УДК 519.2:303.732.4

01.00.00 Физико-математические науки

ПРОВЕРКА СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ ОДНОРОДНОСТИ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ ДВУХ НЕЗАВИСИМЫХ ВЫБОРОК: КРИТЕРИЙ КРАМЕРА-УЭЛЧА ВМЕСТО КРИТЕРИЯ СТЬЮДЕНТА

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор
РИНЦ SPIN-код: 4342-4994
Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru

В настоящее время в большинстве научных, технических и экономических исследований используются статистические методы, разработанные в основном в первой трети XX века. Они составляют содержание распространенных учебников. Однако, математическая статистика бурно развивалась и в последующие 60 с лишним лет. В ряде ситуаций назрела необходимость перехода от классических методов к современным. В качестве примера разобрана задача проверки однородности двух независимых выборок. Рассмотрены условия применимости традиционного метода проверки однородности, основанного на использовании статистики *t* Стьюдента, а также указаны более современные методы. Описана вероятностная модель порождения данных в задаче проверки однородности двух независимых выборок. В терминах этой модели понятие «однородности», т. е. «отсутствие различия», может быть формализовано различными способами. Наивысшая степень однородности достигается, если обе выборки взяты из одной и той же генеральной совокупности (абсолютная однородность). В некоторых случаях целесообразно проверять совпадение некоторых характеристик элементов выборок - математических ожиданий, медиан, дисперсий, коэффициентов вариации и др. (проверка однородности характеристик). Для проверки однородности математических ожиданий часто рекомендуют классический критерий Стьюдента. Предполагают, что выборки взяты из нормальных распределений с одинаковыми дисперсиями. Показано, что для научных технических и экономических данных предпосылки двухвыборочного критерия Стьюдента, как правило, не выполняются. Для проверки однородности математических ожиданий вместо критерия Стьюдента предлагается использовать критерий Крамера-Уэлча. Рассмотрены состоятельные непараметрические критерии Смирнова и Лемана-Розенблатта для проверки абсолютной однородности

UDC 519.2:303.732.4

Physics and mathematical sciences

STATISTICAL HYPOTHESIS TESTING OF HOMOGENEITY OF MATHEMATICAL EXPECTATIONS OF TWO INDEPENDENT SAMPLES: CRAMER-WELCH TEST INSTEAD OF t-TEST

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci., professor
Bauman Moscow State Technical University, Moscow, Russia

Currently, the majority of scientific, technical and economic studies use statistical methods developed mainly in the first third of the XX century. They constitute the content of common textbooks. However, mathematical statistics are rapidly developing in the next 60 years. In some situations there is a need of the transition from classical to modern methods. As an example, we discuss the problem of testing the homogeneity of two independent samples. We have considered the conditions of applicability of the traditional method of testing the homogeneity based on the use of Student's t-statistic, as well as more up-to-date methods. We describe a probabilistic model of generation of statistical data in the problem of testing the homogeneity of two independent samples. In terms of this model the concept of "homogeneity" ("no difference"), can be formalized in different ways. High degree of homogeneity is achieved if the two samples are taken from one and the same population (absolute homogeneity). In some cases it is advisable to testing the coincidence of some characteristics of the elements of the sample - mathematical expectations, medians, variances, coefficients of variation, and others (testing the homogeneity of characteristics). To test the homogeneity of mathematical expectations is often recommended classic t-test. It is believed that the samples taken from a normal distributions with equal variances. It is shown that for scientific, technical and economic data the preconditions of two-sample t-test usually are not performed. To test the homogeneity of mathematical expectations instead of t-test we have offered to use the Cramer-Welch test. We have considered the consistent nonparametric Smirnov and Lehmann-Rosenblatt tests for absolute homogeneity

Ключевые слова: МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, ПРИКЛАДНАЯ СТАТИСТИКА, СТАТИСТИЧЕСКИЕ МЕТОДЫ, НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА, ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ, ДВЕ НЕЗАВИСИМЫЕ ВЫБОРКИ, ПРОВЕРКА ОДНОРОДНОСТИ, ОДНОРОДНОСТЬ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ, КРИТЕРИЙ СТЬЮДЕНТА, КРИТЕРИЙ КРАМЕРА-УЭЛЧА, СОСТОЯТЕЛЬНЫЕ КРИТЕРИИ ОДНОРОДНОСТИ, КРИТЕРИЙ СМИРНОВА, КРИТЕРИЙ ЛЕМАНА-РОЗЕНБЛАТТА

