Математические методы исследования
Mathematical methods of investigation

DOI: https://doi.org/10.26896/1028-6861-2020-86-3-67-76

О МЕТОДАХ ПРОВЕРКИ ОДНОРОДНОСТИ ДВУХ НЕЗАВИСИМЫХ ВЫБОРОК

© **Александр Иванович Орлов**

Московский государственный технический университет им. Н. Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., д. 5; e-mail: prof-orlov@mail.ru

Статья поступила 4 июня 2019 г. Поступила после доработки 5 августа 2019 г. Принята к публикации 20 ноября 2019 г.

Методы проверки однородности двух независимых выборок — классическая область математической статистики. За более чем 111 лет с момента публикации основополагающей статьи Стьюдента разработаны различные критерии проверки статистической гипотезы однородности в различных постановках, изучены их свойства. Настала потребность в упорядочении совокупности найденных научных результатов. Необходим анализ всего многообразия постановок задач проверки статистических гипотез однородности двух независимых выборок, а также соответствующих статистических критериев. Такому анализу посвящена настоящая статья. В ней сведены основные результаты, касающиеся методов проверки однородности двух независимых выборок, и проведено их сравнительное изучение, позволяющее системно анализировать многообразие таких методов в целях выбора наиболее адекватного для обработки конкретных данных. На основе базовой вероятностно-статистической модели сформулированы основные постановки задачи проверки однородности двух независимых выборок. Приведен сравнительный анализ критериев Стьюдента и Крамера – Уэлча, предназначенных для проверки однородности математических ожиданий, обоснована рекомендация по широкому применению критерия Крамера – Уэлча. Из непараметрических методов проверки однородности рассмотрены критерии Вилкоксона, Смирнова, Лемана – Розенблатта. Разобраны два мифа о критерии Вилкоксона. На основе анализа публикаций основоположников показана некорректность термина «критерий Колмогорова – Смирнова». Для проверки абсолютной однородности, т.е. совпадения функций распределения выборок, рекомендовано использовать критерий Лемана – Розенблатта. Рассмотрены актуальные проблемы разработки и применения непараметрических критериев, в том числе различие номинальных и реальных уровней значимости, затрудняющее сравнение критериев по мощности, отмечена необходимость учета совпадений выборочных значений (с точки зрения классической теории математической статистики вероятность совпадений равна 0).

Ключевые слова: проверка статистических гипотез; независимые выборки; однородность характеристик; абсолютная однородность; критерий Крамера – Уэлча; критерий Вилкоксона; критерий Смирнова; критерий Лемана – Розенблатта.

ON THE METHODS OF TESTING THE HOMOGENEITY OF TWO INDEPENDENT SAMPLES

© **Alexander I. Orlov**

Bauman Moscow State Technical University, 5, 2-ya Baumanskaya ul., Moscow, 105005, Russia; e-mail: prof-orlov@mail.ru

Received June 4, 2019. Revised August 5, 2019. Accepted November 20, 2019.

Methods for testing the homogeneity of two independent samples refer to a classic area of mathematical statistics. Various criteria for testing the statistical hypothesis of homogeneity in different statements have been developed and their properties have been studied for more than 110 years since publication of the fundamental Student's article. Nowadays, the streamlining of the totality of gained scientific results has become an urgent problem. It is necessary to analyze the whole variety of problem statements for testing the statistical hypotheses of the homogeneity of two independent samples, as well as the corresponding statistical criteria. Such an analysis is the goal of the article. We summarize the main results regarding the methods for testing the homogeneity of two independent samples and their comparative study which allows system analysis of the diversity of such methods in order to select the most appropriate for processing specific data. The main statements of the problem of testing the homogeneity of two independent samples are formulated using the basic probabilistic-statistical model. A comparative analysis of the Student and Cramer — Welch criteria designed to test the homogeneity of mathematical expectations is presented along with substantiation of the recommendation on the widespread use of the Cramer – Welch criterion. The criteria of Wilcoxon, Smirnov, Lehmann – Rosenblatt are considered among nonparametric methods for testing homogeneity. Two myths about the Wilcoxon criteria are dismantled. Analysis of the publications of the founders revealed the incorrectness of the term "Kolmogorov – Smirnov criterion." To verify the absolute homogeneity, i.e. coincidence of the distribution functions of samples, it is recommended to use the Lehmann – Rosenblatt criterion. The current problems of the development and application of nonparametric criteria are discussed, including the difference between nominal and real significance levels, which complicates comparison of the criteria in power. The necessity of taking into account the coincidence of the sample values (from the view point of the classical theory of mathematical statistics, the probability of coincidences is 0) is marked.

Keywords: testing of statistical hypotheses; independent samples; homogeneity of characteristics; absolute homogeneity; Cramer – Welch test; Wilcoxon test; Smirnov test; Lehmann – Rosenblatt test.

Введение

Проверка однородности двух независимых выборок — классическая область математической статистики. Основополагающей в ней является статья Стьюдента [1]. За более чем 111 лет получены многочисленные результаты различными авторами, в том числе нами. Однако имеется потребность в упорядочении совокупности найденных научных результатов, в том числе опубликованных в журнале «Заводская лаборатория. Диагностика материалов». Необходим анализ всего многообразия постановок задач проверки статистических гипотез однородности двух независимых выборок, а также соответствующих критериев. Такому анализу посвящена данная статья.

Базовая вероятностно-статистическая модель

В математико-статистических терминах постановка задачи такова: имеются две выборки x_1, x_2, ..., x_m и y_1, y_2, ..., y_n, требуется проверить их однородность. Выборка моделируется как совокупность независимых одинаково распределенных числовых случайных величин. Термин «однородность» уточняется ниже.

Понятием, противоположным «однородности», является «различие» (или «наличие эффекта»). Можно переформулировать задачу: требуется проверить, есть ли различие между выборками. Если различия нет, то для дальнейшего изучения в целях практического применения выводов две рассматриваемые выборки часто объединяют в одну.

