

ДИСКУССИОННАЯ
СТАТЬЯ

К ВОПРОСУ ОБ ИСПОЛЬЗОВАНИИ СТАТИСТИЧЕСКОГО АНАЛИЗА
ПРИ ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТОВ
В ОБЛАСТИ ЭКОЛОГИЧЕСКОЙ МИКРОБИОЛОГИИ

© 2019 г. А. Ю. Каллистова^{a, *}, А. Ф. Сабреков^{b, c}, В. М. Гончаров^d,
Н. В. Пименов^a, М. В. Глаголев^{b, c, d, e}

^aИнститут микробиологии им. С.Н. Виноградского, ФИЦ Биотехнологии РАН, Москва, 119071 Россия

^bЮгорский государственный университет, Ханты-Мансийск, 628011 Россия

^cТомский государственный университет, Томск, 634050 Россия

^dМосковский государственный университет им. М.В. Ломоносова, Москва, 119991 Россия

^eИнститут лесоведения РАН, Успенское, Московская обл., 143030 Россия

*e-mail: kallistoanna@mail.ru

Поступила в редакцию 09.11.2018 г.

После доработки 22.11.2018 г.

Принята к публикации 30.11.2018 г.

Статистический анализ является неотъемлемой частью любого эксперимента, поскольку с его помощью исследователь делает выводы из проделанной работы. Для неспециалистов существуют учебники, подробно описывающие алгоритмы проведения статистического анализа; также доступны различные программные продукты. Несмотря на это, статистические методы часто используют неправильно, что приводит к ошибкам и неадекватным выводам. В настоящей статье мы попытались разобраться в основных сложностях, возникающих у исследователя, работающего в области экологической микробиологии, при попытке статистически проанализировать результаты экспериментов. Например, классические параметрические тесты (t -критерий Стьюдента, дисперсионный анализ, коэффициент корреляции Пирсона и т.д.), наиболее часто используемые в современных экспериментальных статьях, применимы только при соблюдении условия нормального распределения случайной величины n независимых измерений. Важно то, что нормальность распределения необходимо доказывать, а доказать его можно лишь при минимальной выборке 20–30 независимых измерений, последнее является весьма критичным для инкубационных, изотопных и молекулярно-биологических экспериментов. При этом семейство нормальных распределений не единственное семейство распределений числовых случайных величин, зависящих от параметров. Более того, реальные распределения обычно отличаются от нормальных, а сами нормальные распределения могут рассматриваться лишь как некоторая аппроксимация. Большое разнообразие параметрических семейств значительно усложняет задачу выбора критериев оценки и статистических тестов для анализа случайной выборки n независимых измерений. Здесь может помочь использование методов непараметрической статистики, которая не зависит от объема выборки и типа распределения, что является важным преимуществом. Однако, как и любой метод анализа, непараметрическая статистика имеет ограничения. Тем не менее, все больше экспериментальных данных, в том числе и в области экологической микробиологии, анализируется в последние годы с помощью непараметрической статистики, что указывает на некую тенденцию вытеснения параметрических методов непараметрическими.

Ключевые слова: статистический анализ, параметрические методы, нормальное распределение, выборка, непараметрические методы, экологическая микробиология, инкубационные эксперименты

DOI: 10.1134/S0026365619020058

Каждый исследователь рано или поздно сталкивается с требованием редактора или рецензента научного журнала провести статистический анализ полученных данных (если этот анализ не был проведен до подачи статьи в журнал). Такое требование вполне обосновано, поскольку статистический анализ играет важную роль в интер-

претации результатов научных экспериментов. Однако неправильное использование статистических методов приводит к ошибкам и неадекватным выводам, которые снижают значимость результатов и подрывают авторитет авторов.

Одной из наиболее распространенных статистических ошибок является применение пара-

метрических¹ тестов к данным, статистический закон распределения которых неизвестен (Schodder et al., 2006; Erceg-Hurn, Mirosevich, 2008; Nahm, 2016). Причина этой ошибки состоит в том, что многие исследователи, в частности микробиологи, непосредственно ставящие эксперименты и обрабатывающие результаты, зачастую недостаточно хорошо знакомы с теоретическими основами алгоритмов расчетов, не вполне понимают пределы применимости теорем, лежащих в основе этих алгоритмов, и не следят за новыми публикациями в области прикладной статистики. Такие исследователи обычно используют то, что имеется в классических учебниках и широко распространенных программных продуктах (Орлов, 2004), а также наиболее часто встречается в статьях по их конкретной тематике. Так, в современных научных статьях, в том числе и в области экологической микробиологии, наиболее часто встречаются такие параметрические тесты, как *t*-критерий Стьюдента (*t*-test), дисперсионный анализ (analysis of variance, ANOVA), метод наименьших квадратов (ordinary least squares regression) и коэффициент корреляции Пирсона (ρ) (Erceg-Hurn, Mirosevich, 2008). Для их использования должны быть выполнены два условия применимости: результаты наблюдений должны иметь нормальные распределения и дисперсии результатов наблюдений в выборках² должны совпадать (Орлов, 2004). Использование параметрических тестов при несоблюдении условий их применимости приводит к неправильному вычислению *p*-значений (уровня значимости), что может привести к

ошибке первого рода – отклонению нулевой гипотезы, в то время, как в действительности эта гипотеза верна (делается вывод о том, что эффект существует, когда его нет). При этом способность обнаружить подлинные эффекты, напротив, значительно снижается. Неверно рассчитанные *p*-значения в сочетании с неточной оценкой величины эффекта (effect size) и доверительных интервалов могут привести к существенным ошибкам в интерпретации данных (Erceg-Hurn, Mirosevich, 2008).

Целью настоящей статьи является обсуждение основных (на наш взгляд) сложностей, возникающих у исследователя, работающего в области экологической микробиологии, при попытке статистически проанализировать результаты своих экспериментов.

