ЗООЛОГИЧЕСКИЙ ЖУРНАЛ, 2023, том 102, № 3, с. 349–353

МЕТОДИКА ЗООЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

УДК 599.322.2;591.431.4;591,543,42

ПРИМЕНЕНИЕ BOOTSTRAP МЕТОДА В АНАЛИЗЕ ПОВЕДЕНЧЕСКИХ ДАННЫХ И МЕТОД ВИЗУАЛИЗАЦИИ ДАННЫХ ВЫБОРА ИЗ ТРЕХ ОБЪЕКТОВ

© 2023 г. А. В. Дудорова[a], *, Д. А. Подгрудков[b, c], **, Е. П. Крученкова[a], ***

[a]*Московский государственный университет имени М.В. Ломоносова, биологический факультет, Москва, 119991 Россия*

[b]*Московский государственный университет имени М.В. Ломоносова, физический факультет, Москва, 119991 Россия*

[c]*Московский государственный университет имени М.В. Ломоносова, Научно-исследовательский институт ядерной физики имени Д.В. Скобельцына, Москва, 119991 Россия*

**e-mail: dudorova.anastasia.2012@post.bio.msu.ru*

***e-mail: d.a.podgrudkov@physics.msu.ru*

****e-mail: ekruster@gmail.com*

Поступила в редакцию 17.06.2022 г.
После доработки 11.08.2022 г.
Принята к публикации 12.08.2022 г.

Предложены применение метода bootstrap-анализа для изучения поведенческих данных и способ представления данных в поведенческих тестах на предпочтение и/или выбор из нескольких вариантов. В качестве примера использования приведены данные по изучению предпочтений у детенышей нильских крыланов (*Rousettus aegyptiacus*). Применение бутстрэп метода позволяет оценить достоверность полученного результата при относительно малой выборке, а предлагаемый новый метод визуализации позволяет наглядно представить данные при не абсолютном выборе.

Ключевые слова: нильский крылан, детеныши, представление данных, поведение

DOI: 10.31857/S0044513423020046, **EDN:** HPGOQP

В современных публикациях преобладают статьи и исследования с хорошо подтвержденными статистически значимыми выводами (Csada et al., 1996). Несомненно, статистика невероятно важный инструмент любого ученого, и биологи не исключение. Однако часто из-за желания представить только хорошо поддерживаемые результаты многие исследования с ограниченной выборкой оказываются неопубликованными (Csada et al., 1996). Сильнее всего из всех биологических дисциплин от этого страдает исследование поведения. Изучать поведение, особенно высших позвоночных, довольно сложно (Мантейфель, 1980; Bart et al., 1998). С одной стороны, даже один тип поведения внутри себя может состоять из многих вариаций, с другой стороны, набрать большое число исследуемых особей среди некоторых видов животных оказывается затруднительно (Попов, 2008). Между тем сила статистической поддержки коррелирует с размером выборки (Jennions, Møller, 2003). В итоге часто возникает ситуация, когда исследователи получают слабую поддержку и не могут однозначно трактовать результаты. Кроме того, всегда есть необходимость оценки достоверности полученного результата. Это возникает во многих разделах наук (а также в экономике, в анализе поведения рынков). Сама оценка достоверности решается строго в рамках математической статистики, однако решения, зачастую, трудоемки. Для этого (оценки достоверности результатов анализа с достаточной точностью на ограниченной выборке) в 1979 г. Брэдли Эфроном был разработан метод бутстрэп анализа (Efron, 1979). Этот метод был обобщен для применения во многих других разделах наук, также его можно применить для анализа поведенческих данных. Но обработать данные – лишь часть дела. Необходимо найти удачное визуальное решение для представления результатов работ перед коллегами, будь то доклад, постер или иллюстрация в публикации. Наглядность крайне важна при обучении и донесении новой информации (Жук и др., 2008). Мы предлагаем рассмотреть один из вариантов применения бутстрэп метода в зоологических исследованиях, посвященных поведению животных. Вместе с ним предлагается визуальный методом представления данных.

