

УДК 57.087.1

<https://doi.org/10.25587/2222-5404-2026-23-1-32-42>

Оригинальная научная статья



Применение бутстрапа в MS Excel для малых выборок на примере временных рядов фотосинтетических данных

А. П. Максимов 

Институт биологических проблем криолитозоны СО РАН, г. Якутск, Российская Федерация

✉ ayal01@yandex.ru

Аннотация

Проблема малой выборки данных (объемом до 30 элементов) и ее статистических оценок в контексте расширения на генеральную совокупность в эколого-биологических исследованиях всегда актуальна и решается по-разному в зависимости от целей и задач. В то же время исследователи нередко просто обходят проблему вследствие неясности применения требуемой статистики на практике и математической сложности для неспециалиста, особенно если данные обрабатываются исключительно в электронной таблице MS Excel без использования макросов, VBA-подпрограмм и надстроек, а также специализированных статистических программ. Поставлена цель получить оценки статистик малых выборок и их генеральных статистических параметров в результате дополнительной генерации выборок средствами MS Excel. Для этого малые выборки размножены классическим бутстрапом, блочным бутстрапом и выборочной генерацией чисел при помощи формул в MS Excel и оценены их эмпирические приближения к истинному распределению с использованием элементарных статистических функционалов (среднее, медиана, стандартное отклонение, процентиль, доверительный интервал). В качестве примера приводится сравнительная оценка указанных функционалов суточного фотосинтеза растения, который является временным рядом с присущими ему нюансами по сравнению с классическим случаем массива случайных и независимых величин. Все предлагаемые вычисления не сложны и представлены в виде комбинированных формул в MS Excel, которые могут быть использованы как в научных, так и в образовательных целях. Показано, что увеличение количества малых выборок упомянутыми выше способами свидетельствует в пользу использования блочного бутстрапа для состоятельной оценки статистических параметров временного ряда.

Ключевые слова: выборка, временной ряд, статистические функции, MS Excel, бутстрап, генерация чисел, боксплот («ящик с усами»), фотосинтез, суточная динамика, экология

Финансирование. Работа выполнена в рамках базового проекта ИБПК СО РАН «Исследование биогеохимических циклов и адаптивных реакций растений субарктических и арктических экосистем северо-востока России», номер государственной регистрации в ЕГИСУ: АААА-А21-121012190034-2

Для цитирования: Максимов А. П. Применение бутстрапа в MS Excel для малых выборок на примере временных рядов фотосинтетических данных. *Вестник СВФУ*. 2026;23(1):32–42. DOI: <https://doi.org/10.25587/2222-5404-2026-23-1-32-42>

Original article

Using bootstrap in MS Excel for small samples with photosynthesis time series data as an example

Ayal P. Maksimov 

Institute for Biological Problems of Cryolithozone SB RAS, Yakutsk, Russian Federation

✉ ayal01@yandex.ru

Abstract

The problem of the small data sample (with the volume of elements up to 30) and its statistical estimates in the context of expansion to the general population in ecological and biological studies is always relevant and is solved differently depending on the goals and objectives. At the same time, researchers often simply bypass the problem because of unclear application of the required statistics in practice and the mathematical complexity for non-specialists, especially if the data is processed exclusively in MS Excel spreadsheet without the use of macros, VBA routines and add-ins, as well as specialized statistical programs. The goal is to obtain estimates of the statistics of small samples and their general statistical parameters as a result of additional sample generation by means of MS Excel. For this purpose, small samples are multiplied by the classic bootstrap, block bootstrap and selective generation of numbers using formulas in MS Excel and their empirical approximations to the true distribution are estimated using elementary statistical functions (mean, median, standard deviation, percentile, confidence interval). As an example, a comparative assessment of the specified functions of daily plant photosynthesis is given, which is a time series with its inherent nuances compared to the classical case of an array of random and independent variables. All the proposed calculations are not complicated and are presented in the form of combined formulas in MS Excel, usable for both scientific and educational purposes. It is shown that an increase in the number of small samples by the above-mentioned methods testifies in favor of using the block bootstrap for a consistent assessment of the statistical parameters of a time series.

