

Проблемы принятия решений на основе *P-VALUE* и уровня значимости в прикладных исследованиях

О.В. Максимова^{1,2)}, В.Л. Шнер²⁾

¹⁾ФГБУ «Институт глобального климата и экологии имени академика Ю.А. Израэля»,
Россия, 107258, Москва, ул. Глебовская, 20Б;

²⁾Университет МИСИС,
Россия, 119049, Москва, Ленинский пр-кт, 4

*Адрес для переписки: o-maximova@yandex.ru

Реферат. Обсуждение корректности интерпретации результатов исследований на основе *p-value* научным сообществом в последние десятилетия явилось предметом интенсивных споров. В работе исследовано два основных вопроса: чем целесообразно руководствоваться при выборе уровня значимости при проверке гипотез и чем качественно отличается проверка гипотез с помощью выбора уровня значимости от использования величины *p-value*. Продемонстрировано, что выбор уровня значимости находится в зоне ответственности исследователя, имеет внестатистические обоснования и зависит от типа исследования. Показано, что алгоритмы принятия решений на основе *p-value* и уровня значимости различаются.

Помимо этого, в работе рассматриваются ошибки качественных выводов при принятии решений на основе как уровня значимости, так и значения *p-value*. Основной вывод, представленный в работе: ошибочным следует признать представление о том, что *p-value* или уровень значимости могут отражать как долгосрочные результаты эксперимента, так и доказательство отдельного результата по выборке. Никакие значения *p-value* и уровня значимости не проверяют и не оценивают вероятность истинности или ложности проверяемой гипотезы.

Для обеспечения более надежных выводов следует:

- проводить верификацию результатов каждый раз с помощью дополнительных опытов,
- учитывать результаты предшествующих исследований,
- понимать рамки ограничений полученных результатов,
- принимать во внимание вариабельность всех процессов.

Ключевые слова. *P-value*, уровень значимости, статистическая значимость, ошибки I и II рода, статистическая гипотеза.

Decision making problems based on p-value and significance level in applied research

O.V. Maksimova^{1), 2)}, V.L. Shper²⁾

¹⁾Yu.A. Izrael Institute of Global Climate and Ecology,
20B, Glebovskaya str., 107058, Moscow, Russian Federation

²⁾ University of Science and Technology (MISIS),
4, Leninsky pr., Moscow, 119049, Russian Federation

*Correspondence address: o-maksimova@yandex.ru

Abstract. The correct interpretation of *p-value*-based research results has been the subject of intense debate in the scientific community in recent decades. This paper examines two questions: what is the appropriate guideline for choosing a significance level when testing hypotheses, and how hypothesis testing using a significance level differs qualitatively from using a *p-value*. It is demonstrated that choosing a significance level is the responsibility of the researcher, has non-statistical justifications, and depends on the type of the study. It is shown that decision-making algorithms based on *p-value* and significance level are different.

Furthermore, the paper examines errors in qualitative inference when making decisions based on both significance level and *p-value*. The main conclusion of this paper is that the idea that a *p-value* or significance level can reflect both the long-term results of an experiment and be the proof of an individual sample result is erroneous. Neither *p-value* nor significance level verifies or evaluates the probability of the truthiness or falseness of the hypothesis being tested. To ensure more reliable conclusions, it is necessary to verify the results every time possible with additional experiments as well as take into account the results of previous studies, understand the limitations of the results found, and take into consideration the variability of all processes.

Keywords. *P-value*, significance level, statistical significance, type I and type II errors, statistical hypothesis.

*Во время экспериментирования начальная гипотеза никогда не
доказывается и не устанавливается, но, возможно, не одобряется.
Можно сказать, что каждый опыт существует лишь для того,
чтобы дать фактам возможность проверить начальную гипотезу.*

Р. Фишер, 1935 г.