Нормальное распределение

Рассмотрим условия применимости перечисленных выше классических параметрических тестов – равенство дисперсий и нормальность распределения. Условие равенства дисперсий нельзя считать выполненным в большинстве исследовательских и практических задач (Орлов, 2004). В любом случае, чтобы быть уверенным в том, что оно выполнено (или не выполнено), его необходимо проверить, что почти никогда не делается в микробиологических работах.

Нормальное (гауссовское) распределение с параметрами сдвига μ и масштаба σ – это такое распределение случайной величины, для которого плотность вероятности составляет:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \quad (1)$$

Соответственно, чтобы найти вероятность того, что случайная величина окажется в интервале между *a* и *b*, необходимо вычислить интеграл от *f*(*x*) в границах от *a* до *b* (Теннант-Смит, 1988). Нормальное распределение обычно обозначается $N(\mu, \sigma)$ или $N(\mu, \sigma^2)$, где σ^2 – дисперсия случайной величины.

Центральная предельная теорема утверждает, что в случае, когда результат измерения складывается под действием многих причин, причем каждая из них вносит лишь малый вклад, а совокупный итог определяется аддитивно (т.е. путем сложения), то распределение результата измерения близко к нормальному. Если причины действуют не аддитивно, а мультипликативно (т.е. действия отдельных причин перемножаются), то распределение случайной величины близко не к нормальному, а к логарифмически нормальному, т.е. не *x*, а $\lg(x)$ имеет приблизительно нормальное распределение (1). Если же нет оснований считать, что действует один из этих двух механиз-

¹ Параметрическая статистика изучает выборки из распределений, описываемых аналитическими формулами с небольшим числом параметров (такие распределения называются параметрическими). Распределение (вероятностей) – функция, которая определяет вероятность того, что случайная величина принимает заданное значение или принадлежит к некоторому заданному интервалу. В параметрических задачах оценивания принимают вероятностную модель, согласно которой результаты наблюдений x_1, x_2, \dots, x_n рассматривают как реализации *n* независимых случайных величин с функцией распределения $F(x; \theta)$. Здесь θ – неизвестный вектор-параметр (фиксированной конечной размерности, в том числе и размерности 1 – тогда θ представляет собой единственное число), лежащий в пространстве параметров Θ , заданном используемой вероятностной моделью. Задача оценивания состоит в определении точечной оценки и доверительных границ (либо доверительной области) для компонентов вектор-параметра θ . Непараметрические статистические методы не опираются на предположения о принадлежности функций распределения результатов измерений к тому или иному параметрическому семейству, т.е. они не требуют, чтобы функция распределения была задана в виде какой-то конкретной аналитической формулы (Орлов, 2004).

² Выборка – совокупность независимых одинаково распределенных случайных элементов, т.е. набор наблюдаемых значений (x_1, x_2, \dots, x_n) или множество объектов, отобранных из изучаемой совокупности, где *n* (sample size) – объем выборки – число наблюдаемых значений, составляющих выборку (Орлов, 2004).

мов формирования итогового результата (или другой вполне определенный механизм), то закон распределения нельзя сказать ничего определенного. Таким образом, в конкретной прикладной задаче нормальность результатов измерений нельзя установить из общих соображений, ее следует проверять с помощью, например, статистических критериев (Орлов, 2004).

К настоящему времени известно довольно много критериев для оценки того, имеет ли случайная выборка, состоящая из n независимых измерений, нормальное распределение. Можно упомянуть графические методы (построение гистограмм, графиков-квантилей), критерии асимметрии и эксцесса, а также критерий Шапиро–Уилка и др. Обзор эффективности различных критериев для проверки нормальности распределения результатов измерений приведен в статье Razali, Wah (2011), см. также фундаментальный справочник А.И. Кобзаря (2006) и учебник Corder, Foreman (2009).

Таким образом, задача исследователя в общем виде сводится к обоснованному выбору метода статистического анализа. Очевидно, что начинать нужно с доказательства нормальности распределения случайной выборки n независимых измерений в каждом конкретном эксперименте. Далее, если данные распределены нормально, использовать классические методы параметрической статистики (например, t -тест, ANOVA и др.). Если нормальность распределения не подтверждается, использовать другие методы. Важно отметить, что, если закон распределения вообще не известен, следует использовать “свободные от распределения” методы непараметрической статистики.

Какие же сложности возникают у исследователя, работающего в области экологической микробиологии, при попытке статистически проанализировать результаты экспериментов?

Во-первых, если результаты наблюдений не имеют нормальные распределения, это еще не значит, что параметрические методы не могут быть использованы для их анализа. Семейство нормальных распределений является наиболее популярным, но не единственным семейством распределений числовых случайных величин, зависящих от параметров. С помощью нормального распределения определяются распределение χ^2 (хи-квадрат), распределение Стюдента и распределение Фишера. Также известны параметрические семейства гамма-распределений (нормальное распределение — предельный случай гамма-распределения), экспоненциальных (экспоненциальные распределения — частный случай распределений Вейбулла–Гнеденко), логнормальных, бета-распределений, распределений Коши, вырожденных распределений, распределений Бернулли, распределений Пуассона, полиномиальных распре-

делений и др. (Орлов, 2004; Боровков, 2010). В прикладных исследованиях также используют такие параметрические семейства распределений, как система кривых Пирсона, ряды Эджворта и Шарлье (Орлов, 2004). Таким образом, разнообразие параметрических семейств распределений числовых случайных величин значительно усложняет задачу выбора критериев оценки и статистических тестов для анализа случайной выборки n независимых измерений.