Keywords: sample, time series, statistical functions, MS Excel, bootstrap, number generation, boxplot, photosynthesis, diurnal dynamics, ecology

Funding. The work was carried out within the framework of a basic project of IBPC SB RAS «Study of biogeochemical cycles and adaptive reactions of plants in subarctic and arctic ecosystems of the northeast of Russia», registration number in USIAS: AAAA-A21-121012190034-2

For citation: Maksimov A. P. Using bootstrap in MS Excel for small samples with photosynthesis time series data as an example. *Vestnik of NEFU*. 2026;23(1):32–42. DOI: <https://doi.org/10.25587/2222-5404-2026-23-1-32-42>

Введение

В силу условий применения многих статистических критериев общепринято, что для репрезентативности выборка в минимальном счете должна насчитывать не менее 20–30 элементов (далее упоминаемых как n) [1–5]. Однако даже тысячный объем выборки не дает оснований достоверно установить вид распределения данных [6–8], а малая выборка с $n=3–5$, по которой есть своя специальная статистика, должна быть не исходной, а пост-анализом некоторого массива значений из ранее уже исследованной выборки с достаточным n [9]. На практике

в экологии и смежных дисциплинах оперируют малым количеством исходных n (в большинстве случаев 5–10), какие методы применять и применять ли вообще в целях нивелирования размера выборки и расширения на генеральную совокупность решает сам исследователь в зависимости от цели и задач.

Для повышения точности оценки данных, вплоть до уровня генеральной совокупности и независимо от вида распределения, вместо ручных статистических анализов малых выборок (до $n=30$) нередко применяют вычислительные возможности компьютера для генерации достаточного количества данных, в частности методами Монте-Карло [10–12], «складного ножа» [13–15], бутстрапа [16–18], в общем смысле именуемых ресамплингом (с английского «re-sampling» – повторный отбор образца). Вопрос полной корректности таких процедур остается открытым [19], но их полезность несомненна, и при подходящих задачах ими следует пользоваться.

Настоящее исследование ставит целью оценку статистик малой выборки и ее генеральных параметров при разных способах искусственного увеличения количества малых выборок средствами MS Excel, стандартно используемой в научной и образовательной среде. Для этого проводится дополнительная компьютерная генерация малых выборок (обозначаемых как m) до надлежащего статистического количества ($m=1000$) с дальнейшей оценкой их эмпирического приближения к истинному распределению при помощи элементарных статистических функционалов (среднее, медиана, стандартное отклонение, процентиль, доверительный интервал).

В качестве объекта анализа рассматривается типичный экофизиологический паттерн – суточная динамика фотосинтеза растения (ассимиляция CO_2), которая является временным рядом с характерным суточным откликом. Требуется рассмотреть влияние разных методов генерации чисел на оценки центральных моментов и интервалов. При этом предполагается нормальное распределение данных, что проверяется определенным статистическим критерием.

Задача состоит в составлении несложных формул в MS Excel для генерации дополнительных выборок разными способами, которые могли бы использоваться неспециалистами в статистике в научных и образовательных целях.

Для озвученных цели и задачи наиболее подходит метод бутстрапа, суть которого состоит в произвольной генерации чисел из исходной измеренной выборки данных с созданием новой искусственной выборки из этих чисел, с последующим возвратом этих чисел в исходную выборку для вывода следующего случайного набора данных. То есть генерируемые числа могут повторяться, и так будет происходить до нужного количества итераций в целях последующей оценки той или иной статистики на основании созданного массива.

Имеются разновидности бутстрапа для разных задач. Классический (простой) бутстрап предусматривает работу с независимыми случайными величинами и в этом смысле не совсем подходит для детерминированного временного ряда с зависимыми переменными, каковой является динамика суточного фотосинтеза, поскольку числа генерируются без учета характера и распределения данных во времени, другими словами, данные классического бутстрапа не моделируют структуру исходных данных временного ряда и результат является несостоятельным.

К примеру, для суточной ассимиляции CO_2 растением простой бутстрап генерирует в числе других невозможные в природе суточные динамики, особенно если в исходной выборке присутствуют отрицательные значения (в результате ночного дыхания растения), соответственно, такие модели не являются адекватными, что нивелирует их ценность. Тем не менее классический бутстрап состоятелен для точечных оценок центральных моментов, так как для этого структура ряда данных не имеет значения.

Поэтому, если нужно учесть временную структуру ряда, классический бутстрап необходимо видоизменить: вводить дополнительные правила генерации чисел или наряду с простым бутстрапом использовать иной способ для корректного набора данных временной серии. Такие алгоритмы бутстрапа специально для временных серий известны [20]: например, блочный [21], решетчатый [22, 23] и локальный [24, 25], которые в математическом изложении сложны для понимания и применения неспециалистами. Наиболее релевантным из них признается блочный метод, который можно реализовать в MS Excel в упрощенном виде по аналогии с классическим бутстрапом, только вместо набора отдельных чисел оперируя блоками из нескольких чисел, выделяемых по тому или иному признаку.

