

И. Г. ШИРОБОКОВ

Музей антропологии и этнографии им. Петра Великого (Кунсткамера) РАН,
Университетская наб., 3, Санкт-Петербург, 199034, Россия
E-mail: ivansmith@bk.ru
ORCID: 0000-0002-3555-7509

ДЕЙСТВИТЕЛЬНО ЛИ АНТРОПОЛОГИ УЧИТЫВАЮТ РАЗМЕР ВЫБОРКИ В КРАНИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ*

АННОТАЦИЯ

Как показывают результаты анонимного опроса, большинство российских физических антропологов полагают, что они учитывают возможное влияние размера выборки на результаты исследования. Анализ более 200 русскоязычных исследований по краниологии, опубликованных с 2000 по 2018 год, свидетельствует о том, что это утверждение справедливо лишь отчасти. Как правило, объектом краниологических исследований становятся выборки небольшого размера. Чем меньше выборка, тем выше вероятность отклонения ее характеристики от реальной характеристики генеральной совокупности, однако на практике чувствительность исследователей к репрезентативности выборки носит ограниченный характер. Определены методы анализа, использование которых сопряжено с риском принятия случайных характеристик в качестве репрезентативных. Показано, что при оценке различий выборочных и ожидаемых значений внутригрупповых дисперсий, кор-

реляций между признаками, а также коэффициентов полового диморфизма авторы в большинстве случаев игнорируют необходимость использования статистических критериев. Некоторые исследователи продолжают применять в анализе индивидуально-типологический подход, игнорируя индивидуальную и эпохальную изменчивость признаков. Предполагаемая причина популярности подхода заключается в том, что он повышает вероятность обнаружения морфологического сходства между различными популяциями (произвольно выделяемыми в них подгруппами). Преобладающая часть связей будет иметь случайный характер, однако другая их часть обнаружит подтверждение со стороны независимых систем признаков. В целом степень уверенности исследователей в результатах анализа лишь отчасти зависит от размера выборки и в значительной степени определяется возможностью интерпретировать данные в известном им контексте.

Ключевые слова: краниология, проблема выборки, исследовательские ошибки, множественные сравнения.

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-39-00138.

I. G. SHIROBOKOV

Peter the Great Museum of Anthropology and Ethnography (the Kunstkamera) of the Russian Academy of Sciences,
Universitetskaya Naberezhnaya, 3, St. Petersburg, 199034, Russia

E-mail: ivansmith@bk.ru

ORCID: 0000-0002-3555-7509

DO ANTHROPOLOGISTS REALLY CONSIDER SAMPLE SIZE IN CRANIOLOGICAL RESEARCH*

ABSTRACT

According to the results of my anonymous survey, most physical anthropologists believe that they take into account the possible biases related to the sample size on the research results. The analysis of more than two hundred Russian-language studies on cranial metric published from 2000 to 2018 shows that this statement is only partially valid. As a rule statistically small samples (median value of 20) frequently become the object of craniological studies. The smaller the sample size the higher the probability of its characteristics to deviate from the real characteristics of the general population. However, the sensitivity of researchers to the possible statistical biases related to the sample size is rather limited. Actually some methods of analysis are defined as associated with high risk of accepting random characteristics as representative ones. In most cases the need of using of

statistical criteria for comparing empirical samples and expected values of intragroup dispersions, correlations between traits, and sexual dimorphism coefficients is ignored by researchers. Some researchers continue to use an individual-typological approach in the analysis of small series and deny the leading role of random individual variability. Probable reason for the popularity of this approach is that it increases the likelihood of finding morphological similarities between compared populations (arbitrarily divided into subgroups). The majority of links will be random but some of them can be supported by independent systems of traits. In general, the degree of researchers' confidence in the results of analysis only partly depends on the sample size and to a large extent on the researcher ability to match these results to the known outer sample context.

Key words: craniology, sampling problem, research bias, multiple comparisons.

ВВЕДЕНИЕ

Осенью 2018 года автором совместно с Е.Н. Учаневой был проведен анонимный интернет-опрос среди российских физических антропологов на тему «Роль субъективных факторов в изучении популяционной антропологии». Отвечая на вопрос о том, учитывают ли они возможное влияние размера выборки

на результаты анализа, 79% опрошенных выбрали вариант «Да, всегда». Еще 16% отметили, что учитывают такое влияние в случае получения сомнительных результатов. Наконец, 15% участников признались, что игнорируют размер выборки. Всего в опросе принял участие 71 человек**. Поскольку в России число физических антропологов, занимающихся популяционными исследованиями, по всей вероят-

* Funding: The reported study was partially supported by RFBR, research project No. 18-39-00138.

** Полные результаты опроса доступны на сайте: <https://kunstkamera.academia.edu/IvanShirobokov>

ности, не превышает 100 человек, полученные данные вполне можно считать репрезентативными. Однако насколько эти результаты соответствуют реальной практике антропологических исследований? И насколько велика доля систематических искажений, при которых случайные отклонения в характеристиках выборок от популяционных интерпретируются как неслучайные?

МАТЕРИАЛЫ И МЕТОДЫ

Для того чтобы ответить на эти вопросы, автором была создана база данных, включающая в себя информацию более чем о 230 публикациях по краниологии, вышедших в свет на русском языке в период с 2000 по 2018 г. Выбор для анализа именно краниологических исследований был обусловлен двумя причинами. Во-первых, эта область антропологии профессионально близка автору. Во-вторых, в России в кругу антропологических дисциплин краниология по-прежнему сохраняет одну из лидирующих позиций по числу посвященных ей публикаций.