Например, в маркетинге важно выделить сегменты потребительского рынка. Если установлена однородность двух выборок мнений потребителей, то возможно объединение сегментов, из которых эти выборки взяты, в один. В дальнейшем это позволит осуществлять по отношению к ним одинаковую маркетинговую политику (проводить одни и те же рекламные мероприятия и т.п.). Если же установлена неоднородность, то поведение потребителей в двух сегментах различно, объединять эти сегменты нельзя и могут понадобиться различные маркетинговые стратегии, своя для каждого из этих сегментов.

Для обоснованного выбора и применения организационно-экономических (эконометрических, статистических) методов необходимо прежде всего построить и обосновать вероятностную модель порождения данных. При проверке однородности двух выборок общепринята модель, в которой x_1, x_2, ..., x_m рассматриваются как результаты m независимых наблюдений некоторой случайной величины X с функцией распределения $F(x)$, неизвестной статистику, а y_1, y_2, ..., y_n — как результаты n независимых наблюдений, вообще говоря, другой случайной величины Y с функцией распределения $G(x)$, также неизвестной статистику. Кроме того, предполагается, что наблюдения в одной выборке не зависят от наблюдений в другой, поэтому выборки и называют независимыми.

Возможность применения модели в конкретной реальной ситуации требует обоснования. Независимость и одинаковая распределенность результатов наблюдений (измерений, испытаний, анализов, опытов, обследований) могут быть установлены исходя из методики проведения конкретных наблюдений или путем проверки статистических гипотез независимости и одинаковой распределенности с помощью соответствующих критериев проверки статистических гипотез [2].

Основные постановки задачи проверки однородности двух независимых выборок

Наивысшая степень однородности (абсолютная однородность) достигается, если обе выборки взяты из одной и той же генеральной совокупности, т.е. справедлива нулевая гипотеза

$$H_0\colon F(x) = G(x) \text{ при всех } x.$$

Отсутствие абсолютной однородности означает, что верна альтернативная гипотеза, согласно которой

$$H_1\colon F(x_0) \neq G(x_0)$$

хотя бы при одном значении аргумента x_0. Если гипотеза H_0 принята, то выборки можно объединить в одну, если нет, то нельзя.

В некоторых случаях целесообразно проверять совпадение не функций распределения, а лишь некоторых характеристик случайных величин X и Y — математических ожиданий, медиан, дисперсий, коэффициентов вариации и др. (т.е. проверять однородность тех или иных характеристик). Например, однородность математических ожиданий означает, что справедлива гипотеза

$$H'_0\colon M(X) = M(Y),$$

где $M(X)$ и $M(Y)$ — математические ожидания случайных величин X и Y, результаты наблюдений над которыми составляют первую и вторую выборки соответственно. Доказательство различия между выборками в рассматриваемом случае — это доказательство справедливости альтернативной гипотезы

$$H'_1\colon M(X) \neq M(Y).$$

Если гипотеза H_0 верна, то и гипотеза H'_0 верна, но из справедливости H'_0, вообще говоря, не следует справедливость H_0. Математические ожидания могут совпадать для различающихся между собой функций распределения. В частности, если в результате обработки выборочных данных принята гипотеза H'_0, то отсюда не следует, что две выборки можно объединить в одну.

Однако в ряде ситуаций целесообразна проверка именно гипотезы H'_0. Например, пусть функция спроса на определенный товар или услугу оценивается путем опроса потребителей (первая выборка) или с помощью данных о продажах (вторая выборка). Тогда маркетологу важно проверить гипотезу об отсутствии систематических расхождений результатов этих двух методов, т.е. гипотезу о равенстве математических ожиданий.

Другой пример — из производственного менеджмента. Пусть изучается эффективность управления бригадами рабочих на предприятии с помощью двух организационных схем, результаты наблюдения — объем производства продукции или услуг на одного члена бригады (производительность), а показатель эффективности организационной схемы — средний (по предприятию) объем производства на одного рабочего. Тогда для сравнения эффективности организационных схем достаточно проверить гипотезу H'_0. Если она принята, то нет оснований заявлять о том, что организационные схемы различаются по эффективности.

Иногда нужно проверить однородность дисперсий. Например, различаются ли два способа измерения по величине случайной ошибки, т.е. по дисперсии случайных погрешностей. В ряде случаев следует проверить, например, однородность коэффициентов вариации.

Проверка однородности характеристик

Наиболее часто рассматривают проверку однородности математических ожиданий.

Традиционный метод проверки однородности (двухвыборочный критерий Стьюдента). Для дальнейшего критического разбора опишем традиционный статистический метод проверки однородности [1], который широко использовали в течение всего XX в. К настоящему времени этот метод устарел, но по традиции встречается в учебной литературе, а потому и продолжает применяться для анализа конкретных данных.

При использовании традиционного метода проверки однородности вычисляют выборочные средние арифметические в каждой выборке

$$\bar{x} = \frac{1}{m}\sum_{1\le i\le m} x_i,\quad \bar{y} = \frac{1}{n}\sum_{1\le i\le n} y_i,$$

затем — выборочные дисперсии

$$s_x^2 = \frac{1}{m-1}\sum_{1\le i\le m}(x_i - \bar{x})^2,\quad s_y^2 = \frac{1}{n-1}\sum_{1\le i\le n}(y_i - \bar{y})^2$$

и статистику Стьюдента

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(m-1)s_x^2 + (n-1)s_y^2}}\sqrt{\frac{mn(n+n-2)}{m+n}}, \qquad (1)$$

на основе которой принимают решение.

По заданному уровню значимости α и числу степеней свободы $(m + n - 2)$ из таблиц распределения Стьюдента (см., например, [2]) находят критическое значение $t_{кр}$. Если $|t| > t_{кр}$, то гипотезу однородности (отсутствия различия) отклоняют, если же $|t| < t_{кр}$, то — принимают. (При односторонних альтернативных гипотезах вместо $|t| > t_{кр}$ проверяют условие $t > t_{кр}$; эту постановку рассматривать не будем, так как в ней нет принципиальных отличий от обсуждаемой здесь.)