Keywords: MATHEMATICAL STATISTICS, APPLIED STATISTICS, STATISTICAL METHODS, NONPARAMETRIC STATISTICS, STATISTICAL HYPOTHESIS TESTING, TWO INDEPENDENT SAMPLES, HOMOGENEITY TESTING, HOMOGENEITY OF MATHEMATICAL EXPECTATIONS, t-TEST, CRAMER-WELCH TEST, CONSISTENT TESTS OF HOMOGENEITY, SMIRNOV TEST, LEHMANN-ROSENBLATT TEST

1. Введение. Задача проверки однородности

Статистические методы применяются в большинстве научных и прикладных работ в области техники и технологии. Среди наиболее популярных - методы проверки однородности двух независимых выборок.

Началом современного этапа теории статистических методов - математической статистики - можно считать основание К. Пирсоном (K. Pearson) в 1900 г. журнала «Biometrika» [1]. В настоящее время в большинстве научных, технических и экономических исследований используются статистические методы, разработанные в основном в первой трети XX века. Объяснение простое – именно они составляют содержание распространенных учебников. Однако математическая статистика бурно развивалась и в последующие 60 с лишним лет [2, 3]. Кроме решения новых задач, изучались свойства традиционных статистических методов, предлагались новые методы для применения в классических постановках. В ряде ситуаций назрела необходимость перехода от классических методов к современным. В качестве примера, разбор которого составляет основное содержание настоящей статьи, рассмотрим задачу проверки однородности двух независимых выборок.

В математико-статистических терминах постановка задачи такова: имеются две независимые выборки x_1, x_2,...,x_m и y_1, y_2,...,y_n (т. е. наборы из m и n действительных чисел), требуется проверить их однородность. Термин «однородность» уточняется ниже.

Противоположным понятием для «однородности» является «различие». Можно переформулировать задачу: требуется проверить, есть ли различие между выборками. Если различия нет, то для дальнейшего изучения выборки часто объединяют.

2. Традиционный метод проверки однородности (критерий Стьюдента)

Для дальнейшего критического разбора опишем традиционный статистический метод проверки однородности. Вычисляют средние арифметические в каждой выборке

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

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, на основе которой принимают решение,

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

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

Рассмотрим условия применимости традиционного метода проверки однородности, основанного на использовании статистики t Стьюдента, а также укажем более современные методы.

3. Вероятностная модель порождения данных

Для обоснованного применения математико-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. Предполагается также, 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, входящих в выборку, могут быть установлены или исходя из методики проведения конкретных наблюдений, или путем проверки статистических гипотез независимости и одинаковой распределенности с помощью соответствующих критериев [4, 5].

Если проведено $(m + n)$ измерений линейных размеров деталей, то описанную выше модель, как правило, можно применять. Если же, например, x_i и y_i - результаты наблюдения одного и того же образца до и после определенного технологического воздействия, то рассматриваемую модель применять нельзя. (В этом случае используют модель т.н. связанных выборок, в которой обычно строят новую выборку $z_i = x_i - y_i$ и используют статистические методы анализа одной выборки, а не двух [6].)

При дальнейшем изложении принимаем описанную выше вероятностную модель двух независимых выборок.

4. Уточнения понятия однородности

Понятие «однородность», т. е. «отсутствие различия», может быть формализовано в терминах вероятностной модели различными способами.

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

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

Отсутствие однородности означает, что верна альтернативная гипотеза, согласно которой

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

хотя бы при одном значении аргумента x_0. Если гипотеза H_0 принята, то выборки можно объединить в одну, если нет - то нельзя. Описанная задача проверки статистических гипотез называется *задачей проверки абсолютной однородности*.