Введение

При проверке статистических гипотез исследователь сталкивается с задачей назначения уровня значимости (вероятности ошибки I рода), т.е. веро-

ятности отклонить гипотезу, которой первоначально он придерживается, если на самом деле эта гипотеза истинна. Одновременно, он всегда рискует совершить ошибку II рода – отвергнуть альтернативную гипотезу, если она истинна. Английский статистик Рональд Фишер, разрабатывавший методы исследований для сельскохозяйственных полевых экспериментов, ввел в 1925 г. порог для уровня значимости, равный 0.05, пытаясь обеспечить баланс между ошибками I и II рода (Fisher, 1925):

«Значение $p = 0.05$ или 1 к 20 удобно брать как предел для оценки отклонения, позволяющий рассматривать его как значимое или нет. Таким образом, отклонения, превышающие стандартные величины отклонений в 2 раза, формально считаются значимыми».

Выбор значения 0.05 был, строго говоря, произвольным. Однако в дальнейшем значение 0.05 получило широкое распространение при проверке статистических гипотез как эталонное. Такую популярность можно объяснить стремлением создания единого стандарта для облегчения сравнения результатов различных исследований, и потребностью в простом и понятном критерии для определения статистической значимости. При этом так часто используемое значение 0.05 в качестве уровня значимости оказалось вырванным из контекста размышлений Рональда Фишера и его научного друга Уильяма Госсета. Известный британский инженер-статистик Уильям Госсет, применявший статистические методы на множестве натурных данных в экспериментальной лаборатории пивоварни Guinness^{1),2)} в Дублине, и переписывавшийся много лет с Р. Фишером и К. Пирсоном, полагал, что концепция статистической значимости сама по себе должна быть неразрывно связана с поставленной практической задачей. В своих письмах к Рональду Фишеру он отмечал, что в зависимости от важности вопроса, стоящего перед исследователем, и особенностей выборки, уровень значимости может выбираться не одинаковым: бывают ситуации, в которых принятое значение 0.1 вполне уместно, а бывают, наоборот, когда значение, к примеру, 0.01 окажется недостаточным для получения достоверных практических выводов (Pearson et al., 1990).

Позднее, Рональд Фишер пришел к выводу, что необходимо отказаться от фиксированных уровней значимости, заменив их системой p -value как вероятности полученного по выборке результата или более экстремального, если первоначальная гипотеза верна (Fisher, 1956). Однако, и эту идею с течением времени стали воспринимать как некий эталонный подход к проверке статистических гипотез, широко распространенный как в учебном процессе, так и в прикладных исследованиях. Многие журналы стали принимать к публикации лишь те статьи, в которых присутствует p -value, которое как бы автоматически обосновывает научные результаты и выводы авторов (Шпер, 2021). И тогда термин «статистическая значимость» стал восприниматься авторами и рецензентами журналов как эквивалент научной доказательности

1) <https://www.guinness.com> (дата обращения 09.06.2025).

2) <https://priceconomics.com/the-guinness-brewer-who-revolutionized-statistics/> (дата обращения 09.06.2025).

выводов. Поэтому новая волна обсуждения корректности интерпретации результатов исследований на основе p -value всколыхнула научное сообщество в последние десятилетия и явилась предметом интенсивных споров (Goodman, 1999; Wasserstein, 2016; Amrhein, 2019; Шпер, 2021).

Исследуем два основных вопроса:

- чем целесообразно руководствоваться, выбирая уровень значимости при проверке статистических гипотез;
- чем качественно отличается проверка гипотез с помощью выбора уровня значимости от использования величины p -value.

Методы и материалы

Опишем понятие величины p -value при проверке статистических гипотез.

Пусть исследователь придерживается некоторой гипотезы H_0 (нулевая/основная гипотеза) (рис. 1). Ее проверка не может быть реализована без выдвижения альтернативной гипотезы H_1 , которой будет следовать эксперт при отвержении основной гипотезы H_0 . Чаще всего, следуя структуре большинства статистических критериев, основная гипотеза выдвигается простой, а альтернативная – сложной. Это связано с тем, что в основе критериев проверки основной гипотезы лежат законы распределения с фиксированными значениями параметров.