Для каждой публикации, включенной в базу данных, был определен набор количественных и дискретных признаков, характеризующих материалы, методы и результаты исследования. Учитывались размер исследуемой выборки (суммарная численность, а также численность мужских и женских черепов), статистические методы, использованные при проведении внутри- и межгруппового анализов, а также наличие и общий характер выводов (заключения об однородности или неоднородности антропологического состава, о различиях в характеристиках мужчин и женщин, предположения о родственных связях с другими группами населения, и др.)*. Для получения сопоставимых данных не рассматривались публикации единичных черепов, проблемные исследования, а также исследования, выполненные в рамках «неклассических»

программ признаков или основанные на сильно редуцированных наборах признаков стандартной программы.

В данной работе представлены первые результаты анализа, выполненного при помощи простых методов описательной и непараметрической статистики в пакете программ «STATISTICA 12.0». Частоты использования конкретных статистических методов вычислялись относительно общего числа публикаций, частоты конкретных вариантов интерпретации результатов — относительно общего количества выборок (которое превышает число публикаций и насчитывает около 300 групп).

Важным условием анализа было разделение описательной и аналитической (интерпретационной) частей в исследованиях. Например, в случаях, когда индивидуальные морфологические описания черепов, составляющих некоторую выборку, заметно различаются между собой, такие расхождения не рассматривались как свидетельствующие о неоднородности выборки, если автор не писал об этом напрямую или не придавал морфологическим характеристикам статуса разных антропологических типов или рас. Аналогичным образом, в случаях, когда обобщенные морфологические описания мужской и женской частей выборки различаются между собой, они не интерпретировались как факт расхождений в антропологическом составе мужчин и женщин, только если на это не указывал сам автор исследования или это однозначно не следовало из интерпретации данных. Такое разделение является важным для оценки влияния размера выборки именно на результаты исследований. Очевидно, например, что сама по себе вероятность обнаружения заметных морфологических расхождений между характеристиками мужских и женских черепов тем выше, чем ниже их численность и чем выше ожидаемая внутригрупповая изменчивость анализируемых признаков. Однако для анализа значимым

* Полный список проанализированных публикаций и база данных находятся в свободном доступе на сайте: <https://kunstkamera.academia.edu/IvanShirobokov>

Таблица 1. Медианы и среднее число черепов в краниологических выборках разных эпох (по материалам публикаций 2000–2018 гг.)

Table 1. Medians and average size of craniological samples from different periods (publications database from the years 2000 to 2018)

Период	Медиана			Среднее число		
	всех черепов	мужских	женских	всех черепов	мужских	женских
Неолит-энеолит	11	6	7	16	10	7
Эпоха бронзы	10	5	4	16	10	8
Эпоха железа	21	13	8	33	20	15
Средневековье	22	16	11	35	21	17
Новое время	26	17	11	42	25	19

является лишь то, склонен ли сам исследователь рассматривать выявленные различия в качестве неслучайных или же не придает им большого значения.

По тому же принципу рассчитывался процент использования конкретных методов анализа. Например, практически все публикации содержат табличный материал, включающий в себя, помимо средних значений признаков, их стандартные отклонения и число наблюдений. Но если в публикации не уделялось внимание обсуждению изменчивости признаков внутри выборки, при оценке относительной доли работ, в которых проводился анализ показателей внутригрупповой изменчивости, она не учитывалась.

РЕЗУЛЬТАТЫ

При условии случайного характера формирования выборки (то есть равной вероятности отбора в нее каждого из элементов генеральной совокупности), чем меньше ее размер, тем выше вероятность отклонения выборочной характеристики от характеристики генераль-

ной совокупности*. Именно с маленькими выборками, как правило, имеют дело краниологи. Средняя численность краниологической серии составляет: всего 33 черепа (20 мужских и 15 женских)**. Причем число наблюдений для конкретных признаков почти всегда меньше числа черепов, а распределение выборок по объему заметно отличается от нормального: маленькие выборки количественно преобладают, поэтому медиана меньше средней (20, 13 и 9 черепов соответственно). Чем древнее эпоха, к которой относится выборка, тем меньше медиана и среднее число черепов, при этом число женских систематически уступает числу мужских (табл. 1). По всей видимости, реальные различия в размерах относящихся к разным эпохам выборок даже выше полученных. Как уже указывалось выше, в анализе не учитывались публикации единичных черепов, которые преимущественно относятся к ранним периодам.

Рассмотрим наиболее распространенные приемы анализа выборок, ожидаемое влияние размера на результаты которых особенно велико. В 35% публикаций (64% от числа публикаций, где использовался хотя бы один статистический

* Вопрос о том, являются ли краниологические выборки действительно случайными, требует отдельного детального рассмотрения. Для целей данного исследования достаточно допущения о том, что выборки не обнаруживают некоторого общего систематического отклонения от характеристик популяций, из которых извлечены.

** Общее число черепов, без разделения по полу, не совпадает с суммарной численностью мужских и женских черепов выборки, поскольку не каждая из них включает в себя черепа обоего пола.

метод внутригруппового анализа) исследователи прибегают к анализу *показателей внутригрупповой изменчивости*.