В некоторых случаях целесообразно проверять не совпадение функций распределения, а совпадение некоторых характеристик случайных величин X и Y - математических ожиданий, медиан, дисперсий, коэффициентов вариации и др. В таких случаях говорят о *проверке однородности* соответствующих *характеристик*. Например, *однородность математических ожиданий* означает, что справедлива гипотеза

$$H'_0 : M(X) = M(Y),$$

где $M(X)$ и $M(Y)$ - математические ожидания случайных величин X и Y, результаты наблюдений над которыми составляют первую и вторую выборки соответственно. Различие между выборками в рассматриваемом случае - это справедливость альтернативной гипотезы

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

Если гипотеза H_0 верна, то и гипотеза H'_0 верна, но из справедливости H'_0 не следует справедливость H_0. 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 концентрация SiO_2 в мартеновском шлаке определяется весовым (первая выборка) или фотоколориметрическим (вторая выборка) методами. Тогда важно проверить гипотезу об отсутствии систематических расхождений результатов весового и фотоколориметрического методов [4], т.е. гипотезу о равенстве математических ожиданий. Другой пример – из медицины труда. Пусть изучается эффективность лечения определенного профессионального заболевания двумя препаратами; результаты наблюдения - число дней нетрудоспособности, а показатель эффективности лечения - среднее число дней нетрудоспособности на одного больного. Тогда для сравнения эффективности препаратов достаточно проверить гипотезу H'_0.

5. Классические условия применимости критерия Стьюдента

Пусть выполнены два классических условия применимости критерия Стьюдента, основанного на использовании статистики t, заданной формулой (1):

а) результаты наблюдений имеют нормальные распределения:

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

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:

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

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

$$H''_0 : m_1 = m_2,$$

а *обе* альтернативные гипотезы H_1 и H'_1 сводятся к гипотезе

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

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 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.

6. О проверке условия нормальности

Априори нет оснований предполагать нормальность распределения результатов научных, технических или экономических наблюдений. Следовательно, нормальность надо проверять. Разработано много статистических критериев для проверки нормальности распределения результатов наблюдений [4]. 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. Так, 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 [7]. В большинстве научных, технических или экономических исследований число наблюдений существенно меньше.

Есть и одна общая (для всех видов измерений) причина отклонений от нормальности: любой результат наблюдения записывается конечным (обычно не более 2 - 5) количеством цифр, а с математической точки зрения вероятность такого события равна 0 для любой непрерывной функции распределения [8].

Из сказанного выше следует, что распределение результатов научных, технических или экономических наблюдений практически всегда более или менее отличается от нормального. Более подробно это утверждение обосновано в [9].

7. Последствия нарушения условия нормальности

Если условие а) не выполнено, то распределение статистики t не является распределением Стьюдента. Однако при справедливости гипотезы H'_0 и условия б) распределение 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 традиционный метод (критерий Стьюдента) можно использовать для проверки гипотезы H'_0 при больших объемах выборок. При этом вместо таблиц распределения Стьюдента достаточно пользоваться таблицами стандартного нормального распределения $\Phi(x)$.

Сформулированное в предыдущем абзаце утверждение справедливо для любых функций распределения $F(x)$ и $G(x)$ таких, что $M(X) = M(Y)$, $D(X) = D(Y)$ и выполнены некоторые внутриматематические условия, обычно считающиеся справедливыми в реальных задачах [10 -

12]. Если же $M(X) \neq M(Y)$, то при больших объемах выборок

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

где

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

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

8. О проверке условия равенства дисперсий

Иногда условие б) вытекает из методики получения результатов наблюдений, например, когда с помощью одного и того же прибора 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 основании 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, от которой неизбежны отклонения (см. выше), причем в отличие от t-критерия его распределение сильно меняется при малых отклонениях от нормальности [10, 13]. Кроме того, F-критерий отвергает гипотезу $D(X) = D(Y)$ лишь при большом различии выборочных дисперсий. Так, для данных [4] 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 на 1% уровне значимости. Следовательно, при проверке однородности применение F-критерия для предварительной проверки равенства дисперсий нецелесообразно.

К такому же выводу с точки зрения теории математической статистики пришли авторы статьи [14].

Итак, в большинстве научных, технических и экономических задач условие б) нельзя считать выполненным, а проверять его нецелесообразно.

9. Последствия нарушения условия равенства дисперсий

Если объемы выборок m и n велики, то распределение статистики t описывается с помощью только математических ожиданий $M(X)$ и $M(Y)$, дисперсий $D(X)$, $D(Y)$ и отношения объемов выборок:

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

где a_{mn} определено формулой (3),

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

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. Если объемы выборок мало различаются, то b_{mn} близко к 1.