В литературе зачастую описывается только приведенный выше алгоритм. Этого недостаточно для квалифицированного анализа статистических данных. Рассмотрим условия применимости традиционного метода проверки однородности, основанного на использовании статистики t Стьюдента, а также обсудим современные методы проверки однородности двух выборок.

Классические условия применимости критерия Стьюдента. Согласно математико-статистической теории, должны быть выполнены два классических условия применимости критерия Стьюдента, основанного на использовании статистики t, заданной формулой (1):

а) результаты наблюдений имеют нормальные распределения

$$F(x) = N(x; m_1, \sigma_1^2),\ G(x) = N(x; m_2, \sigma_2^2)$$

с математическими ожиданиями m_1 и m_2 и дисперсиями σ_1^2 и σ_2^2 в первой и во второй выборках соответственно;

б) дисперсии результатов наблюдений в первой и второй выборках совпадают, т.е.

$$D(X) = \sigma_1^2 = D(Y) = \sigma_2^2.$$

Если условия а) и б) выполнены, то нормальные распределения $F(x)$ и $G(x)$ отличаются только математическими ожиданиями, а поэтому обе гипотезы — H_0 и H'_0 — сводятся к гипотезе

$$H''_0\colon m_1 = m_2,$$

а обе альтернативные гипотезы — H_1 и H'_1 — к гипотезе

$$H''_1\colon m_1 \neq m_2.$$

Если условия а) и б) выполнены, то статистика t при справедливости H''_0 имеет распределение Стьюдента с $(m + n - 2)$ степенями свободы. Только в этом случае описанный выше традиционный метод обоснован безупречно. Если хотя бы одно из условий а) и б) не выполнено, то нет никаких оснований считать, что статистика t имеет распределение Стьюдента, поэтому применение традиционного метода, строго говоря, не обосновано. Обсудим возможность проверки этих условий и последствия их нарушений.

Имеют ли результаты наблюдений нормальное распределение? Как подробно показано в литературе (см., например, [3]), априори нет оснований предполагать нормальность распределения результатов экономических, технико-экономических, технических, медицинских и иных наблюдений. Следовательно, нормальность надо проверять (или использовать непараметрические методы анализа данных). Разработано много статистических критериев для проверки нормальности распределения результатов наблюдений [2]. Однако проверка нормальности — более сложная и трудоемкая статистическая процедура, чем проверка однородности (как с помощью статистики t Стьюдента, так и с использованием непараметрических критериев, рассматриваемых ниже).

Для достаточно надежного установления нормальности требуется весьма большое число наблюдений. 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 [4]. В большинстве технических, экономических, медицинских и иных исследований число наблюдений существенно меньше.

Есть и еще одна общая причина отклонений от нормальности: любой результат наблюдения записывается конечным (обычно 2 – 5) количеством цифр, а с математической точки зрения вероятность такого события равна 0. Точнее, для случайной величины с непрерывной плотностью распределения вероятность попадания в счетное множество рациональных чисел равна 0. Следовательно, при статистической обработке данных в организационно-экономических исследованиях распределение результатов наблюдений практически всегда более или менее отличается от нормального распределения.

Последствия нарушения условия нормальности. Если условие а) не выполнено, то распределение статистики t не является распределением Стьюдента. Однако можно показать, используя Центральную предельную теорему теории вероятностей и теоремы о наследовании сходимости [4, гл. 4], что при справедливости гипотезы H''_0 и условия б) распределение статистики t при росте объемов выборок приближается к стандартному нормальному распределению $\Phi(x) = N(x; 0,1)$. К этому же распределению приближается распределение Стьюдента при возрастании числа степеней свободы. Другими словами, несмотря на нарушение условия нормальности критерий Стьюдента можно использовать (при определенных условиях!) для проверки гипотезы H'_0 при больших объемах выборок. При этом вместо таблиц распределения Стьюдента достаточно пользоваться таблицами стандартного нормального распределения $\Phi(x)$. Это утверждение справедливо для любых функций распределения $F(x)$ и $G(x)$

таких, что $M(X) = M(Y)$, $D(X) = D(Y)$ и выполнены некоторые условия, обычно считающиеся справедливыми в реальных задачах [5]. Если же $M(X) \neq M(Y)$, то нетрудно вычислить, что при больших объемах выборок

$$P(t \le x) \approx \Phi(x - a_{mn}), \qquad (2)$$

где

$$a_{mn} = \frac{\sqrt{mn}[M(X) - M(Y)]}{\sqrt{mD(X) + nD(Y)}}. \qquad (3)$$

Формулы (2) – (3) позволяют приближенно вычислять мощность t-критерия (точность возрастает при увеличении объемов выборок m и n).

О проверке условия равенства дисперсий. Иногда условие б) следует из методики получения результатов наблюдений, например, когда с помощью одного и того же прибора или одинаковой методики m раз измеряют характеристику первого объекта и n раз — второго, а параметры распределения погрешностей измерения при этом не меняются. Однако ясно, что в постановках большинства исследовательских и практических задач нет оснований априори предполагать равенство дисперсий.

Целесообразно ли проверять равенство дисперсий статистическими методами, например, как это иногда предлагают, с помощью F-критерия Фишера? Этот критерий основан на нормальности распределений результатов наблюдений. А от нормальности неизбежны отклонения (см. выше). Причем хорошо известно, что в отличие от t-критерия распределение F-критерия Фишера сильно меняется при малых отклонениях от нормальности [6]. Кроме того, F-критерий отвергает гипотезу $D(X) = D(Y)$ лишь при большом различии выборочных дисперсий. Так, для данных [2] о двух группах результатов химических анализов отношение выборочных дисперсий равно 1,95, т.е. существенно отличается от 1. Тем не менее гипотеза о равенстве теоретических дисперсий принимается при использовании F-критерия на 1 %-ном уровне значимости. Следовательно, при проверке однородности применение F-критерия для предварительной проверки равенства дисперсий в целях обоснования возможности использования критерия Стьюдента нецелесообразно.

