

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ И ЕГО ЗНАЧЕНИЕ В ЭКОЛОГИЧЕСКОЙ АНТРОПОЛОГИИ. ЧАСТЬ 1

Л.К. Гудкова

МГУ имени М.В. Ломоносова, НИИ и Музей антропологии, Москва

Статья посвящена изучению соизменчивости количественных признаков и представляет собой продолжение работ автора по теме приспособительной изменчивости, которая является основным содержанием физиологической, или экологической, антропологии. В первой части статьи обсуждаются история вопроса; статистические и биологические особенности корреляционного анализа. Во второй части основное внимание будет уделено его значению в изучении адаптированности популяций современного человека.

Корреляционный анализ, несмотря на кажущуюся многим антропологам тривиальность понятия и рутинность метода, используется, к сожалению, не всегда в соответствии с правилами и без попыток биологической трактовки получаемых результатов. Поэтому в статье в сжатой форме и с соответствующими ссылками дается необходимая информация о корреляции, которая, в первую очередь, будет полезна исследователям, не имеющим специального антропологического образования. Рассматривается только внутригрупповая прямолинейная связь двух количественных признаков с непрерывной изменчивостью.

Несмотря на то, что корреляционный метод получил широкое применение, вопрос о причинно-следственной обусловленности корреляций остается дискуссионным. Для правильного проведения корреляционного анализа и последующей причинной интерпретации полученных результатов важно иметь четкое представление о цели исследования [Малиновский, 1948].

В заключении сформулированы правила, которые следует соблюдать во избежание некорректного использования метода.

Ключевые слова: физиологическая, экологическая, антропология; корреляционный анализ, коэффициент корреляции

Введение

Статья посвящена изучению соизменчивости количественных признаков и представляет собой логическое продолжение моих работ по приспособительной изменчивости, которая является важнейшим содержанием физиологической, или экологической, антропологии [Гудкова, 2013а; 2013б; 2014]. Методология исследований по названной дисциплине базируется на принципах системности и целостности. Так как свойство целого зависит не только от особенностей составляющих частей, а главным образом от связей между этими составляющими – поиск системных и межсистемных корреляций служит основным инструментом холистического подхода. В рамках популяционной физиологии человека (одного из разделов физиологической антропологии), наряду с другими проблемами, изучается физиологический статус популяций как целостной структуры, определяемой устойчивыми

связями; а также анализируются морфофизиологические и другие межсистемные корреляции. Лабильность физиологических характеристик и относительная стабильность морфологических позволяет выяснять, как проявляется целостность организма отдельных индивидов на популяционном уровне в группах, находящихся на разных этапах адаптационного процесса и в кратковременных стрессовых ситуациях. Таким образом, изучение коррелятивной изменчивости способствует синтетическому взгляду на адаптивное состояние популяции, на возможности ее приспособительной изменчивости [Гудкова, 2008; 2010].

Корреляционный метод как таковой в работах по экологической антропологии не рассматривался, а лишь использовался в качестве одного из статистических приемов, к сожалению, зачастую не всегда в соответствии с правилами и без попыток биологической трактовки получаемых данных. В связи с этим, несмотря на кажущуюся многим ан-

тропологам тривиальность понятия и рутинность метода, считаю полезным в сжатой форме предоставить читателям необходимую информацию по теме коррелятивной изменчивости. Поэтому в первой части статьи обсуждаются история вопроса, статистические и биологические особенности анализа, а также даются примеры некорректных подходов к использованию корреляционного метода. Во второй части основное внимание будет уделено значению метода в экологической антропологии и его роли в изучении адаптированности популяций современного человека.

Исторические и теоретические аспекты учения о корреляции

В ряде публикаций, имеющих отношение к проблеме корреляции, даются сведения о возникновении понятия и о его дальнейшем развитии в науке. В отечественной антропологии наиболее значимой в историческом и теоретическом аспектах работой такого плана является статья Я.Я. Рогинского «Закономерности связей между признаками в антропологии» [Рогинский, 1962]. Статья состоит из четырех частей, которые называются: корреляция в общей морфологии, корреляция в антропологии, некоторые задачи изучения корреляций в антропологии и некоторые проблемы общетеоретического значения. Труд выдающегося ученого не утратил своей актуальности и современным антропологам, начинающим осваивать корреляционный анализ, следует в первую очередь обратиться к изучению статьи Я.Я. Рогинского.