Так, для данных [4] имеем $b^*_{mn}= 0{,}987$, где b^*_{mn} - оценка b_{mn}, полученная заменой в формуле (5) теоретических дисперсий на выборочные.

10. 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 научных, технических и экономических 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. При значительном различии объемов и дисперсий выборок распределение статистики Стьюдента значительно отличается от распределения Стьюдента. Вместе с тем влияние отклонений распределений элементов выборок от нормальности при росте объемов выборок уменьшается и в асимптотике исчезает.

В соответствии с выводами методологии статистических методов [8] предполагаем, что элементы выборок имеют математические ожидания и дисперсии. По нашей экспертной оценке, выход за пределы этого предположения имеет лишь внутриматематический интерес.

11. Критерий Крамера - Уэлча равенства математических ожиданий

Вместо критерия Стьюдента предлагаем для проверки H'_0 использовать критерий Крамера - Уэлча, основанный на статистике

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

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 [12, 15] вытекает, что при росте объемов выборок распределение статистики T сходится к стандартному нормальному распределению с математическим ожиданием 0 и дисперсией 1. Итак, при справедливости H'_0 и больших объемах выборок распределение статистики T приближается с помощью стандартного нормального распределения $\Phi(x)$, из таблиц которого предлагаем брать критические значения.

Критерий на основе статистики (6) извлечен нами из классической монографии Г. Крамера [16], признанной отечественной вероятностно-статистической научной школой базовым учебником по математической статистике. Критерием Крамера - Уэлча он назван нами в статьях [17, 18], в честь шведа Г. Крамера и англичанина Б.Л. Уэлча, который изучал распределение статистики T для выборок из нормальных распределений [19, 20]. Работы Уэлча продолжила В.И. Пагурова [21].

При $m = n$, как следует из формул (1) и (6), статистики критериев Стьюдента и Крамера - Уэлча совпадают, $t = T$. При $m \neq n$ этого равенства нет, статистики критериев Стьюдента и Крамера - Уэлча различны. В частности, под корнем в знаменателе перед выборочной дисперсией s_x^2 в (1) стоит множитель (m - 1), а в (6)- множитель n.

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

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

где

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

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

Из сказанного выше следует, что применение критерия Крамера - Уэлча не менее обосновано, чем применение критерия Стьюдента. Дополнительное преимущество - 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 нами совместно с Ю.Э. Камнем и Я.Э. Камнем методом статистических испытаний (Монте-Карло). Рассмотрены различные варианты функций распределения $F(x)$ и $G(x)$. Результаты показывают, что даже при таких небольших объемах выборок точность аппроксимации предельным стандартным нормальным распределением вполне удовлетворительна. Подробные результаты этого исследования будут опубликованы в одной из дальнейших статей.

Исходя из сказанного выше, представляется целесообразным во всех тех случаях, когда в настоящее время используется критерий Стьюдента, заменить его на критерий Крамера - Уэлча. Конечно, такая замена потребует переделки ряда нормативно-технических и методических документов, исправления учебников и учебных пособий для вузов.

12. Непараметрические методы проверки однородности

В большинстве научных, технических и экономических задач представляет интерес не проверка равенства математических ожиданий, а обнаружение различия генеральных совокупностей, из которых извлечены выборки, т.е. проверка гипотезы H_0. Методы проверки гипотезы H_0 позволяют обнаружить не только изменение математического ожидания, 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. Поэтому для проверки H_0 следует использовать методы, пригодные при любом виде $F(x)$ и $G(x)$, т.е. непараметрические методы [22]. (Термин «непараметрический метод» означает, что при использовании этого метода нет необходимости предполагать, что функции распределения результатов наблюдений принадлежат какому-либо определенному параметрическому семейству.)

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. [4, 5, 23]. Распределения статистик всех этих критериев при справедливости H_0 не зависят от конкретного вида совпадающих функций распределения $F(x) \equiv G(x)$. Следовательно, таблицами точных и предельных (при больших объемах выборок) распределений статистик этих критериев и их процентных точек [4, 23] можно пользоваться при любых непрерывных функциях распределения результатов наблюдений.

13. Каким из непараметрических критериев пользоваться?

Как известно [13], для выбора одного из нескольких критериев необходимо сравнить их мощности, определяемые видом альтернативных гипотез. Сравнению мощностей критериев посвящена обширная литература [24].

