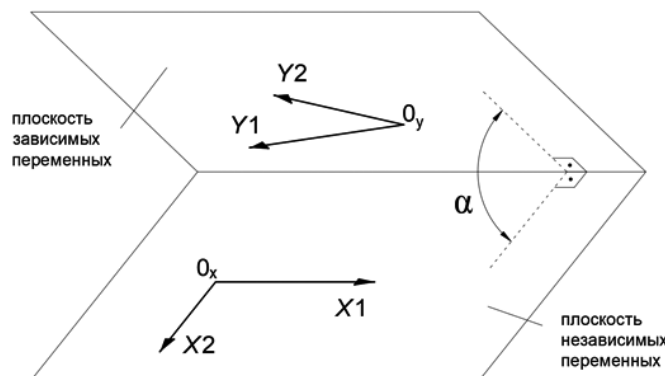


Рис. 5. Смысл обобщённого коэффициента множественной корреляции

как вектор, а набор независимых переменных рассмотреть как набор векторов, образующих некоторый базис, определяющий гиперплоскость, то множественный коэффициент корреляции будет в точности равен косинусу угла между упомянутым вектором и гиперплоскостью. Соответственно, можно сделать обобщения этого коэффициента корреляции по аналогии с обобщениями, сделанными для коэффициента парной корреляции. Так на рисунке 4 векторы X_1 и X_2 составлены из отклонений значений независимых переменных от их средних значений. Вектор Y составлен из отклонений значений зависимой переменной от среднего значения

Дальнейшим обобщением множественного коэффициента корреляции является не просто корреляция между зависимой переменной и набором независимых переменных, а распространение этого коэффициента на корреляцию между набором независимых переменных и целым набором зависимых переменных. При этом геометрический смысл такого коэффициента корреляции будет прежним – это косинус угла между двумя гиперплоскостями, определяемыми, соответственно, независимыми и зависимыми переменными (рис. 5).

Выводы

Проведённое исследование позволило установить геометрический смысл как парного коэффициента линейной корреляции Пирсона, так и множественного коэффициента корреляции. Показано, что коэффициент корреляции представляет собой косинус угла между вектором зависимой переменной и вектором независимой переменной (при парной корреляции) или косинус угла между вектором зависимой переменной и гиперплоскостью независимых переменных (при множественной корреляции).

Сделаны широкие обобщения коэффициента корреляции. В том числе введено понятие множественной корреляции между наборами зависимых и независимых переменных.

Приведены наглядные примеры, когда сделанные обобщения оказываются полезными при решении различных исследовательских задач и позволяют избежать оценочных ошибок.

Список источников

1. Гмурман, В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман ; 9-е изд. – Москва : Высшая школа, 2003. – 479 с. – Текст : непосредственный.
2. Фёзстер, Э. Методы корреляционного и регрессионного анализа / Э. Фёзстер, Б. Рёнтц. – Москва : Финансы и статистика, 1983. – 302 с. – Текст : непосредственный.
3. Шилов, Г.Е. Математический анализ. Специальный курс / Г.Е. Шилов. – Москва : Физматгиз, 1961. – 436 с. – Текст : непосредственный.
4. Каазик, Ю.Я. Математический словарь / Ю.Я. Каазик. – Москва : Физматлит, 2007. – 334 с. (в переводе). – ISBN 978-5-9221-0847-8. – Текст : непосредственный.

References

1. Gmurman V.E. Teoriya veroyatnostei i matematicheskaya statistika [Probability Theory and Mathematical Statistics]. Moscow, Vysshaya shkola Publ., 2003, 479 p. (In Russ.)
2. Fezster E., Rents B. Metody korrelyatsionnogo i regressionnogo analiza [Methods of Correlation and Regression Analysis]. Moscow, Finansy i statistika Publ., 1983, 302 p. (In Russ.)
3. Shilov G.E. Matematicheskii analiz. Spetsial'nyi kurs [Mathematical Analysis. Special Course]. Moscow, Fizmatgiz Publ., 1961, 436 p. (In Russ.)
4. Kaazik, Yu.Ya. Matematicheskii slovar' [Mathematical Dictionary] (translated). Moscow, Fizmatlit Publ., 2007, 336 p. ISBN 978-5-9221-0847-8. (In Russ.)