УДК 519.2:303.732.4

01.00.00 Физико-математические науки

РЕАЛЬНЫЕ И НОМИНАЛЬНЫЕ УРОВНИ ЗНАЧИМОСТИ ПРИ ПРОВЕРКЕ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

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

При проверке статистических гипотез критические значения часто указывают для априорно фиксированных (номинальных) уровней значимости. В качестве таковых, обычно используются значения из тройки чисел 0,01, 0,05, 0,1, к которым иногда добавляют еще несколько: 0,001, 0,005, 0,02 и др. Однако, для статистик с дискретными функциями распределения, к которым, в частности, относятся все непараметрические статистические критерии, реальные уровни значимости могут и не совпадать с номинальными, отличаться в разы. Под реальным уровнем значимости понимается максимально возможный уровень значимости дискретной статистики, не превосходящий заданный номинальный уровень значимости (т.е при переходе к следующему по величине возможному значению дискретной статистики соответствующий уровень значимости оказывается больше заданного номинального). В статье рассмотрено различие между номинальными и реальными уровнями значимости на примере непараметрических критериев проверки однородности двух независимых выборок. Изучены двухвыборочный критерий Вилкоксона, критерий Ван-дер-Вардена, двухвыборочный двухсторонний критерий Смирнова, критерий знаков, критерий серий (Вольфовица). Рассчитаны реальные уровни значимости этих критериев для номинального уровня значимости 0,05. Проведено изучение мощности перечисленных статистических критериев методом Монте-Карло. Основной вывод: при малых объемах выборок использовать номинальные уровни значимости вместо реальных уровней значимости для дискретных статистик недопустимо

Ключевые слова: МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА, ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ, ДВЕ НЕЗАВИСИМЫЕ ВЫБОРКИ, ПРОВЕРКА ОДНОРОДНОСТИ, УРОВЕНЬ ЗНАЧИМОСТИ, КРИТЕРИЙ ВИЛКОКСОНА, КРИТЕРИЙ ВАН-ДЕР-ВАРДЕНА, КРИТЕРИЙ СМИРНОВА, КРИТЕРИЙ ЗНАКОВ, КРИТЕРИЙ

UDC 519.2:303.732.4

Physics and mathematical sciences

REAL AND NOMINAL SIGNIFICANCE LEVELS IN STATISTICAL HYPOTHESIS TESTING

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

In the statistical hypothesis testing, critical values often point to a priori fixed (nominal) significance levels. As such, typically researcher uses the values of three numbers 0.01, 0.05, 0.1, to which may be added a few levels: 0.001, 0.005, 0.02, and others. However, for the statistics with discrete distribution functions, which, in particular, include all nonparametric statistical tests, the real significance levels may be different from the nominal, differ at times. Under the real significance level we refer to the highest possible significance level of discrete statistics, not exceeding a given nominal significance level (ie, the transition to the next highest possible value corresponding discrete statistical significance level is greater than a predetermined nominal). In the article, we have discussed the difference between nominal and real significance levels on the example of nonparametric tests for the homogeneity of two independent samples. We have also studied two-sample Wilcoxon test, the criterion of van der Waerden, Smirnov two-sample two-sided test, sign test, runs test (Wolfowitz) and calculated the real significance levels of the criteria for nominal significance level of 0.05. The study of the power of these statistical tests is accomplished by means of Monte Carlo method. The main conclusion: the use of nominal significance levels instead of real significance levels for discrete statistics is inadmissible for small sample sizes

Keywords: MATHEMATICAL STATISTICS, NONPARAMETRIC STATISTICS, STATISTICAL HYPOTHESIS TESTING, TWO INDEPENDENT SAMPLES, HOMOGENEITY TESTING, SIGNIFICANCE LEVEL WILCOXON TEST, VAN DER WAEDEN TEST, SMIRNOV TEST, SIGN TEST, RUNS TEST

СЕРИЙ

1. Введение. Реальные и номинальные уровни значимости

Во многих монографиях, справочниках и таблицах (например, [1 - 3]) при рассмотрении задач проверки статистических гипотез критические значения статистик критериев указаны для априорно фиксированных (номинальных в терминологии [4]) уровней значимости α_H. В качестве таковых обычно используются значения из тройки чисел 0,01, 0,05, 0,1, к которым иногда добавляют еще несколько: 0,001, 0,005, 0,02 и др.