Оценка признаков, как правило, проводится путем сравнения выборочных значений стандартных отклонений признаков с некоторыми ожидаемыми величинами (определяемыми в соответствии с табличными значениями [Алексеев, Дебец 1964]). Большинство участников опроса (76%) полагают, что универсального способа оценки неоднородности выборки не существует. Тем не менее на практике именно анализ внутригрупповой изменчивости признаков является одним из наиболее распространенных способов проведения такой оценки. Превышение ожидаемого значения (или верхнего предела ожидаемых значений) показателя часто рассматривается исследователями как возможное свидетельство гетерогенности выборки. В абсолютном большинстве случаев такое сравнение проводится путем простого сопоставления выборочных и ожидаемых показателей, без учета возможных случайных отклонений. Только в 15% случаев от числа публикаций, в которых анализировалась внутригрупповая изменчивость признаков, проводилась оценка равенства дисперсий при помощи F-критерия, и ни в одной из них не использовались поправки на множественные сравнения. При этом по результатам оценки изменчивости признаков в 74% случаев выборка была признана авторами неоднородной.

Кроме того, во всех случаях, когда авторы указывали, что прибегали к анализу дисперсий при помощи F-критерия, в выборке была выявлена повышенная изменчивость признаков, а в публикации содержалось заключение о неоднородности ее состава. Такое положение заставляет предположить, что в реальности тестирование используется краниологами чаще, чем оговаривается в публикациях. Отрицательные результаты тестирования не приводятся в тексте, и исследователь либо воздерживается от вывода о неоднородности группы, либо делает его, опираясь на другие статистические методы или

визуальные представления о морфологической неоднородности.

В 9% публикаций (в 17% случаев от числа публикаций, где использовался хотя бы один статистический метод внутригруппового анализа) авторы прибегали к оценке *внутригрупповых корреляций* между признаками. Лишь в 14% таких случаев проводился анализ статистической значимости различий между наблюдаемыми и ожидаемыми коэффициентами корреляции (как правило, использовалась матрица из: [Ефимова 1991]). Как и в случае с анализом внутригрупповой изменчивости признаков, при проведении корреляционного анализа никогда не используются поправки на множественные сравнения. В целом в 86% случаев по результатам корреляционного анализа выборка была признана авторами неоднородной.

Возможно, исследователи в принципе чаще прибегают к различным методам внутригруппового анализа в случаях, когда подозревают, что выборку характеризует высокая неслучайная морфологическая вариабельность. Безусловно, игнорирование статистических критериев не говорит о том, что в действительности во всех этих случаях выборки являются гомогенными, а выводы исследователей ошибочны. Скорее это свидетельствует о ненадежности используемой аргументации. При таком подходе к анализу дисперсий и корреляций высока вероятность ошибки первого рода: случайные отклонения чаще признаются неслучайными и рассматриваются как свидетельство неоднородности группы. Часть таких отклонений, безусловно, обусловлена небольшим размером выборок, что игнорируется исследователями.

Те же замечания относятся к оценке различий между мужской и женской частями выборки при помощи *коэффициентов полового диморфизма* (КПД). Этот прием анализа используется в каждой пятой публикации (21%). Как правило, анализ сводится к сопоставлению выборочных и ожидаемых значений КПД (определяемых в соответствии с табличными значениями [Алексеев, Дебец 1964]). Чаще всего он используется для

оценки однородности выборки, хотя корректность приема для решения этой задачи является по меньшей мере спорной [Евтеев 2008: 164]. Но даже если отбросить сомнения в его корректности, логику принимающих его исследователей не всегда можно назвать последовательной. Если повышенные КПД рассматриваются в качестве доказательства гетерогенности выборки, то пониженные значения признака трактуются некоторыми авторами как свидетельство морфологического сходства мужчин и женщин и однородности выборки [Бабаков и др. 2001; Харламова, Галеев 2013]. Однако очевидно, что если существует некая общая для всего человечества норма различий в размерах мужских и женских черепов, то при численности, достаточной для выводов об их нарушении (например, обусловленном разной антропологической принадлежностью части мужчин и женщин), разброс величин КПД относительно ожидаемого среднего значения будет симметричным. В рамках такого подхода направление отклонений само по себе не имеет значения для заключения об однородности, хотя оно и будет важным для оценки конкретных различий между морфологическими характеристиками предковых групп, обусловивших расхождения между мужскими и женскими черепами.

Нельзя утверждать, что степень уверенности исследователей никак не зависит от размера выборки. Если принимать в расчет только публикации, авторы которых прибегали к анализу КПД, то окажется, что средняя численность выборок, антропологический состав мужчин и женщин в которых, по мнению авторов, не идентичен, выше, чем когда исследователь воздерживается от такого заключения (38 и 26 черепов соответственно, $p=0.026$, U -критерий Манна-Уитни). Тем не менее из 44% случаев, когда величина КПД рассматривалась как одно из свидетельств различий в антропологическом составе мужчин и женщин, ни в одном не использовались какие-либо статистические критерии.

В 7% публикаций (в 38% публикаций, в которых так или иначе учитывались значения коэф-

фициентов полового диморфизма) размеры женских черепов трансформировались в мужские путем перемножения на стандартные величины КПД. Пересчет проводится при небольшом объеме выборки с целью повышения ее репрезентативности или для получения данных по отсутствующим в серии мужским черепам. Средний объем выборок, для которых проводилась трансформация значений признаков, заметно меньше того, для которых расчет КПД проводился без пересчетов размеров женских черепов (12 и 39 черепов соответственно при медианах 10 и 29, U -тест Манна-Уитни, $p=0.000018$).