Во-вторых, математики уже довольно давно говорят о том, что реальные распределения независимых измерений практически всегда отличаются от включенных в классические параметрические семейства и уж, по крайней мере, в большинстве случаев отличаются от нормальных (Орлов, 2004). Это проявляется в том, что среди наблюдений, если их несколько десятков (или сотен), находится несколько таких, которые нельзя получить из гауссовского закона — слишком далеко отстоят они от центра распределения (и центра выборки). Присутствие таких “выбросов”, если их не исключить, может сильно исказить результаты статистической обработки (Тюрин, Шмерлинг, 2004). Здесь важно четко понимать, что исключать “выбросы” из выборки можно только в том случае, если они являются результатом грубых ошибок. Для этого пользуются специально разработанными критериями, а оставшуюся выборку (без “выбросов”), если она имеет нормальное распределение, подвергают далее параметрической статистической обработке, опираясь на нормальный закон. Общепринятая процедура исключения грубых ошибок описана, например, в книге Л.З. Румшицкого (1971). Она состоит в том, что, используя эмпирическую оценку (s) среднеквадратической ошибки, а также “выскакивающее” (x^*) и среднее по выборке (m) значения, рассчитывают величину $t = |x^* - m|/s$, которую сравнивают с критическими значениями при заданной надежности вывода; для критических значений существуют подробные таблицы — см., например, табл. III (Румшицкий, 1971). Если t оказывается достаточно велико, то x^* признается грубой ошибкой и удаляется из выборки (т.е. дальнейший статистический анализ ведется уже без x^*). Но поскольку весь анализ на наличие выброса велся в предположении справедливости нормального распределения, то x^* можно отбросить только в том случае, если выборка действительно удовлетворяет этому распределению.

Если же “выбросы” — это реальные измерения, характеризующиеся совершенно другим законом распределения, отличным от нормального (например, законом с “тяжелыми хвостами”, таким, как степенное распределение), то отбрасывать их нельзя, а статистическую обработку полной выборки (с “выбросами”) необходимо проводить в соответствии с этим иным распределением.

Проиллюстрируем сказанное на конкретном примере: мы измеряем интенсивность поглощения метана образцами почвы, и количество образцов (повторностей) в выборке составляет 20. Поясним, что удельный поток метана (j , мкмоль $\text{м}^{-2} \text{ч}^{-1}$) линейно зависит от коэффициента диффузии в почве (D , $\text{см}^2 \text{с}^{-1}$) (Born et al., 1990; Dörr et al., 1993; Глаголев, Филиппов, 2011). Максимальная скорость поглощения автотрофной почвой метана из атмосферы (т.е. окисление атмосферного метана аэробными метанотрофными бактериями) лимитируется именно диффузией (Striegl, 1993; Potter et al., 1996). При этом коэффициент диффузии газа в почве определяется пористостью аэрации почвы (P_a), согласно зависимости (Глаголев, Смагин, 2006) с учетом погрешности коэффициента диффузии (Глаголев и соавт., 2017):

$$D(P_a) = 0.0046P_a^3 + 0.047P_a^2 - 0.0878P_a + 0.19. \quad (2)$$

Допустим, что некий параметр (например, пористость аэрации) варьирует в наших образцах в соответствии с нормальным законом распределения, как это обычно принимается для случайных погрешностей. Допустим, что для пористости аэрации $\mu = 10\%$ и $\sigma = 5\%$. При помощи MATLAB-функции `normrnd` мы сгенерировали 20 случайных чисел, соответствующих пористости аэрации отдельных образцов почвы и распределенных в соответствии с $N(10, 5)$: 7.837, 1.672, 10.627, 11.438, 4.268, 15.955, 15.946, 9.812, 11.637, 10.873, 9.067, 13.629, 7.058, 9.318, 10.57, 15.334, 10.296, 9.522, 5.8383, 20.916. Проверка, выполненная при помощи MATLAB-функции `lillietest` (Lilliefors test), показала: статистическая гипотеза о том, что вышеприведенная выборка удовлетворяет нормальному распределению, не может быть отвергнута (с надежностью не менее 95%). Для приведенной выборки получим следующие значения D по формуле (2): 4.6027, 0.1961, 10.0855, 12.2181, 1.029, 29.4366, 29.3923, 8.1989, 12.782, 10.7048, 6.6867, 19.3689, 3.529, 7.1742, 9.9453, 26.4802, 9.2891, 7.5868, 2.1948, 61.0065. Поскольку интенсивность поглощения метана пропорциональна коэффициенту диффузии, то приведенные значения могут рассматриваться как скорость потребления метана образцами почвы (в условных единицах), т.е. те значения, которые получит исследователь в эксперименте (“выборка D”). Если проверить “выборку D” с помощью MATLAB-функции `lillietest`, то статистическая гипотеза о том, что эта выборка удовлетворяет нормальному распределению должна быть отвергнута (с надежностью не менее 95%), т.е. тест покажет, что выборка не удовлетворяет нормальному распределению. Но исследователь не знает, что так и должно быть, поскольку ему неизвестно, что вариабельность элементов вы-

борки вызвана различной пористостью аэрации образцов. Поэтому, мысль исследователя устремится в тривиальном направлении: нет ли среди измерений грубой ошибки? Если провести анализ “выборки D” на наличие “выброса”, то легко можно установить, что значение 61.0065, намного превышающее любое другое значение в выборке, якобы является “выбросом” (причем надежность вывода в данном случае составляет не меньше 99%). Исходя из этого, исследователь, отбросит данное подозрительное значение и подвергнет исправленную выборку (содержащую теперь лишь 19 значений, т.е. все значения “выборки D”, кроме значения 61.0065) тесту на нормальность, убедившись при этом, что теперь выборка удовлетворяет нормальному распределению (с надежность данного вывода не менее 95%).