Однако ясно, что для дискретных статистик (т.е. статистик с дискретными функциями распределения) к которым, в частности, относятся все непараметрические статистические критерии [5, 6], реальные уровни значимости α_P могут и не совпадать с номинальными. Под α_P понимается максимально возможный уровень значимости дискретной статистики, не превосходящий заданный номинальный α_H (т.е при переходе к следующему по величине возможному значению дискретной статистики соответствующий уровень значимости оказывается больше заданного номинального). Поэтому в лучших таблицах [5, 6] для ограниченных объемов выборок (2 - 100) табулируются точные распределения дискретных статистик. Для каждой конкретной статистики реальный уровень значимости α_P - функция от объемов выборок $n = (n_1, \ldots, n_t)$, т.е. $\alpha_P = \alpha_P(n)$. (Здесь t – число выборок, по которым рассчитывается значение статистики; рассматриваем в основном случай двух выборок, т.е. $t = 2$).

Аналогичная проблема имеется при использовании программных продуктов. При разработке новых статистических программных продуктов необходимо выбрать режим использования реальных и номинальных уровней значимости. Поскольку это не вызывает недоразумений, будем обсуждать таблицы, предназначенные для принятия решений при проверке статистических гипотез.

В одних таблицах приведены α_P [5, 6], в других - нет [1 - 3]. Возникает естественный вопрос: с чем это связано? Очевидно, есть две возможности. Либо в работах [1 - 3] нарушена культура табулирования, либо реальные α_P и номинальные α_H уровни значимости практически совпадают для всех n. Продемонстрируем, что по крайней мере для некоторых статистик выполнено первое из этих двух утверждений.

2. Критерий серий

В качестве примера рассмотрим критерий серий (Вольфовица) проверки однородности двух независимых выборок. Статистика этого критерия V – это число серий, т.е. частей общего вариационного ряда двух выборок, каждая из которых состоит из элементов одной выборки. При справедливости нулевой гипотезы о тождестве функций распределения, соответствующих двум независимым выборкам объемов n_1 и n_2, известно точное распределение [5, табл.6.7]

$$P(V = r \mid n_1, n_2) = \begin{cases} \dfrac{2C_{n_1-1}^{k-1} C_{n_2-1}^{k-1}}{C_{n_1+n_2}^{n_1}}, & \text{если } r = 2k, \\ \dfrac{C_{n_1-1}^{k} C_{n_2-1}^{k-1} + C_{n_1-1}^{k-1} C_{n_2-1}^{k}}{C_{n_1+n_2}^{n_1}}, & \text{если } r = 2k+1, \end{cases}$$

где r = 2, 3, …, $2n_1$ при $n_1 = n_2$ и r = 2, 3, …, $2n_1 + 1$ при $n_1 < n_2$ (без ограничения общности можно принять, что объем первой выборки не превосходит объема второй выборки, т.е. $n_1 \le n_2$).

Несложный расчет для номинального уровня значимости α_H = 0,05 показывается, что

при $n_1 = n_2 = 6$ реальный уровень значимости α_P = 0,0260;

при $n_1 = n_2 = 8$ реальный уровень значимости α_P = 0,0178;

при $n_1 = n_2 = 10$ реальный уровень значимости α_P = 0,0370;

при $n_1 = n_2 = 12$ реальный уровень значимости α_P = 0,0190.
Таким образом, для рассматриваемых объемов выборок реальный уровень значимости в 2 - 3 раза меньше, чем номинальный. Это, очевидно, необходимо учитывать при интерпретации результатов анализа реальных статистических данных.

3. Непараметрические критерии проверки однородности двух независимых выборок

Проблема изучения соотношения реальных (истинных) и номинальных уровней значимости была поставлена нами в статье [4] на примере непараметрических критериев проверки однородности двух независимых выборок. В табл. 1, построенной в [4] на основе таблиц, приведенных в [5 - 7] для ряда непараметрических критериев проверки однородности двух независимых выборок, приведены реальные уровни значимости $\alpha_P(n)$ для номинального уровня значимости α_H = 0,05 и объемов выборок $n_1 = n_2 = 6, 8, 10, 12$. Проанализированы пять критериев.

1. Двухвыборочный критерий Вилкоксона, являющийся линейной функцией от критерия Манна-Уитни и подробно рассмотренный в статьях [8, 9]. Напомним, что статистика Вилкоксона S - это сумма рангов элементов первой выборки

$$S = R_1 + R_2 + ... + R_{n_1}$$

в общем вариационном ряду, построенном по объединенной выборке, включающей в себя все элементы обеих выборок (без ограничения общности можно принять, что объем первой выборки не превосходит объема второй выборки, т.е. $n_1 \leq n_2$).

