

**НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ
ХАРАКТЕРИСТИК РАСПРЕДЕЛЕНИЙ
ВЕРОЯТНОСТЕЙ**

**NONPARAMETRIC ESTIMATION OF
CHARACTERISTICS OF PROBABILITY
DISTRIBUTIONS**

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*

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

Статья посвящена непараметрическому точечному и интервальному оцениванию характеристик распределения вероятностей (математического ожидания, медианы, дисперсии, среднего квадратического отклонения, коэффициента вариации) по выборке результатов измерений. Выборочные значения рассматриваются как реализации независимых одинаково распределенных случайных величин с произвольной функцией распределения, имеющей нужное число моментов. Непараметрические процедуры анализа данных сопоставляются с параметрическими, основанными на предположении о том, что выборочные значения имеют нормальное распределение. Точечные оценки строятся очевидным образом – используют выборочные аналоги теоретических характеристик. Интервальные оценки основаны на асимптотической нормальности выборочных моментов и функций от них. Непараметрические асимптотические доверительные интервалы получены в результате применения специальной технологии вывода асимптотических соотношений прикладной статистики. Эта технология в качестве первого шага использует многомерную центральную предельную теорему, примененную к сумме векторов, координаты которых – степени исходных случайных величин. Второй шаг – преобразование предельного многомерного нормального вектора с целью получения интересующего исследователя вектора. При этом используются соображения линеаризации и отбрасываются бесконечно малые величины. Третий шаг – строгое обоснование полученных результатов на стандартном для асимптотических математико-статистических рассуждений уровне. При этом обычно приходится использовать необходимые и достаточные условия наследования сходимости. Статья содержит 10 числовых примеров. Исходные данные - сведения о наработке 50 резцов до предельного состояния. Использование методов, разработанных в предположении нормальности распределения, может привести к заметно искаженным выводам в ситуации, когда гипотеза нормальности не выполнена. Практические рекомендации таковы: при анализе реальных данных следует использовать непараметрические

The article is devoted to the nonparametric point and interval estimation of the characteristics of the probabilistic distribution (the expectation, median, variance, standard deviation, variation coefficient) of the sample results. Sample values are regarded as the implementation of independent and identically distributed random variables with an arbitrary distribution function having the desired number of moments. Nonparametric analysis procedures are compared with the parametric procedures, based on the assumption that the sample values have a normal distribution. Point estimators are constructed in the obvious way - using sample analogs of the theoretical characteristics. Interval estimators are based on asymptotic normality of sample moments and functions from them. Nonparametric asymptotic confidence intervals are obtained through the use of special output technology of the asymptotic relations of Applied Statistics. In the first step this technology uses the multidimensional central limit theorem, applied to the sums of vectors whose coordinates are the degrees of initial random variables. The second step is the conversion limit multivariate normal vector to obtain the interest of researcher vector. At the same considerations we have used linearization and discarded infinitesimal quantities. The third step - a rigorous justification of the results on the asymptotic standard for mathematical and statistical reasoning level. It is usually necessary to use the necessary and sufficient conditions for the inheritance of convergence. This article contains 10 numerical examples. Initial data - information about an operating time of 50 cutting tools to the limit state. Using the methods developed on the assumption of normal distribution, it can lead to noticeably distorted conclusions in a situation where the normality hypothesis failed. Practical recommendations are: for the analysis of real data we should use nonparametric confidence limits

доверительные границы

Ключевые слова: МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, ПРИКЛАДНАЯ СТАТИСТИКА, СТАТИСТИЧЕСКИЕ МЕТОДЫ, НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА, ТОЧЕЧНОЕ ОЦЕНИВАНИЕ, ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ, МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ, МЕДИАНА, ДИСПЕРСИЯ, СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ, КОЭФФИЦИЕНТ ВАРИАЦИИ, ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

Keywords: MATHEMATICAL STATISTICS, APPLIED STATISTICS, STATISTICAL METHODS, NONPARAMETRIC STATISTICS, POINT ESTIMATION, CONFIDENCE INTERVALS, EXPECTATION, MEDIAN, VARIANCE, STANDART DEVIATION, VARIATION COEFFICIENT, LIMIT THEOREMS

1. Введение

Рассмотрим несколько типовых задач анализа числовых данных, часто встречающихся при применении статистических методов в различных областях научных исследований и отраслях народного хозяйства. В настоящей статье выборка моделируется как совокупность независимых одинаково распределенных *числовых* случайных величин.