Альтернативно можно улучшить малую выборку уплотнением значений, то есть генерацией дополнительных чисел между значениями исходных данных, что даст увеличение массива с сохранением структуры суточной динамики.

В настоящей статье рассматриваются упомянутые выше подходы в целях несложной работы с ними в MS Excel неспециалистами в статистике:

- классический бутстрап (КБ) выборки для точечных оценок центральных параметров;
- блочный бутстрап (ББ) выборки для характеристики временного ряда как основной метод;
- выборочная генерация чисел (ВГЧ) между значениями в выборке как альтернативный вариант для характеристики временного ряда.

Материалы и методы

Объектами (материалами) анализа служат три малые выборки данных ($n=10$) по фотосинтезу березы в Центральной Якутии (вид *Betula platyphylla*) в начале, середине и конце вегетационного периода, а именно усредненные по времени замеры суточные значения ассимиляции CO_2 (A или фотосинтез, в микромолях $\text{м}^{-2} \text{с}^{-1}$) 04.06.2023, 11.07.2023 и 28.08.2023 по часам суток 0, 3, 6, 7, 8, 9, 12, 15, 18, 21. Почасовой интервал с 6 до 9 ч связан с тем, что, по многолетним данным, именно в это время суток наблюдается максимальный фотосинтез, остальные точки можно замерять с большим интервалом в 2–3 ч для получения достоверного суточного паттерна. Измерения проводились на научной лесной станции ИБПК СО РАН «Спасская Падь» ($62^{\circ}15'$ с.ш., $129^{\circ}37'$ в.д.) в типичной таежной экосистеме, инфракрасным газоанализатором LCi (ADC Bioscientific, СК).

Получены следующие массивы данных фотосинтеза (в микромоль $\text{м}^{-2} \text{с}^{-1}$) в соответствии с указанными выше датами и временными интервалами суток: $\{-0,93; 0,01; 7,60; 8,42; 8,51; 7,22; 7,16; 5,60; 5,10; 0,57\}$, $\{-0,55; -0,56; 11,13; 12,85; 12,29; 10,65; 6,36; 5,73; 1,65; 0,35\}$, $\{-0,52; -0,40; 1,66; 8,92; 11,60; 11,75; 8,53; 7,48; 4,63; -0,64\}$.

Оцениваемые статистики: средняя арифметическая μ как несмещенная оценка математического ожидания, стандартное отклонение средней CO как мера рассеяния, медиана Me (50-й процентиль) для сравнения с μ и косвенной оценки распределения, доверительный интервал средней арифметической (ДИС), квантили $q_{0,025}$ (2,5-й процентиль) и $q_{0,975}$ (97,5-й процентиль) для оценки размаха. Статистики вычисляются по исходной выборке (В) и по результатам классического бутстрапа (КБ), блочного бутстрапа (ББ), выборочной генерации чисел (ВГЧ) встроенными функциями MS Excel, использование которых общеизвестно и не вызывает трудностей.

Классический бутстрап с возвратом чисел реализован через комбинированную формулу в MS Excel «=ИНДЕКС(массив;СЛУЧМЕЖДУ(1;ЧСТРОК(массив));номер_столбца)». Данная формула возвращает случайное число (которое может повторяться) из столбца исходной выборки, которое генерируется до получения

новой выборки с $n=10$ и далее в формульном виде размножается до нужного количества m новых выборок.

Для правильной и удобной работы с исходными выборками рекомендуется свести их в одну таблицу по столбцам, присвоить этой таблице имя и далее работать уже с именем как с единым массивом всех выборок, выбирая ту или иную выборку для бутстрапа только указанием номера столбца этой выборки в формуле, а не изменяя всю формулу бутстрапа для каждой выборки раз за разом.

Для блочного бутстрапа используется та же формула, что и для классического, только меняются аргументы в «(1;ЧСТРОК(массив))», куда вписываются номера строк начала и конца того или иного блока значений внутри столбца выборки. Выбор блоков осуществлялся по степени близости величин чисел выборки друг к другу, в итоге получилось по 4 блока в каждой выборке. Далее каждая полученная формульная выборка размножается до нужного их количества.

ВГЧ между значениями выборки для повышения их плотности осуществлена по формуле в MS Excel «=СЛЧИС()*(большее_число-меньшее_число)+меньшее_число», выбирая соответствующие соседние значения для получения дополнительных значений между ними (интервал – каждые полчаса измерений). Формульный столбец далее также размножается до нужного количества.