Итак, было выполнено 20 измерений потребления метана. Получившаяся выборка оказалась несколько подозрительной: во-первых, она не соответствовала нормальному распределению и, во-вторых, одно из измерений столь сильно отличалось от остальных, что заставило предположить наличие грубой ошибки. Действительно, общепринятый анализ для выявления грубых ошибок показал, что это измерение является ошибкой, и удаление данного измерения из выборки привело к тому, что она стала удовлетворять нормальному распределению. Следовательно, для дальнейшего анализа нужно использовать эту редуцированную выборку. Для какого дальнейшего анализа? Это зависит от целей исследователя. Если, например, он хочет посчитать среднюю интенсивность потребления метана данной почвой, то сможет вычислить среднее арифметическое, которое в данном случае окажется равным чуть больше 11.

На самом деле выборка действительно не соответствует нормальному распределению, но она и не должна ему соответствовать, поскольку величина пористости аэрации, распределенная по нормальному закону, была преобразована по нелинейной формуле (2), в результате чего тип распределения, разумеется, изменился. А раз так, то ничего подозрительного в значении 61.0065 нет. И не следует в данном случае проводить тест на “выброс” и отбрасывать значение 61.0065, исходя из предположения о нормальности распределения, поскольку на самом деле распределение отличается от нормального. А если значение 61.0065 не отбрасывать, то среднее арифметическое по выборке получится равным ~ 14 , а вовсе не 11 (погрешность около 20%!).

Приведенный пример не является умозрительным или слишком частным. В области потребления и эмиссии метана почвами специально выполнены работы с целью идентификации законов распределений, и к настоящему времени ясно, что нормальное распределение является скорее

исключением, чем правилом (Паников, 1995; Веретенникова, Глаголев и соавт., 2007, 2012; Glagolev et al., 2011; Дюкарев, 2014 и др.). Подавляющее большинство измерений эмиссии метана из различных болот Западной Сибири весьма сильно отличаются от нормального распределения, а многие — вообще от какого-либо из числа наиболее популярных в математической статистике распределений. В связи с этим, в созданной для Западной Сибири “стандартной модели” эмиссии не делалось попыток подогнать данные распределения к известным, а просто использовались (в виде сплайн-аппроксимаций) эмпирические распределения, построенные по реальным данным полевых измерений (Глаголев и соавт., 2012).

Если говорить не о болотах, а о других природных объектах, в которых идут процессы метаногенеза и метанотрофии, то и там наблюдались распределения эмиссии, отличающиеся от нормального. Сабреков и соавт. (Sabrekov et al., 2017) обнаружили лог-нормальное распределение эмиссии метана из озер средней тайги Западной Сибири, а для озер южной тайги — степенное распределение (это — распределение с очень “тяжелым хвостом”). К сожалению, большинство авторов не проводили аккуратного статистического исследования типа распределений, но опубликованные ими графики плотностей распределения достаточно красноречивы. В частности, распределения, похожие на лог-нормальные, обнаружены при исследовании эмиссии метана из “грязевых микровулканов” в пойме р. Мухриной (в окрестностях г. Ханты-Мансийск) (Белова и соавт., 2013; Oshkin et al., 2014). Аналогичные распределения (а также распределения с еще более “тяжелыми хвостами”, чем у лог-нормальных, и мультимодальные распределения) обнаружены при измерении эмиссии метана в поймах рек Западной Сибири (Terentieva et al., 2018). Вышеприведенный эмпирический материал вполне соответствует современным теоретическим представлениям. Существующие математические модели, например, изложенные в статьях (Сабреков и соавт., 2014; Зинченко, 2017; Cao et al., 1996; Bohn et al., 2013; Xu et al., 2016 и др.), отражают нелинейные связи между эмиссией метана и факторами внешней среды, такими как температура, осадки, уровень болотных вод. Предположим, что какой-то определенный параметр внешней среды имеет нормальное распределение — часто это действительно так, см., например, метеорологические параметры для заповедника “Малая Сосьва” за несколько десятилетий (Таланова, 2018). Если мы при этом воздействуем нелинейным оператором на такое распределение, то оно утратит свою нормальность.

Может показаться, что, если нормальных и других параметрических распределений в природе почти не существует, тогда вообще непонятен смысл использования параметрических методов

для статистического анализа. Однако мы хотели бы предостеречь читателя от столь радикального взгляда. Нормальные распределения могут рассматриваться как некоторая аппроксимация, хотя никогда нет полного совпадения с реальностью (Орлов, 2004). Но если статистические тесты показывают, что на достаточно высоком уровне значимости наше распределение не отличается от нормального (или какого-то другого известного распределения), то на этом уровне значимости мы можем делать содержательные выводы из статистического анализа выборки, проведенного на основе предположения о ее соответствии данному распределению.

К сожалению, более важным (и более пессимистичным) является следующий вопрос: как относиться к результатам огромного количества уже опубликованных исследований, в которых выводы сделаны на основании статистического анализа с использованием классических параметрических тестов? Ведь в подавляющем большинстве случаев нормальность выборки не проверялась. Общий ответ здесь очевиден: с сугубой осторожностью! Конечно, ставить крест на всех этих результатах не хотелось бы, но и доверять им нет никаких оснований. Однако если есть возможность обратиться к исходным материалам (даже если они не были опубликованы, подчас, оказывается возможным получить их, связавшись с авторами работы), то следует проверить их соответствие тому параметрическому закону, на основании которого делались выводы в публикации.