Термин «корреляция» (от позднелат. *correlatio*, что дословно означает «соотношение») введен в науку французским зоологом Ж. Кювье. Занимаясь систематикой животных, Ж. Кювье в своих сравнительно-анатомических построениях исходил из принципа соотношения частей организма. В одном из произведений он сформулировал его таким образом: «Всякое организованное существо представляет нечто целое, единую и замкнутую систему, части которой взаимно соответствуют; ни одна из этих частей не может изменяться без того, чтобы не изменились другие...» [Cuvier, 1815. Цит. по: Филипченко, 2011, с. 121]. Декларируемая ученым идея целостности имеет прямое отношение к физиологической антропологии, равно как и его взгляды на приспособленность организма к окружающей среде, на его с ней единство. Однако, не будучи эволюционистом, Ж. Кювье все связи считал функциональными. Это ошибочное представление с постепенным развитием взглядов различных ученых на корреляцию, конечно же, со временем менялось. Так, соотечественник и идейный про-

тивник Ж. Кювье зоолог Э. Жоффруа Сент-Илер в своем учении о единстве плана строения всех животных сформулировал четыре правила, два из которых (принцип связей и закон равновесия органов) относились непосредственно к теме корреляции [Geoffroy Saint-Hilaire, 1822. Цит. по Филипченко, 2011]. Позднее в эволюционной морфологии явление уравнивания органов получило название морфогенетических корреляций [Амлинский, 1955. Цит. по: Рогинский, 1962].

Первым, кто заявил об эволюционном значении корреляций, был английский биолог Ч. Дарвин. Труды знаменитого ученого имеют особое значение для экологической антропологии, так как Ч. Дарвин постоянно обращал внимание на соотношение изменчивость и корреляции в развитии различных признаков целостного организма в процессе его приспособления к внешней среде. Рассуждая о причинах, «приводящих к многочисленным и таинственным случаям корреляции, которых мы совершенно не понимаем» [Дарвин, 1939, с. 427–428. Цит. по: Шмальгаузен, 1982.], Ч. Дарвин, пытаясь объяснить их причинность, отмечал, что чаще всего она скрыта от нас и, без сомнения, различна в разных случаях [Филипченко, 2011]. Таким образом, уже в XIX веке возникает вопрос о причинно-следственной обусловленности связей, дискуссии по поводу которой продолжаются до настоящего времени.

Таковы, в общих чертах, были теоретические представления о коррелятивной изменчивости до начала прошлого столетия. Крайне важно иметь в виду, что научные изыскания того периода ограничивались организменным уровнем. Переход на популяционный уровень стал возможным лишь с открытием математических приемов определения степени корреляционной зависимости, что также произошло в XIX веке. С XX века тема «коррелятивная изменчивость» начинает разрабатываться по двум направлениям. С одной стороны, в аспекте теоретической биологии обсуждается исключительно эволюционное значение корреляций; с другой стороны, с использованием биометрических методов анализируются всевозможные корреляции в самых разных отраслях науки. В рамках данной статьи не представляется возможным охватить все грани проблемы, поэтому рассматриваться будут труды только тех авторов, которые имеют прямое отношение к вопросам экологической антропологии.

В 1938 г. вышла в свет книга И.И. Шмальгаузена «Организм как целое в индивидуальном и историческом развитии», в которой впервые были обозначены основные направления разработки проблемы целостности и проанализирована роль корреляции в обеспечении целостности организма [Гудкова, 2010]. «В индивидуальном развитии од-

новременно с расчленением организма происходит и усложнение системы корреляций, объединяющей развивающийся организм в одно целое» [Шмальгаузен, 1982, с. 25]. И.И. Шмальгаузен выделяет три вида корреляций: геномные (генетические), морфогенетические и эргонтические (от греч. слова «работа»). Второй тип корреляций выражает морфофизиологическую целостность развивающегося организма, а эргонтические корреляции, формирующие окончательные соотношения в форме и связях, являются ясным выражением морфофизиологической целостности сформированного организма. Развитие организма определяется как взаимодействиями частей внутри него, так и его постоянными взаимодействиями со средой. Поэтому в случае разлада корреляционной системы организма можно говорить о нарушении его адаптивных реакций в зависимости от факторов внешней среды [Шмальгаузен, 1982]. Следует отметить, что статистические корреляции И.И. Шмальгаузен, будучи теоретиком-эволюционистом, рассматривал исключительно как метод исследования, не имеющий отношения к понятиям теоретической биологии. Однако, с точки зрения антропологии, корреляции, изучаемые биометрией, играют важную научную роль, и анализ соизменчивости различных признаков имеет и теоретическое, и практическое значение [Рогинский, 1962].

О широте исследований по вычислению корреляций, проводимых в Институте антропологии Московского университета, можно прочесть в указанной выше статье Я.Я. Рогинского [Рогинский, 1962]. Из работ, рассматриваемых Я.Я. Рогинским, в формате предлагаемой мною статьи имеет смысл остановиться на публикациях В.В. Бунака и А.А. Малиновского, посвященных конституциологии. Надо сказать, что в антропологии при изучении конституции главенствующее значение всегда придавалось проблеме целостности и методам ее изучения. В.В. Бунак указывает на необходимость включения функциональных признаков в конституциональную типологию (что и было осуществлено впоследствии в работах по физиологической антропологии), так как особенности метаболизма относятся к наиболее важным свойствам, характеризующим жизнедеятельность организма в целом. Выдающийся ученый в статье «Нормальные конституционные типы в свете данных о корреляции отдельных признаков» пишет, что «самое существенное в установлении конституционных типов заключается в установлении связи признаков» [Бунак, 1940, с. 59]. В.В. Бунак выделял исторические корреляции и механо-онтогенетические. Первые, считал он, выявляются при межгрупповых сравнениях, не выражают никаких физиологических зависимостей и не имеют отношения к понятию кон-

ституции. Относительно вторых В.В. Бунак полагал, что они в самых различных группах всегда сохраняют единство направления. Ошибочность этого утверждения позже была показана А.А. Малиновским [Малиновский, 1945, 1948].