Выбирается статистический критерий, позволяющий проверить, согласуются ли выборочные данные с основной гипотезой. Критерий формирует статистику (функцию выборочных данных), которая имеет распределение F_0 , если придерживаться гипотезы H_0 . Согласно выбранному критерию, рассчитывается наблюдаемое значение критерия $x_{\text{набл}}$. Исходя из закона распределения F_0 , для него есть значения, более и менее вероятные. Тогда при условии истинности гипотезы H_0 на рис. 1 вероятность $P(X \geq x_{\text{набл}})$ задает значение p -value. В общем случае можно сказать, что p -value количественно определяет вероятность наблюдения полученных результатов (или более экстремальных), если нулевая гипотеза верна. То есть это то значение уровня значимости, которое дала выборка (эксперимент).

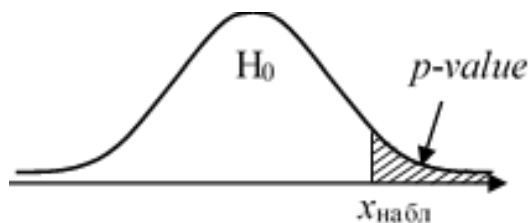


Рисунок 1. Графическая интерпретация величины p -value ($x_{\text{набл}}$ – наблюдаемое значение выбранной статистики, H_0 – основная гипотеза, которой придерживается исследователь)

Figure 1. Graphical interpretation of the p -value ($x_{\text{набл}}$ is the observed value of the selected statistic, H_0 is the main hypothesis that the researcher adheres to)

Результаты и обсуждение

Рассмотрим, с какой целью исследователь делает выборки и какие. Ответ на этот вопрос позволит разобраться с *первой задачей* исследования: чем целесообразно руководствоваться при выборе уровня значимости для проверки статистической гипотезы?

Научные исследования можно отнести к перечислительным или аналитическим (Адлер, 2020). Они преследуют совершенно разные цели: перечислительные исследования направлены на оценку параметров данной конкретной системы на основе результатов, как правило, случайного выборочного наблюдения, а аналитические направлены, в первую очередь, на предсказание поведения данных в будущем, т.е. прогноз. В первом случае изучаются данные в статике (изучается генеральная совокупность), во втором – данные в динамике (ход реального процесса).

В перечислительных исследованиях (и иногда в аналитических, если нас интересует выявление взаимосвязей некоторых параметров или их сравнение), какими бы небольшими не были рассчитываемые по выборке статистики, вопрос их значимости имеет значение (Rummel, 1976). Например, в экологических исследованиях часто ставится задача именно выявления факторов, влияющих на состояние окружающей среды, а не составления многопараметрических моделей с целью прогноза (Кухта и др., 2022; Кухта и др., 2024). Это связано с тем, что клубок взаимосвязей в природе очень многогранен и запутан, и многофакторные математические модели, как правило, в таких исследованиях очень сложны и рассматриваются редко. В таких случаях обнаруженные, к примеру, статистически значимые корреляции величиной 0.4 или 0.5 могут оказаться полезны, поскольку помогают верифицировать общие закономерности развития экосистем. Например, в работе (Кухта и др., 2022) на основе значимых на уровне 0.1 корреляций верифицируется выдвигаемая на основе многолетних наблюдений гипотеза о лимитирующей роли количества осадков для роста сосны обыкновенной в разных типах местообитаний, а в работе (Кухта и др., 2024) верифицируется гипотеза о влиянии выпадений соединений азота в фоновом районе загрязнения воздуха за счет трансграничного переноса на рост древостоев, где с помощью значимых корреляций подтверждается, что в бедной ресурсами среде обитания поступления в почву азота стимулируют рост деревьев в качестве биогенных элементов.