Выборка

И наконец, остается, наверное, самый важный вопрос для микробиолога, занимающегося постановкой инкубационных экспериментов: каким конкретно должно быть минимальное значение *n* независимых измерений для того, чтобы корректно провести статистический анализ, например, доказать нормальность распределения? В литературе часто встречается термин “большая” и “малая” выборка (big and small sample size). “Выборка считается “большой”, если точность статистического правила оказывается достаточной для поставленных целей исследования. Понятие “большая выборка” не вполне четкое и зависит от целей исследования и выбранных математических средств” (Тюрин, Шмерлинг, 2004). “Для того чтобы гарантировать, что функция распределения результатов наблюдений отличается от некоторой нормальной не более чем на 0.01 (при любом значении аргумента), требуется порядка 2500 наблюдений” (Орлов, 2004). Однако на современном этапе развития микробиологии (да и биологии вообще, если не включать в нее медицинскую биологию) требование 99% надежности, по-видимому, следует признать избыточным. По

крайней мере, так обстоит дело в наше время — часто в биологических статьях ограничиваются уровнем надежности 95%, а в статистических справочниках (см., например, Румшицкий, 1971; Дмитриев, 1995; Кобзарь, 2006; Бостанджиян, 2013) таблицы различных тестов и критериев часто начинаются вообще с надежности 90% или даже еще меньших. Поэтому в учебниках встречаются не столь строгие минимальные значения n для статистического анализа данных с использованием параметрических тестов — 20–30 независимых измерений на группу (Pett, 1997; Warner, 2012), однако при этом, по-видимому, не ставится требование отличия на 0.01.

Объем выборки $n \geq 20$, широко используемый в клинических и физических исследованиях, встречается и в работах по экологической микробиологии. Допустим, что объектом исследований является микробное сообщество воды и осадков озера. В водоемах активность и состав микробных сообществ варьируют по глубине (в пределах 1 м для водной толщи и 1 см для осадков), а также различается в зависимости от точки отбора образца (прибрежная зона или середина озера, мелководный или глубоководный участок и т.п.). Рутинно определяемые с помощью датчиков физико-химические параметры (например, удельная электропроводность, мутность, соленость, температура, рН, концентрация растворенного кислорода и др.), а также потоки метана могут быть измерены в $n \geq 20$ независимых точках на одной и той же глубине по всей площади озера и в дальнейшем экстраполированы на озеро в целом. Тогда можно рассчитать стандартное отклонение, проверить нормальность распределения и далее использовать параметрические или непараметрические тесты для сравнения данного конкретного озера с другими озерами по тем же параметрам и глубине или для изучения временной динамики исследуемого параметра (месячная, сезонная, годовая динамика).

Если речь идет об изменении параметра с глубиной, то для корректного проведения статистического анализа необходимо сделать ≥ 20 независимых измерений (“повторностей”) каждого исследуемого параметра на одной конкретной глубине и сравнить с таким же количеством повторностей на другой, и так далее, в зависимости от количества исследуемых экспериментальных точек по профилю. Такого “большого” числа повторностей не найдешь ни в одной работе, описывающей инкубационные (влияние разных субстратов, акцепторов электронов на потенциальную активность и состав сообщества), изотопные (определение скоростей микробных процессов) или молекулярно-биологические (определение видового состава микробного сообщества) исследования. Существует стандартное требование: инкубационные эксперименты должны проводиться как минимум

в трех повторностях. В реальных лабораторных условиях $n = 3$ зачастую является также и максимумом. Такая “малая” выборка адекватно воспринимается людьми, лично знакомыми со спецификой постановки инкубационных (изотопных, молекулярно-биологических) экспериментов с участием природных микробных сообществ. Для неспециалистов приведем простой пример: требуется сравнить потенциальную метаногенную активность озерных осадков, отобранных по профилю с трех глубин (0–5, 5–10 и 10–15 см) в одной экспериментальной точке. Для этого необходимо провести лабораторную инкубацию образца осадков с каждой исследуемой глубины в строго анаэробных, близких к *in situ* условиях с использованием наиболее распространенных субстратов метаногенеза. Количество инкубационных вариантов для образца осадка с одной глубины составляет 13 (12 субстратов и 1 контроль). Эксперименты для каждого субстрата и контроля ставятся в 3-х повторностях ($n = 3$), следовательно, необходимо поставить 39 анаэробных инкубаций (флаконов) для образца осадка с одной глубины. Таким образом, для сравнения 3-х глубин требуется 117 флаконов. Каково будет количество флаконов для $n = 20$, чтобы корректно проверить имеет ли случайная выборка n независимых измерений нормальное распределение? Следует также отметить, что параллельно ставится еще примерно столько же инкубаций для определения других процессов, например, аэробного или анаэробного окисления метана.

В большинстве изученных нами статей по близкой тематике авторы проводили инкубационные эксперименты с $n = 2–6$. Когда выборка слишком мала, распределение неизвестно, даже если вы действительно хотите доказать его нормальность или если авторы подобных работ уверенно пишут об использовании параметрических тестов для анализа своих данных. Мы не будем приводить ссылки на конкретные работы, поскольку критика результатов чужих экспериментов не входит в задачи данной статьи. Мы хотим подчеркнуть, что специфика инкубационных экспериментов такова, что зачастую довольно проблематично поставить такое количество повторностей, которое требуется для корректного статистического анализа с использованием классических параметрических методов.

Что же делать, если выборка слишком мала, а требования журнала не позволяют обойти статистический анализ? Можно предложить два способа решения данной проблемы. Во-первых, поставить один инкубационный вариант (например, взять один субстрат из списка) в $n \geq 20$ повторностях и доказать для этого варианта, что распределение нормальное (или какое-то иное). Для других вариантов с $n = 3$ допустить, что распределение аналогично испытанному с $n = 20$. Данный способ,

Таблица 1. Статистический анализ данных с использованием непараметрических vs. параметрических тестов (Corder, Foreman, 2009)