Классический и блочный бутстрапы проведены в размерности $10(n)*1000(m)=10000$ чисел, ВГЧ – с 10 исходных точек каждые 1–3 часа измерений до 33 сгенерированных чисел по 1000 повторов (всего $43(n)*1000(m)=43000$ чисел) с получасовым интервалом (получасовой интервал общепринят как временной стандарт для представления данных потоков вещества и энергии).

После получения нужного количества сгенерированных массивов каждой выборки ($m=1000$) в них вычисляются указанные выше статистики, после чего выводятся общие средние по всем выборкам (то есть средние средних) для их оценки в результате генераций чисел и для оценки самих статистических функций в предположении, что эмпирическим приближением в результате генераций чисел получены их достоверные значения в генеральной совокупности. Каждая тысяча генерации выборок дополнительно повторена 10 раз при помощи функции в MS Excel «Данные → Анализ «Что если» → Таблица данных», которая удобна для автоматизации формульной работы с изменениями данных.

Для графического представления и сравнения статистических функционалов использован встроенный тип графика в MS Excel с версии 2016 «Ящик с усами» (или боксплот), автоматически рассчитывающий из исходных данных для графика μ , Me (или квартиль Q_2), квартили Q_1 и Q_3 (25-й и 75-й процентиля соответственно) минимальное и максимальное значение, выбросы.

Нормальность распределений данных выборок и сгенерированных массивов средних оценивалась критерием Жака-Бера (JB) [26] как одним из наиболее простых и одновременно мощных критериев. Критерий JB основан на анализе асимметрии и эксцесса, имеет формулу $n*(S^2/6+K^2/24)$, где S и K – асимметрия и эксцесс массивов, определяемые встроенными функциями MS Excel; точная p -значимость критерия JB определяется по формуле в MS Excel «=ХИ2.РАСП.ПХ(значение_ JB ;2)» с критическим уровнем $\alpha=0,05$. Итоговая комбинированная формула в MS Excel с выводом только конечного результата – точной p -значимости критерия JB – вводится в ячейку таблицы как «=ХИ2.РАСП.ПХ(СЧЁТ(выборка)*(СКОС(выборка)^2/6+(ЭКССЕСС(выборка)^2/24));2)».

Нормальность распределения в некоторых случаях дополнительно проверялась другими критериями при помощи статистической программы PAST [27].

Результаты

Итоговые усредненные точечные значения статистик исходных выборок и результаты генерации их дополнительных массивов до $m=1000$ с последующим усреднением в 10-кратных прогонах в виде значений средних арифметических статистических функций сведены в таблице. Оценки самих основных статистических функций, необходимых и достаточных для рассмотрения общих тенденций (μ , CO , $ДИС$), представлены в виде боксплотов на рисунке.

Таблица

Статистики исходных (В) и сгенерированных выборок (КБ, ББ, ВГЧ) с их процентными соотношениями по суточному фотосинтезу *Betula platyphylla*

Table

Statistics of initial (B) and generated samples (CB, BB, SNG) with their percent ratios on diurnal photosynthesis of *Betula platyphylla*

	В	КБ	ББ	ВГЧ	% КБ от В	% ББ от В	% ВГЧ от В
μ	5,41	4,95	5,41	4,86	-8%	0%	-10%
CO	4,72	3,40	4,71	3,93	-28%	0%	-17%
$ДИС(t)$	3,37	2,43	3,37	1,21	-28%	0%	-64%
Me	6,16	5,90	6,08	5,38	-4%	-1%	-13%
$q0,025$	-0,63	-0,15	-0,57	-0,62	-77%	-9%	-2%
$q0,975$	10,98	8,30	10,84	10,91	-24%	-1%	-1%
$p JB$	0,46	0,03	0,12	0,38	-94%	-74%	-17%

Примечание: представлены усредненные значения фотосинтеза (ассимиляция CO_2 в микромолях $m^{-2} c^{-1}$) трех выборок (от 04.06.2023, 11.07.2023, 28.08.2023) и результатов генерации чисел по всем этим датам (каждая – по 10 прогонов с $m=1000$ выборок), с процентными сравнениями полученных оценок. Проценты изменений величин между исходными и конечными числами вычислены в MS Excel как «=ячейка_результата/ячейка_исходного_значения-1». Обозначения: В – выборки, КБ – классический бутстрап, ББ – блочный бутстрап, ВГЧ – выборочная генерация чисел, μ – средняя арифметическая, CO – стандартное отклонение средней арифметической, $ДИС(t)$ – доверительный интервал средней арифметической по t -статистике, Me – медиана (соответствует 50-му процентилу), $q0,025$ и $q0,975$ – 2,5-й и 97,5-й процентиля, отражающие 95%-й размах данных, $p JB$ – точное p -значение критерия Жака-Бера на нормальность распределения (для сгенерированных данных представлен для массива средних арифметических)