Однако, если нам нужны результаты для составления прогнозов, то, к примеру, корреляция выборочных данных с модельными, даже равная 0.8, может оказаться недостаточной, так как в таком случае процент объясненной дисперсии составляет лишь 64%, и остается высокая доля не связанных с выбранной моделью дисперсий (36%). Или другой пример: вряд ли кого-то удовлетворит прогноз, что дневная температура воздуха по городу в определенный день составит от 5°C до 19°C. Но еще сложнее, если полученные результаты должны стать основой для принятия управленческих решений. В таком случае приемлемым может быть только высокий процент объясненной дисперсии (как правило, 90% и более), а не просто статистически значимая

корреляция выборочных данных с математической моделью (с модельной ситуацией можно ознакомиться, к примеру, в работе Rummel, 1976). Таким образом, статистическая значимость, например, коэффициента корреляции, еще не означает сильную взаимосвязь изучаемых признаков, позволяющую составить надежный прогноз. Одной из задач, где значимых корреляций недостаточно для прогноза, является выбор глобальной климатической модели, наиболее точно воспроизводящей климат некоторого региона: от того, насколько точно выбранная модель будет объяснять вклад дисперсии натуральных данных, может зависеть социально-экономический путь развития региона (Klimont et al., 2017; Максимова, 2025).

Приведем пример, когда вне зависимости от выбранного уровня значимости можно получить ложные умозаключения. Ситуация касается сравнения двух выборок, относящихся к разным отрезкам времени, что требует как глубокого знания условий отбора, так и иногда специальных методов и подходов статистического анализа (Максимова, 2025). Опишем пример, приведенный в работе (Шпер, 2021). Пусть на предприятии уровень брака продукции за какой-то предыдущий временной промежуток был равен 2%, и пусть недавно открылось производство на новой площадке. Естественно, есть желание убедиться, что качество не пострадало. Для этого делается новая выборка, в которой обнаруживается 4.2% бракованных изделий. На основе этих данных нам нужно ответить на вопрос: изменилось ли качество продукции? С математической точки зрения на основании анализа выборок необходимо решить задачу: значимо статистически различие между 4.2 и 2% или нет? Основная проблема принятия решения по результатам проверки такой гипотезы заключается в том, что выводы делаются, предполагая сохранение условий эксперимента, что подразумевает неизменность всей системы, в рамках которой получены наши эмпирические данные. Однако, выполнение этого условия маловероятно, так как мы должны убедиться, что новые изделия сделаны из таких же материалов, точно такого же качества, на том же оборудовании, теми же операторами, при тех же прочих условиях окружающей среды, причем были отобраны теми же способами, что и все предыдущие, точно так же измерены и т. д. (Шпер, 2021). Поэтому к выводам, сделанным с помощью некоторого статистического критерия при выбранном уровне значимости, вне зависимости от его величины необходимо относиться с большой осторожностью, изучив предварительно всю систему во времени или использовать дополнительные инструменты для сравнения (Максимова, 2025).

Подведем итоги:

Выбор уровня значимости при проверке статистической гипотезы зависит от типа исследования. Для перечислительных исследований можно ориентироваться на традиционно используемые значения; для аналитических исследований это решение, которое должен принимать исследователь, исходя из сущности данных и важности прогноза. К традиционным значениям 0.1 и/или 0.05 в этих случаях следует относиться с осторожностью. Никакие сколь угодно малые значения выбранных уровней значимости не дают никаких точных гарантий.

Уровень ошибки принятия гипотезы выбирается исследователем и имеет внестатистические обоснования. Интуиция и знание процесса/совокупности при выборе уровня значимости и интерпретации результатов являются первоочередными: Рональд Фишер в беседе с Крамером отметил «*Меня иногда обвиняют в интуиции как в преступлении*», имея ввиду, что интуиция в выборе подхода исследования (математической модели) вырабатывается на основе предварительного тщательного знакомства с выборкой (Шейнин, 2008).