МАТЕРИАЛ И МЕТОДИКА

Описание бутстрэп метода

Метод бутстрэп анализа позволяет оценить параметры экспериментальной выборки (дисперсию, среднее и пр.) путем искусственного увеличения размера выборки. Исследователи используют результаты вычислений по синтетическим выборкам, собранным из исходных реальных данных, и предполагают, что синтетические выборки являются выборками с такой же функцией распределения, как и исходная выборка (Efron, 1979). При этом анализируется большое число "искусственных" выборок, называемых бутстреп-выборками. Обычно случайным образом генерируется несколько тысяч выборок. Для создания искусственной выборки ($y_1, y_2, .., y_n$) из исходной ($x_1, x_2, .., x_m$) выберем на первом шаге случайным образом один из элементов исходной выборки (пусть x_i) и запишем его первым элементом искусственной выборки (y_1). Затем вернем этот элемент (x_i) обратно в исходную выборку. Далее выберем второй элемент для искусственной выборки (y_2) снова из полной исходной выборки ($x_1, x_2, .., x_m$). Пусть это элемент (x_j), при этом j может быть равно i, т.е. любой элемент исходной выборки может повториться несколько раз в искусственной выборке. Повторим описанную процедуру случайного выбора n раз (до желаемой длины выборки).

Таким образом, бутстрэп предполагает случайный выбор с возвращением, т.е. выбранные элементы исходных данных возвращается в выборку и далее могут быть снова выбраны. На каждом шаге мы выбираем элемент исходной выборки с вероятностью $1/m$. В итоге имеем новую выборку в n элементов, набранную из m элементов исходной выборки. При этом n может быть как больше, так и меньше m, элементы исходной выборки могут повторяться в новой неоднократно либо могут вообще в нее не попасть. Таких выборок генерируется необходимое количество по желанию исследователя, в данном случае мы остановились на 100000, что легко достижимо для современных компьютеров.

На основе собранных бутстрэпом выборок можно проводить различный анализ данных: считать среднее, вычислять ошибку и прочее, все то, что можно и планировалось сделать на оригинальных данных. Стоит заметить, что бутстрэп метод реализован во многих пакетах анализа данных (MATLAB, Statistica, R, PopTools для Microsoft Exel), однако нами метод был реализован самостоятельно в собственной программе, чтобы избежать скрытых настроек и неконтролируемых параметров в этих программах. Математическая реализация осуществлялась строго по оригинальной статье Эфрона.

Описание метода визуализации ("треугольные графики")

Данный метод визуализации был применен нами в ситуации, когда исследуемое животное совершало выбор из трех вариантов. Выбор не был абсолютным (т.е. животное не выбирало постоянно и однозначно один вариант, а имело набор сочетаний всех трех вариантов, предоставленных для выбора). При этом необходимо было наиболее наглядно представить данные и проверить гипотезу о том, что у животного в исследовании есть предпочтение какого-либо из предоставляемых вариантов.

Нами был выбран равносторонний треугольник, где каждая из вершин была возможным вариантом выбора. Внутри треугольника отмечали точки, соответствующие пропорции выбора вариантов. Чем чаще животное совершало определенный выбор, тем ближе точка оказывалась к одной из вершин. Если животное совершало всегда один и тот же выбор, то точка оказывалась в соответствующей вершине треугольника. Данный метод является обобщением общепринятого представления выбора из двух вариантов для выбора из трех (стоит заметить, что метод можно обобщить и на большое число выборов, хотя это отразится на наглядности). Традиционная оценка для выбора из двух вариантов это $N_1/(N_1 + N_2)$, где N_1, N_2 количество выборов первого и второго варианта соответственно. При появлении третьего варианта выбора картина становится сложнее. Можно рассматривать пары выборов и сравнивать их между собой, но это неудобно. Однако при рассмотрении только двух выборов, они оказываются равноудалены один от другого. Точка безразличия оказывается посередине. При добавлении третьего выбора он также должен оказаться равноудален от двух других выборов. Визуально это ложится в равносторонний треугольник, где вершины – три варианта выбора. Точка безразличия оказывается в центре треугольника. Положение (координаты в треугольнике со стороной 1) точки выбора вычисляется следующим образом:

$$x = \frac{\sqrt{3}}{2}\frac{N_1}{N_1 + N_2 + N_3},$$
$$y = \frac{1}{2}\frac{N_2 - N_3}{N_2 + N_3}\left(1 - \frac{2x}{\sqrt{3}}\right),$$