Итак, в большинстве технических, экономических, медицинских и иных задач условие б) нельзя считать выполненным, а проверять его перед проверкой однородности нецелесообразно.

Последствия нарушения условия равенства дисперсий. Если объемы выборок m и n велики, то можно показать, что распределение статистики t описывается с помощью только математических ожиданий $M(X)$ и $M(Y)$, дисперсий $D(X)$, $D(Y)$ и отношения объемов выборок, а именно,

$$P(t \le x) \approx \Phi(b_{mn}x - a_{mn}), \qquad (4)$$

где a_{mn} определено формулой (3);

$$b_{mn}^2 = \frac{\lambda D(X) + D(Y)}{D(X) + \lambda D(Y)}, \ \lambda = \frac{m}{n}. \qquad (5)$$

Если $b_{mn} \neq 1$, то распределение статистики t отличается от распределения, заданного формулой (2), полученной в предположении равенства дисперсий. Когда $b_{mn} = 1$? В двух случаях — при $m = n$ и при $D(X) = D(Y)$. Таким образом, при больших и равных объемах выборок требовать выполнения условия б) нет необходимости. Кроме того, ясно, что если объемы выборок мало различаются, то b_{mn} близко к 1. Так, для данных [2] о двух группах результатов химических анализов имеем $b_{mn}^* = 0{,}987$, где b_{mn}^* — оценка b_{mn}, полученная заменой в формуле (5) теоретических дисперсий на их выборочные оценки.

Область применимости традиционного метода проверки однородности с помощью критерия Стьюдента. Подведем итоги рассмотрения t-критерия. Он позволяет проверять гипотезу H_0' о равенстве математических ожиданий, но не гипотезу H_0 о том, что обе выборки взяты из одной и той же генеральной совокупности. Классические условия применимости критерия Стьюдента в подавляющем большинстве технических, экономических, медицинских и иных задач не выполнены. Тем не менее при больших и примерно равных объемах выборок его можно применять. При конечных объемах выборок традиционный метод носит неустранимо приближенный характер.

Критерий Крамера – Уэлча равенства математических ожиданий. Вместо критерия Стьюдента для проверки H_0' целесообразно использовать критерий Крамера – Уэлча, основанный на статистике

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s_x^2}{m} + \dfrac{s_y^2}{n}}} = \frac{\sqrt{mn}(\bar{x} - \bar{y})}{\sqrt{ns_x^2 + ms_y^2}}. \qquad (6)$$

Критерий Крамера – Уэлча имеет прозрачный смысл — разность выборочных средних арифметических для двух выборок делится на естественную оценку среднего квадратического отклонения этой разности. Естественность указанной оценки состоит в том, что неизвестные статистику дисперсии заменены их выборочными оценками. Из многомерной Центральной предельной теоремы и из теорем о наследовании сходимости [4, гл. 4] следует, что при росте объемов выборок распределение статистики T Крамера – Уэлча сходится к стандартному нормальному распределению с математическим ожиданием 0 и дисперсией 1. Эта сходимость установлена Г. Крамером [7]. Термин «статистика Крамера – Уэлча» введен нами в [8].

Итак, при справедливости H'_0 и больших объемах выборок распределение статистики T приближается к стандартному нормальному распределению $\Phi(x)$, из таблиц которого и следует брать критические значения. При $m = n$, как следует из формул (1) и (6), $t = T$. При $m \neq n$ этого равенства нет. В частности, при s_x^2 в формуле (1) стоит множитель $(m - 1)$, а в формуле (6) — множитель n.

Если $M(X) \neq M(Y)$, то при больших объемах выборок

$$P(T \leq \mathrm{X}) \approx \Phi(x - c_{mn}), \qquad (7)$$

где

$$c_{mn} = \frac{\sqrt{mn}[M(X) - M(Y)]}{\sqrt{nD(X) + mD(Y)}}. \qquad (8)$$

При $m = n$ или $D(X) = D(Y)$, согласно формулам (3) и (8), $a_{mn} = c_{mn}$, в остальных случаях равенства нет.

Из асимптотической нормальности статистики T, формул (7) и (8) следует, что правило принятия решения для критерия Крамера – Уэлча выглядит так:

если $|T| \leq \Phi^{-1}(1 - \alpha/2)$, то гипотеза однородности (равенства) математических ожиданий принимается на уровне значимости α;

если $|T| > \Phi^{-1}(1 - \alpha/2)$, то гипотеза однородности (равенства) математических ожиданий отклоняется на уровне значимости α.

В прикладной статистике наиболее часто применяется уровень значимости $\alpha = 0{,}05$. Тогда значение модуля статистики T Крамера – Уэлча надо сравнивать с граничным значением $\Phi^{-1}(1 - \alpha/2) = 1{,}96$.

Из сказанного выше следует, что при анализе организационно-экономических данных более обосновано применение критерия Крамера – Уэлча, чем критерия Стьюдента. Дополнительное преимущество критерия Крамера – Уэлча по сравнению с критерием Стьюдента — не требуется равенства дисперсий $D(X) = D(Y)$. Распределение статистики T не является распределением Стьюдента, однако и распределение статистики t, как показано выше, не является таковым в реальных ситуациях.

Распределение статистики T при объемах выборок $m = n = 6, 8, 10, 12$ и различных функциях распределений выборок $F(x)$ и $G(x)$ изучено методом статистических испытаний (Монте-Карло). Результаты (частично опубликованы в статьях [9, 10]) показывают, что даже при таких небольших объемах выборок точность аппроксимации предельным стандартным нормальным распределением вполне удовлетворительна. Поэтому представляется целесообразным заменить критерий Стьюдента, предлагаемый к использованию в устаревших литературных источниках, на критерий Крамера – Уэлча. Конечно, такая замена потребует переработки ряда нормативно-технических и методических документов, исправления учебников и учебных пособий для вузов.