Тип анализа	Непараметрический тест	Параметрическая альтернатива
Сравнение двух зависимых выборок	Критерий Вилкоксона для зависимых выборок (или критерий знаковых рангов Вилкоксона)	<i>t</i> -Критерий для зависимых выборок
Сравнение двух независимых выборок	Критерий Манна–Уитни–Уилкоксона (U-критерий Манна–Уитни, критерий суммы рангов Уилкоксона)	<i>t</i> -Критерий для независимых выборок
Сравнение трех и более зависимых выборок	Критерий Фридмана	Дисперсионный анализ с повторными измерениями
Сравнение трех и более независимых выборок	Критерий Краскела–Уоллиса	Однофакторный дисперсионный анализ
Сравнение нечисловых (в номинальной шкале) данных	Критерий Хи-квадрат, точный тест Фишера	Отсутствует
Сравнение двух упорядоченных (ранжированных) переменных	Ранговая корреляция Спирмена	Корреляция Пирсона
Сравнение двух выборок, когда одна переменная образует шкалу с двумя дискретными градациями признака	Точечная бисериальная корреляция	Корреляция Пирсона
Сравнение двух выборок, когда одна переменная образует шкалу с двумя непрерывными градациями признака	Бисериальная корреляция	Корреляция Пирсона
Проверка случайности выборки	Критерий Вальда–Вольфовица (критерий серий)	Отсутствует

но с меньшим объемом выборки, уже описан в публикациях. Стоит, однако, отметить, что, если результаты независимых измерений, полученные для одного варианта, имеют нормальное (или другое) распределение, это не обязательно означает, что результаты независимых измерений, полученные для остальных вариантов, будут иметь такое же распределение. Во-вторых, можно использовать непараметрические тесты, позволяющие проанализировать “малую” выборку независимых измерений, распределение которой неизвестно.

Непараметрические методы

Принципы непараметрической (независящей от параметров) статистики подробно изложены в ряде учебников и статей (например, Pett, 1997; Тюрин, Шмерлинг, 2004 (и ссылки там же); Egseg-Hurn, Mirosevich, 2008; Corder, Foreman, 2009; Nahm, 2016). Непараметрическая статистика позволяет выполнить те же задачи, что и параметрическая, т.е. провести анализ случайной выборки *n* независимых измерений (табл. 1). При этом непараметрическая статистика не зависит от объема

выборки и типа распределения, что является важным преимуществом. Критерии непараметрической статистики основаны не на самих наблюдениях, а, чаще всего, на их рангах, т.е. на номерах, которые присваиваются наблюдениям при их расстановке в определенном порядке, обычно в порядке возрастания. Все возможные последовательности рангов имеют равные вероятности для появления, поэтому все статистики, основанные на рангах, свободны от распределения. Для ранговых методов, в отличие от параметрических, наличие небольшого числа пусть даже очень больших “выбросов” влияет на результат незначительно, поэтому они более устойчивы к выбросам, засорениям и прочим несовершенствам статистического материала. Помимо ранговых, созданы и другие непараметрические системы обработки, например, основанные на знаках наблюдений или остатков (Тюрин, Шмерлинг, 2004).

Следует отметить, что так же, как и любой метод анализа, непараметрическая статистика имеет ряд ограничений. Например, переход от точных значений исходных наблюдений к рангам сопровождается некоторой потерей информации (Тю-

рин, Шмерлинг, 2004). Кроме того, непараметрические методы дают ограниченную (по сравнению с параметрическими методами) информацию, и ее сложнее интерпретировать, при этом для “большой” выборки усложняются расчеты (Nahm, 2016). Использование методов непараметрического анализа снижает риск ошибочных выводов, поскольку эти методы не делают никаких предположений о распределении, но при этом могут иметь более низкую статистическую мощность. Иначе говоря, непараметрические методы “всегда применимы, но не всегда эффективны”, в то время как параметрические методы “не всегда применимы, но эффективны всегда, когда применимы” (Nahm, 2016). Следует отметить, что “малый” объем выборок также является одним из ключевых ограничений для применения непараметрических критериев. Например, при сравнении медиан двух выборок ранговые непараметрические методы сортируют значения переменной в каждой из выборок и сравнивают их друг с другом. При этом величина различий между значениями в выборках никак не учитывается. Для “малых” выборок это приводит к тому, что минимальный достижимый уровень значимости при проверке гипотезы о равенстве медианы двух выборок оказывается довольно высоким, часто выше традиционного значения 0.05, которое принимается за границу, разделяющую совпадающие и отличающиеся медианы выборок. Например, если мы сравниваем с помощью двустороннего критерия Манна–Уитни–Уилкоксона медианы двух выборок объемом 3 каждая, то различия между ними, какими бы они ни были, не могут быть установлены с уровнем значимости менее 0.10 (Shieh et al., 2006). В случае одностороннего теста наименьший уровень значимости для этого примера будет 0.05. Использование *t*-критерия для сравнения (с предварительным преобразованием, например, логарифмическим, если данные распределены не в соответствии с нормальным распределением) может быть более предпочтительным в данном случае, поскольку в рамках его применения минимальный достижимый уровень значимости отсутствует.

В последние годы, однако, все больше экспериментальных данных, в том числе и в области экологической микробиологии, анализируется с помощью методов непараметрической статистики, что указывает на некую тенденцию вытеснения параметрических методов непараметрическими. Возможно, что именно непараметрическая статистика является единственным адекватным способом анализа результатов инкубационных экспериментов. В заключение следует отметить, что в лабораторных физиолого-биохимических экспериментах к перечисленным сложностям добавляется еще и изменчивость тест-объектов, в результате чего могут формироваться неоднород-

ные выборки, к которым статистические методы в принципе неприменимы. В этих случаях решающую роль в обеспечении достоверности различий играет стандартизация процессов хранения культур и приготовления посевного материала.

Работа выполнена при финансовой поддержке Российского научного фонда, грант № 16-14-10201.