где N_{1-3} – абсолютное значение выбора каждого варианта. Точка безразличия оказывается в координатах $(1/2; \sqrt{3}/6)$.

Положение точек в пространстве можно анализировать: расположение, плотность группировки и пр. В результате такой график позволяет проверять различные гипотезы.

Описание животных и эксперимента

Описанная комбинация методов обработки данных была применена нами для анализа поведения при взаимодействиях взрослых самцов и детенышей нильских крыланов (*Rousettus aegyptiacus*). Работа проводилась на базе Московского зоопарка, в искусственной колонии, созданной более 20 лет назад. В нашем эксперименте участвовали три детеныша (с индивидуальными номерами 116, 141, 142) и три взрослых самца нильского крылана (№ 86, 91, 53). Нами было отмечено, что детеныши (в отсутствие матери, которая улетает кормиться) часто собираются возле взрослых самцов. При этом каждый детеныш имел свои предпочтения по выбору самца. Наблюдения проводились визуально 14 дней. Ежедневная длительность наблюдений составляла 3 ч. Суммарный объем наблюдений составил 42 ч. Методом мгновенной выборки один раз в течение 5 мин оценивалось положение каждого детеныша относительно каждого из взрослых самцов. Если расстояние между самцом и детенышем было меньше половины длины туловища детеныша, то детеныш отмечался “в контакте с самцом”, если расстояние было больше – “вне контакта”. Таким образом, для каждого дня наблюдений получено 36 положений для каждого детеныша и самца. Для каждого дня наблюдений для каждого самца и детеныша были вычислены количество мгновенных выборок положения “в контакте” и доля этого состояния в отношении к общему числу мгновенных выборок. Время, проведенное детенышем около самца, вычислялось не из общего времени наблюдений, а из времени, когда детеныш не висел на матери.

Построение “треугольных графиков”

Используя собранные данные, мы применили наш метод визуализации – треугольный график. На основе абсолютных значений, полученных по каждому дню, строился треугольный график, где вершинами были обозначены все три самца группы. Чем больше детеныш проводит времени возле конкретного самца, тем ближе к одной из вершин оказывается точка. Для каждой точки были определены координаты в формате (x, y) при использовании абсолютного значения времени, проведенного детенышем с каждым из самцов за день наблюдений (см. описание вверху). Расстояние между самцами (длина стороны треугольника) было принято за 100. Для каждого детеныша было определено среднее значение предпочтения (x_s, y_s) в течение 14 дней наблюдений, причем выбор строился на количестве времени, проведенного с каждым самцом за весь период наблюдения.

При наличии большого объема данных можно провести все вычисления на них и не применять дополнительно бутстрэп метод. В нашем же случае оригинальных данных было маловато и, для того чтобы понять, что данных по каждому детенышу достаточно для однозначных выводов, был применен бутстрэп анализ.

Применение бутстрэп анализа

Из экспериментальных данных были сгенерированы (см. выше) по 100000 выборок для каждого из детенышей. В исходную выборку включены элементы “вне контакта”, т.е. исходная выборка по каждому детенышу: 14 дней по 36 наблюдений (наблюдения ведись по 3 ч), итого 504 элемента. В каждой из 100000 сгенерированных выборок также было по 504 элемента.