Что касается значения работ А.А. Малиновского в изучении соизменчивости признаков, то достаточно вспомнить его кандидатскую диссертацию, защищенную в 1935 г. Она была посвящена конституциям человека – корреляции элементарных признаков. Позже исследователь напишет следующее: «...для учения о целом организме существенно не классифицировать конституции по какому либо признаку, но выявлять те связи, которые при этом обнаруживаются и комплексность охвата здесь нужна чрезвычайная – от морфологии скелета до процессов, протекающих в высших отделах нервной системы, ибо учение о конституции изучает именно организм в целом» [Малиновский, 1945, с. 236].

Я.Я. Рогинский чрезвычайно высоко оценивал антропологические работы А.А. Малиновского. «Заслуга Малиновского, пишет он, заключалась прежде всего в глубине, точности и разносторонности анализа статистической корреляции в учении о конституции» [Рогинский, 1962, с. 20]. Надо сказать, что заслуги ученого велики во многих дисциплинах (от практической медицины до теоретической биологии), однако больше всего в научном мире ценятся его теоретические достижения. Экологическим антропологам, например, будет полезно знать, что ему принадлежит представление об «экологическом напряжении», смысл которого состоит в повышенном требовании к проявлению физиологических функций организма в меняющихся условиях внешней среды. Небезынтересно узнать также, что последний период своей научной деятельности А.А. Малиновский посвятил развитию и синтезу идей «Общей теории систем» и «Тектологии». Комплексный и системный подходы, математические методы – все это позволяло ему воспринимать «объект познания и в единстве как целое, и в динамике как изменяющееся» [Клебанер, 2009, с. 74].

В заключение раздела уточним классификацию корреляций, используемую в антропологии. Уже «начиная с беглых замечаний Дарвина и кончая исследованиями современных морфологов, проходит одна основная линия, разделяющая корреляции на исторические и морфофизиологические» [Рогинский, 1962, с. 21]. В настоящее время «историческими» корреляции называют потому, что совместная встречаемость средних значений разных признаков может отражать этническую историю населения. В антропологии эти связи чаще называют – межгрупповые, а морфо-

физиологические – внутригрупповые. Связь между наборами средних величин, полученных по многим выборкам, принципиально отличается от внутригрупповой. Обычно межгрупповые корреляции оказываются неустойчивыми и меняются от одного набора выборок к другому [Дерябин, 2007].

Статистические свойства корреляционного анализа

Прежде чем перейти к дальнейшему изложению следует конкретизировать, что обсуждаться будет только внутригрупповая прямолинейная связь двух количественных признаков с непрерывной изменчивостью. Вне всякого сомнения, данная ниже информация антропологам известна. Однако мне кажется целесообразным в сжатой и лишенной математических выкладок форме сконцентрировать в одной статье биометрические аспекты корреляционного анализа.

Для ознакомления с начальным периодом становления биометрических методов изучения коррелятивной изменчивости можно воспользоваться изданной в 1928 г. работой Ю.А. Филипченко «Изменчивость и методы ее изучения» [Филипченко, 2011]. Ученый пишет, что впервые точное определение степени корреляционной зависимости осуществил английский исследователь Ф. Гальтон, предложивший для этого особый графический метод [Galton, 1888, 1889. Цит. по: Филипченко, 2011]. Этот прием давал возможность оценивать величину связи между двумя рядами признаков. Величина получила название коэффициента корреляции, обозначаемого со времен Ф. Гальтона символом «г», и стала общепринятым показателем степени коррелятивной изменчивости. Ф. Гальтон применял вычисление корреляции в антропометрии, а также при изучении наследственности. К. Пирсон, продолживший работу своего учителя, считал, что исследование Ф. Гальтона «открыло новую эру в статистике, что был создан “новый орган” для изучения природы как органической, так и неорганической» [Pearson, 1930. Цит. по: Канаев, 1972, с. 99]. Однако графический метод широкого распространения не получил и коэффициент корреляции впоследствии начали вычислять по способу, предложенному вышеупомянутым английским математиком К. Пирсоном и шотландским статистиком Д.У. Юлом [Pearson, 1896; Yule, 1897. Цит. по: Филипченко, 2011], которые, как замечает Ю.А. Филипченко, не создали в данном случае ничего нового. Они лишь воспользовались формулой, представленной в 1846 г. французским исследователем О. Бравэ, правда, придав ей более удобный вид.