2. Критерий Ван-дер-Вардена [5, 7], представляющий собой дальнейшее развитие (модификацию) критерия Вилкоксона и предназначенный для анализа выборок, распределение которых близко к нормальному. Статистика X критерия Ван-дер-Вардена имеет вид

$$X = \Phi^{-1}\left\{\frac{R_1}{n_1+n_2+1}\right\} + \Phi^{-1}\left\{\frac{R_2}{n_1+n_2+1}\right\} + \ldots + \Phi^{-1}\left\{\frac{R_{n_1}}{n_1+n_2+1}\right\},$$

где $\Phi^{-1}(p)$ есть квантиль порядка p стандартного нормального распределения $\Phi(x)$ с математическим ожиданием 0 и дисперсией 1, т.е. $\Phi^{-1}(p)$ - обратная функция к $\Phi(x)$.

3. Двухвыборочный двухсторонний критерий Смирнова однородности двух независимых выборок, рассмотренный в [5, 10]. Он основан на использовании разности эмпирических функций распределения $F_{n_1}(x)$ и $G_{n_2}(x)$, построенных по первой и второй выборкам соответственно. Термин «двухсторонний» означает, что берется супремум модуля этой разности. Статистика двухвыборочного двухстороннего критерия Смирнова

$$D = D(n_1, n_2) = \sup_x \left| F_{n_1}(x) - G_{n_2}(x) \right|$$

в случае равенства объемов выборок $n_1 = n_2$ принимает значения, кратные $1/n_1$, поскольку только такие значения принимают эмпирические функции распределения $F_{n_1}(x)$ и $G_{n_2}(x)$, а потому рассматриваемая статистика имеет ровно (n_1+1) возможных значений.

4. Критерий знаков Z используют в случае равенства объемов выборок $n_1 = n_2$. Статистика этого критерия равна числу положительных разностей X_k - Y_k элементов двух выборок с одинаковыми номерами. При справедливости нулевой гипотезы статистика Z имеет биномиальное распределение B(1/2; n_1), а потому имеет (n_1+1) возможных значений.

5. Критерий серий (Вольфовица) V, о котором шла речь выше в разделе 2 настоящей статьи. Число его возможных значений не превосходит $2n_1$.

Таблица 1. Реальные уровни значимости $\alpha_P(n)$

для номинального уровня значимости α_H = 0,05

Наименование и обозначение критерия	Объемы выборок $n_1 = n_2$				Примечания и ссылки
	6	8	10	12	
Вилкоксона S	0,0320	0,0400	0,0480	0,0420	[6, с.280-281], [7, с.418]
Ван-дер-Вардена X	0,0498	0,0498	0,0500	0,0500	Рассчитано по методике [7, с.249-250]
Смирнова D	0,0044	0,0372	0,0246	0,0158	[5, с.412], [6, с.406-427]
Знаков Z	0,0312	0,0078	0,0214	0,0386	[6, с.273-274]
Вольфовица (серий) V	0,0260	0,0178	0,0370	0,0190	Рассчитано по методике [5, с.91-92]

Анализ содержания табл. 1 подтверждает предположение о существенности отличия реальных уровней значимости $\alpha_P(n)$ от номинальных уровней значимости α_H.

4. Обсуждение полученных результатов

Предположим теперь, что, несмотря на установленные отличия, мы используем при проверке гипотезы однородности таблицы [1 - 3], в которых указаны $\alpha_H > \alpha_P$, а не α_P. Это приводит к снижению мощности критерия по сравнению с соответствующим рандомизированным критерием, обеспечивающим равенство α_P и α_H.