Теперь перейдем к рассмотрению *второй задачи* исследования. Чем качественно отличается проверка гипотез с помощью выбора уровня значимости от использования величины p -value? Несмотря на сходство структуры показателей уровня значимости и p -value, принятие решений на основе этих двух характеристик различается. При использовании p -value самое распространенное заблуждение заключается в том, что величина $(1 - p\text{-value})$ рассматривается как вероятность того, что основная гипотеза верна. Такой подход может привести к большому числу ложных заключений относительно дальнейших, основанных на этом выводе, решений. На самом же деле p -value показывает лишь вероятность наблюдаемого в выборке, или более экстремального, значений, если основная гипотеза истинна (рис. 1). Проверка гипотез на основе использования p -value – это практический подход, статистические свойства которого выводятся из бесконечной генеральной совокупности, применяемый к любому единичному эксперименту. Теория проверки гипотез при заданном уровне значимости предполагает следование правилу поведения, которое при повторении экспериментов гарантирует исследователю, что он может ошибаться, но при этом достаточно редко. Проверка гипотез сама по себе не предоставляет доказательств истинности или ложности конкретного суждения, как отмечал Рональд Фишер (Fisher, 1935). Путаница между подходами возникает из-за того, что критическую область, задаваемую леммой Неймана-Пирсона, можно определить в терминах p -value (Neyman, Pearson, 1933). Какая гипотеза наиболее правдоподобна: для которой полученное значение $p\text{-value} = 0.01$ или для которой $p\text{-value} = 0.05$? Ответа на этот вопрос не существует, так как величина p -value привязана к выборке, и, тем самым, обладает свойством случайности. Поэтому сравнение выдвигаемых гипотез на основе величины p -value некорректно. В такой постановке задачи меньшее значение p -value указывает лишь на то, что исследуемая гипотеза с меньшей вероятностью вызвана простой случайностью выборочных данных. Более того, служит заблуждением тот факт, что вероятность p -value сохранится при повторном эксперименте.

Опишем еще одну трудность, возникающую при проверке гипотез с использованием сравнения уровней значимости или при его выборе. Действительно ли гипотеза, принятая на уровне значимости 0.01, более вероятна, чем принятая на уровне значимости 0.05? Конечно, ответ отрицателен; и для такого ответа есть две причины. Первая причина связана, в первую очередь, с аналитическими исследованиями: если процесс рассматривается в долгосрочной перспективе, то он может меняться (как было отмечено выше), и более вероятной может быть не та гипотеза, что была принята первоначально. Вто-

рая причина связана с известным фактом из общего курса математической статистики: меньший уровень значимости (вероятность ошибки I рода) вызывает, как правило, симметричное увеличение вероятности ошибки II рода. Опишем этот факт подробнее: если мы отвергаем H_0 , мы можем отвергнуть ее и в том случае, когда она истинна; если мы принимаем H_0 , мы можем принять ее и в том случае, когда она ложна. Эти два источника ошибок редко можно полностью устранить. Этот факт ясно демонстрирует иллюстрация на рис. 2 при проверке гипотезы $H_0: M(X) = m_0$ против альтернативной $H_1: M(X) = m_1$. Таким образом, с уменьшением уровня значимости, возникает ловушка, связанная с увеличением вероятности не принять верную альтернативную гипотезу H_1 (рис. 2). В некоторых практических случаях важнее избежать первого типа ошибок, а в других – второго. Известна историческая проблема, рассмотренная Лапласом – что серьезнее: осудить невинного или оправдать виновного? (Neyman, Pearson, 1933). Ответ на этот вопрос будет зависеть от последствий каждого типа ошибок: будет ли наказание в виде смертной казни или штрафа; какова опасность для общества освобожденных преступников и т.п.?

Более того, если будет выбран уровень значимости слишком маленький (почти ноль), то наш используемый статистический критерий будет «статься» по всем выборкам принимать основную гипотезу H_0 .

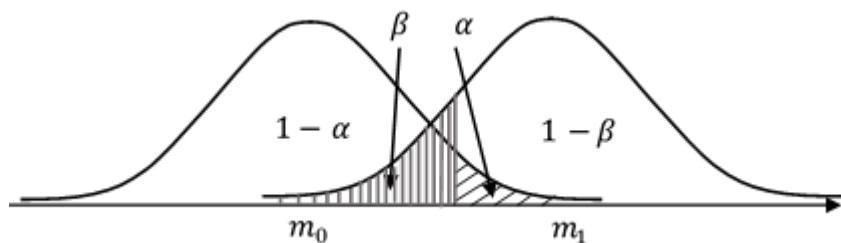


Рисунок 2. Геометрическая интерпретация вероятностей ошибок I рода (α) и II рода (β) при проверке гипотезы $H_0: M(X) = m_0$ против альтернативной гипотезы $H_1: M(X) = m_1$