Применение корреляционного метода позволяет установить «степень зависимости одного вариационного ряда от изменений другого. В этих случаях говорят о коррелятивной зависимости, или о коррелятивной изменчивости, двух вариационных рядов или двух представленных ими особенностей организма» [Филипченко, 2011, с. 120].

В пятой главе книги Ю.А. Филипченко, в сущности, даны все необходимые правила использования корреляционного метода и с тех пор, как будет ясно из последующего текста статьи, ничего не изменилось. Так, коррелятивную зависимость около 0,50 исследователь считает средней. Он полагает, что определение степени коррелятивной зависимости с достаточной долей достоверности, возможно, не менее чем для 30 наблюдений: меньшие цифры дают уже ненадежные результаты. «Ошибка коэффициента зависит всецело от числа особей сравниваемых рядов и от величины самого коэффициента. Чем выше последний, тем меньше ошибка, и наоборот, конечно, при одном и том же числе особей» [Филипченко, 2011, с. 135].

Крайне ценные выводы делаются ученым при анализе результатов, получаемых при смешении нескольких выборок. Он приводит пример изучения им корреляционной зависимости между различными особенностями сортов мягкой пшеницы и задается вопросом – от чего же зависит столь заметное различие между корреляцией внутри отдельных линий пшеницы и корреляцией в смешанном материале. Ю.А. Филипченко приходит к заключению, что «здесь индивидуальная изменчивость совершенно не причем, а все дело в групповой изменчивости, во взаимном отношении линий друг к другу, в их расположении внутри общей корреляционной таблицы...» [Филипченко, 2011, с. 142]. Он указывает, что корреляции в отдельных выборках, складываясь друг с другом, могут усиливать общую величину связи. В других случаях они могут взаимно погашать друг друга и корреляция в смешанном материале исчезает. Например, корреляция между длиной и шириной зерна в отдельных выборках (у автора – в биотипах) колеблется от 0,20 до 0,80. В смешанных популяциях, составленных из разных биотипов, коэффициент корреляции колеблется от -0,79 до +0,85. «Внутрибиотипические корреляции обуславливаются согласованной нормой реагирования двух особенностей в ответ на воздействия окружающей среды и стоят в самой тесной связи с индивидуальной изменчивостью. В популяционных корреляциях к этому примешивается другой, более сильный момент; взаимное отношение биотипов друг к другу, и нередко этот момент резко меняет характер внутрибиотипических корреляций при их суммировании. Словом, популяционные корреляции

тесно связаны с групповой изменчивостью» [Филипченко, 2011, с. 145]. Я намеренно привела эту большую цитату в надежде более корректного отношения антропологов к объединению выборок с целью увеличения численности. Для исследователей, работающих в сфере экологической антропологии, полезным будет еще один пример. Ссылаясь на работы Матвеева [Матвеев, 1928. Цит. по: Филипченко, 2011] о корреляционных зависимостях некоторых признаков у льна, Ю.А. Филипченко пишет, что в чистых линиях этого растения некоторые из внутрибиотипических корреляций довольно постоянны, другие же в зависимости от изменения внешних условий меняются очень сильно: от $-0,92$ до $+0,97$ (рис. 1). Наблюдаемое явление объясняется тем, что раз «корреляция между двумя признаками в чистой линии зависит от согласованной нормы реагирования двух особенностей в ответ на воздействия окружающей среды, то вполне возможно, что самый характер реагирования этих особенностей может быть различным в зависимости от изменения условий...» Это обстоятельство «не следует упускать из виду при работе методом корреляции во избежание возможных при этом грубых ошибок» [Филипченко, 2011, с. 146].

В современной науке понятие коэффициента корреляции используется очень широко, в связи с чем, методы его вычисления и описание правил его применения можно найти в большом количестве учебников и книг по вариационной статистике и биометрии [Например: Рокицкий, 1961; Плохинский, 1970; Лакин, 1990]. Со второй половины прошлого столетия появилось много переводной литературы по статистике [Например: Фишер, 1958; Тернер, 1976; Эренберг, 1981; Аффифи, Эйзен, 1982; Кимбл, 1982]. Однако в настоящее время антропологам нет особой необходимости обращаться к источникам разных авторов, так как нашим замечательным, рано ушедшим коллегой, В.Е. Дерябиным было написано и издано несколько специальных биометрических руководств.

В первую очередь В.Е. Дерябин рассматривает корреляционный анализ с позиции системного подхода: для понимания двух переменных как единой системы следует анализировать взаимосвязанность их изменчивости. В самом общем случае автор определяет взаимосвязанность вариации двух переменных «как наличие изменений вероятности появления некоторого варианта одной из них в зависимости от того, какое значение приняла – другая» [Дерябин, 2007, с. 155]. Простейшей и наиболее понятной является так называемая функциональная связь, когда каждому определенному значению одной переменной соответствует вполне определенное значение –

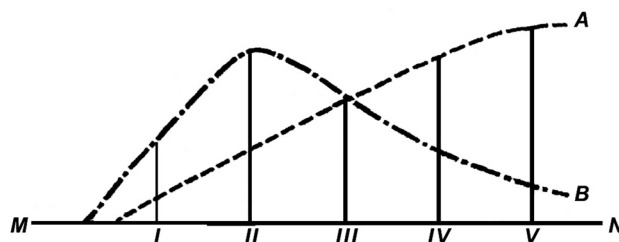


Рис. 1. Схема изменения коррелятивной зависимости двух признаков А и В при изменении внешних факторов от М к N. По Матвееву [цит. по: Филипченко, 2011, с. 145, рис. 32]

другой, однако в реальной ситуации такой феномен маловероятен.