Разъяснение. Поясним, что такое рандомизированный критерий. Пусть Y – статистика некоторого статистического критерия, принимающая дискретные значения, числа a и b, где $a < b$ - два соседних значения этой статистики, такие, что $P(Y > b) < \alpha_H$ и $P(Y > a) > \alpha_H$ (вероятности взяты в предположении справедливости нулевой гипотезы). Если критическое значение критерия равно b, т.е. нулевая гипотеза принимается при $Y \leq b$, то $\alpha_P = P(Y > b) < \alpha_H$. Если же критическое значение равно следующему возможному (при движении в сторону уменьшения) значению a, т.е. нулевая гипотеза принимается при $Y \leq a$, то $\alpha_P = P(Y > a) > \alpha_H$. Рандомизированный критерий получим, если при $Y = b$ в некоторой доле p случаев будем принимать нулевую гипотезу, а в остальных случаях – альтернативную. Поскольку

$$P(Y = b) = P(Y > a) - P(Y > b),$$

то (реальный) уровень значимости рандомизированного критерия равен

$$(1 - p)\ P(Y = b) + P(Y > b) = (1 - p)\ P(Y > a) + p\ P(Y > b).$$

Ясно, что при соответствующем выборе параметра рандомизации p уровень значимости рандомизированного критерия совпадет с заданным номинальным уровнем α_H.

Для малых объемов выборок (2 - 20 элементов) понижение мощности из-за того, что $\alpha_H > \alpha_P$, может быть существенным. Для иллюстрации этого в табл. 2 приведены результаты моделирования наиболее употребительных (согласно [5]) критериев проверки однородности двух независимых выборок.

5. Изучение мощности статистических критериев методом Монте-Карло

Моделируются выборки одинакового объема из нормальных законов распределения с математическими ожиданиями m_1 и m_2 и дисперсиями σ_1^2 и σ_2^2. Номинальный уровень значимости, определяющий конкретные критические значения для критериев, принят равным α_H = 0,05. Мощность критерия определяется моделированием N = 5000 пар выборок. При использовании N = 5000 моделируемых пар выборок среднее квадратическое отклонение оценок мощности $\sigma_M \leq 0{,}0223$ (при $M \geq$ 0,95 имеем $\sigma_M \leq 0{,}01$).

Изучены критерии Вилкоксона *S*, Вольфовица *V*, Ван-дер-Вардена *X*, Смирнова *D*. Критерий Стьюдента *t* (см. например, [5]), как равномерно наиболее мощный в классе нормальных законов распределения, приведен для сравнительной оценки мощности рассматриваемых непараметрических критериев. (Моделирование и расчеты, приведенные в настоящей статье, выполнены Ю.Э. Камнем и Я.Э. Камнем [4], за что автор им искренне благодарен.)

Таблица 2. Мощности статистических критериев при α_H = 0,05

Номер экспери-мента	Объем выборок $n_1 = n_2$	Параметры				Мощность M статистического критерия				
		m_1	m_2	σ_1^2	σ_2^2	S	V	X	D	t
1	6	0	1	1	1	0,318	0,006	0,298	0,238	0,396
2	8	0	1	1	1	0,452	0,104	0,426	0,068	0,484
3	10	0	1	1	1	0,520	0,180	0,534	0,116	0,598
4	12	0	1	1	1	0,632	0,076	0,618	0,462	0,682
5	6	0	2	1	1	0,828	0,308	0,808	0,716	0,904
6	8	0	2	1	1	0,958	0,510	0,954	0,458	0,976
7	10	0	2	1	1	0,984	0,704	0,990	0,632	0,988
8	12	0	2	1	1	0,996	0,568	0,996	0,978	0,998

Замечание. Приведенные в табл. 2 значения мощностей критериев интересны нам с точки зрения обсуждения их зависимости от различия реальных и номинальных уровней значимости. При этом необходимо подчеркнуть, что эти значения зависят от предположений, принятых при моделировании. Так, критерии Вилкоксона и Ван-дер-Вардена «настроены» на использование в случае распределений, близких к нормальному семейству. При проверке гипотезы о совпадении функций распределения двух независимых выборок из логистического распределения с альтернативой сдвига критерий Вилкоксона является асимптотически оптимальным. А в случае выборок из нормального распределения аналогичным свойством обладает критерий Ван-дер-Вардена, причем известно, что семейства нормальных и логистических распределений весьма близки – расстояние Колмогорова между ними не превышает 0,01 (см. по вопросам асимптотической оптимальности непараметрических критериев монографии [11 - 13]). Поэтому нет ничего удивительного в том, что мощности критериев Вилкоксона и Ван-дер-Вардена близки к оптимуму в случае нормального распределения – к мощности критерия Стьюдента. При этом мощности критериев Смирнова и особенно критерия Вольфовица заметно меньше. Однако для выборок из других распределений (например, распределений Вейбулла-Гнеденко или гамма-распределений) ситуация иная – критерий Смирнова, как показывает компьютерное моделирование, оказывается более мощным, чем критерии Вилкоксона и Ван-дер-Вардена. Более того, критерий Смирнова – состоятельный, т.е. позволяет отклонить любую конкретную альтернативу (при соответствующих объемах выборок), а критерии Вилкоксона и Ван-дер-Вардена не являются состоятельными, некоторых альтернатив они «не чувствуют». Поэтому вполне обоснованной является рекомендация о широком использовании состоятельных критериев Смирнова и типа омега-квадрат (Лемана-Розенблатта), данная в [14]. Что же касается критерия серий (Вольфовица), то из-за его отрицательных свойств (выраженной дискретности, низкой мощности) он в настоящее время выходит из употребления при анализе реальных данных, несмотря на прозрачность определения.