Анализ конкретных данных обычно начинают с получения "статистического портрета", т.е. с оценивания основных характеристик распределения [1]. Настоящая статья посвящена непараметрическому точечному и интервальному оцениванию характеристик распределения (математического ожидания, медианы, дисперсии, среднего квадратического отклонения, коэффициента вариации) по выборке результатов измерений. Выборочные значения рассматриваются как реализации независимых одинаково распределенных случайных величин с произвольной функцией распределения, имеющей нужное число моментов. Непараметрические процедуры анализа данных сопоставляются с параметрическими (ср. [2]).

Существенная часть алгоритмов статистического анализа данных исходит из предположения о нормальности распределения результатов наблюдений. Между тем специально проведенные исследования (сводка дана, например, в [3] и [4, разд. 2.1]) показывают, что распределения погрешностей физических измерений, как правило, отличны от

нормальных. Из-за отклонений от нормальности свойства алгоритмов могут в одних случаях измениться сравнительно слабо, как при проверке гипотезы однородности математических ожиданий для выборок равного объема (см. [4, разд.5.2], [5]), но иногда изменения таковы, что алгоритмы из научных переходят в эвристические. Например, свойства алгоритмов отбраковки выбросов (резко выделяющихся наблюдений) крайне неустойчивы по отношению к отклонениям от нормальности: если зафиксировать правило отбраковки, то крайне неустойчив уровень значимости, а если зафиксировать уровень значимости, то крайне неустойчиво критическое значение (см.[4, разд.4.2], [6]). Поэтому Российской академией статистических методов в 1998 г. выдвинута задача изучения влияния отклонения от нормальности на свойства всех практически используемых алгоритмов статистического анализа.

Одна из основных задач в области статистических методов – оценивание по выборочным данным характеристик генеральной совокупности, таких, как математическое ожидание, медиана, дисперсия, среднее квадратическое отклонение, коэффициент вариации. Точечные оценки строятся очевидным образом – используют выборочные аналоги теоретических характеристик. Для получения интервальных оценок приходится использовать асимптотическую нормальность выборочных моментов и функций от них.

Пусть исходные данные – это выборка x_1, x_2, \dots, x_n , где n – объем выборки. В вероятностной модели выборочные значения x_1, x_2, \dots, x_n рассматриваются как реализации независимых одинаково распределенных случайных величин X_1, X_2, \dots, X_n с общей функцией распределения $F(x) = P(X_i < x)$, $i = 1, 2, \dots, n$. Поскольку функция распределения произвольна (с точностью до условий регулярности типа существования моментов), то рассматриваемые задачи доверительного оценивания характеристик распределения являются *непараметрическими*. Существование моментов

является скорее математическим ограничением, чем реальным, поскольку практически все реальные статистические данные финитны (т.е. ограничены сверху и снизу, например, шкалой прибора). Для простоты изложения примем это предположение финитности, из которого вытекает существование теоретических моментов любого порядка.

В дальнейшем изложении будут использоваться выборочное среднее арифметическое

$$\bar{X} = (X_1 + X_2 + \dots + X_n) / n,$$

выборочная дисперсия

$$s_0^2 = \{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2\} / (n - 1),$$

выборочное среднее квадратическое отклонение s_0 (квадратный корень из выборочной дисперсии) и некоторые другие выборочные характеристики, которые введем позже.

2. Точечное и интервальное оценивание математического ожидания.

Точечной оценкой для математического ожидания в силу закона больших чисел является выборочное среднее арифметическое \bar{X} . В некоторых случаях могут быть использованы и другие оценки. Например, если известно, что распределение симметрично относительно своего центра, то центр распределения является не только математическим ожиданием, но и медианой, а потому для его оценки можно использовать выборочную медиану.

Нижняя доверительная граница для математического ожидания имеет вид

$$\bar{X} - U(p) s_0 / n^{1/2},$$

где: p – доверительная вероятность (истинное значение математического ожидания находится между нижней доверительной границей и верхней

доверительной границей с вероятностью, асимптотически равной доверительной);

$U(p)$ – число, заданное равенством $\Phi(U(p)) = (1 + p)/2$, где $\Phi(x)$ – функция стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1. Например, при $p = 95\%$ (т.е. при $p = 0,95$) имеем $U(p) = 1,96$. Функция $U(p)$ имеется в большинстве литературных источников по теории вероятностей и математической статистике (см., например, [7]).