Note: the averaged values of photosynthesis (CO_2 -assimilation in micromoles $m^{-2} s^{-1}$) of three samples (on 06.04.2023, 07.11.2023 and 08.28.2023) and the results of generating numbers (10 runs for each, with $m=1000$ samples) for all the dates are presented, with percentage comparison scores. The percentage changes in values between the initial and final numbers are calculated in MS Excel as «=result_cell/initial_value_cell-1». Designations: B – samples, КБ – classic bootstrap, ББ – block bootstrap, ВГЧ – selective number generation, μ – arithmetic mean, CO – standard deviation of the arithmetic mean, $ДИС(t)$ – confidence interval of the arithmetic mean according to t -statistics, Me – median (corresponds to the 50th percentile), $q0.025$ and $q0.975$ – 2.5th and 97.5th percentiles, reflecting the 95% range of the data, $p JB$ – exact p -value of the Jarque-Bera test for the normality of distribution (for the generated data $p JB$ is presented for an array of arithmetic means)

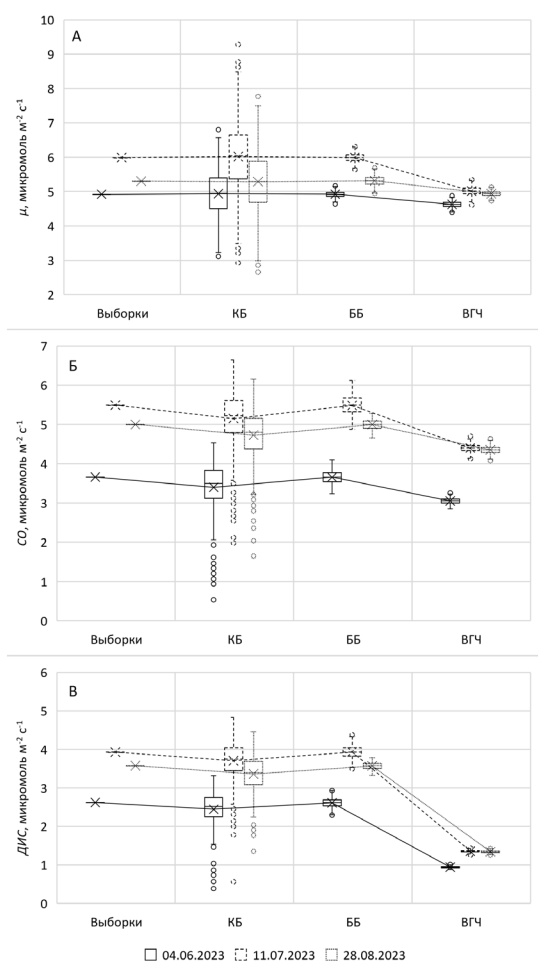


Рис. Боксплоты («ящики с усами») средних арифметических выборок и сгенерированных моделями массивов
Fig. Boxplots («box and whiskers plot») of arithmetic means of samples and model-generated arrays

Примечание: график А – средние арифметические, график Б – стандартные отклонения средних, график В – доверительные интервалы средних. Обозначения: КБ – классический бутстрап, ББ – блочный бутстрап, ВГЧ – выборочная генерация чисел. Кружки над/под черточками усов – выбросы; верхняя черточка усов – максимальное значение из массива данных; верхний край бокса – 3-й квартиль (3/4 или 75%); линия в боксе – медиана (2-й квартиль, 1/2 или 50%); крестик в центре бокса – средняя арифметическая; линии между крестиками соединяют соответствующие средние массивов; нижний край бокса – 1-й квартиль (1/4 или 25%); нижняя черточка усов – минимальное значение из массива данных

Note: graph А – arithmetic means, graph Б – standard deviations of means, graph В – confidence intervals of means. Designations: КБ – classic bootstrap, ББ – block bootstrap, ВГЧ – selective number generation. Circles above/below the whiskers are outliers; the top bar of the whiskers is the maximum value from the data array; the upper edge of the box is the 3rd quartile (3/4 or 75%); the line in the box is the median (2nd quartile, 1/2 or 50%); the cross in the center of the box is the arithmetic mean; the lines between the crosses connect the corresponding means of the arrays; the lower edge of the box is the 1st quartile (1/4 or 25%); the bottom line of the whiskers is the minimum value from the data array

Обсуждение

Общие точечные оценки статистик показывают, что значения для ББ практически идентичны с таковыми выборок (В). Оценки КБ имеют незначительное отклонение для μ и Me , заметное снижение по CO , $ДИС$ и значительное сужение диапазона квантилей, особенно по нижним 2,5%. Наблюдается ощутимое расхождение результатов ВГЧ с таковыми выборок, за исключением μ и квантилей, при этом ВГЧ скорее сходится к КБ, нежели к ББ.