Непараметрические методы проверки однородности

В большинстве управленческих, технических, экономических, медицинских и иных задач анализа данных в прикладных исследованиях представляет интерес не проверка равенства математических ожиданий или иных характеристик распределения, а обнаружение различия генеральных совокупностей, из которых извлечены выборки, т.е. проверка гипотезы H_0. Методы проверки гипотезы H_0 должны позволять обнаружить не только изменение математического ожидания или дисперсии, но и любые иные изменения функции распределения результатов наблюдений при переходе от одной выборки к другой (увеличение разброса, появление асимметрии и т.д.). Как установлено выше, методы, основанные на использовании статистик t Стьюдента и T Крамера – Уэлча, не позволяют проверять гипотезу H_0. Априорное предположение о принадлежности функций распределения $F(x)$ и $G(x)$ к какому-либо определенному параметрическому семейству (например, семействам нормальных, логарифмически нормальных, распределений Вейбулла – Гнеденко, гамма-распределений и др.) обычно нельзя достаточно надежно обосновать [3, 4]. Поэтому для проверки H_0 следует использовать методы, пригодные при любом виде $F(x)$ и $G(x)$, т.е. непараметрические методы.

Термин «непараметрический метод» означает, что при использовании этого метода нет необходимости предполагать, что функции распределения результатов наблюдений принадлежат каким-либо определенным параметрическим семействам. Современное представление о методах непараметрической статистики дано в статье [11].

Для проверки гипотезы H_0 разработано много непараметрических методов — критерии Смирнова, типа омега-квадрат (Лемана – Розенблатта), Вилкоксона (Манна – Уитни), Ван-дер-Вардена, Сэвиджа, хи-квадрат и др. [2, 12, 13]. Распределения статистик всех этих критериев при справедливости H_0 не зависят от конкретного вида совпадающих функций распределения

$F(x) \neq G(x)$. Следовательно, таблицами точных и предельных (при больших объемах выборок) распределений статистик этих критериев и их процентных точек [2, 13] можно пользоваться при любых непрерывных функциях распределения результатов наблюдений.

Какой из непараметрических критериев применять? Как известно [8], для выбора одного из нескольких критериев необходимо сравнить их мощности, определяемые видом альтернативных гипотез. Сравнению мощностей критериев посвящена обширная литература.

Хорошо изучены свойства критериев при альтернативной гипотезе сдвига

$$H_{1c}: G(x) = F(x - d),\ d \neq 0.$$

Критерии Вилкоксона, Ван-дер-Вардена и ряд других ориентированы для применения именно в этой ситуации. Если m раз измеряют характеристику одного объекта и n раз — другого, а функция распределения погрешностей измерения произвольна, но не меняется при переходе от объекта к объекту (это более жесткое требование, чем условие равенства дисперсий), то рассмотрение гипотезы H_{1c} оправдано. Однако в большинстве прикладных исследований нет оснований считать, что функции распределения, соответствующие выборкам, различаются только сдвигом.

Полагаем, что ради адекватности математической модели практической ситуации в качестве альтернативной гипотезы надо рассматривать H_1 (отсутствие абсолютной однородности), а не H_{1c} (наличие сдвига).

Двухвыборочный критерий Вилкоксона. Этот критерий (в литературе его называют также критерием Манна – Уитни), как показано в [14], предназначен для проверки гипотезы

$$H_{0m}: P(X < Y) = 1/2,$$

где X — случайная величина, распределенная как элементы первой выборки, а Y — случайная величина, распределенная как элементы второй выборки. Альтернативой является отрицание H_{0m}: вероятность $P(X < Y)$ отлична от 0,5. Это — непараметрическая гипотеза. Но из нее не следует, что функции распределения двух выборок совпадают. Обратное, конечно, верно: если X и Y одинаково распределены, то $P(X < Y) = 1/2$, т.е. медиана распределения разности $X - Y$ равна 0.

Критерий Вилкоксона — один из самых известных инструментов непараметрической статистики (наряду с критериями на основе статистик типа Колмогорова – Смирнова, омега-квадрат и коэффициентами ранговой корреляции). Свойствам этого критерия и таблицам его критических значений уделяется место во многих монографиях и статьях по математической и прикладной статистике (см., например, [2, 4, 12, 13]). Однако в литературе имеются и неточные утверждения относительно возможностей критерия Вилкоксона. Так, отдельные авторы полагают, что с его помощью можно обнаружить любое различие между функциями распределения $F(x)$ и $G(x)$. По мнению других, этот критерий нацелен на проверку равенства медиан распределений, соответствующих выборкам. И то и другое неверно, как подробно показано в [14].

При альтернативной гипотезе, когда функции распределения выборок $F(x)$ и $G(x)$ не совпадают, распределение статистики Вилкоксона зависит от величины $a = P(X < Y)$. Если a отличается от 1/2, то мощность критерия Вилкоксона стремится к 1 и он отличает нулевую гипотезу $F \equiv G$ от альтернативной. Если же $a = 1/2$, то это не всегда имеет место. При справедливости альтернативной гипотезы сдвига мощность стремится к 1, различие медиан также всегда обнаруживается.

Состоятельные критерии проверки абсолютной однородности двух независимых выборок

Естественно потребовать, чтобы рекомендуемый для массового использования в прикладных исследованиях критерий однородности был состоятельным. Это значит, что для любых отличных друг от друга функций распределения $F(x)$ и $G(x)$ (другими словами, при справедливости альтернативной гипотезы H_1) вероятность отклонения гипотезы H_0 должна стремиться к 1 при увеличении объемов выборок m и n. Из перечисленных выше критериев однородности состоятельными являются только критерии Смирнова и типа омега-квадрат.