Предостерегая от некорректного использования коэффициента корреляции, В.Е. Дерябин указывает на необходимость учитывать некоторые его специальные свойства. Одним из таких свойств является зависимость от ошибок измерений анализируемых признаков. Если признаки измерены со случайными ошибками, то коэффициент корреляции «окажется меньшим по сравнению с тем, которое можно было бы получить при вычислениях по значениям признаков, измеренным без сильного влияния ошибок измерений. Таким образом, наличие в вариации признаков независимого случайного компонента, связанного с ошибками измерений, приводит к уменьшению коэффициента корреляции» [Дерябин, 2007, с. 185].

Для интерпретации величины коэффициента корреляции В.Е. Дерябин рекомендует использовать данные Л.А. Сошниковой с соавторами. В этой работе предлагается следующая градация значений коэффициента: при $0.1 < r < 0.3$ – теснота связи слабая; при $0.3 < r < 0.7$ – средняя, и лишь при $r > 0.7$ – сильная [Сошникова с соавт., 1999]. Далее В.Е. Дерябин указывает на часто встречающуюся ошибку, возникающую при изучении корреляционных связей. Например, при достоверности полученного коэффициента нередко делается вывод о существовании сильной связи, что якобы позволяет надежно предсказывать значение одного признака по величине другого. Однако «неслучайность величины коэффициента корреляции свидетельствует лишь о существовании связи. Возможность предсказаний значений одного признака по величине другого показателя опирается на величину r^2 – коэффициента детерминации» [Дерябин, 2007, с. 218]. Автор приводит пример, когда при численности выборки в 400 наблюдений между двумя признаками обнаружена связь равная 0,2. В этом случае нельзя по значению одной переменной судить о величине другой, так как коэффициент детерминации будет 0,04, то есть

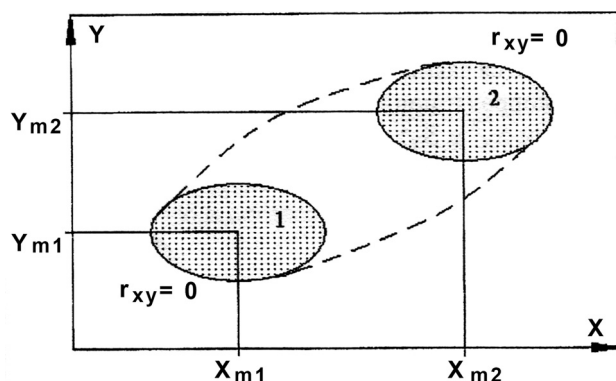


Рис. 2. Появление корреляции вследствие неоднородности материала в ситуации, когда в действительности $r=0$ [цит. по: Дерябин, 2007, с. 186, рис. 56]

объясняемая доля дисперсии составит всего лишь 4%. «Иными словами, 96% вариации любого из двух признаков никак не связаны с изменчивостью другого показателя» [Дерябин, 2007, с. 218]. В.Е. Дерябин рассматривает зависимость точности величины коэффициента корреляции и от объема выборки.

Важной особенностью коэффициента является также вариабельность его значений по причине неоднородности анализируемого материала. Приведенный автором рисунок (рис. 2) хорошо иллюстрирует и представленные на странице 30 примеры из работы Ю.А. Филипченко [Филипченко, 2011]. На рисунке изображена ситуация, когда внутри двух выборок корреляция признаков X и Y отсутствует. «Однако при рассмотрении их как одной единой группы индивидов, корреляционное поле обнаруживает наличие положительной связи, которая возникает только из-за того, что средние величины X_m и Y_m различаются. В тех случаях, когда внутри групп имеется определенная коррелированность признаков, межгрупповая неоднородность заметно исказит ее» [Дерябин, 2007, с. 185].

К вопросу о значении коэффициента корреляции

Проблема причинно-следственной обусловленности корреляций возникла в XIX в. и до настоящего времени остается предметом дискуссий. Ю.А. Филипченко, например, пишет, что при вычислении степени коррелятивной зависимости причинность связи вообще не обсуждается, так как «совершенно безразлично, является ли изменение одного признака причиной изменения другого, или же оба они варьируют вместе под влиянием изменений какой-нибудь третьей общей причины»

[Филипченко, 2011, с. 120]. Важно лишь выразить степень зависимости в изменениях двух или более рядов переменных определенным числовым образом. Именно это представляет практическую ценность метода, а «все дальнейшие выводы приходится делать уже иным путем, и учение о корреляции не может быть ответственно за их правильности или ошибочности» [Филипченко, 2011, с. 140]. Такого же мнения придерживаются биометрики. Так, П.Ф. Рокицкий полагает, что достоверная корреляция свидетельствует о сопряженности в вариации двух признаков, но вывод о наличии причинной зависимости между ними делать неправильно. Тем не менее, автор не исключает существование какой-то причины, влияющей на оба признака, установить которую можно исключительно с помощью биологического анализа [Рокицкий, 1961].