В целом КБ выдает систематический понижающий тренд, что связывается с большим включением в массивы отрицательных значений в силу природы КБ по сравнению с другими типами генерации чисел. Колебания результатов по ВГЧ относительно выборок определенно связаны с измененным количеством n , особенно по $ДИС(t)$, где n – один из прямых ключевых параметров расчета.

Нормальность распределения не отвергается на достаточно высоких уровнях точного p -значения кроме КБ, однако и для ББ p -значение JB обнулялось или находилось в пограничной области (0,03-0,05) в некоторых случаях прогонов. В первую очередь надо отметить относительно высокие значения p (0,44-0,50) выборок по тесту JB , что дополнительно подтверждено и другими критериями, кроме выборки от 04.06.2023, где часть тестов – Шапиро-Уилка и Андерсона-Дарлинга – отвергают нулевую гипотезу с $p=0,036$. Однако частотные статистические диаграммы выборок (не приведены) далеки от гауссовской, а по мере генерации новых выборок p -значение существенно снижается, вплоть до обнуления (за исключением ВГЧ). Данное обстоятельство свидетельствует о важности критического подхода к вопросу о нормальности распределения в частности и к параметрической оценке данных в целом.

Значение имеют характер и масштаб данных, а именно их определенный временной вид с большим размахом величин и наличием отрицательных и нулевых значений в начале и конце цикла. Центральные предельные теоремы, обосновывающие применение нормального распределения, в числе условий применения подразумевают независимость и примерно одинаковый масштаб величин, что в случае суточной динамики фотосинтеза не соблюдается. В частности, в случае КБ по мере хаотического набора чисел из выборки усиливаются несоответствия масштабов величин и структур временных серий, что приводит к нарушениям симметрии распределения для критерия JB и редукции p -значения критерия. Таким образом, если работа с исходной малой выборкой как с данными с нормальным распределением формально-статистически достаточно обоснована, то бутстрапы с $m=1000$ не подтверждают это в полной мере. Тест же с увеличением m до 100 000 приводит к уверенному отклонению нулевой гипотезы о нормальности ($p\ JB=0,00095$).

При сравнении разбросов статистических функций виден разительный контраст между КБ и другими моделями генераций чисел (рис.), с тяжелым нижним хвостом в распределениях по КБ. Интервалы параметров КБ чрезвычайно широки в силу безусловного характера генерации, не учитывающего структуру временного ряда, и такие размахи несостоятельны для достоверных оценок статистических функционалов. Соответственно, массивы КБ не могут рассматриваться как адекватная замена данных временной серии, которые могли бы быть получены в случае их реальных измерений.

В отличие от КБ, модели ББ и ВГЧ сохраняют временную структуру исходных данных и показывают адекватные величины и разброс параметров, при этом ВГЧ выдает относительно неточные оценки, что выражено сдвигом вниз на рисунке. Разрешение модели ВГЧ можно еще увеличить частотами в десятки минут и меньше, но для суточной динамики фотосинтеза это не имеет практического и

научного смысла. Помимо всего, ВГЧ относительно трудоемка для реализации в MS Excel – необходимо осуществить большее количество операций по сравнению с ББ и не ошибиться в моментах, связанных с простой внимательностью. Тем не менее ВГЧ заслуживает большего внимания как метод, показывающий наименьший разброс статистик.

Таким образом, наилучшее соответствие с исходными выборками и лучшая оценка статистических функций получены в результате ББ, и именно этот метод рекомендуется к применению в целях эмпирического приближения к генеральной совокупности как наиболее адекватный метод для временных рядов, что также подтверждено математическими исследованиями в целом [20, 28]. По возможности желательно ввести в модель ББ рекомендуемые специалистами поправки на рецентрирование, смещение и стационарность [28], но для задач эколого-биологической направленности это излишне и предлагаемая упрощенная формульная работа в MS Excel вполне достаточна.