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 научных, технических и экономических исследований, в частности при анализе данных об измерениях концентрация SiO_2 в мартеновском шлаке весовым или фотоколориметрическим методами, нет оснований считать, что функции распределения, соответствующие выборкам, различаются только сдвигом.

В соответствии с теорией математической статистики естественно потребовать, чтобы рекомендуемый для массового использования в технических и технико-экономических исследованиях критерий однородности был состоятельным [13]. Это значит, что для любых отличных друг от друга функций распределения $F(x)$ и $G(x)$ (другими словами, при справедливости H_1) вероятность отклонения гипотезы H_0 должна стремиться стремилась к 1 при увеличении объемов выборок m и n. Из перечисленных выше критериев состоятельными являются только критерии Смирнова и омега-квадрат. В частности, критерий Вилкоксона не позволяет отвергнуть гипотезу H_0 для таких функций распределения $F(x)$ и $G(x)$, что

$$\int_{-\infty}^{\infty} F(x)dG(x) = 0.5\,. \qquad (9)$$

Если $F(x)$ и $G(x)$ не совпадают, но удовлетворяют соотношению (9), то при больших m и n гипотеза H_0 принимается столь же часто, как и в случае совпадающих $F(x) \equiv G(x)$. Подробнее это свойство критерия Вилкоксона разобрано в статьях [25, 26].

Проведенное нами совместно с Ю.Э.Камнем и Я.Э.Камнем исследование мощности (методом статистических испытаний) первых четырех из перечисленных выше критериев (при различных вариантах функций распределения $F(x)$ и $G(x)$ подтвердило преимущество критериев Смирнова и омега-квадрат и при объемах выборок 6 - 12.

14. Критерий Смирнова однородности двух выборок

Он предложен членом-корреспондентом АН СССР Н.В. Смирновым в 1939 г. [27]. Единственное ограничение - функции распределения $F(x)$ и $G(x)$ должны быть непрерывными. Напомним, что согласно Л.Н. Большеву и Н.В. Смирнову [4] 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 [4] и по результатам сравнения принимают или отклоняют гипотезу H_0. Практически значение статистики $D_{m,n}$ рекомендуется [4] вычислять по формулам

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

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

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

где $x'_1 < x'_2 < \ldots x'_m$ - элементы первой выборки $x_1, x_2, \ldots, x_m$, переставленные в порядке возрастания, а $y'_1 < y'_2 < \ldots y'_n$ - элементы второй выборки $y_1, y_2, \ldots, y_n$, также переставленные в порядке возрастания (вероятность совпадения каких-либо элементов выборок равна 0 в силу непрерывности функция распределения $F(x)$ и $G(x)$).

Разработаны алгоритмы и программы для ЭВМ, позволяющие рассчитывать точные распределения, процентные точки и достигаемый уровень значимости для статистики Смирнова [4, 28, 29], разработаны подробные таблицы [30].

Однако у критерия Смирнова есть и недостатки. Его распределение сосредоточено в сравнительно небольшом числе точек, поэтому функция распределения растет большими скачками. В результате не удается выдержать заданный уровень значимости, реальный уровень значимости может в несколько раз отличаться от номинального [31].

15. Критерий типа омега-квадрат (Лемана - Розенблатта)

Статистика критерия типа омега-квадрат (о термине см. [32]) для проверки однородности двух независимых выборок имеет вид:

$$A = \frac{mn}{m+n}\int_{-\infty}^{\infty}(F_m(x) - G_n(x))^2\, dH_{m+n}(x) ,$$

где $H_{m+n}(x)$ – эмпирическая функция распределения, построенная по объединенной выборке,

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

Статистика A типа омега-квадрат была предложена Э. Леманом в 1951 г. [33], изучена М. Розенблаттом в 1952 г. [34] и другими исследователями. Поэтому мы называем основанный на ней критерий однородности двух независимых выборок *критерием Лемана - Розенблатта*. Статистика A является непараметрической и ранговой [22], поскольку она зависит лишь от рангов элементов двух выборок в объединенной выборке:

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

где r_i - ранг x'_i и s_j - ранг y'_j в общем вариационном ряду, построенном по объединенной выборке [4].

16. Рекомендации по выбору критерия однородности

Состоятельные критерии Смирнова и Лемана - Розенблатта для проверки абсолютной однородности двух независимых выборок, включая алгоритмы расчетов, рассмотрены в [35].