Несмотря на то, что корреляционный метод получил широкое применение, ряд исследователей из-за проблем с причинной интерпретацией сомневаются в его практической ценности. Использование коэффициентов корреляции для сравнения различных наборов данных и для получения прогнозов считается рискованным, потому что они характеризуют лишь степень тесноты и направление взаимосвязи между переменными и не дают никакой информации о природе этой взаимосвязи [Эренберг, 1981; Кимбл, 1982]. Существует опасность спутать понятие статистической зависимости с более фундаментальным понятием причинной связи и поэтому качественный анализ причин, лежащих в основе получаемых математических связей, находится за пределами биометрии [Тёрнер, 1976; Пасекон, 2005]. Для анализа причинности, считают биометрики, важно иметь в виду определение категории причинности, природу этого философского понятия [Кимбл, 1982]. С точки зрения философии причинная цепь является непрерывной, так как каждое составляющее этой цепи одновременно и причина, и следствие. По мере необходимости из причинной цепи «выделяется какое-то звено, рассматриваемое как причина, и другое, связанное с ним и во времени за ним следующее, рассматриваемое как следствие» [Философский энциклопедический словарь, 1999, с. 365]. Действительно, ощущение этой непрерывности может вызывать трудности в интерпретации получаемых корреляций и все-таки коэффициент корреляции среди прочих статистических показателей наиболее соответствует биологическим задачам [Фишер, 1958]. Статистические корреляции улавливают «фактические отношения признаков», которые представляют собой «единство связи и независимости между признаками» и «могут служить основой причинного биологического анализа» [Малиновский, 1948, с. 140–141].

Для экологической антропологии крайне важное значение имеют труды А.А. Малиновского. Надеюсь, что каждый научный сотрудник, занимающийся экологической антропологией, ознакомится с оригинальными статьями выдающегося ученого [Малиновский, 1945; 1948], постараюсь наикратчайшим образом изложить главные идеи его работ, касающиеся вопросов корреляции признаков.

Корреляция признаков основана на том, что каждый представляет собой функцию нескольких независимых переменных, но при этом хотя бы одна независимая переменная (общий аргумент) «является общей для всех коррелированных признаков, и этим обуславливается параллелизм их изменений» [Малиновский, 1948, с. 144]. Другие независимые переменные (частные аргументы), не одинаковы для разных признаков и заставляют их меняться вне связи с остальными признаками. Наиболее простые случаи, где признаки зависят от одного общего аргумента, А.А. Малиновский называет элементарной корреляцией. Чем больше их изменчивость будет зависеть от общего аргумента и чем меньше от частных, тем больше будет величина корреляции. Однако многие свойства организма связаны между собою не одним общим аргументом, а одновременно несколькими, так как в живом организме практически все явления постоянно варьируются под влиянием многочисленных факторов, часть которых всегда остается неясной. Таким образом, общие причинностные влияния могут выступать в виде единого комплекса или в виде нескольких слабо связанных факторов вариации. В первом случае корреляция может называться одноосновой, во втором – многоосновой.

Вопреки мнению ряда авторов А.А. Малиновский считает, что значение корреляций не зависит от их величины, так как с исследовательской точки зрения достоверные малые корреляции по своему значению практически не уступают большим корреляциям. Поэтому не размеры той или иной связи могут прояснить поставленную проблему, а возможность понять чисто качественную сторону, на которую указывает принцип корреляции. Более того, по общебиологическим соображениям иногда даже статистически недостоверные корреляции можно считать вполне реальными. Поэтому надо анализировать не отдельные коэффициенты, а всю корреляционную матрицу, так как всякая связь дает дополнительную информацию для выяснения общего аргумента, его физиологической сущности.

А.А. Малиновский опровергает ошибочное утверждение относительно физиологической корреляции, которая, в отличие от исторической, должна сохранять один и тот же характер в разных популяциях. В случае самых разных обстоя-

тельств в популяциях может доминировать то одна, то другая причина вариации признаков, и тогда могут меняться от популяции к популяции не только размеры, но и направление корреляций. Однако, с другой стороны, обсуждая элементарные корреляции, он отмечает их устойчивость, а именно, их свойство сохранять стабильность связей в разных популяциях и условиях.