Представляется сомнительным, что такая процедура повышает репрезентативность выборки, если принять во внимание, что в каждой третьей (34%) из опубликованных за последние 20 лет выборок, по мнению авторов, антропологический состав мужчин и женщин в той или иной степени различается (не является идентичным). Кроме того, сами по себе коэффициенты полового диморфизма имеют высокую межгрупповую вариабельность [Аксянова, Евтеев 2009]. Следовательно, можно предполагать, что перевод при помощи средних КПД размеров женских черепов в мужские, увеличивая размер выборки, в лучшем случае не меняет величину случайных отклонений признаков от реальной характеристики популяции, а в наиболее вероятном — приводит к еще большим ее искажениям.

При проведении внутригруппового анализа исследователи часто прибегают к *методу главных компонент (ГК)* — в 21% от общего числа публикаций и 39% от доли публикаций, в которых использовался хотя бы один статистический метод. Обычно используются данные по мужским сериям, реже проводится отдельный анализ мужских и женских черепов. Средняя численность мужских черепов в выборках, обработанных при помощи метода главных компонент, заметно выше, чем в случаях, когда метод не использовался (28 и 18 черепов соответственно при медианах 21 и 11, U -критерий Манна-Уитни, $p=0.0003$). Но достаточна ли она велика

для проведения анализа? Поскольку обычно в анализ ГК включается набор из 10–14 краниометрических признаков (в отдельных случаях — до 40), в соответствии с рекомендациями В.Е. Дерябина минимальная численность серии для корректного проведения расчетов должна составлять не менее 21–28 черепов [Дерябин 2005: 128]. Следовательно, более чем в половине случаев использования метода ГК краниологические выборки не удовлетворяют этим условиям. С другой стороны, по всей видимости, эти условия все же не являются необходимыми. Для расчета ГК число наблюдений необязательно должно превышать число признаков, поскольку главные компоненты могут быть получены без вычисления корреляционной или ковариационной матрицы, путем разложения по сингулярным числам [Ефимов, Ковалева 2008: 30].

Более существенной в рассматриваемом контексте является проблема применения анализа главных компонент для межгрупповых сопоставлений. В этом качестве ГК используются в 22% публикаций (31% от числа работ, в которых использовался хотя бы один статистический метод для межгруппового анализа). Проблема заключается в том, что при использовании анализа главных компонент для сравнения краниологических характеристик групп полностью игнорируются внутригрупповая изменчивость и корреляции между анализируемыми признаками [Дерябин 1998]. Иногда указывается, что в ряде краниологических исследований была показана применимость метода для межгрупповых сопоставлений, однако в действительности во всех таких случаях обоснование корректности метода подменяется демонстрацией возможности получить интерпретируемые результаты. Помимо игнорирования внутригрупповых корреляций, метод ГК, как правило применяемый к наборам среднегрупповых значений признаков, не позволяет учитывать размер выборок. Используя анализ главных компонент для межгрупповых сопоставлений и оценивая близость выборок на графике, исследователи неосознанно исходят из того, что характеристики всех серий

в равной степени репрезентативны, а полученные показатели межгрупповой изменчивости неслучайны.

В максимальном выражении нечувствительность к размеру выборки проявляет себя при *индивидуально-типологическом подходе*. Критика типологического подхода в России имеет давнюю историю, но ее основные положения можно свести к трем пунктам, доказывающим, что последовательные типологисты вынуждены игнорировать: во-первых, нормальную внутригрупповую изменчивость краниометрических признаков; во-вторых, эпохальную изменчивость признаков и их локальных комплексов; в-третьих, независимое наследование признаков. Тем не менее, как показывают материалы, положенные в основу данного исследования, даже в XXI в. выводы 13% публикаций по краниологии в той или иной степени опираются на индивидуальные расовые характеристики черепов. Средний объем проанализированных при помощи типологического подхода выборок заметно меньше, чем в случаях, когда такой подход не применялся, — 18 против 34 (U-критерий Манна-Уитни, $p=0.000085$). Причем в 72% случаях применение подхода привело авторов к заключению о неоднородности выборки, что в полтора раза выше доли аналогичных выводов, сделанных без обращения к типологическим конструкциям (51%, $p=0.027$). Отчасти эти различия могут объясняться тем, что исследователи прибегают к индивидуальной расовой идентификации в случаях высокой морфологической варибельности черепов, действительно обусловленной сложной историей формирования конкретной популяции. Но в значительной степени они все же, по всей вероятности, определяются склонностью интерпретировать внутригрупповую морфологическую изменчивость в качестве неслучайной. Тот известный факт, что более 80% всей изменчивости краниометрических показателей приходится на внутривнутрипопуляционный уровень [Козинцев 2016; Relethford 2002], типологистами просто игнорируется.

ОБСУЖДЕНИЕ

Даже по результатам этого краткого обзора становится очевидным, что на практике размер выборки далеко не всегда учитывается при проведении краниологических исследований (табл. 2). Это не означает, что исследователи полностью игнорируют фактор размера, но показывает, что наше понимание репрезентативности является ограниченным — заключение, которое хорошо согласуется с наблюдениями, касающимися как специалистов, занятых в других областях науки, в том числе профессиональных статистиков, так и всех людей [Канеман и др. 2018].