В отличие от критерия Смирнова, для критерия типа омега-квадрат Лемана - Розенблатта нет выраженного эффекта различия между номинальными и реальными уровнями значимости [31]. 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.

17. О внедрении современных методов прикладной статистики в практику технических и технико-экономических исследований

Даже из проведенного выше разбора лишь одной из типичных статистических задач - задачи проверки однородности двух независимых выборок - можно сделать вывод о целесообразности организации работ по критическому анализу сложившейся в научных, технических и экономических исследованиях практики статистической обработки данных и по внедрению накопленного арсенала современных методов прикладной статистики [7]. По нашему мнению, широкого внедрения заслуживают, в частности, методы многомерного статистического анализа, планирования эксперимента, статистики объектов нечисловой природы. Очевидно, рассматриваемые работы должны быть плановыми, организационно оформленными, проводиться мощными самостоятельными организациями и подразделениями [36, 37]. Целесообразно создание службы статистических консультаций в системе научно-исследовательских учреждений и вузов технического и организационно-экономического профиля, идею которой в свое время пропагандировал В.В. Налимов [38].

Литература

1. Орлов А.И. Основные этапы становления статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 97. С. 73–85.
2. Орлов А.И. Современная прикладная статистика // Заводская лаборатория. Диагностика материалов. 1998. Т.64. № 3. С. 52–60.
3. Орлов А.И. Основные черты новой парадигмы математической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. № 90. С. 45–71.
4. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики / 3-е изд. – М.: Наука, 1983. – 474 с.
5. Гаек Я., Шидак З. Теория ранговых критериев / Пер. с англ. - М.: Наука, 1971. – 376 с.
6. Орлов А.И. Методы проверки однородности связанных выборок // Заводская лаборатория. Диагностика материалов. 2004. Т. 70. № 7. С. 57–61.
7. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.
8. Орлов А.И. О методологии статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 104. С. 53–80.
9. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. Диагностика материалов. 1991. Т.57. № 7. С. 64–66.
10. Шеффе Г. Дисперсионный анализ / Пер. с англ. / 2-е изд. - М.: Наука, 1980. - 512 с.
11. Елисеев В.Г. Асимптотическое разложение для распределения статистики критерия проверки гипотезы о равенстве средних // Прикладная статистика. Ученые записки по статистике, т.45. - М.: Наука, 1983. – С. 265–267.
12. Орлов А.И. Теоретические инструменты статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 101. С. 253 – 274.
13. Боровков А.А. Математическая статистика. - М.: Наука, 1984. – 472 с.
14. Славова В.В., Чибисов Д.М. О предварительном тестировании в задаче Беренса-Фишера // Обозрение прикладной и промышленной математики. 2007. Т.16. Вып.2. С. 210–225.
15. Орлов А.И. Устойчивость в социально-экономических моделях. – М.: Наука, 1979, - 296 с.