Верхняя доверительная граница для математического ожидания имеет вид

$$\bar{X} + U(p) s_0 / n^{1/2} .$$

Выражения для верхней и нижней доверительных границ получены с помощью центральной предельной теоремы теории вероятностей и теоремы о наследовании сходимости (см., например, [8]). Они являются асимптотическими, т.е. становятся тем точнее, чем больше объем выборки. В частности, вероятность попадания истинного значения математического ожидания в интервал между нижней и верхней доверительными границами асимптотически приближается к доверительной вероятности. Но при конечном объеме выборки может незначительно отличаться от нее. Это – недостатки непараметрического подхода. Достоинством же является то, что его можно применять всегда, когда случайная величина имеет математическое ожидание и дисперсию, что в силу финитности (ограниченности шкал) имеет быть практически всегда в реальных ситуациях.

Сопоставим с параметрическим подходом. Обычно в таких случаях предполагают нормальность результатов наблюдений (которой, как уже было отмечено, практически никогда нет). Тогда формулы нижней и верхней доверительных границ для математического ожидания имеют похожий вид, только вместо $U(p)$ стоят квантили распределения

Стьюдента. Как известно, при росте объема выборки квантили распределения Стьюдента сходятся к соответствующим квантилям стандартного нормального распределения, так что при больших объемах выборок оба подхода дают близкие результаты. Классические доверительные интервалы несколько длиннее, поскольку квантили распределения Стьюдента больше квантилей стандартного нормального распределения, хотя это различие и невелико.

Таблица 1

Наработка резцов до предельного состояния (до отказа), ч

№ п/п	Наработка, ч	№ п/п	Наработка, ч	№ п/п	Наработка, ч
1	9	18	47,5	35	63
2	17,5	19	48	36	64,5
3	21	20	50	37	65
4	26,5	21	51	38	67,5
5	27,5	22	53,5	39	68,5
6	31	23	55	40	70
7	32,5	24	56	41	72,5
8	34	25	56	42	77,5
9	36	26	56,5	43	81
10	36,5	27	57,5	44	82,5
11	39	28	58	45	90
12	40	29	59	46	96
13	41	30	59	47	101,5
14	42,5	31	60	48	117,5
15	43	32	61	49	127,5
16	45	33	61,5	50	130
17	46	34	62		

Пример 1. Рассмотрим данные о наработке резцов до отказа, приведенные в табл. 1. Для них объем выборки $n = 50$, выборочное среднее арифметическое $\bar{X} = 57,88$ (это и есть точечная оценка для математического ожидания), выборочная дисперсия $s_0^2 = 663,00$.

Следовательно, выборочное среднее квадратическое отклонение $s_0 = \sqrt{663,00} = 25,75$ и согласно приведенным выше формулам при доверительной вероятности $p = 0,95$ нижняя доверительная граница для математического ожидания такова:

$$57,88 - 1,96 \times 25,75 / \sqrt{50} = 57,88 - 7,14 = 50,74,$$

а верхняя доверительная граница есть $57,88 + 7,14 = 65,02$.

Если заранее известно, что результаты наблюдения имеют нормальное распределение, то нижняя и верхняя доверительная границы для математического ожидания определяются по формулам

$$\bar{X} - t(p, n-1) s_0 / \sqrt{n}, \quad \bar{X} + t(p, n-1) s_0 / \sqrt{n}$$

соответственно. Эти формулы отличаются от предыдущих тем, что квантиль нормального распределения $U(p)$ заменена на аналогичную квантиль распределения Стьюдента с $(n - 1)$ степенью свободы. Другими словами, $t(p, n-1)$ – это число, заданное равенством $ST_{n-1}(p) = (1 + p)/2$, где $ST_{n-1}(x)$ – функция распределения Стьюдента с $(n - 1)$ степенью свободы.

Для доверительной вероятности $p = 0,95$ при объеме выборки $n = 50$ согласно [7] имеем квантиль распределения Стьюдента $t(p, n-1) = 2,0096$. Следовательно, нижняя доверительная граница для математического ожидания такова:

$$57,88 - 2,0096 \times 25,75 / \sqrt{50} = 57,88 - 7,32 = 50,56,$$

а верхняя доверительная граница есть $57,88 + 7,32 = 65,20$. Таким образом, длина доверительного интервала увеличилась с 14,28 до 14,64, т.е. на 2,5%.