Точечные оценки результатов бутстрапирования варьируют от 0 до существенных 28% разницы по сравнению с исходными выборками, т. е. нельзя пренебрегать практической значимостью бутстрапа, и его применение рекомендуется. При этом необходимо отметить, что результаты довольно изменчивы при каждом прогоне моделей, особенно сильно по КБ, в то время как для ББ и ВГЧ результаты варьируют незначительно. Реальная разница результатов выявляется при рассмотрении не точечных моментов, а разброса величин самих статистических функционалов, в первую очередь квантилей, что и является определяющим. Поэтому при оценках статистик следует не ограничиваться точечными средними, а обязательно смотреть также на их разброс – в MS Excel для этого предусмотрен наглядный и быстрый способ в виде построения боксплотов со встроенными вычислениями необходимых и достаточных функций, что и было использовано в настоящем исследовании.

Подводя итог, следует резюмировать следующее:

а) влияние условия нормальности распределения или ее отсутствия на оценки статистических параметров не очевидно. Более того, с ростом m (количество генерируемых выборок) нормальность распределения уверенно отклоняется;

б) КБ временных рядов представляет крайне широкий размах параметров, что не позволяет считать их состоятельными оценками. При этом точечные центральные значения статистик хорошо удерживаются КБ, и его можно абсолютно корректно использовать для массивов случайных и независимых величин;

в) ББ и ВГЧ с сохранением структуры временного ряда данных представляют существенно компактные диапазоны параметров распределения по сравнению с КБ, при этом ВГЧ выдает относительно неточную оценку;

г) практическая значимость точечных оценок параметров сильно варьирует между вариантами бутстрапов, в связи с чем необходимо проводить интервальные оценки интересующих статистических функций.

Заключение

Показано, что методом бутстрапа можно относительно легко пользоваться на уровне несложных формул и встроенных средств электронной таблицы MS Excel без надстроек, макросов, VBA-подпрограмм и без использования специальных сторонних программ, что полезно для неспециалистов в статистике и студентов. Могут быть использованы разные варианты метода, каждый из которых имеет свою сферу применения:

– для временных рядов (и иных массивов с определенной структурой) необходимо использовать блочный бутстрап;

– для массивов данных без определенной структуры (случайные независимые величины) – классический бутстрап;
– выборочная генерация чисел может использоваться для временных рядов в сравнительных целях с блочным бутстрапом, но требует дальнейшего уточнения для полноценного применения.

Имеет значение методологический аспект статьи, состоящий в том, что исследователю важно понимать суть производимых им вычислений и используемых критериев, и непосредственная работа с формулами, особенно с комбинированными, в MS Excel позволяет этому закрепиться в навыках. Рекомендуется использование встроенных средств MS Excel, значительно облегчающих рутинные операции, в частности, автоматизации формул для их многократного повторения («Данные → Анализ «Что если» → Таблица данных») и боксплота («ящика с усами») с автоматическим вычислением статистических параметров.

Л и т е р а т у р а / References

1. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. Abington: Routledge; 1988:590.
2. Hair JF., Black WC., Babin BJ., et al. Multivariate Data Analysis. 7th Edition. New Jersey: Pearson Prentice Hall; 2010:816.
3. Moore DS., McCabe GP., Craig BA. Introduction to the Practice of Statistics. 10th Edition. New York: WH. Freeman; 2021:784.
4. Schochet PZ., Chiang HS. Do Experiments and Quasi-Experiments Give Different Results? Final Report. National Center for Education Evaluation and Regional Assistance. Institute of Education Sciences. U.S. Department of Education. NCEE Report 2010-4007. 2010;1-86.
5. Green SB. How many subjects does it take to do a regression analysis? Multivariate Behavioral Research. 1991;26(3):499-510.
6. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? Заводская лаборатория. 1991;57(7):64-66.
- Orlov AI. Is the distribution of observation results often normal? Zavodskaya laboratoriya. 1991;57(7):64-66 (in Russ.).
7. Орлов А.И. Прикладная статистика. Учебник. 2-е изд. Москва: Экзамен; 2007:672.
- Orlov AI. Applied statistics. Textbook. 2nd ed. Moscow: Examen; 2007:672 (in Russ.).
8. Селезнев В.Д., Денисов К.С. Исследование свойств критериев согласия функций распределения данных с гауссовой методом Монте-Карло для малых выборок. Заводская лаборатория. 2005;71(1):68.
- Seleznev VD, Denisov KS. Study of properties of agreement criteria of data distribution functions with Gaussian Monte-Carlo method for small samples. Zavodskaya laboratoriya. 2005;71(1):68 (in Russ.).
9. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов исследований. Ленинград: Энергоатомиздат; 1985:248.
- Novitskiy PV, Zograf IA. Estimation of errors in research results. Leningrad: Energoatomizdat; 1985:248 (in Russ.).
10. Metropolis N., Ulam S. The Monte-Carlo Method. Journal of the American Statistical Association. 1949;44(247):335-341.
11. Соболев И.М. Метод Монте-Карло. Популярные лекции по математике. Москва: Наука; 1972;46:64.
- Sobol IM. Monte-Carlo Method. Popular Lectures on Mathematics. Moscow: Nauka; 1972;46:64 (in Russ.).
12. Москвитина А.В., Феоктистова М.В., Толстых О.Д. Реализация метода Монте-Карло при исследовании случайных процессов. Молодая наука Сибири: электрон. науч. журн. 2020;3(9).
- Moskvitina AV, Feoktistova MV, Tolstykh OD. Implementation of the Monte-Carlo method in research random process. Molodaya nauka Sibiri: ehlektronnyj nauchnyj zhurnal = Young science of Siberia: electronic scientific journal 2020;3(9) (in Russ.).