Figure 2. Geometric interpretation of the probabilities of errors of the first kind (α) and second kind (β) when testing the hypothesis $H_0: M(X) = m_0$ against the alternative hypothesis $H_1: M(X) = m_1$

Какой выход из такой ситуации? Во-первых, математическая статистика в таких случаях может лишь показать, как контролировать и минимизировать риск ошибок, а вот как именно следует установить баланс между ошибками I и II рода, должен определить исследователь, отталкиваясь от знания ситуации (Fisher, 1925; Neyman, Pearson, 1933). Во-вторых, необходимо верифицировать каждый возможный раз дополнительными опытами результат, чтобы дать фактам возможность проверить начальную гипотезу, и при этом учитывать опыт предшествующих исследований (Fisher, 1935; Шпер, 2021). Также, если есть несколько критериев проверки гипотезы H_0 , то выбирается тот, который при заданном уровне значимости минимизирует вероятность ошибки II рода, т.е. задает минимальную вероятность принятия H_0 , когда истинной гипотезой является H_1 (т.е. более мощный критерий). И, самое важное, следует признать ограниченность полученных результатов, принять существова-

ние вариабельности всех процессов и отказаться от принятия решений вида «истина – ложь», основанных на каких-либо произвольных пороговых значениях (Шпер, 2021).

Подведем итоги:

Сходство использования показателей уровня значимости и *p-value* обусловлено леммой Неймана-Пирсона, однако, алгоритм принятия решений на основе этих двух характеристик различается: *p-value* предполагает получение выводов о единичном эксперименте из бесконечной генеральной совокупности (соответственно, *p-value* нельзя рассматривать как вероятность того, что основная гипотеза не верна), а работа с уровнем значимости предполагает следование правилам проверки гипотез.

Вопрос о том, какой из двух гипотез больше доверять на основе полученных значений *p-value* (или выбранных уровней значимости), не корректен с точки зрения подхода проверки гипотез.

Заключение

В работе представлено исследование двух важных вопросов, возникающих при проверке статистических гипотез: выбор уровня значимости и различие в проверке с помощью уровня значимости и с использованием величины *p-value*. Продемонстрировано на примерах, что выбор уровня значимости при проверке статистических гипотез лежит на плечах исследователя, имеет внестатистические обоснования и зависит от типа исследования. Показано, что алгоритмы принятия решений на основе *p-value* и уровня значимости различаются.

В ходе исследования также обсуждаются заблуждения, с которыми сталкиваются исследователи в прикладных областях при статистической проверке научных результатов:

- *p-value* или уровень значимости α – это вероятность того, что нулевая гипотеза не верна;
- маленькое значение *p-value* или уровня значимости α доказывает истинность принятой нулевой гипотезы,
- $(1 - p-value)$ или $(1 - \alpha)$ – это вероятность того, что альтернативная гипотеза верна;
- высокие значения $(1 - p-value)$ или $(1 - \alpha)$ – это вероятность того, что результаты сохранятся при повторении эксперимента,
- α – это вероятность того, что результат был обусловлен ошибкой выборки.

В работе продемонстрировано, что все вышенаписанные заключения ложны.

Самый важный вывод, который следует из исследования: ошибочным следует признать представление о том, что одно число (*p-value* или уровень значимости) может отражать как долгосрочные результаты эксперимента, так и доказательство отдельного результата по выборке.

Никакие значения *p-value* и уровня значимости не проверяют и не оценивают вероятность истинности или ложности гипотезы. Согласие или несогла-

ние опытных данных с нулевой гипотезой никак не влияет на ее истинность или ложность (основная гипотеза изначально истинна или ложна, но, как правило, это исследователю не известно). Поэтому никакие однозначные выводы по результатам расчета p -value делать нельзя. Для обеспечения надежности выводов необходимо проводить верификацию результатов каждый раз дополнительными опытами и учитывать результаты предшествующих исследований, понимать рамки ограничений своих научных исследований и выводов, и принимать во внимание вариабельность всех процессов.