Согласно расчетам, проведенным для проверки согласия с параметрическим семейством распределений методом моментов в [4, разд. 4.1] и [9], рассматриваемые данные согласуются с гамма-распределением, а не с нормальным распределением, поэтому использование распределения Стьюдента для получения доверительных границ некорректно.

Иногда рекомендуют сначала проверить нормальность результатов наблюдений, а потом, в случае принятия гипотезы нормальности, рассчитывать доверительные границы с использованием квантилей распределения Стьюдента. Однако проверка нормальности - более сложная статистическая процедура, чем оценивание математического ожидания. Кроме того, применение одной статистической процедуры, как правило, нарушает предпосылки следующей процедуры, в частности, независимость результатов наблюдений (см., например, [4, разд. 4.3] и [10]). Поэтому цепочка статистических процедур, следующих друг за другом, как правило, образует статистическую технологию, свойства которой неизвестны на современном уровне развития статистических методов.

Итак, только непараметрическую статистическую процедуру следует применять для анализа реальных данных. Как правило, встречающиеся на практике распределения не являются нормальными, а потому использование квантилей распределения Стьюдента неправомерно.

3. Точечное и интервальное оценивание медианы

Естественной точечной оценкой для медианы является выборочная медиана.

Пример 2. Для данных о наработке резцов до отказа (табл. 1) объем выборки – четное число, поэтому выборочной медианой является полусумма 25-го и 26-го членов вариационного ряда, т.е. $(56 + 56,5)/2 = 56,25$.

Чтобы построить доверительные границы для медианы, по доверительной вероятности p находят $U(p)$. Затем вычисляют натуральное число

$$C(p) = [n/2 - U(p)n^{1/2}/2],$$

где $[.]$ – знак целой части числа. Нижняя доверительная граница для медианы имеет вид (при $C(p) \geq 1$; если $p = 0,95$ и $U(p) = 1,96$, то $C(p) \geq 1$ при $n \geq 8$)

$$X(C(p)),$$

где $X(i)$ – член с номером i вариационного ряда, построенного по исходной выборке (т.е. i -я порядковая статистика).

Верхняя доверительная граница для медианы имеет вид

$$X(n + 1 - C(p)).$$

Теоретическое основание для приведенных доверительных границ содержится в литературе по порядковым статистикам (см., например, монографию [11, с.68]).

Пример 3. Для данных о наработке резцов до отказа $n = 50$. Рассмотрим как обычно, доверительную вероятность $p = 0,95$. Тогда

$$C(p) = [50/2 - 1,96 \sqrt{50} / 2] = [18,07] = 18.$$

Следовательно, нижней доверительной границей является $X(18) = 47,5$, а верхней доверительной границей $X(50 + 1 - 18) = X(33) = 61,5$.

Поскольку в случае нормального распределения медиана совпадает с математическим ожиданием, то каких-либо специальных способов ее оценивания в классическом случае нет.

4. Точечное и интервальное оценивание дисперсии

Точечной оценкой дисперсии является выборочная дисперсия s_0^2 . Эта оценка - несмещенная и состоятельная. Доверительные границы находятся с помощью величины

$$d^2 = (m_4 - ((n - 1) / n)^4 s_0^4) / n,$$

где m_4 - выборочный четвертый центральный момент, т.е.

$$m_4 = \{(X_1 - \bar{X})^4 + (X_2 - \bar{X})^4 + \dots + (X_n - \bar{X})^4\} / n.$$

Нижняя доверительная граница для дисперсии такова:

$$s_0^2 - U(p)d,$$

где s_0^2 - выборочная дисперсия; $U(p)$ - квантиль нормального распределения порядка $(1+p)/2$, а d - положительный квадратный корень из величины d^2 , введенной выше.