Проверка гипотезы о наличии предпочтения

По 100000 сгенерированным выборкам было определено среднее количество подсадок детеныша к каждому из самцов. Рассчитанные по бутстрепу средние отличались от средних по исходной выборке менее чем на 0.1%. Также для каждой из этих сгенерированных выборок были посчитаны x_s, y_s (всего 100000 точек) и нанесены на треугольный график, таким образом было получено распределение возможных результатов. Плотность этого распределения имела ярко выраженный максимум. Вокруг этого максимума были оценены размеры областей, в которые укладываются 50, 90, 95 и 99% всех сгенерированных выборок.

В данном случае эти области были построены следующим методом. Набор теоретических координат x_s, y_s был переведен в карту плотностей $n_{i,j}$. Карта плотностей состояла из ячеек 1 на 1 – бинов. Таким образом, $n_{i,j}$ – количество точек, где координаты (x_s, y_s) удовлетворяют условиям: $x_i < x_s \le x_{i+1}$ и $y_j < y_s \le y_{j+1}$, при этом $x_{i+1} = x_i + \Delta x$ и $y_{j+1} = y_j + \Delta y$. Δx, Δy – шаг по сетке, когда мы строили карту плотностей (y нас по 1), x_1, $y_1 = 0$.

Вычисление размера областей, куда укладывается какая-то часть из выборки, можно делать разными методами. Мы считали таким образом: находили бин (т.е. ячейку), где $n_{i,j}$ наибольшее. Далее проверялись все прилегающие бины, находился бин с наибольшим значением. Этот бин прибавлялся к максимальному бину, формируя область максимума, и исключался из общего набора. Снова определялся в общем наборе бин с максимальным n_{ij} из числа прилегающих к полученному пятну, прибавлялся к пятну и исключался из общего набора и т.д. На каждом шаге в этой области вычислялась доля событий, лежащих в этой области относительно всех событий (посадок). Процедура повторялась до тех пор, пока не были определены области, в которые укладывались 50, 90, 95 и 99% от всех посадок.

Полученные этим методом результаты приведены на рис. 1. На том же графике: крест, выделенный синим цветом, обозначает точку безразличия (точка, в которой нет предпочтения ни од-