СПИСОК ЛИТЕРАТУРЫ

- Белова С.Э., Ошкин И.Ю., Глаголев М.В., Лапшина Е.Д., Максюттов Ш.Ш., Дедыш С.Н. Метанотрофные бактерии грязевых микровулканов в поймах северных рек // Микробиология. 2013. Т. 82. № 6. С. 732–740.
- Belova S.E., Oshkin I.Yu., Glagolev M.V., Lapshina E.D., Maksyutov Sh.Sh., Dedys S.N. Methanotrophic bacteria in cold seeps of the floodplains of northern rivers // Microbiology (Moscow). 2013. V. 82. P. 743–750.
- Боровков А.А. Математическая статистика: Учебник. 4-е изд., стер. СПб.: Издательство “Лань”, 2009. 704 с.
- Бостанджиян В.А. Пособие по статистическим распределениям. Черноголовка: Ред.-издат. отд. ИПХФ РАН, 2013. 1060 с.
- Веретенникова Е.Э., Дюкарев Е.А. Эмиссия метана торфяными залежами олиготрофных болот южно-таежной подзоны Западной Сибири // Торфяники Западной Сибири и цикл углерода: прошлое и настоящее: Материалы Четвертого международного полевого симпозиума (Новосибирск, 4–17 августа 2014 г.) / Под ред. Титляновой А.А. и Дергачевой М.И. Томск: Изд-во Том. ун-та, 2014. С. 157–159.
- Глаголев М.В., Смагин А.В. Количественная оценка эмиссии метана болотами: от почвенного профиля — до региона (к 15-летию исследований в Томской области) // Доклады по экологическому почвоведению. 2006. Т. 3. № 3. С. 75–114.
- Глаголев М.В., Головацкая Е.А., Шнырев Н.А. Эмиссия парниковых газов на территории Западной Сибири // Сибирский экологический журнал. 2007. Т. 14. С. 197–210.
- Glagolev M.V., Golovatskaya E.A., Shnyrev N.A. Greenhouse gas emission in West Siberia // Contemporary Problems of Ecology. 2008. V. 1. P. 136–146.
- Глаголев М.В., Филиппов И.В. Инвентаризации поглощения метана почвами // Динамика окружающей среды и глобальные изменения климата. 2011. Т. 2. № 2 (4). С. 1.
- Glagolev M.V., Filippov I.V. Inventory of soil methane consumption // Environmental Dynamics and Global Climate Change. 2011. V. 2. № 2(4). EDCCrev0002.
- Глаголев М.В., Сабреков А.Ф., Клептова И.Е., Филиппов И.В., Лапшина Е.Д., Мачида Т., Максюттов Ш.Ш. Эмиссия метана из болот подтайги Западной Сибири (к развитию “стандартной модели”) // Почвоведение. 2012. № 10. С. 1077–1088.
- Glagolev M.V., Sabrekov A.F., Kleptova I.E., Filippov I.V., Lapshina E.D., Machida T., Maksyutov Sh.Sh. Methane emission from bogs in the subtaiga of Western Siberia: the development of standard model // Euras. Soil Sci. 2012. V. 45. P. 947–957.