Верхняя доверительная граница для дисперсии имеет вид

$$s_0^2 + U(p)d.$$

При выводе приведенных соотношений используется асимптотическая нормальность выборочной дисперсии, установленная, например, в монографии [12, с.419]. Соответственно, непараметрический доверительный интервал является асимптотическим. В классическом случае точечная оценка имеет тот же вид, а вот доверительные границы находят с помощью квантилей распределения хи-квадрат с числом степеней свободы, на 1 меньшим объема выборки. В случае нормального распределения четвертый момент в 3 раза больше квадрата дисперсии, а потому можно оценить d^2 как $2s_0^4/n$. Это дает быстрый способ для интервальной оценки дисперсии в нормальном случае.

Пример 4. Для данных о наработке резцов до отказа объем выборки $n = 50$, выборочная дисперсия $s_0^2 = 663,00$, четвертый выборочный момент $m_4 = 1702050,71$. Поэтому

$$d^2 = (1702050,71 - ((50 - 1) / 50)^4 663,00^2) / 50 = 25932,13.$$

Тогда $d = 161,03$. Для доверительной вероятности $p = 0,95$ нижняя доверительная граница для дисперсии случайной величины такова:

$$663,00 - 1,96 \times 161,03 = 663,00 - 315,63 = 347,37,$$

а верхняя доверительная граница для дисперсии есть $663,00 + 315,63 = 978,63$.

Пример 5. В случае нормального распределения с целью быстрого получения доверительного интервала величина d^2 оценивается как

$$(2s_0^4) / n = (2 \times 663,00^2) / 50 = 17582,76,$$

а потому $d = 132,6$. Для доверительной вероятности $p = 0,95$ нижняя доверительная граница для дисперсии заменяется на

$$663,00 - 1,96 \times 132,6 = 663,00 - 259,90 = 403,10,$$

а верхняя доверительная граница – на $663,00 + 259,90 = 922,9$.

Сужение границ для дисперсии вполне естественно. Данные о наработке резцов до предельного состояния (т.е. до отказа) соответствуют гамма-распределению, а это распределение является асимметричным, с «тяжелым» правым «хвостом». Последнее означает, что плотность убывает заметно медленнее, чем для нормального распределения. Как следствие, четвертый момент заметно больше, чем для нормального распределения с теми же математическим ожиданием и дисперсией. А потому больше и параметр d . Из проведенных расчетов видно, что использование алгоритмов расчетов, соответствующих нормальному распределению, в ситуации, когда распределение результатов наблюдений существенно отличается от нормального, может привести к заметному искажению статистических выводов.

Пример 6. В классическом случае нормального распределения исходят из того, что величина $(n - 1) s_0^2 / \sigma^2$ имеет распределение хи-квадрат с $(n - 1)$ степенью свободы. Для доверительной вероятности $p = 0,95$ следует рассмотреть неравенство

$$31,555 < (n - 1) s_0^2 / \sigma^2 < 70,222,$$

справедливое с вероятностью 0,95, поскольку

$$F(31,555) = 0,025, F(70,222) = 0,975,$$

где $F(x)$ – функция хи-квадрат распределения с 49 степенями свободы. Следовательно, нижняя доверительная граница для дисперсии нормально распределенной случайной величины такова:

$$(n - 1) s_0^2 / 70,222 = (49 \times 663,00) / 70,222 = 462,63,$$

а верхняя доверительная граница есть

$$(n - 1) s_0^2 / 31,555 = (49 \times 663,00) / 31,555 = 1029,54.$$

Полученный доверительный интервал не является симметричным относительно точечной оценки. Нижняя доверительная граница больше, чем в примерах 4 и 5, но и верхняя доверительная граница тоже больше. Несимметричность доверительного интервала в примере 6 приводит к тому, что его трудно сопоставить с симметричными интервалами примеров 4 и 5. Что же касается практических рекомендаций, то они однозначны: обычно нет основания считать, что результаты измерений имеют нормальное распределение, поэтому при анализе реальных данных надо пользоваться непараметрическими методами.

5. Точечное и интервальное оценивание среднего квадратического отклонения

Точечной оценкой является выборочное среднее квадратическое отклонение, т.е. неотрицательный квадратный корень из выборочной дисперсии. Дисперсия рассматриваемой случайной величины - выборочного среднего квадратического отклонения s_0 – оценивается как дробь

$$d^2 / (4s_0^2).$$

Нижняя доверительная граница для среднего квадратического отклонения исходной случайной величины имеет вид

$$s_0 - U(p)d / (2s_0),$$

где s_0^2 – выборочная дисперсия, $U(p)$ – квантиль нормального распределения порядка $(1+p)/2$, а d – положительный квадратный корень из величины d^2 , введенной выше при оценивании дисперсии.