Некоторые факты однозначно свидетельствуют о большей осторожности исследователей при работе с маленькими выборками. В каждой четвертой публикации (26%) содержатся оговорки о небольшом размере серии и возможных случайных отклонениях результатов анализа. Средний объем «неоднородных» (40 черепов) и «однородных» (37) выборок вдвое выше, чем выборки с неопределенной характеристикой (22). Отчасти имеет свое оправдание и тот факт, что для проверки равенства ожидаемых и выборочных значений дисперсий, коэффициентов

корреляций, КПД, формы распределения статистические критерии используются крайне редко. Предположим, что исследователи использовали бы F-критерий в каждом случае анализа внутригрупповой изменчивости признаков. Если исходить из медианных значений выборки, то при формальном максимально допустимом уровне ошибки ($\alpha=0.05$) в случае анализа большинства мужских серий (не более 13 черепов) для отклонения нулевой гипотезы выборочная дисперсия признака должна превышать ожидаемую величину более чем в 1.75 раза, а для женских (не превышает 9 черепов) — более чем в 1.94 раза. С условием использования поправки Бонферрони для множественных сравнений и указанных медиан при анализе набора из 10 признаков, для отклонения нулевой гипотезы у мужчин выборочная дисперсия должна превышать ожидаемую более чем в 2.36 раза, у женщин — в 2.75 раза. Исходя из этого можно предположить, что небольшой объем выборок настолько повышает вероятность ошибок второго рода, что лишает смысла процедуру тестирования. Возможно, по этой причине некоторые краниологи больше опираются на свою интуицию, чем на значения признаков и результаты конкретных тестов.

Таблица 2. Проявление нечувствительности к размеру выборки при использовании различных методов исследования (по материалам публикаций 2000–2018 гг.)

Table 2. Cases of the using methods demonstrated insensitivity of researchers to sample size (publications database from the years 2000 to 2018)

Метод исследования	Публикации с использованием метода, %	Случаи нечувствительности к размеру выборок при использовании метода, %
Анализ внутригрупповой дисперсии признаков	35	74
Анализ внутригрупповых корреляций признаков	9	86
Использование метода главных компонент для внутригруппового анализа	21	?
Использование метода главных компонент для межгруппового анализа	22	100
Сравнение мужских и женских черепов при помощи коэффициентов полового диморфизма (КПД)	21	44
Трансформация размеров женских черепов в мужские при помощи КПД	7	100
Индивидуально-типологический подход	13	100

Один из альтернативных подходов к анализу, снимающих проблему множественных сравнений в краниологии, описан в предыдущей статье автора [Ширококов 2018]. Он заключается в оценке соответствия морфометрической характеристики выборки как целого некоторым ожидаемым параметрам при условии проверки справедливости нескольких гипотез о происхождении исследуемой группы.

Логично предположить, что если исследователи работают с небольшими выборками и при этом не всегда учитывают возможные случайные отклонения признаков, время от времени (и даже довольно часто) они должны приходиться к ошибочным заключениям. Почему же антропологи успешно получают «хорошие» результаты, часто согласующиеся с выводами предшественников, а также историческими и археологическими гипотезами об особенностях формирования тех или иных групп населения?

Ответ на этот вопрос заключается в том, что на самом деле в большинстве случаев мы не можем провести строгую верификацию получаемых результатов. Предполагаемое исследователем соответствие результатов изучения выборки некоторым представлениям об изучаемой реальности не является доказательством того, что, во-первых, выборка репрезентативна, то есть ее характеристика соответствует реальной характеристике популяции, из которой она извлечена, и, во-вторых, результаты анализа действительно подтверждают справедливость конкретной гипотезы, а не согласуются с некоторым множеством альтернативных, которые оказались за пределами поля зрения исследователя.

Как показывают специальные исследования, если характеристику выборки нельзя назвать нерепрезентативной (при условии, что известна характеристика группы, из которой она извлече-

на), то при ее оценке испытуемые не принимают во внимание размер выборки. Но как только результат воспринят как нерепрезентативный, то размер начинает играть ведущую роль в оценке выборки [Канеман и др. 2018: 97–98]. Это позволяет предполагать, что в краниологических исследованиях существует фактор, влияющий на уверенность исследователя в надежности полученных результатов в большей степени, чем размер выборки. Весьма вероятно, что этим фактором является именно интерпретируемость — возможность осмыслить результаты анализа в известном нам контексте*. У нас отсутствуют данные о реальной репрезентативности выборки, но мы исходим из предположения, что репрезентативность положительно связана с размером выборки. У нас отсутствует объективный способ верификации результатов, и его отчасти замещает интерпретируемость. Косвенно об этом свидетельствует то, что хотя оговорки о недостаточной для надежных выводов численности черепов чаще встречаются в публикациях, посвященных анализу маленьких выборок, значительная часть последних таких оговорок не содержит. Более того, в целом размер выборок в публикациях, авторы которых воздерживаются от содержательных выводов по результатам межгрупповых сопоставлений, не имеет систематических отклонений от размера выборок в публикациях, выводы в которых не ограничиваются просто указанием на степень выявленного сходства между некоторыми группами**.

Как уже указывалось во введении, 16% участников интернет-опроса указали, что учитывают возможное влияние размера выборки при получении сомнительных результатов. В строгом смысле такой подход ничем не лучше, чем игнорирование размера выборки, поскольку даже существенные случайные отклонения вы-

* Очевидно, что ни размер выборки, ни интерпретируемость результатов не являются единственными возможными факторами; важными могут быть квалификация конкретного исследователя, использование в публикации результатов анализа других независимых систем признаков.

** Под «содержательными выводами» в данном случае условно подразумеваются заключения автора публикации о происхождении исследуемой группы, ее родственных связях с другими популяциями, пришлом или автохтонном характере формирования и т. д.