- Глаголев М.В., Сабреков А.Ф., Терентьева И.Е.* Ответ А.В. Смагину: IV. Поверхностная диффузия или случайный шум? // Динамика окружающей среды и глобальные изменения климата. 2017. Т. 8. № 1. С. 55–65.
- Glagolev M.V., Sabrekov A.F., Terentjeva I.E.* Reply to A.V. Smagin: IV. Surface diffusion or random noise? // Environmental Dynamics and Global Climate Change. 2017. V. 8. № 1. P. 55–65.
- Дмитриев Е.А.* Математическая статистика в почвоведении. М.: Изд-во МГУ, 1995. 320 с.
- Зинченко А.В.* Модель гумификации и минерализации органических веществ в почве и ее использование для расчета составляющих углеродного баланса болотных экосистем // Динамика окружающей среды и глобальные изменения климата. 2017. Т. 8. № 2. С. 3–17.
- Zinchenko A.V.* Model of soil organic matter humification and mineralization and its application for calculation of peatland ecosystems carbon budget characteristics // Environmental Dynamics and Global Climate Change. 2017. V. 8. № 2. P. 3–17.
- Кобзарь А.И.* Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.
- Орлов А.И.* Прикладная статистика. Учебник. М.: Издательство “Экзамен”, 2004. 656 с.
- Паников Н.С.* Таежные болота – глобальный источник атмосферного метана? // Природа. 1995. № 6. С. 14–25.
- Румицкий Л.З.* Математическая обработка результатов эксперимента. М.: Наука, 1971. 192 с.
- Сабреков А.Ф., Глаголев М.В., Клепцова И.Е., Мачида Т., Максюттов Ш.Ш.* Эмиссия метана из болотных комплексов тайги Западной Сибири // Почвоведение. 2014. № 1. С. 58–70.
- Sabrekov A.F., Glagolev M.V., Kleptsova I.E., Machida T., Maksyutov S.S.* Methane emission from mires of the West Siberian taiga // Euras. Soil Sci. 2013. V. 46. № 12. P. 1182–1193.
- Таланова Г.И.* Климат заповедника “Малая Сосьва”: многолетние данные // Динамика окружающей среды и глобальные изменения климата. 2018. Т. 9. № 1. С. 22–45.
- Talanova G.I.* Climate of the reserve “Malaya Sosva”: long-term material // Environmental Dynamics and Global Climate Change. 2018. V. 9. № 1. P. 22–45.
- Теннант-Смит Дж.* Бейсик для статистиков. М.: Мир, 1988. 208 с.
- Тюрин Ю.Н., Шмерлинг Д.С.* Непараметрические методы статистики // Социология. 4М. 2004. № 18. С. 154–166.
- Bohn T.J., Podest E., Schroeder R., Pinto N., McDonald K.C., Glagolev M., Filippov I., Maksyutov S., Heimann M., Chen X., Lettenmaier D.P.* Modelling the large-scale effects of surface moisture heterogeneity on wetland carbon fluxes in the West Siberian Lowland // Biogeosci. 2013. V. 10. P. 6559–6576.
- Born M., Dörr H., Levin I.* Methane consumption in aerated soils of the temperate zone // Tellus. 1990. V. 42B. P. 2–8.
- Cao M., Marshall S., Gregson K.* Global carbon exchange and methane emissions from natural wetlands: Application of a process-based model // J. Geophys. Res. 1996. V. 101. № D9. P. 14399–14414.
- Corder G.W., Foreman D.I.* Nonparametric statistics for non-statisticians: a step-by-step approach. Hoboken, New Jersey: John Wiley & Sons, Inc., 2009. 247 p.
- Dörr H., Katruff L., Levin I.* Soil texture parameterization of the methane uptake in aerated soils // Chemosphere. 1993. V. 26. P. 697–713.
- Erceg-Hurn D.M., Mirosevich V.M.* Modern robust statistical methods an easy way to maximize the accuracy and power of your research // Am. Psychologist. 2008. V. 63. P. 591–601.
- Glagolev M., Kleptsova I., Filippov I., Maksyutov S., Machida T.* Regional methane emission from West Siberia mire landscapes // Environ. Res. Lett. 2011. V. 6. № 4. 045214.
- Nahm F.S.* Nonparametric statistical tests for the continuous data: the basic concept and the practical use // Korean J. Anesthesiol. 2016. V. 69. P. 8–14.
- Oshkin I.Y., Wegner C.-E., Lüke C., Glagolev M.V., Filippov I.V., Pimenov N.V., Liesack W., Dedysh S.N.* Gammaproteobacterial methanotrophs dominate cold methane seeps in floodplains of West Siberian rivers // Appl. Environ. Microbiol. 2014. V. 80. P. 5944–5954.
- Pett M.A.* Nonparametric statistics in health care research: statistics for small samples and unusual distributions // Thousand Oaks, CA: SAGE Publications, 1997. 307 p.
- Potter C.S., Davidson E.A., Verchot L.V.* Estimation of global biogeochemical controls and seasonality in soil methane consumption // Chemosphere. 1996. V. 32. P. 2219–2246.
- Razali N.M., Wah Y.B.* Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests // J. Statist. Model. Analyt. 2011. V. 2. P. 21–33.
- Sabrekov A.F., Runkle B.R.K., Glagolev M.V., Terentjeva I.E., Stepanenko V.M., Kotsyurbenko O.R., Maksyutov S.S., Pokrovsky O.S.* Variability in methane emissions from West Siberia’s shallow boreal lakes on a regional scale and its environmental controls // Biogeosci. 2017. V. 14. P. 3715–3742.
- Schoder V., Himmelmann A., Wilhelm K.P.* Preliminary testing for normality: some statistical aspects of a common concept // Clin. Exp. Dermatol. 2006. V. 31. P. 757–761.
- Shieh G., Jan S., Randles R.* On power and sample size determinations for the Wilcoxon–Mann–Whitney test // Nonparametric Statistics. 2006. V. 18. P. 33–43.
- Striegl R.G.* Diffusional limits to the consumption of atmospheric methane by soils // Chemosphere. 1993. V. 26. P. 715–720.
- Terentjeva I.E., Sabrekov A.F., Ilysov D., Ebrahimi A., Glagolev M.V., Maksyutov S.* Highly dynamic methane emission from the West Siberian boreal floodplains // Wetlands. 2018. <https://org/10.1007/s13157-018-1088-4>.
- Warner R.M.* Applied statistics: from bivariate through multivariate techniques. 2 edn. Thousand Oaks, CA: SAGE Publications, 2012. 1208 p.
- Xu X., Yuan F., Hanson P.J., Wullschlegel S.D., Thornton P.E., Riley W.J., Song X., Graham D.E., Song C., Tian H.* Reviews and syntheses: Four decades of modeling methane cycling in terrestrial ecosystems // Biogeosci. 2016. V. 13. P. 3735–3755.

On the Application of Statistical Analysis for Interpretation of Experimental Results in Environmental Microbiology

A. Yu. Kallistova^{1,*}, A. F. Sabrekov^{2,3}, V. M. Goncharov⁴, N. V. Pimenov¹, and M. V. Glagolev^{2,3,4,5}

¹*Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, 119071 Russia*

²*Yugra State University, Khanty-Mansiisk, 628011 Russia*

³*Tomsk State University, Tomsk, 634050 Russia*

⁴*Lomonosov Moscow State University, Moscow, 119991 Russia*

⁵*Institute of Forest Science, Russian Academy of Sciences, Uspenskoe, Moscow oblast, 143030 Russia*

**e-mail: kallistoanna@mail.ru*

Received November 9, 2018

Revised November 22, 2018

Accepted November 30, 2018

Statistical analysis is an integral part of every experiment, since it helps researchers to make conclusions out of the work done. Textbooks with detailed description of the algorithms for statistical analysis exist, as well as diverse software packages. Statistical methods are, however, often applied incorrectly, which leads to erroneous and inadequate conclusions. The present article is an attempt to determine the main problems emerging in the course of statistical analysis of experimental results in the field of environmental microbiology. For instance, the classical parametric tests, which are most often used in experimental articles (*t*-test, analysis of variance, Pearson correlation coefficient, etc.), are applicable only when a random variable of *n* independent observations has normal distribution. Importantly, the normality of the distribution must be proved, and it is possible to do only for minimal sample size of 20–30 independent observations per group. The latter is crucial for the incubation, isotope, and molecular biological experiments. The family of normal distributions is not the only family of parameter-dependent (parametric) distributions of random variables. Moreover, real-life distributions usually differ from the normal ones, and normal distribution may be considered only as a certain approximation. High diversity of parametric families complicates the choice of criteria and statistical tests for analysis of a set of data collected from *n* independent observations. Nonparametric statistics may be of help, since it is free of sample size and distribution requirements, although, as any method of analysis, it also has limitations. An increasing number of experimental data, including those in the field of environmental microbiology, are nowadays analyzed by using nonparametric statistics, which indicates a certain tendency for substitution of parametric methods by nonparametric ones.

Keywords: statistical analysis, parametric methods, normal distribution, sample size, nonparametric methods, environmental microbiology, incubation experiments