Верхняя доверительная граница для среднего квадратического отклонения исходной случайной величины имеет вид

$$s_0 + U(p)d / (2s_0).$$

Пример 7. Для данных о наработке резцов до отказа точечной оценкой для среднего квадратического отклонения является $s_0 = \sqrt{663,00} = 25,75$.

При доверительной вероятности $p = 0,95$ нижняя доверительная граница такова:

$$25,75 - 1,96 \times 161,03 / (2 \times 25,75) = 25,75 - 6,13 = 19,62.$$

Соответственно верхняя доверительная граница симметрична нижней относительно точечной оценки и равна $= 25,75 + 6,13 = 31,88$.

Правила интервального оценивания для среднего квадратического отклонения получены из аналогичных правил для оценивания дисперсии с помощью метода линеаризации (см., например, [8]). Доверительный интервал - симметричный, непараметрический и асимптотический.

Есть и другой способ доверительного оценивания. Поскольку среднее квадратическое отклонение – это квадратный корень их дисперсии, то доверительные границы можно получить, извлекая квадратные корни из одноименных границ для дисперсии.

Пример 8. Для данных о наработке резцов до отказа при доверительной вероятности $p = 0,95$ согласно примеру 4 доверительный интервал для дисперсии – это $[347,37; 978,63]$. Извлекая квадратные корни, получаем доверительный интервал $[18,64; 31,28]$ для среднего квадратического отклонения, соответствующий тому же значению доверительной вероятности. Он не является симметричным относительно

точечной оценки. Его длина 12,64 несколько больше длины симметричного доверительного интервала 12,26 в примере 7.

Подход, основанный на гипотезе нормальности распределения результатов наблюдения, связан с использованием распределения хи-квадрат и сводится к извлечению квадратных корней из доверительных границ для дисперсии.

Пример 9. Формально применяя классический подход к данным о наработке резцов до отказа, исходим из доверительного интервала для дисперсии [462,63; 1029,54], соответствующего доверительной вероятности $p = 0,95$. Извлекая квадратные корни, находим доверительный интервал для среднего квадратического отклонения [21,51; 32,09]. Как и следовало ожидать, длина этого несимметричного интервала 10,58 меньше длины непараметрического доверительного интервала, равной 12,68.

6. Точечное и интервальное оценивание коэффициента вариации

Коэффициент вариации $V = \sigma/M(X)$ широко используется при анализе конкретных технических, экономических, социологических, медицинских и иных данных (поскольку они, как правило, положительны). Точечной оценкой теоретического коэффициента вариации V является выборочный коэффициент вариации

$$V_n = s_0 / \bar{X}.$$

Согласно [13, 14] дисперсия выборочного коэффициента вариации состоятельно оценивается с помощью вспомогательной величины

$$D^2 = (V_n^4 - V_n^2 / 4 + m_4 / (4s_0^2 \bar{X}^2) - m_3 / \bar{X}^3) / n,$$

где \bar{X} – выборочное среднее арифметическое, s_0^2 – выборочная дисперсия,

m_3 – выборочный третий центральный момент, т.е.

$$m_3 = \{(X_1 - \bar{X})^3 + (X_2 - \bar{X})^3 + \dots + (X_n - \bar{X})^3\} / n,$$

m_4 - выборочный четвертый центральный момент (см. выше), V_n – выборочный коэффициент вариации, n - объем выборки.

Нижняя доверительная граница для (теоретического) коэффициента вариации исходной случайной величины имеет вид

$$V_n - U(p) D,$$

где V_n – выборочный коэффициент вариации, $U(p)$ – квантиль нормального распределения порядка $(1+p)/2$ (как и ранее), D – положительный квадратный корень из величины D^2 , введенной выше.