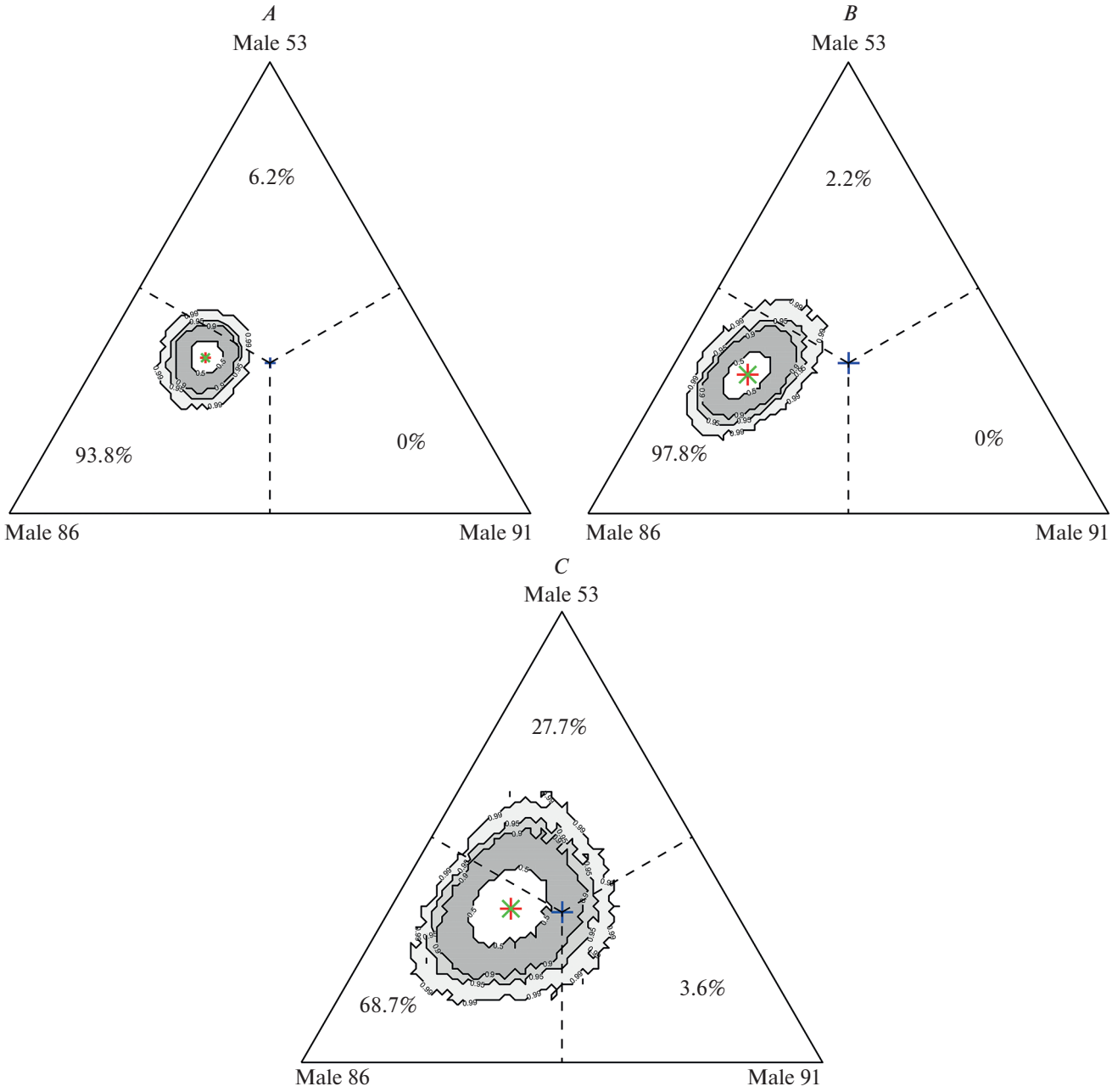

Рис. 1. Индивидуальные предпочтения детенышей в выборе самцов: *A* – детеныш 116, *B* – детеныш 141, *C* – детеныш 142. Бутстрэп анализ. Приведены области, в которые укладываются 50, 90, 95 и 99% всех сгенерированных выборок. Синий крест на графике обозначает точку безразличия (точка, в которой нет предпочтения ни одного из самцов); зеленым крестом обозначена точка с координатами (x_s, y_s), полученная на основе реальных значений; красным – средняя точка предпочтения по бутстрэп методу. Пунктирными линиями обозначены границы областей, принадлежащих каждому из самцов.

ного из самцов и которая равноудалена от вершин); зеленым цветом – точку с координатами (x_s, y_s), вычисленную на основе реальных значений, (как описано выше); красным цветом – точку предпочтения, причем средние значения получены бутстрэп методом (см. выше). Пунктирными линиями обозначены границы областей, принадлежащих каждому из самцов.

Для того чтобы узнать, есть ли достоверное предпочтение какого-либо самца, было проведено следующее. Для каждой из трех областей, ограниченных вершиной треугольника и двумя пунктирными перпендикулярами, было посчитано количество точек, которое здесь оказывается (из 100000). Далее численные значения были переведены в проценты. По этим данным можно судить