борки от реальной характеристики популяции в случае соответствия результатов ожидания остаются просто незамеченными. Кроме того, весьма вероятно, что результаты, полученные при анализе маленькой выборки, согласуются не с одной, а с несколькими альтернативными (не совместимыми между собой) гипотезами о формировании исследуемой популяции. Следовательно, они не должны использоваться для избирательной поддержки только той из них, которую исследователь считает справедливой. Однако именно это нередко и происходит, и результаты опроса тому подтверждение.

Из перечисленных замечаний можно вывести основные причины сохранения популярности типологического подхода среди некоторых антропологов. Исходная посылка всякого краниологического исследования заключается в том, что степень морфологического сходства между сериями положительно коррелирует со степенью генетического родства между ними. Однако хотя родство предполагает сходство, само по себе сходство по некоторому набору признаков вполне может оказаться случайным. Потенциальный круг морфологически сходных серий можно значительно расширить, если выделить внутри анализируемых выборок морфологические варианты (типы, комплексы), например, при помощи кластеризации выборок и особенно на основе собственных субъективных представлений о сходстве типов. Выделение последних позволяет «подтвердить» наличие родственных связей между различными группами населения, археологическими культурами, эпохами и территориями в разных случаях:

- там, где они действительно существуют и корректный статистический анализ исходных выборок позволяет сделать предположение об их присутствии;
- там, где они действительно существуют, но статистический анализ исходных выборок не позволяет их выявить (например, из-за существенного случайного отклонения значений признаков в выборке);
- там, где сходство является случайным.

Даже если большинство выявленных связей будет ложным, некоторые из них обнаружат соответствие в данных независимых систем признаков (в том числе археологических и генетических). Следовательно, они могут быть использованы для избирательной поддержки конкретных гипотез о родстве некоторых групп населения. Соответствие родственных связей некоторым независимым данным в свою очередь укрепит точку зрения исследователя об эффективности типологического подхода. Таким образом, при нестрогом подходе к анализу высокая внутригрупповая изменчивость краниометрических признаков парадоксальным образом играет на руку типологистам. Неслучайно в некоторых публикациях обоснованием для использования индивидуально-типологического метода является именно небольшое число черепов, составляющих выборку, потенциально наименее репрезентативную, — тактически такой подход оправдывает себя, так как увеличивает возможности для интерпретаций материалов.

Возможно, в рамках индивидуально-типологического подхода корректным можно было бы считать утверждение о том, что данные краниологического анализа не противоречат некоторым гипотезам о формировании исследуемой популяции, но уж никак не подтверждают или тем более не доказывают ее справедливость. Однако любому человеку, в том числе исследователю-антропологу, свойственна такая черта, как предвзятость подтверждения — сознательная или бессознательная склонность фокусироваться на данных, согласующихся с его точкой зрения или известной ему информацией.

Влияние на исследователей контекстной информации и предвзятость подтверждения в популяционной антропологии, насколько известно автору, не становились предметом специального исследования. Однако существуют исследования, посвященные влиянию такой информации на заключения, принимаемые судебными экспертами [Dror, Charlton 2006; Kassin et al. 2013]. Специальные исследования показывают, что

при наличии контекстной информации оценки пола и возраста, принимаемые специалистами в области судебной медицины и физической антропологии, подвержены смещению в сторону ожидаемых характеристик. А предоставление информации об антропологической принадлежности умершего человека до проведения экспертизы может повлиять на оценку расовой принадлежности его скелета специалистами. Причем даже в случаях, когда информация является абсолютно некорректной, а в ее отсутствие определение расовой принадлежности у экспертов не вызывает сомнений, ее предоставление способно изменить точку зрения части исследователей или заставить их отказаться от проведения идентификации [Nakhaeizadeh et al. 2014].

От предвзятого отношения к анализу данных не свободны и те авторы, которые ищут их в работах других исследователей. В антропологии показательным тому примером является история о проверке исследования американского антрополога XIX в. Сэмюэля Мортон, предпринятая выдающимся американским биологом Стивеном Гулдом. Последний пришел к выводу, что Мортон из-за своих расовых предрассудков искажил результаты вычислений объема черепа у представителей разных рас, исходя из большей развитости интеллекта (и соответственно объема мозга) у европеоидов. Однако недавние исследования показали, что хотя Мортон и совершал некоторые ошибки, но отклонения в измерениях черепов не являются систематическими и не могут быть интерпретированы как допущенные в пользу представителей какой-либо из рас. Более того, систематические ошибки были обнаружены в расчетах самого Стивена Гулда, абсолютно убежденного, что взгляды Мортон повлияли на проведение исследования, но не сумевшего преодолеть собственную предвзятость в отношении последнего [Lewis et al. 2011].

Влияние предвзятости подтверждения было выявлено и при подготовке данного исследования. В первом варианте результатов анализа обнаружено, что средний объем выборок в случаях, когда авторы пишут о разном проис-

хождении мужчин и женщин, на статистически значимом уровне меньше, чем в случае, когда авторы от таких выводов воздерживаются. Различия в размерах выборок интерпретировались как ожидаемое автором свидетельство того, что краниологи не осознают, насколько серьезное влияние случайные отклонения выборочных значений признаков в мужской и женской частях выборки могут оказывать на интерпретацию результатов.