Литература

Адлер, Ю. П. (2020) Зачем нужны выборки? *Контроль качества продукции*, № 8, с. 46-49.

Кухта, А.Е., Максимова, О.В. (2022) Воздействие климатических факторов вегетационного периода на приросты сосны обыкновенной в Среднем Поволжье и на побережье Белого моря, *Метеорология и гидрология*, № 1, с. 72-83, doi 10.52002/0130-2906-2022-1-72-83.

Кухта, А.Е., Максимова, О.В., Махрова, Т.Г., Громов, С.А. (2024) Отклик приростов сосны обыкновенной *Pinus sylvestris* бореальных фитоценозов на трансграничное загрязнение, *Проблемы региональной экологии*, № 6, с. 26-34, doi 10.24412/1728-323X-2024-2-26-34.

Максимова, О.В. (2025) Сравнение временных рядов в прикладных исследованиях: близость, синхронность и корреляция, *Экологический мониторинг и моделирование экосистем*, № 1-2, с. 46-64, doi 10.24412/ 2782-3237-2025-1-2-46-64.

Шейнин, О.Б. (2008) *Статьи по истории теории вероятностей и статистики*, Берлин: NG Verlag, 258 с.

Шпер, В.Л. (2021) Что такое p -значение, и почему вокруг него столько шума? *Контроль качества продукции*, №3, с. 37-43.

Amrhein, V., Greenland, S., McShane, B. (2019) Scientists rise up against statistical significance, *Nature*, no. 567, pp. 305-307, doi: 10.1038/d41586-019-00857-9.

Fisher, R.A. (1925) *Statistical Methods for Research Workers*, Edinburgh, UK: Oliver and Boyd.

Fisher, R.A. (1935) *The Design of Experiments* (9th ed.), Macmillan, 250 p.

Fisher, R.A. (1956) *Statistical Methods and Scientific Inference*, New York, Hafner Press, 175 p.

Goodman, S.N. (1999) Toward Evidence-Based Medical Statistics. The P value fallacy, *Annals of Internal Medicine*, no. 130(12), p. 995, doi 10.7326/0003-4819-130-12-199906150-00008.

Klimont, Z., Hoglund-Isaksson, L., Heyes, C., Rafaj, P., Schopp, W., Cofala, J., Purohit, P., Borken-Kleefeld, J., Kupiainen, K., Kiesewetter, G., Winiwarter, W.,

Amann, M., Zhao, B., Wang, S., Bertok, I., and Sander, R. (2017) *Global scenarios of air pollutants and methane, 1990-2050*, ACP.

Neyman, J, Pearson, E. (1933) On the problem of the most efficient tests of statistical hypotheses, *Philos Trans R Soc Lond A.*, no. 231, pp. 289-337, doi: 10.1098/rsta.1933.0009.

Pearson, E.S., Plackett, R.L., Barnard, G.A. (1990) *A Statistical Biography of William Sealy Gosset*, New York, Clarendon Press, 150 p.

Rummel, R.J. (1976) *Understanding Correlation*, Honolulu, Hawaii, Department of Political Science, University of Hawai'i.

Wasserstein, R.L., Lazar, N.A. (2016) The ASA Statement on p-Values: Context, Process, and Purpose, *The American Statistician*, no. 70(2), pp. 129-133, doi 10.1080/00031305.2016.1154108.

Reference

Adler, Yu. P. (2020) Zachem nuzhny vyborki? [Why Samples Are Needed? Product Quality Control], *Kontrol' kachestva produktsii*, 8, pp. 46-49.

Kukhta, A.Ye., Maksimova, O.V. (2022) Vozdeystviye klimaticheskikh faktorov vegetatsionnogo perioda na prirosty sosny obyknovennoy v Srednem Povolzh'ye i na poberezh'ye Belogo moray [he Impact of Climatic Factors of the Growing Season on Scots Pine Growth in the Middle Volga Region and on the White Sea Coast], *Meteorology and Hydrology*, 1, pp. 72-83.