о предпочтении или, наоборот, об избегании какого-либо из самцов.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Бутстрэп анализ позволяет оценить достоверность полученного результата при относительно малой выборке. Согласно полученным нами данным, детеныши нильских крыланов достоверно имеют предпочтение, избегают одного самца и чаще выбирают других. Детеныши №№ 116 и 141 избегали самца № 91 (0% подсадок) и тяготели к контакту с № 86 самцом (93.8% для детеныша № 116 и 97.8% для детеныша № 141 соответственно). Детеныш № 142 показывал тенденцию к выбору самца № 86 (68.7%), но оказывался самым непривередливым, проводя время в контакте и с самцом № 91, и с самцом № 53. Причины такого предпочтения, вероятно, связаны с индивидуальными чертами самца № 86, поскольку все детеныши, независимо друг от друга, показали схожее предпочтение. Т.е. они способны различать не только свою мать, но и других членов группы. Бутстреп анализ в сочетании с предложенным визуальным представлением позволяет легко работать дальше с данными, делать выводы и оформлять их для публикации в различных изданиях. В качестве ремарки хочется отметить, что данные "треугольные графики" визуально сходны с фазовыми треугольными диаграммами (Dhoot et al., 2018), однако метод их построения и область применения совершенно различны. Также, при желании, метод можно применить и на большее (чем три) число вершин (вариантов), но если четыре варианта можно представить пирамидой, то пять вариантов потребуют уже четырехмерного представления. При этом нарушится наша концепция о том, что данный метод применяется для наглядности. Таким образом, для малого набора данных можно провести анализ и проверить достоверность полученного результата. Подробно все ограничения бутстрэп метода приведены в оригинальной статье (Efron, 1979), отдельно хочется подчеркнуть, что выборка должна быть однородна и не должна содержать заведомо сильно отличающихся случаев нехарактерного поведения. Также не должно быть заведомо известного исследователю изменения модели поведения в ходе сбора данных для одного пула, из которого потом будет формироваться искусственная выборка.

СПИСОК ЛИТЕРАТУРЫ

Жук Ю.А., Куликов Л.В., Пономарев Д.А., 2008. Экспериментальное исследование применения мультимедийной наглядности при обучении студентов химии // Вестник Санкт-Петербургского университета. Социология. № 3. С. 374–379.

Мантейфель Б.П., 1980. Экология поведения животных. Пер. с англ. М.: Мир. 220 с.

Попов Р.С., 2008. Руководство по научным исследованиям в зоопарках. Методические рекомендации по этологическим наблюдениям за млекопитающими в зоопарках. М.: Московский зоопарк. 165 с.

Bart J., Notz W.J., Fligner M.A., Notz W.I., 1998. Sampling and statistical methods for behavioral ecologists. Cambridge University Press. 330 p.

Csada R.D., James P.C., Espie R.H., 1996. The "file drawer problem" of non-significant results: does it apply to biological research? // Oikos. V. 76. P. 591–593.

Dhoot A.S., Naha A., Priya J., Xalxo N., 2018. Phase Diagrams for Three Component Mixtures in Pharmaceuticals and its Applications // Journal of Young Pharmacists. V. 10. P. 132–137.

Efron B., 1979. Bootstrap Methods: Another Look at the Jackknife // Ann. Statist. V. 7. P. 1–26.

Jennions M.D., Møller A.P., 2003. A survey of the statistical power of research in behavioral ecology and animal behavior // Behavioral Ecology. V. 14. P. 438–445.

BOOTSTRAP METHOD APPLICATION IN BEHAVIORAL DATA ANALYSIS AND THE DATA VISUALIZATION METHOD OF CHOICE FROM THREE OBJECTS

A. V. Dudorova[1, *], D. A. Podgrudkov[2, 3, **], E. P. Kruchenkova[1, *]**

[1]Lomonosov Moscow State University, Faculty of Biology, Moscow, 119991 Russia

[2]Lomonosov Moscow State University, Faculty of Physics, Moscow, 119991 Russia

[3]Lomonosov Moscow State University, Scobeltsyn Nuclear Physics Research Institute, Moscow, 119991 Russia

**e-mail: dudorova.anastasia.2012@post.bio.msu.ru*

***e-mail: d.a.podgrudkov@physics.msu.ru*

****e-mail: ekruster@gmail.com*

Both bootstrap analysis method and data presentation in behavioral tests are suggested to transcribe preference choice and/or multiple choices, the study of the Egyptian fruit bat pup (*Rousettus aegyptiacus*) preferences serving as an example. The use of the bootstrap method allows for the reliability of a given result from poor data to be evaluated. The new method of visualization proposed allows to clearly present data at a complex choice.

Keywords: Egyptian fruit bat, pup, data presentation, behavior