6. Выводы

Рассмотрения настоящей статьи позволяют сделать следующие выводы (см. также [4]).

1. При создании методик, таблиц, программных продуктов необходимо соблюдать определенную культуру табулирования. В качестве положительных примеров можно указать работы [5, 6].

2. При малых объемах выборок использовать номинальные уровни значимости α_H вместо реальных уровней значимости α_P для дискретных статистик недопустимо.

3. При конечных объемах выборок выбор того или иного критерия с дискретной статистикой должен сопровождаться исследованием влияния варьирования уровня значимости на качественную интерпретацию результатов проверки гипотез. В частности, выбор одного из двух конкурирующих непараметрических критериев K_1 и K_2 прежде всего должен зависеть от априорного выбора исследователем реального уровня значимости α_{P1} или α_{P2}, соответствующего первому критерию K_1 или второму K_2, в качестве номинального уровня значимости α_H.

Последний (третий) вывод демонстрирует сложность сравнения критериев с дискретными статистиками между собой, поскольку точки скачков распределений их статистик не совпадают. Следовательно, в отличие от критериев с непрерывными статистиками нельзя выбрать единый фиксированный уровень значимости и сравнить свойства критериев при этом уровне значимости.

В заключение отметим, что для любого критерия проверки статистических гипотез реальный уровень значимости приближается к номинальному при безграничном возрастании объемов выборок, т.е. $\alpha_P(n) \to \alpha_H$ при $\min(n_1,...,n_t) \to \infty$. Поэтому для прикладных исследований значительный интерес представляет определение верхней оценки скорости сходимости $\alpha_P(n)$ к α_H. Соответствующие теоретические результаты для критериев проверки однородности двух независимых или связанных выборок можно получить, основываясь на оценках скорости сходимости в принципе инвариантности [15, гл.4]. Некоторые оценки приведены в [16, гл.2]. Скорость сходимости также может быть оценена методом статистических испытаний (Монте-Карло). Пример подобного исследования подробно рассмотрен в [17 - 18] в ходе обсуждения проблем вероятностно-статистического моделирования помех, создаваемых электровозами.

В настоящей статье затронута лишь небольшая часть непараметрических методов анализа числовых статистических данных [19, 20]. В частности, обратим внимание на непараметрические оценки плотности, которые используются для описания данных, проверки однородности, в задачах восстановления зависимостей и других областях эконометрики [21, 22]. Непараметрические оценки плотности рассмотрены в [15, раздел 5.6].

Литература

1. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. – М.: Мир, 1982. – 488 с.
2. Гублер Е.В., Генкин А.А. Применение критериев непараметрической статистики в медико-биологических исследованиях. – Л.: Медицина, 1973. – 144 с.
3. Ивченко Г.И., Медведев Ю.И. Математическая статистика. – М.: Высшая школа, 1984. – 248 с.
4. 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.
5. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. - 416 с.
6. Холлендер М., Вульф Д. Непараметрические методы статистики. – М.: Финансы и статистика, 1983. - 518 с.
7. Ван-дер-Варден Б.Л. Математическая статистика. – М.: ИЛ, 1960. – 434 с.
8. Орлов А.И. Какие гипотезы можно проверять с помощью двухвыборочного критерия Вилкоксона? // Заводская лаборатория. Диагностика материалов. 1999. Т.65. №1. С.51-55.
9. Орлов А.И. Двухвыборочный критерий Вилкоксона – анализ двух мифов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 104. С. 91 – 111.
10. 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. 32-45.
11. Кендалл М. Дж., Стьюарт А. Статистические выводы и связи. – М.: Наука, 1973. – 900 с.
12. Кокс Д.Р., Хинкли Д.В. Теоретическая статистика. – М.: Мир, 1978. – 560 с.

13. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
14. Орлов А.И. Состоятельные критерии проверки абсолютной однородности независимых выборок // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №11. С.66-70.
15. Орлов А.И. Прикладная статистика. – М. : Экзамен, 2006. – 671 с.
16. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. – 296 с.
17. Карякин Р.Н., Орлов А.И., Адамов С.Ю. Вероятностная теория высших гармоник помех, создаваемых электровозами // Прикладной многомерный статистический анализ. Ученые записки по статистике, т.33. - М.: Наука, 1978. - С. 376-380.
18. Орлов А.И. Вероятностно-статистическое моделирование помех, создаваемых электровозами // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 225 – 238.
19. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239 – 269.
20. Орлов А.И. Структура непараметрической статистики (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 62-72.
21. Орлов А.И. Оценки плотности распределения вероятностей в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 99. С. 15–32.
22. Орлов А.И. Предельные теоремы для ядерных оценок плотности в пространствах произвольной природы // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 108. С. 316 – 333.

References

1. Afifi A., Jejzen S. Statisticheskij analiz. Podhod s ispol'zovaniem JeVM. – M.: Mir, 1982. – 488 s.
2. Gubler E.V., Genkin A.A. Primenenie kriteriev neparametricheskoj statistiki v mediko-biologicheskih issledovanijah. – L.: Medicina, 1973. – 144 s.
3. Ivchenko G.I., Medvedev Ju.I. Matematicheskaja statistika. – M.: Vysshaja shkola, 1984. – 248 s.
4. 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.
5. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoj statistiki. – M.: Nauka, 1983. - 416 s.
6. Hollender M., Vul'f D. Neparametricheskie metody statistiki. – M.: Finansy i statistika, 1983. - 518 s.
7. Van-der-Varden B.L. Matematicheskaja statistika. – M.: IL, 1960. – 434 s.
8. Orlov A.I. Kakie gipotezy mozhno proverjat' s pomoshh'ju dvuhvyborochnogo kriterija Vilkoksona? // Zavodskaja laboratorija. Diagnostika materialov. 1999. T.65. №1. S.51-55.
9. Orlov A.I. Dvuhvyborochnyj kriterij Vilkoksona – analiz dvuh mifov //

Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 104. S. 91 – 111.

10. 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. 32-45.

11. Kendall M. Dzh., St'juart A. Statisticheskie vyvody i svjazi. – M.: Nauka, 1973. – 900 s.

12. Koks D.R., Hinkli D.V. Teoreticheskaja statistika. – M.: Mir, 1978. – 560 s.

13. Nikitin Ja.Ju. Asimptoticheskaja jeffektivnost' neparametricheskih kriteriev. - M.: Nauka, 1995. - 240 s.

14. Orlov A.I. Sostojatel'nye kriterii proverki absoljutnoj odnorodnosti nezavisimyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. 2012. T.78. №11. S.66-70.

15. Orlov A.I. Prikladnaja statistika. – M. : Jekzamen, 2006. – 671 s.

16. Orlov A.I. Ustojchivost' v social'no-jekonomicheskih modeljah. - M.: Nauka, 1979. – 296 s.

17. Karjakin R.N., Orlov A.I., Adamov S.Ju. Verojatnostnaja teorija vysshih garmonik pomeh, sozdavaemyh jelektrovozami // Prikladnoj mnogomernyj statisticheskij analiz. Uchenye zapiski po statistike, t.33. - M.: Nauka, 1978. - S. 376-380.

18. Orlov A.I. Verojatnostno-statisticheskoe modelirovanie pomeh, sozdavaemyh jelektrovozami // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 106. S. 225 – 238.

19. Orlov A.I. Sovremennoe sostojanie neparametricheskoj statistiki // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 106. S. 239 – 269.

20. Orlov A.I. Struktura neparametricheskoj statistiki (obobshhajushhaja stat'ja) // Zavodskaja laboratorija. Diagnostika materialov. 2015. T.81. №7. S. 62-72.

21. Orlov A.I. Ocenki plotnosti raspredelenija verojatnostej v prostranstvah proizvol'noj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 99. S. 15–32.

22. Orlov A.I. Predel'nye teoremy dlja jadernyh ocenok plotnosti v prostranstvah proizvol'noj prirody // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 108. S. 316 – 333.