Kukhta, A.Ye., Maksimova, O.V., Makhrova, T.G., Gromov, S.A. (2024) Otklik prirostov sosny obyknovennoy Pinus sylvestris boreal'nykh fitotsenozov na transgranichnoye zagryazneniye [Response of Scots pine Pinus sylvestris growth in boreal phytocenoses to transboundary pollution], *Problemy regional'noy ekologii*, 6, pp. 26-34.

Maksimova, O.V. (2025) Sravneniye vremennykh ryadov v prikladnykh issledovaniyakh: blizost', sinkhronnost' i korrelyatsiya [Response of Scots pine Pinus sylvestris growth in boreal phytocenoses to transboundary pollution], *Ekologicheskiy monitoring i modelirovaniye ekosistem*, 1-2, pp. 46-64.

Sheynin, O.B. (2008) *Stat'i po istorii teorii veroyatnostey i statistiki* [Articles on the History of Probability Theory and Statistics], Berlin, NG Verlag, 258 p.

Shper, V.L. (2021) Chto takoye p-znachenije, i pochemu vokrug nego stol'ko shuma? [What is a p-value, and why is there so much fuss about it?], *Kontrol' kachestva produktsii*, 3, pp. 37-43

Amrhein, V., Greenland, S., McShane, B. (2019) Scientists rise up against statistical significance, *Nature*, no. 567, pp. 305-307, doi: 10.1038/d41586-019-00857-9.

Fisher, R.A. (1925) *Statistical Methods for Research Workers*, Edinburgh, UK: Oliver and Boyd.

Fisher, R.A. (1935) *The Design of Experiments* (9th ed.), Macmillan, 250 p.

Fisher, R.A. (1956) *Statistical Methods and Scientific Inference*, New York, Hafner Press, 175 p.

Goodman, S.N. (1999) Toward Evidence-Based Medical Statistics. 1. The P value fallacy, *Annals of Internal Medicine*, no. 130(12), p. 995, doi 10.7326/0003-4819-130-12-199906150-00008.

Klimont, Z., Hoglund-Isaksson, L., Heyes, C., Rafaj, P., Schopp, W., Cofala, J., Purohit, P., Borken-Kleefeld, J., Kupiainen, K., Kieseewetter, G., Winiwarter, W., Amann, M., Zhao, B., Wang, S., Bertok, I., and Sander, R. (2017) *Global scenarios of air pollutants and methane 1990-2050*, ACP.

Neyman, J, Pearson, E. (1933) On the problem of the most efficient tests of statistical hypotheses, *Philos Trans R Soc Lond A.*, no. 231, pp. 289-337, doi 10.1098/rsta.1933.0009.

Pearson, E.S., Plackett, R.L., Barnard, G.A. (1990) *Student. A Statistical Biography of William Sealy Gosset*, New York, Clarendon Press, 150 p.

Rummel, R.J. (1976) *Understanding Correlation*, Honolulu, Hawaii, Department of Political Science, University of Hawai'i.

Wasserstein, R.L., Lazar, N.A. (2016) The ASA Statement on p-Values, Context, Process, and Purpose, *The American Statistician*, no. 70(2), pp. 129-133, doi 10.1080/00031305.2016.1154108.

Статья поступила в редакцию (Received): 31.10.2026.

Статья доработана после рецензирования (Revised): 07.11.2026.

Для цитирования / For citation:

Максимова, О.В., Шпер, В.Л. (2026) Проблемы принятия решений на основе P-VALUE и уровня значимости в прикладных исследованиях, *Экологический мониторинг и моделирование экосистем*, т. XXXVII, № 1-2, с. 9-20, doi: 10.21513/0207-2564-2026-1-2-9-20.

Maksimova, O.V., Shper, V.L. (2026) Decision making problems based on p-value and significance level in applied research, *Environmental Monitoring and Ecosystem Modelling*, vol. XXXVII, no. 1-2, pp. 9-20, doi: 10.21513/0207-2564-2026-1-2-9-20.