Критерий Смирнова однородности двух независимых выборок. Он был предложен членом-корреспондентом АН СССР Н. В. Смирновым в 1939 г. (см. справочник [2]). Единственное ограничение — функции распределения $F(x)$ и $G(x)$ должны быть непрерывными. Согласно Л. Н. Большеву и Н. В. Смирнову [2] значение эмпирической функции распределения в точке x равно доле результатов наблюдений в выборке, меньших x. Критерий Смирнова основан на использовании эмпирических функций распределения $F_m(x)$ и $G_n(x)$, построенных по первой и второй выборкам соответственно. Значение статистики Смирнова

$$D_{m,n} = \sup_x |F_m(x) - G_n(x)|$$

сравнивают с соответствующим критическим значением (см., например, [2]) и по результатам сравнения принимают или отклоняют гипотезу H_0 о совпадении (однородности) функций распределения. Практически статистику $D_{m,n}$ рекомендуется [2] вычислять по формулам

$$D^{+}_{m,n} = \max_{1\le r\le n}\left[\frac{r}{n} - F_m(y'_r)\right] = \max_{1\le s\le m}\left[G_n(x'_s) - \frac{s-1}{m}\right], \quad (9)$$

$$D^{-}_{m,n} = \max_{1\le r\le n}\left[F_m(y'_r) - \frac{r-1}{n}\right] = \max_{1\le s\le m}\left[\frac{s}{m} - G_n(x'_s)\right], \quad (10)$$

$$D_{m,n} = \max(D^{+}_{m,n}, D^{-}_{m,n}), \quad (11)$$

где $x'_1 < x'_2 < ... < x'_m$ — элементы первой выборки $x_1, x_2, ..., x_m$, переставленные в порядке возрастания, а $y'_1 < y'_2 < ... < y'_n$ — элементы второй выборки $y_1, y_2, ..., y_n$, также переставленные в порядке возрастания. Поскольку функции распределения $F(x)$ и $G(x)$ предполагаются непрерывными, то вероятность совпадения каких-либо выборочных значений равна 0. Рекомендации по нахождению численного значения статистики Смирнова на основе выборочных данных приведены в статье [15].

Разработаны алгоритмы и компьютерные программы, позволяющие рассчитывать точные распределения, процентные точки и достигаемый уровень значимости для двухвыборочной статистики Смирнова, составлены подробные таблицы (см., например, методику [16], содержащую описание алгоритмов, тексты программ и таблицы критических значений).

Однако у критерия Смирнова есть и недостатки. Его распределение сосредоточено в сравнительно небольшом числе точек. Ясно, что принимаемые этой статистикой значения пропорциональны величине $1/L$, где L — наименьшее общее кратное объемов выборок m и n. Поэтому функция распределения растет большими скачками. Как следствие, не удается выдержать заданный уровень значимости. Реальный (другими словами, истинный) уровень значимости может значительно, даже в несколько раз отличаться от номинального (подробному обсуждению неклассического феномена существенного отличия реальных уровней значимости непараметрических критериев от номинальных посвящена работа [9]).

При больших объемах выборок можно воспользоваться доказанной Н. В. Смирновым в 1939 г. теоремой: в случае совпадения непрерывных функций распределения элементов двух независимых выборок

$$\lim_{m\to\infty, n\to\infty} P\left\{\sqrt{\frac{mn}{m+n}} D_{m,n} < y\right\} = K(y),$$

где $K(y)$ — функция распределения Колмогорова, заданная формулой

$$K(y) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2y^2).$$

Критерий типа омега-квадрат (Лемана – Розенблатта). Статистика критерия типа омега-квадрат для проверки однородности двух независимых выборок имеет вид

$$A = \frac{mn}{m+n}\int_{-\infty}^{\infty}(F_m(x) - G_n(x))^2\,\mathrm{d}H_{m+n}(x), \quad (12)$$

где $H_{m+n}(x)$ — эмпирическая функция распределения, построенная по объединенной выборке. Легко видеть, что

$$H_{m+n}(x) = \frac{m}{m+n}F_m(x) + \frac{n}{m+n}G_n(x).$$

Статистика A типа омега-квадрат была предложена Э. Леманом в 1951 г., изучена М. Розенблаттом в 1952 г., а затем и другими исследователями. Она зависит лишь от рангов элементов двух выборок в объединенной выборке. Данная статистика представляется в виде (см., например, [2])

$$A = \frac{1}{mn(m+n)}\left[m\sum_{i=1}^{m}(r_i - i)^2 + n\sum_{j=1}^{n}(s_j - j)^2\right] - \frac{4mn-1}{6(m+n)},$$

где r_i — ранг x'_i и s_j — ранг y'_j в общем вариационном ряду, построенном по объединенной выборке.

Правила принятия решений при проверке однородности двух выборок на основе статистик Смирнова и типа омега-квадрат, т.е. таблицы критических значений в зависимости от уровней значимости и объемов значимости, приведены, например, в [2]. При достаточно больших объемах выборок правило принятия решения формулируется просто: если наблюдаемое значение статистики меньше соответствующего квантиля предельного распределения, гипотеза однородности принимается, в противном случае — отклоняется.

Известно [4], что

$$\lim_{m\to\infty, n\to\infty} P(A < x) = a_1(x)$$

(в обозначениях [2]), где $a_1(x)$ — предельная функция распределения классической статистики омега-квадрат (Крамера – Мизеса – Смирнова), используемой для проверки согласия эмпирического распределения с заданным теоретическим.

Рекомендации по выбору критерия однородности. Для критерия типа омега-квадрат (Лемана – Розенблатта) нет выраженного эффекта различия между номинальными и реальными уровнями значимости. Поэтому мы рекомендуем для проверки однородности функций распределения (гипотеза H_0) применять статистику A типа омега-квадрат. Если методическое, табличное или программное обеспечение для статистики Лемана – Розенблатта отсутствует, рекомендуем использовать критерий Смирнова. Для проверки однородности математических ожиданий (гипотеза H'_0) целесообразно применять критерий Крамера – Уэлча. По нашему мнению, статистики Стьюдента, Вилкоксона и др. допустимо использовать лишь в отдельных частных случаях, рассмотренных выше.