Для корректного проведения корреляционного анализа и последующей трактовки полученных результатов, прежде чем перейти к выполнению задачи выявления зависимости (или связи) между двумя соответствующими парами переменных, важно иметь четкое представление о цели исследования. Неопределенная постановка цели «затрудняет правильное использование метода статистических корреляций. Это сказывается как в представлении о самодовлеющем значении чисто формальных зависимостей, так и в неправильном подборе материала для исследования и недостаточного использования его» [Малиновский, 1948, с. 171]. Неясность в постановке вопроса приводит к неудовлетворительным, иногда даже к взаимно противоречащим объяснениям. «Целью исследования... должен быть общий аргумент – то связующее звено, которое и обуславливает корреляции признаков» [Малиновский, 1948, с. 162]. Причинный анализ позволит отделить истинные связи от фиктивных и найти «реальные общие аргументы, связывающие признаки в системы корреляций» [Малиновский, 1948, с. 171].

Заключение

Все вышеизложенное позволяет сформулировать правила, которые следует соблюдать при использовании корреляционного метода.

В первую очередь, исследователю необходимо иметь четкую цель и мотивировать ее для выполнения предполагаемой работы. В основе ее должен лежать поиск общего аргумента, обуславливающего корреляцию признаков. От постановки цели зависит и выбор материала, и его дальнейшая разработка. Вводить в программу исследования соизменчивости признаков переменные, между которыми не может быть разумной связи, по меньшей мере, нецелесообразно.

Желательно, чтобы численность изучаемой выборки включала не менее 30 наблюдений. Меньшие численности дают ненадежные результаты. Недопустимо для увеличения численности смешивать различные выборки, потому что неоднородность анализируемого материала приводит к существенной вариабельности значений коэффициента корреляции.

При сборе материала важно соблюдать точность измерений, лабораторных определений различных показателей и т.д., так как величина коэффициента корреляции зависит от ошибок измерений анализируемых признаков.

Для обсуждения нужно представлять всю корреляционную матрицу. Даже если сам автор по какой-либо причине не рассматривает все полученные корреляции, наличие полной матрицы может исключить вопросы, возникающие у заинтересованных читателей. Для экологической антропологии обсуждение всей матрицы имеет принципиальное значение, так как логический подход к анализу всех корреляций способствует решению проблемы адаптированности популяций. Ложные корреляции, существующие между частью признака и его общей величиной, рассматривать не подлежат.

При предсказательной трактовке величины коэффициентов корреляций необходимо иметь в виду долю объясняемой дисперсии (коэффициент детерминации), которая составляет « r^2 ».

При соблюдении вышеназванных правил можно с достаточной уверенностью интерпретировать получаемые результаты, в основе причинного анализа которых всегда лежит поиск общих аргументов, связывающих признаки в системы корреляций.

В антропологии существует большое число публикаций, где используется корреляционный анализ. Примером ответственного подхода к теме может служить статья В.Е. Дерябина «О корреляциях между некоторыми продольными и поперечными размерами тела» [Дерябин, 1975]. Начиная с теоретического обоснования цели исследования через математически и биологически выверенный анализ материала, автор приходит к выводам, которые, естественно, уже не могут вызывать никаких сомнений.

К сожалению, имеются публикации совершенно иного характера – достаточно просмотреть номера последних лет нашего антропологического журнала. Например, не предоставляется вся корреляционная матрица; без должной аргументации объединяются коэффициенты корреляций; коэффициенты обозначаются неверными символами и т.д. В результате делаются необоснованные выводы, не подкрепленные наличием устойчивых связей и без необходимого биологического анализа, но, тем не менее, заявляющие о полученных «закономерностях». Серьезные ошибки допущены в работе, имеющей непосредственное отношение к экологической антропологии. Так, в пятой главе монографии «Антропоэкология Центральной Азии» [Антропоэкология... 2005], озаглавленной «Некоторые физиологические характеристики центрально-азиатских популяций», читаем следующее: «Многочисленность

наших данных предоставляет редкую возможность проанализировать взаимозависимость этих признаков у коренного населения Центральной Азии» [Антропоэкология... 2005, с. 107]. «Многочисленность» достигнута недопустимым объединением разных выборок из коренного населения Алтае-Саянского нагорья и Монголии. Можно было, опираясь на результаты канонического или кластерного анализа, хотя бы распределить уникальный материал по пяти группам, но с необходимыми комментариями относительно, скажем, сходства средних величин коррелируемых признаков в объединяемых выборках. Однако даже этого не произошло. Поэтому в матрице имеются значения, не поддающиеся никакой разумной интерпретации, и поэтому в очередной раз мы читаем: «Позволим себе ограничиться краткой характеристикой корреляционной матрицы, ибо ее функциональное осмысление – дело будущего» [Антропоэкология... 2005, с. 108]. О каком будущем может идти речь, когда, например, величина ложной (!) корреляции уровней общего белка и альбумина (признака, составляющего более половины общего белка) и у мужчин, и у женщин получилась отрицательной (!) и равной всего лишь 0,11.

В завершение приведу слова из статьи Н.Н. Богданова, посвященной 100-летию со дня рождения А.А. Малиновского, «...доверять результатам математических расчетов, в частности, пресловутым корреляциям, можно лишь тогда, когда хоть скольконбудь понятен их биологический смысл! Сколько же научного хлама не произвели бы на свет современные деятели от науки, прислушайся они к завету старого ученого!» [Богданов, 2009, с. 76].