16. Крамер Г. Математические методы статистики / Пер. с англ. / 2-е изд. - М.: Мир, 1975. – 648 с.
17. Орлов А.И. О применении статистических методов в медико-биологических исследованиях // Вестник Академии медицинских наук СССР. 1987. № 2. С. 88–94.
18. Орлов А.И. О проверке однородности двух независимых выборок // Заводская лаборатория. Диагностика материалов. 2003. Т.69. № 1. С. 55–60.
19. Welch B.L. The significance of the difference between two means when the population variances are unequal // Biometrika. 1938. V.29. Pp. 350–362.
20. Welch B.L. The generalization of Student's problem when several different population variances are involved // Biometrika. 1947. V.34. Pp. 28–35.
21. Пагурова В.И. О сравнении средних значений в двух нормальных выборках // Теория вероятностей и ее применения. 1968. Т.13. № 3. С. 561–568.
22. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239–269.
23. Холлендер М., Вулф Д.А. Методы непараметрической статистики / Пер. с англ. - М.: Финансы и статистика, 1983. – 518 с.
24. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. – М.: Наука, 1995. – 238 с.
25. Орлов А.И. Какие гипотезы можно проверять с помощью двухвыборочного критерия Вилкоксона? // Заводская лаборатория. Диагностика материалов. 1999. Т.65. № 1. С. 51-55.
26. Орлов А.И. Двухвыборочный критерий Вилкоксона – анализ двух мифов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 104. С. 91 – 111.
27. Смирнов Н.В. Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках // Бюллетень МГУ. 1939. Т. 2. № 2. С.3-14.
28. Орлов А.И., Орловский И.В. Оценка остаточного члена порядка n^{-2} для функции распределения двухвыборочной статистики Смирнова // Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. – Пермь: Изд-во Пермского государственного университета, 1978. С.100-109.
29. Орлов А.И. Методы оценки близости допредельных и предельных распределений статистик // Заводская лаборатория. Диагностика материалов. 1998. Т.64. №5. С. 64-67.
30. Методика. Проверка однородности двух выборок параметров продукции при оценке ее технического уровня и качества / Орлов А.И., Фомин В.Н., Черномордик О.М. и др. – М.: ВНИИСтандартизации, 1987. – 116 с.
31. Камень Ю.Э., Камень Я.Э., Орлов А.И. Реальные и номинальные уровни значимости в задачах проверки статистических гипотез // Заводская лаборатория. Диагностика материалов. 1986. Т. 52. № 12. С. 55-57.
32. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 97. С. 31-45.
33. Lehmann E.L. Consistency and unbiasedness of certain nonparametric tests / Annals of Mathematical Statistics. 1951. V.22. № 1. Pp.165-179.
34. Rosenblatt M. Limit theorems associated with variants of the von Mises statistic // Annals of Mathematical Statistics. 1952. V.23. № 4. Pp. 617-623.
35. Орлов А.И. Состоятельные критерии проверки абсолютной однородности независимых выборок // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №11. С. 66-70.

36. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов. // Заводская лаборатория. Диагностика материалов. 1992. Т.58. №1. С. 67-74.
37. Орлов А.И. Проблемы внедрения математических и инструментальных методов контроллинга // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 107. С. 1017 – 1048.
38. Налимов В.В. Теория эксперимента. - М.: Наука, 1971. – 208 с.

References

1. Orlov A.I. Osnovnye jetapy stanovlenija statisticheskih metodov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 97. S. 73–85.
2. Orlov A.I. Sovremennaja prikladnaja statistika // Zavodskaja laboratorija. Diagnostika materialov. 1998. T.64. № 3. S. 52–60.
3. Orlov A.I. Osnovnye cherty novoj paradigmy matematicheskoj statistiki // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. № 90. S. 45–71.
4. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoj statistiki / 3-e izd. – M.: Nauka, 1983. – 474 s.
5. Gaek Ja., Shidak 3. Teorija rangovyh kriteriev / Per. s angl. - M.: Nauka, 1971. – 376 s.
6. Orlov A.I. Metody proverki odnorodnosti svjazannyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. 2004. T. 70. № 7. S. 57–61.
7. Orlov A.I. Prikladnaja statistika. - M.: Jekzamen, 2006. - 671 s.
8. Orlov A.I. O metodologii statisticheskih metodov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 104. S. 53–80.
9. Orlov A.I. Chasto li raspredelenie rezul'tatov nabljudenij javljaetsja normal'nym? // Zavodskaja laboratorija. Diagnostika materialov. 1991. T.57. № 7. S. 64–66.
10. Sheffe G. Dispersionnyj analiz / Per. s angl. / 2-e izd. - M.: Nauka, 1980. - 512 s.
11. Eliseev V.G. Asimptoticheskoe razlozhenie dlja raspredelenija statistiki kriterija proverki gipotezy o ravenstve srednih // Prikladnaja statistika. Uchenye zapiski po statistike, t.45. - M.: Nauka, 1983. – S. 265–267.
12. Orlov A.I. Teoreticheskie instrumenty statisticheskih metodov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 101. S. 253 – 274.
13. Borovkov A.A. Matematicheskaja statistika. - M.: Nauka, 1984. – 472 s.
14. Slavova V.V., Chibisov D.M. O predvaritel'nom testirovanii v zadache Berensa-Fishera // Obozrenie prikladnoj i promyshlennoj matematiki. 2007. T.16. Vyp.2. S. 210–225.
15. Orlov A.I. Ustojchivost' v social'no-jekonomicheskih modeljah. – M.: Nauka, 1979, - 296 s.
16. Kramer G. Matematicheskie metody statistiki / Per. s angl. / 2-e izd. - M.: Mir, 1975. – 648 s.
17. Orlov A.I. O primenenii statisticheskih metodov v mediko-biologicheskih issledovanijah // Vestnik Akademii medicinskih nauk SSSR. 1987. № 2. S. 88–94.
18. Orlov A.I. O proverke odnorodnosti dvuh nezavisimyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. № 1. S. 55–60.
19. We1sh V.L. The significance of the difference between two means when the population variances are unequal // Biometrika. 1938. V.29. Pp. 350–362.