Однако после повторного анализа публикаций, проведенного автором и Е.Н. Учаневой, определения выводов были пересмотрены. Оказалось, что в некоторых публикациях, якобы сохранивших заключение о разном происхождении мужчин и женщин, в действительности исследователи были более осторожными в формулировках и указывали, что антропологический состав мужской и женской частей выборки нельзя считать идентичным или что они включают в себя разные доли общих компонентов. Корректировка данных привела к исчезновению неслучайных различий между выборками. При этом существенные различия между размерами выборок, антропологический состав мужской и женской частей которых, по мнению авторов, не идентичен, выявлены не были. Размещение в свободном доступе данных, использованных при подготовке этого исследования, позволяет всем желающим попытаться самостоятельно оценить влияние ошибок различного рода как на его результаты, так и на выводы учтенных в нем публикаций. В дальнейшем предполагается значительное расширение набора анализируемых признаков с целью выявления общих закономерностей в логике краниологических исследований и более подробной оценке влияния на их результаты используемых методов анализа.

ЗАКЛЮЧЕНИЕ

Из-за специфики объекта исследований краниологи не могут сами устанавливать размер выборки, предположительно достаточный для соблюдения условий анализа и получения неко-

торых ожидаемых результатов, как это принято делать при подготовке современных естественно-научных и социологических исследований. Однако они работают с материалом, вписанным в определенный археологический, исторический и антропологический контекст, и часто заранее знают, какие из возможных результатов являются наиболее вероятными (ожидаемыми), а какие — совершенно неправдоподобными. Размер выборки лишь отчасти определяет степень уверенности исследователей в надежности заключений. В условиях невозможности установления репрезентативности выборки ее замещает

интерпретируемость результатов. Случайные отклонения выборочных значений признаков от реальных популяционных характеристик, более вероятные при низкой численности серии, могут периодически оставаться незамеченными или даже получать правдоподобное объяснение в известном контексте. Поскольку значительная часть исследований основана на анализах малых выборок, в строгом смысле результаты часто корректнее описывать как не противоречащие конкретной гипотезе (гипотезам) формирования исследуемых групп населения, нежели как доказывающие ее справедливость.

СПИСОК ЛИТЕРАТУРЫ

- Аксянова, Евтеев 2009. Аксянова Г.А., Евтеев А.А. Межгрупповая изменчивость полового диморфизма в строении черепа у коренного населения Северной и Центральной Азии // ЭО. — 2009. — №1. — С.61–78.
- Алексеев, Дебец 1964. Алексеев В.П., Дебец Г.Ф. Краниометрия. Методика антропологических исследований. — М.: Наука, 1964. — 128 с.
- Бабаков и др. 2001. Бабаков О., Рыкушина Г.В., Дубова Н.А., Васильев С.В., Пестряков А.П., Ходжайов Т.К. Антропологическая характеристика населения, захороненного в некрополе Гонур-Депе // Саррианиди В.И. Некрополь Гонура и иранское язычество. — М.: Мир-Медиа, 2001. — С. 105–132.
- Дерябин 1998. Дерябин В.Е. О методиках многомерного таксономического анализа в антропологии. Канонический анализ против главных компонент // Вестник антропологии. — Вып. 4. — 1998. — С. 30–67.
- Дерябин 2005. Дерябин В.Е. Краткий справочник по решению типовых задач биометрической обработки антропологических данных. Депонент МГУ (М.). — [Б. м.: б. и.], 2005. — 247 с.
- Евтеев 2008. Евтеев А.А. Проблема полового диморфизма в краниологии. Дис. ... канд. биол. наук. — М., 2008. — 230 с.
- Ефимов, Ковалева 2008. Ефимов В.М., Ковалева В.Ю. Многомерный анализ биологических данных. — СПб.: ВИЗР РАСХН, 2008. — 87 с.
- Ефимова 1991. Ефимова С.Г. Палеоантропология Поволжья и Приуралья. — М.: Изд-во МГУ, 1991. — 95 с.
- Канеман и др. 2018. Канеман Д., Словик П., Тверски А. Принятие решений в неопределенности. Правила и предубеждения. — Харьков: Гуманитарный центр, 2018. — 536 с.
- Козинцев 2016. Козинцев А.Г. О некоторых аспектах статистического анализа в краниометрии // Радловский сборник: научные исследования и музейные проекты МАЭ РАН в 2015 году. — СПб.: МАЭ РАН, 2016. — С. 381–390.
- Харламова, Галеев 2013. Харламова Н.В., Галеев Р.М. Население Иркутска XVIII–XIX вв. по данным краниологии // Известия Иркутского государственного университета. Серия «Геоархеология. Этнология. Антропология». — 2013. — №1 (2). — С. 230–243.
- Широбоков 2018. Широбоков И.Г. Вероятностный подход к оценке гипотез и проблема множественных сравнений в краниологии // Кунсткамера. — 2018. — № 1. — С. 87–96.
- Dror, Charlton 2006. Dror I.E., Charlton D. Why experts make errors // Journal of Forensic Identification. — 2006. — Vol. 56 (4). — P. 600–616.
- Kassin et al. 2013. Kassin S.M., Dror I.E., Kukucka J. The forensic confirmation bias: Problems, perspectives, and proposed solutions // Journal of Applied Research in Memory and Cognition. — 2013. — Vol. 2 — No 1. — P. 42–52.
- Lewis et al. 2011. Lewis J.E., DeGusta D., Meyer M.R., Monge J.M., Mann A.E., Holloway R.L. The mismeasure of science: Stephen Jay Gould versus Samuel George Morton on skulls and bias //

PLoS Biology. — 2011. — Vol. 9 (7). — <https://doi.org/10.1371/annotation/138c7c99-249f-432c-a4f7-2993b7b87c0a>