13. Quenouille MH. Approximate Tests of Correlation in Time-Series. Journal of the Royal Statistical Society: Series B. 1949;11(1):68-84.
14. Tukey JW. Bias and Confidence in Not-Quite Large Samples. Annals of Mathematical Statistics. 1958;29(2):614.
15. Efron B. The Jackknife, the Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics. Monograph. Philadelphia: SIAM; 1982(38):85.
16. Efron B. Bootstrap methods: Another look at the jackknife. Ann. Statist. 1979;7(1):1-26.
17. Davison A., Hinkley D. Bootstrap Methods and Their Application. Cambridge: Cambridge University Press; 1997:582.
18. Chernick MR. Bootstrap methods: a guide for practitioners and researchers. Hoboken: Wiley; 2007:384.
19. Шитиков В.К. Использование рандомизации и бутстрапа при обработке результатов экологических наблюдений. Принципы экологии. 2012;1(1):4-24.
- Shitikov VK. Use of randomization and bootstrap at processing the results of ecological observations. Principles of ecology. 2012;1(1):4-24 (in Russ.).
20. Бюльман П. Бутстрап-схемы для временных рядов. Квантиль. 2007;(3):37-56.
- Bühlmann P. Bootstrap schemes for time series. Quantile. 2007;(3):37-56 (in Russ.).
21. Künsch HR. The Jackknife and the Bootstrap for General Stationary Observations. Annals of Statistics. 1989;17:1217-1261.
22. Grenander U. Abstract Inference. Hoboken: Wiley; 1981:526.
23. Kreiss J-P. Bootstrap procedures for AR(∞)-processes. In: Jöckel KH., Rothe G., Sendler W. (eds). Bootstrapping and Related Techniques. Lecture Notes in Economics and Mathematical Systems. Springer, Berlin, Heidelberg; 1992;376:107-113.
24. Neumann M., Kreiss J-P. Regression-Type Inference in Nonparametric Autoregression. Annals of Statistics. 1998;26(4):1570-1613.
25. Paparoditis E., Politis DN. The Local Bootstrap for Kernel Estimators under General Dependence Conditions. Annals of the Institute of Statistical Mathematics. 2000;52:139-159.
26. Jarque CM., Bera AK. A Test for Normality of Observations and Regression Residuals. International Statistical Review. 1987;55(2):163-172.
27. Hammer Ø., Harper DAT., Ryan PD. PAST: Paleontological Statistics software package for education and data analysis. Palaeontologia Electronica. 2001;4(1):1-9.
28. Анатольев С. Основы бутстрапирования. Квантиль. 2007;(3):1-12.
- Anatolyev S. The basics of bootstrapping. Quantile. 2007;(3):1-12 (in Russ.).

Сведения об авторе

МАКСИМОВ Айаал Петрович – инженер-исследователь, ФГБУН «Институт биологических проблем криолитозоны СО РАН», г. Якутск, Российская Федерация, <https://orcid.org/0000-0003-4761-6956>, WoS Researcher ID: AAA-7513-2022, Scopus Author ID: 8529446900, РИНЦ SPIN-код: 3272-2334, e-mail: ayal01@yandex.ru

About the author

Ayal P. MAKSIMOV – engineer researcher, IBPC SB RAS, Yakutsk, Russian Federation, <https://orcid.org/0000-0003-4761-6956>, WoS Researcher ID: AAA-7513-2022, Scopus Author ID: 8529446900, RISC SPIN-code: 3272-2334, e-mail: ayal01@yandex.ru

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов

Conflict of interest

The author declares no conflict of interest

Поступила в редакцию / Submitted 14.10.2025

Поступила после рецензирования / Revised 03.11.2025

Принята к публикации / Accepted 19.01.2026