20. We1sh V.L. The generalization of Student's problem when several different population variances are involved // Biometrika. 1947. V.34. Pp. 28–35.
21. Pagurova V.I. O sravnenii srednih znachenij v dvuh normal'nyh vyborkah // Teorija verojatnostej i ee primenenija. 1968. T.13. № 3. S. 561–568.
22. Orlov A.I. Sovremennoe sostojanie neparametricheskoj statistiki // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 106. S. 239–269.
23. Hollender M., Vulf D.A. Metody neparametricheskoj statistiki / Per. s angl. - M.: Finansy i statistika, 1983. – 518 s.
24. Nikitin Ja.Ju. Asimptoticheskaja jeffektivnost' neparametricheskih kriteriev. – M.: Nauka, 1995. – 238 s.
25. Orlov A.I. Kakie gipotezy mozhno proverjat' s pomoshh'ju dvuhvyborochnogo kriterija Vilkoksona? // Zavodskaja laboratorija. Diagnostika materialov. 1999. T.65. № 1. S. 51-55.
26. Orlov A.I. Dvuhvyborochnyj kriterij Vilkoksona – analiz dvuh mifov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 104. S. 91 – 111.
27. Smirnov N.V. Ocenka rashozhdenija mezhdu jempiricheskimi krivymi raspredelenija v dvuh nezavisimyh vyborkah // Bjulleten' MGU. 1939. T. 2. № 2. S.3-14.
28. Orlov A.I., Orlovskij I.V. Ocenka ostatochnogo chlena porjadka n-2 dlja funkcii raspredelenija dvuhvyborochnoj statistiki Smirnova // Statisticheskie metody ocenivanija i proverki gipotez. Mezhvuzovskij sbornik nauchnyh trudov. – Perm': Izd-vo Permskogo gosudarstvennogo universiteta, 1978. S.100-109.
29. Orlov A.I. Metody ocenki blizosti dopredel'nyh i predel'nyh raspredelenij statistik // Zavodskaja laboratorija. Diagnostika materialov. 1998. T.64. №5. S. 64-67.
30. Metodika. Proverka odnorodnosti dvuh vyborok parametrov produkcii pri ocenke ee tehnicheskogo urovnja i kachestva / Orlov A.I., Fomin V.N., Chernomordik O.M. i dr. – M.: VNIIStandartizacii, 1987. – 116 s.
31. Kamen' Ju.Je., Kamen' Ja.Je., Orlov A.I. Real'nye i nominal'nye urovni znachimosti v zadachah proverki statisticheskih gipotez // Zavodskaja laboratorija. Diagnostika materialov. 1986. T. 52. № 12. S. 55-57.
32. Orlov A.I. Neparametricheskie kriterii soglasija Kolmogorova, Smirnova, omega-kvadrat i oshibki pri ih primenenii // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 97. S. 31-45.
33. Lehmann E.L. Consistency and unbiasedness of certain nonparametric tests / Annals of Mathematical Statistics. 1951. V.22. № 1. Pp.165-179.
34. Rosenblatt M. Limit theorems associated with variants of the von Mises statistic // Annals of Mathematical Statistics. 1952. V.23. № 4. Pp. 617-623.
35. Orlov A.I. Sostojatel'nye kriterii proverki absoljutnoj odnorodnosti nezavisimyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. 2012. T.78. №11. S. 66-70.
36. Orlov A.I. O sovremennyh problemah vnedrenija prikladnoj statistiki i drugih statisticheskih metodov. // Zavodskaja laboratorija. Diagnostika materialov. 1992. T.58. №1. S. 67-74.
37. Orlov A.I. Problemy vnedrenija matematicheskih i instrumental'nyh metodov kontrollinga // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 107. S. 1017 – 1048.
38. Nalimov V.V. Teorija jeksperimenta. - M.: Nauka, 1971. – 208 s.