Библиография

- Антропоэкология Центральной Азии. М., 2005.
 Афифи А., Эйзен С. Статистический анализ. М.: Мир, 1982.
 Бунак В.В. Нормальные конституционные типы в свете данных о корреляции отдельных признаков // Ученые записки МГУ, 1940. Вып. 34. С. 59–101.
 Богданов Н.Н. Научный наставник // Природа, 2009. № 8. С. 75–78.
 Гудкова Л.К. Популяционная физиология человека. М.: ЛКИ, 2008.
 Гудкова Л.К. Проблема целостности в физиологической антропологии // Вестник Московского университета. Серия XXIII. Антропология, 2010. № 3. С. 16–24.
 Гудкова Л.К. Физиологическая антропология // Вестник Московского университета. Серия XXIII. Антропология, 2013а. № 1. С. 52–61.
 Гудкова Л.К. Изменчивость как понятие и как основное содержание физиологической (экологической) антропологии. Часть I // Вестник Московского университета. Серия XXIII. Антропология, 2013б. № 3. С. 4–14.

- Гудкова Л.К. Изменчивость как понятие и как основное содержание физиологической (экологической) антропологии. Часть II // Вестник Московского университета. Серия XXIII. Антропология, 2014. № 4. С. 4–17.
- Дерябин В.Е. О корреляциях между некоторыми продольными и поперечными размерами тела // Вопросы антропологии, 1975. Вып. 50. С. 165–178.
- Дерябин В.Е. Курс лекций по элементарной биометрии для антропологов. Биологический факультет МГУ. М. 2007.
- Канаев И.И. Френсис Гальтон. Л.: Наука, 1972.
- Клебанер В.С. «Он человек был, человек во всем» // Природа, 2009. № 8. С. 66–74.
- Кимбл Г. Как правильно пользоваться статистикой. М.: Финансы и статистика, 1982.
- Лакин Г.Ф. Биометрия. М.: Высшая школа, 1990.
- Малиновский А.А. Физиологические источники корреляции в строении человеческого организма // Журнал общей биологии, 1945. Т. VI. № 4. С. 235–255.
- Малиновский А.А. Элементарные корреляции и изменчивость человеческого организма // Труды Института цитологии, гистологии и эмбриологии, 1948. Т. 2. Вып. 1. С. 136–198.
- Пасеков В.П. О теоретических проблемах биометрического и причинного подходов в популяционных исследованиях. М.: ВЦРАН, 2005.
- Плохинский Н.А. Биометрия. М., 1970.
- Рогинский Я.Я. Закономерности связей между признаками в антропологии // Советская этнография, 1962. № 5. С. 15–29.
- Рокицкий П.Ф. Основы вариационной статистики для биологов. Минск, 1961.
- Сошникова Л.А., Тамашевич В.Н., Уебе Г., Шеффер М. Многомерный статистический анализ в экономике. М.: Юнити, 1999.
- Тёрнер Д. Вероятность, статистика и исследование операций. М.: Статистика, 1976.
- Филипченко Ю.А. Изменчивость и методы ее изучения. М.: ЛКИ, 2012.
- Философский энциклопедический словарь. М., 1999.
- Фишер Р.А. Статистические методы для исследователей. М.: Гос. стат. изд-во, 1958.
- Шмальгаузен И.И. Организм как целое в индивидуальном и историческом развитии. М., 1982.
- Эренберг А. Анализ и интерпретация статистических данных. М.: Финансы и статистика, 1981.

Контактная информация:

Гудкова Людмила Константиновна: e-mail: lkgoodkova@bk.ru.

THE CORRELATION ANALYSIS AND ITS SIGNIFICANCE IN ECOLOGICAL ANTHROPOLOGY. PART 1

L.K. Goodkova

Lomonosov Moscow State University, Institute and Museum of Anthropology, Moscow

The article is aimed at studying covariability of quantitative traits and represents a sequel of works on adaptive variability, which represents the main subject of study in physiological and ecological anthropology. In the first part of the article the history of the problem and the statistical and biological aspects of correlation analysis are discussed. The second part will focus on the value of the method in studying adaptability of modern human populations.

Correlation analysis, despite the seeming simplicity of the concept and conventionalism of the method, is not always used in accordance with the rules and without attempts for biological interpretation. Therefore the article succinctly and with appropriate references provides the necessary information on the subject of correlative variability, which primarily will be useful for researchers who do not have special anthropological education. Only the intra-group rectilinear connection of two quantitative traits with continuous variability is discussed.

Though the correlation analysis has been widely used, the question of causal conditionality of correlation is controversial. For the correct performance of correlation analysis and following interpretation of the results obtained, it is important to be clear about the objective of the study.

In conclusion rules are formulated, which should be followed in order to avoid incorrect application of the method.

Keywords: ecological, physiological, anthropology; correlation analysis; correlation coefficient