Проблемы разработки и применения непараметрических критериев

Обсудим четыре проблемы, связанные с разработкой и применением непараметрических критериев.

О критериях Колмогорова и Смирнова. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат, разработанные в 1930-е годы, являются основой непараметрической математической статистики. Они продолжают быть образцом при разработке новых методов и широко применяются на практике. Используются термины «статистики типа Колмогорова – Смирнова», «статистики типа омега-квадрат». В литературе и в описании программных продуктов встречается некорректный термин «статистика Колмогорова – Смирнова». Иногда некорректно пишут, например, о применении критерия Колмогорова для проверки согласия с семейством нормальных распределений. В последнем случае речь идет отнюдь не только о терминологии — случаются грубые ошибки при применении непараметрических критериев математической статистики, приводящие к неправильным управленческим решениям [17].

В литературе, особенно переводной, иногда используют термин «критерий Колмогорова – Смирнова» по отношению к процедурам проверки непараметрических статистических гипотез, в частности, проверки однородности двух независимых выборок. Однако анализ публикаций академика А. Н. Колмогорова (1903 – 1987 гг.) и члена-корреспондента АН СССР Н. В. Смирнова (1900 – 1966 гг.), проведенный в статье [17], свидетельствует о том, что такого критерия не существует. У А. Н. Колмогорова и Н. В. Смирнова нет совместных работ, они никогда не изучали одновременно один и тот же статистический критерий. Право на существование имеет лишь термин «критерий типа Колмогорова – Смирнова». Так, критерий проверки однородности двух независимых выборок, основанный на формулах (9) – (11), — это критерий Смирнова. Можно сказать, что это «критерий типа Колмогорова – Смирнова». Критерий (12) — это критерий Лемана – Розенблатта. Можно сказать, что это «критерий типа омега-квадрат».

Предельная теория непараметрических статистик. Предельные (при безграничном росте объемов выборок) распределения непараметрических статистик, предназначенных для проверки однородности двух независимых выборок, находились различными методами. Один из наиболее общих — «принцип инвариантности» [18]. Нами разработан метод аппроксимации ступенчатыми функциями, с его помощью получен ряд необходимых и достаточных условий [19].

Анализ совпадающих наблюдений. Обычно предполагается, что функции распределения $F(x)$ и $G(x)$ непрерывны и строго возрастают. Из непрерывности этих функций следует, что с вероятностью 1 все $m + n$ результатов наблюдений различны. При рассмотрении реальных статистических данных иногда наблюдаются совпадения результатов наблюдений, но сам факт их наличия — свидетельство нарушений предпосылок только что описанной базовой математической модели. В статье [20] мы предлагаем вероятностно-статистическую модель, объясняющую появление совпадений и дающую алгоритмы анализа совпадений. Эта модель основана на предположении о появлении совпадений данных в результате «слипания» мало различающихся результатов наблюдений. Мы предлагаем добавить малую поправку к каждому элементу совпадающей группы результатов наблюдений и в результате получить выборку без совпадений, для которой рассчитать значение ранговой статистики. Рассмотрев различные варианты поправок, получаем «облако» значений ранговой статистики. Анализ этого «облака» позволяет получить статистические выводы.

Проверка однородности связанных выборок. Не всегда можно применять методы проверки однородности двух независимых выборок объемов m и n соответственно. Если проведено $(m + n)$ измерений объемов продаж в $(m + n)$ торговых точках, то эта модель, как правило, адекватна. Если же, например, x_i и y_i — объемы продаж одного и того же товара до и после определенного рекламного воздействия, то рассматриваемую модель применять нельзя (при этом $m = n$), поскольку очевидно, что анализируемые объемы продаж определяются не только и не столько рекламным воздействием, сколько особенностями конкретной торговой точки (ее расположением, продолжительностью работы, репутацией и т.д.). В подобных случаях используют модель связанных выборок. В ней обычно строят новую выборку $z_i = x_i - y_i$ и используют статистические методы анализа одной выборки, а не двух. Методы проверки однородности для связанных выборок рассмотрены в [21]. Они во многом аналогичны соответствующим методам проверки однородности двух независимых выборок.

Заключение

В статье приведены основные научные результаты, касающиеся методов проверки однородности двух независимых выборок. Они позволяют системно анализировать многообразие таких методов в целях выбора наиболее адекватного для обработки конкретных данных. Подробности и примеры можно найти в литературных источниках, большинство из которых легко доступны.