- Nakhaeizadeh et al. 2014. Nakhaeizadeh S., Dror I.E., Morgan R.M. Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias // *Science & Justice*. — 2014. — Vol. 54. — No 3. — P. 208–214.
- Relethford 2002. Relethford J. Apportionment of global human genetic diversity based on craniometrics and skin color // *American Journal of Physical Anthropology*. — 2002. — Vol. 118. — No 4. — P. 393–398.
-
- ## REFERENCES
- Aksianova, G.A., Evteev, A.A., Mezhruppovaia izmenchivost' polovogo dimorfizma v stroenii cherepa u korennoho naseleniia Severnoi i Tsentral'noi Azii [Variability of the Sexual Dimorphism in the Crania Among Indigenous Populations of Northern and Central Asia], *Etnograficheskoe obozrenie*, 2009, no. 1, pp. 61–78, (in Russian).
- Alekseev, V.P., Debets, G.F., *Kraniometriia. Metodika antropologicheskikh issledovaniĭ* [Cranio-metry. Methods of Anthropological Investigations], Moscow: Nauka, 1964, 128 p., (in Russian).
- Babakov, O., Rykushina, G.V., Dubova, N.A., Vasil'ev, S.V., Pestriakov, A.P., Khodzhaiov, T.K., Antropologicheskaiia kharakteristika naseleniia, zakhoronennogo v nekropole Gonur-Depe [Anthropological Characteristic of Population Buried at Gonur-Depe Necropolis], in: Sarianidi V.I. *Nekropol' Gonura i iranskoe iazychestvo*, Moscow: Mir-Media, 2001, pp. 105–132, (in Russian).
- Deriabin, V.E., O metodikakh mnogomernogo taksonomicheskogo analiza v antropologii. Kanonicheskii analiz protiv glavnykh component [On Methods of Multivariate Taxonomic Analysis in Anthropology: Canonical Analysis Versus Principal Components], *Vestnik antropologii*, 1998, vol. 4, pp. 30–67, (in Russian).
- Deriabin, V.E., *Kratkii spravocnik po resheniiu tipovykh zadach biometricheskoi obrabotki antropologicheskikh dannyykh* [A Brief Guide to Solving Typical Problems of Biometric Processing of Anthropological Data], Moscow: VINITI, 2005, 247 p., (in Russian).
- Dror I.E., Charlton, D., Why Experts Make Errors, *Journal of Forensic Identification*, 2006, vol. 56 (4), pp. 600–616.
- Evteev, A.A., *Problema polovogo dimorfizma v kranologii* [Issue of Sexual Dimorphism in Craniology], *Dissertatsiia na soiskanie uchenoi stepeni kandidata biologicheskikh nauk*, Moscow, 2008, 230 p., (in Russian).
- Efimov, V.M., Kovaleva, V.Iu., *Mnogomernyi analiz biologicheskikh dannyykh* [Multivariate Analysis of Biological Data], St. Petersburg: VIZR RASKhN, 2008, 87 p., (in Russian).
- Efimova, S.G., *Paleoantropologiya Povolzh'ia i Priural'ia* [Paleoanthropology of Volga and Pre-Urals Region], Moscow: MGU, 1991, 95 p., (in Russian).
- Kaneman, D., Slovik, P., Tverski, A., *Priniatie reshenii v neopredelennosti. Pravila i predubezhdeniia* [Judgement Under Uncertainty: Heuristics and Biases], Kharkiv: Gumanitarnyi tsentr, 2018, 536 p., (in Russian).
- Kassin S.M., Dror, I.E., Kukucka, J., The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions, *Journal of Applied Research in Memory and Cognition*, 2013, vol. 2, no. 1, pp. 42–52.
- Kozintsev, A.G., O nekotorykh aspektakh statisticheskogo analiza v kranimetrii [On Some Aspects of Statistical Analysis in Craniometry], *Radlovskii sbornik. Nauchnye issledovaniia i muzeinye proekty MAE RAN v 2015 godu*, St. Petersburg: MAE RAS, 2016, pp. 381–390, (in Russian).
- Kharlamova, N.V., Galeev, R.M., Naselenie Irkutskia XVI-II-XIX vv. po dannym kranologii [Irkutsk Population of XVIII-XIX Centuries by Craniology Data], *Izvestiia Irkutskogo gosudarstvennogo universiteta. Seriya "Geoarkheologiya. Etnologiya. Antropologiya"*, 2013, vol. 1 (2), pp. 230–243, (in Russian).
- Lewis, J.E., DeGusta, D., Meyer, M.R., Monge, J.M., Mann, A.E., Holloway, R.L., The Mismeasure of Science: Stephen Jay Gould Versus Samuel George Morton on Skulls and Bias, *PLoS Biology*, 2011, vol. 9 (7), (Online), Available from: <https://doi.org/10.1371/annotation/138c7c99-249f-432c-a4f7-2993b7b87c0a> (Accessed 15.05.2019).
- Nakhaeizadeh, S., Dror, I.E., Morgan, R.M., Cognitive Bias in Forensic Anthropology: Visual Assessment of Skeletal Remains is Susceptible to Confirmation

- Bias, *Science & Justice*, 2014, vol. 54, issue 3, pp. 208–214.
- Relethford, J., Apportionment of Global Human Genetic Diversity Based on Craniometrics and Skin Color, *American Journal of Physical Anthropology*, 2002, vol. 118 (4), pp. 393–398.
- Shirobokov, I.G., Veroiatnostnyi podkhod k otsenke gipotez i problema mnozhestvennykh sravnenii v kraniologii [Probabilistic Approach to Evaluating Hypotheses and Multiple Comparisons Problem in Craniology], *Kunstkamera*, 2018, vol. 1, pp. 87–96, (in Russian).