ЛИТЕРАТУРА

1. **Student.** The probable error of a mean / Biometrika. 1908. N 6(1). P. 1 – 25.
2. **Большев Л. Н., Смирнов Н. В.** Таблицы математической статистики. — М.: Наука, 1983. — 416 с.
3. **Орлов А. И.** Распределения реальных статистических данных не являются нормальными / Научный журнал КубГАУ. 2016. № 117. С. 71 – 90.
4. **Орлов А. И.** Прикладная статистика. — М.: Экзамен, 2006. — 671 с.
5. **Орлов А. И.** Проверка статистической гипотезы однородности математических ожиданий двух независимых выборок: критерий Крамера – Уэлча вместо критерия Стьюдента / Научный журнал КубГАУ. 2015. № 110. С. 197 – 218.
6. **Боровков А. А.** Математическая статистика. Изд. 4-е. — М.: Лань, 2010. — 704 с.
7. **Крамер Г.** Математические методы статистики. Изд. 3-е стереотипн. — М. – Ижевск: НИЦ «Регулярная и хаотическая динамика», 2003. — 648 с.
8. **Орлов А. И.** О применении статистических методов в медико-биологических исследованиях / Вестник Академии медицинских наук СССР. 1987. № 2. С. 88 – 94.
9. **Орлов А. И.** Реальные и номинальные уровни значимости при проверке статистических гипотез / Научный журнал КубГАУ. 2015. № 114. С. 42 – 54.
10. **Орлов А. И.** Метод статистических испытаний в прикладной статистике / Заводская лаборатория. Диагностика материалов. 2019. Т. 85. № 5. С. 67 – 79.
11. **Орлов А. И.** Современное состояние непараметрической статистики / Научный журнал КубГАУ. 2015. № 106. С. 239 – 269.
12. **Гаек Я., Шидак З.** Теория ранговых критериев / Пер. с англ. — М.: Наука, 1971. — 376 с.
13. **Холлендер М., Вульф Д.** Непараметрические методы статистики. — М.: Финансы и статистика, 1983. — 518 с.
14. **Орлов А. И.** Двухвыборочный критерий Вилкоксона — анализ двух мифов / Научный журнал КубГАУ. 2014. № 104. С. 91 – 111.
15. **Орлов А. И.** Состоятельные критерии проверки абсолютной однородности независимых выборок / Заводская лаборатория. Диагностика материалов. 2012. Т. 78. № 11. С. 66 – 70.
16. Методика. Проверка однородности двух выборок параметров продукции при оценке ее технического уровня и качества / Орлов А. И., Миронова Н. Г., Фомин В. Н., Черномордик О. М. — М.: ВНИИСтандартизации, 1987. — 116 с.
17. **Орлов А. И.** Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении / Научный журнал КубГАУ. 2014. № 97. С. 32 – 45.
18. **Биллингсли П.** Сходимость вероятностных мер. — М.: Наука, 1977. — 353 с.
19. **Орлов А. И.** Предельная теория непараметрических статистик / Научный журнал КубГАУ. 2014. № 100. С. 31 – 52.
20. **Орлов А. И.** Модель анализа совпадений при расчете непараметрических ранговых статистик / Заводская лаборатория. Диагностика материалов. 2017. Т. 83. № 11. С. 66 – 72.
21. **Орлов А. И.** О проверке однородности связанных выборок / Научный журнал КубГАУ. 2016. № 123. С. 708 – 726.

REFERENCES

1. **Student.** The probable error of a mean / Biometrika. 1908. N 6(1). P. 1 – 25.
2. **Bol'shev L. N., Smirnov N. V.** Tables of mathematical statistics. — Moscow: Nauka, 1983. — 416 p. [in Russian].
3. **Orlov A. I.** Distributions of real statistical data are not normal / Nauch. Zh. KubGAU. 2016. N 117. P. 71 – 90 [in Russian].
4. **Orlov A. I.** Applied statistics. — Moscow: Ékzamen, 2006. — 671 p. [in Russian].
5. **Orlov A. I.** Statistical hypothesis testing of homogeneity of mathematical expectations of two independent samples: the Cramer – Welch instead of the Student's test / Nauch. Zh. KubGAU. 2015. N 110. P. 197 – 218 [in Russian].
6. **Borovkov A. A.** Mathematical statistics. 4th edition. — Moscow: Lan', 2010. — 704 p. [in Russian].
7. **Kramer G.** Mathematical methods of statistics. 3rd edition. — Moscow – Izhevsk: NIC "Regulyarnaya i khaoticheskaya dinamika", 2003. — 648 p. [in Russian].
8. **Orlov A. I.** On the application of statistical methods in biomedical research / Vestn. AMN SSSR. 1987. N 2. P. 88 – 94 [in Russian].
9. **Orlov A. I.** Real and nominal significance levels in statistical hypothesis testing / Nauch. Zh. KubGAU. 2015. N 114. P. 42 – 54 [in Russian].
10. **Orlov A. I.** Statistical simulations method in applied statistics / Zavod. Lab. Diagn. Mater. 2019. Vol. 85. N 5. P. 67 – 79 [in Russian].
11. **Orlov A. I.** Current status of nonparametric statistics / Nauch. Zh. KubGAU. 2015. N 106. P. 239 – 269 [in Russian].
12. **Gaek Ya., Shidak Z.** Theory of rank tests. — Moscow: Nauka, 1971. — 376 p. [Russian translation].
13. **Khollender M., Vul'f D.** Nonparametric statistical methods. — Moscow: Finansy i statistika, 1983. —- 518 p. [in Russian].
14. **Orlov A. I.** Two-sample Wilcoxon test — analysis of two myths / Nauch. Zh. KubGAU. 2014. N 104. P. 91 – 111 [in Russian].
15. **Orlov A. I.** Consistent tests of absolute homogeneity of independent samplings / Zavod. Lab. Diagn. Mater. 2012. Vol. 78. N 11. P. 66 – 70 [in Russian].
16. Guidelines. Testing the homogeneity of two samples of product parameters when estimating its technical level and quality / Orlov A. I., Mironova N. G., Fomin V. N., Chernomordik O. M. — Moscow: VNIIStandartizatsii, 1987. — 116 p.
17. **Orlov A. I.** Nonparametric goodness-of-fit Kolmogorov, Smirnov, omega-square tests and the errors in their application / Nauch. Zh. KubGAU. 2014. N 97. P. 32 – 45 [in Russian].
18. **Billingsli P.** Convergence of probability measures. — Moscow: Nauka, 1977. — 353 p. [in Russian].
19. **Orlov A. I.** Limit Theory of Nonparametric Statistics / Nauch. Zh. KubGAU. 2014. N 100. P. 31 – 52 [in Russian].
20. **Orlov A. I.** The Model of Coincidence Analysis in the Calculation of Nonparametric Rank Statistics / Zavod. Lab. Diagn. Mater. 2017. Vol. 83. N 11. S. 66 – 72 [in Russian].
21. **Orlov A. I.** Testing of homogeneity of paired samples / Nauch. Zh. KubGAU. 2016. N 123. P. 708 – 726 [in Russian].