Верхняя доверительная граница для (теоретического) коэффициента вариации исходной случайной величины имеет вид

$$V_n + U(p) D.$$

Как и в предыдущих случаях, доверительный интервал - непараметрический и асимптотический. Он получен в результате применения специальной технологии вывода асимптотических соотношений прикладной статистики [8]. Эта технология в качестве первого шага использует многомерную центральную предельную теорему, примененную к сумме векторов, координаты которых – степени исходных случайных величин. Второй шаг – преобразование предельного многомерного нормального вектора с целью получения интересующего исследователя вектора. При этом используются соображения линеаризации и отбрасываются бесконечно малые величины. Третий шаг – строгое обоснование полученных результатов на стандартном для асимптотических математико-статистических рассуждений уровне. При этом обычно приходится использовать необходимые и достаточные условия наследования сходимости (см., например, [8]). Именно таким образом получены приведенные выше результаты для выборочного коэффициента вариации. Формулы оказались существенно более сложными, чем в предыдущих случаях. Это объясняется тем, что выборочный коэффициент вариации - функция двух выборочных

моментов, а ранее рассматривались либо выборочные моменты поодиночке, либо функция от одного выборочного момента - выборочной дисперсии.

Пример 10. Для данных о наработке резцов до отказа выборочное среднее арифметическое $\bar{X} = 57,88$, выборочная дисперсия $s_0^2 = 663,00$, выборочное среднее квадратическое отклонение $s_0 = 25,75$, выборочный третий центральный момент $m_3 = 14927,91$, выборочный четвертый центральный момент $m_4 = 1702050,71$. Следовательно, выборочный коэффициент вариации таков:

$$V_n = 25,75 / 57,88 = 0,4449.$$

Рассчитаем значение вспомогательной величины

$$\begin{aligned} D^2 &= ((0,4449)^4 - (0,4449)^2/4 + 1702050,71 / (4 \times 663,00 \times (57,88)^2) - \\ &- 14927,91 / (57,88)^3) / 50 = (0,0392 - 0,0495 + 0,1916 - 0,0770) / 50 = \\ &= 0,1043 / 50 = 0,002086. \end{aligned}$$

Следовательно, $D = 0,04567$. При доверительной вероятности $p = 0,95$ нижняя доверительная граница для теоретического коэффициента вариации имеет вид

$$0,4449 - 1,96 \times 0,04567 = 0,4449 - 0,0895 = 0,3554,$$

а верхняя доверительная граница такова:

$$0,4449 + 0,0895 = 0,5344.$$

Среди классических результатов математической статистики, основанных на гипотезе нормальности результатов наблюдений, нет методов нахождения доверительных границ для коэффициента вариации, поскольку задача построения таких границ не выражается в терминах обычно используемых распределений, например, распределений Стьюдента, Фишера и хи-квадрат.

Примеры применения доверительных границ для коэффициентов вариации при решении прикладных задач приведены, например, в наших работах, посвященных анализу технических характеристик и показателей

качества эластомерных (резиботехнических) материалов и изделий [13, 14].

Итак, сформулированы правила непараметрического оценивания обычно используемых характеристик распределения случайной величины. Эти правила основаны на асимптотических результатах теории вероятностей и математической статистики. Использование методов, разработанных в предположении нормальности распределения, может привести к заметно искаженным выводам в ситуации, когда гипотеза нормальности не выполнена. Практические рекомендации таковы: при анализе реальных данных следует использовать непараметрические доверительные границы.

Новая парадигма математических методов исследования [15 - 18] требует перехода от параметрических статистических методов к непараметрическим. Непараметрическая статистика - одна из точек роста современной прикладной статистики [19 - 21]. Непараметрические статистические методы являются важной составной частью перспективных математических и инструментальных методов контроллинга [22, 23].

Литература

1. Орлов А.И. Вероятность и прикладная статистика: основные факты: справочник. – М.: КноРус, 2010. – 192 с.
2. Орлов А.И. Современное состояние непараметрической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 106. С. 239–269.
3. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. Диагностика материалов. 1991. Т.57. №7. С.64-66.
4. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
5. Орлов А.И. О проверке однородности двух независимых выборок // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №1. С.55-60.
6. Орлов А.И. Неустойчивость параметрических методов отбраковки резко выделяющихся наблюдений. // Заводская лаборатория. Диагностика материалов. 1992. Т.58. №7. С.40-42.

7. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. - 416 с.
8. Орлов А.И. Теоретические инструменты статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 101. С. 253–274.
9. Орлов А.И. Метод моментов проверки согласия с параметрическим семейством распределений // Заводская лаборатория. Диагностика материалов. 1989. №10. С.90-93.
10. Орлов А.И. Проблема множественных проверок статистических гипотез // Заводская лаборатория. Диагностика материалов. 1996. Т.62. №5. С.51-54.
11. Холлендер М., Вулф Д. Непараметрические методы статистики. – М.: Финансы и статистика, 1983. - 518 с.
12. Боровков А.А. Математическая статистика. – М.: Наука, 1984. - 472 с.
13. Орлов А.И., Друянова Г.Б. Непараметрическое оценивание коэффициентов вариации технических характеристик и показателей качества // Надежность и контроль качества. 1987. №7. С.10-16.
14. Орлов А.И. Комментарий к заметке Ф.В. и В.Ф. Залесских «Об относительных ошибках двух или нескольких выборочных средних» // Заводская лаборатория. Диагностика материалов. 1989. Т.55. №3. С.101-102.
15. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. 2012. Том 78. №1, часть I. С.87-93.
16. Орлов А.И. Основные черты новой парадигмы математической статистики // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2013. – № 90. С. 45-71.
17. Орлов А.И. Новая парадигма математических методов экономики // Экономический анализ: теория и практика. 2013. № 36 (339). – С.25–30.
18. Орлов А.И. Новая парадигма анализа статистических и экспертных данных в задачах экономики и управления // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 98. С. 1254 – 1260.
19. Орлов А.И. Современная прикладная статистика // Заводская лаборатория. Диагностика материалов. 1998. Т.64. №3. С. 52-60.
20. Орлов А.И. Некоторые нерешенные вопросы в области математических методов исследования // Заводская лаборатория. Диагностика материалов. 2002. Т.68. №3. С.52-56.
21. Орлов А.И. Точки роста статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 136– 62.
22. Орлов А.И., Фалько С.Г. Экономико-математические методы в контроллинге // Экономическая теория, прикладная экономика и хозяйственная практика: проблемы эффективного взаимодействия: материалы Всероссийской научно-практической конференции. - Ярославль: ЯрГУ, 2006. – С.104-106.
23. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф. С.Г. Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с.

References

1. Orlov A.I. Verojatnost' i prikladnaja statistika: osnovnye fakty: spravocchnik. – М.: KnoRus, 2010. – 192 s.

2. Orlov A.I. Sovremennoe sostojanie neparametricheskoj statistiki // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 106. S. 239–269.
3. Orlov A.I. Chasto li raspredelenie rezul'tatov nabljudenij javljaetsja normal'nym? // Zavodskaja laboratorija. Diagnostika materialov. 1991. T.57. №7. S.64-66.
4. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie : uchebnik : v 3 ch. Ch.3. Statisticheskie metody analiza dannyh. - M.: Izd-vo MGTU im. N.Je. Baumana, 2012. - 624 s.
5. Orlov A.I. O proverke odnorodnosti dvuh nezavisimyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. №1. S.55-60.
6. Orlov A.I. Neustojchivost' parametricheskikh metodov otrakovki rezko vydelyajushhihsja nabljudenij. // Zavodskaja laboratorija. Diagnostika materialov. 1992. T.58. №7. S.40-42.
7. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoj statistiki. – M.: Nauka, 1983. - 416 s.
8. Orlov A.I. Teoreticheskie instrumenty statisticheskikh metodov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 101. S. 253–274.
9. Orlov A.I. Metod momentov proverki soglasija s parametricheskim semejstvom raspredelenij // Zavodskaja laboratorija. Diagnostika materialov. 1989. №10. S.90-93.
10. Orlov A.I. Problema mnozhestvennyh proverok statisticheskikh gipotez // Zavodskaja laboratorija. Diagnostika materialov. 1996. T.62. №5. S.51-54.
11. Hollender M., Vulf D. Neparametricheskie metody statistiki. – M.: Finansy i statistika, 1983. - 518 s.
12. Borovkov A.A. Matematicheskaja statistika. – M.: Nauka, 1984. - 472 s.
13. Orlov A.I., Drujanova G.B. Neparametricheskoe ocenivanie koeficientov variacii tehniceskikh harakteristik i pokazatelej kachestva // Nadezhnost' i kontrol' kachestva. 1987. №7. S.10-16.
14. Orlov A.I. Kommentarij k zametke F.V. i V.F. Zalesskikh «Ob odnositel'nyh oshibkah dvuh ili neskol'kih vyborochnyh srednih» // Zavodskaja laboratorija. Diagnostika materialov. 1989. T.55. №3. S.101-102.
15. Orlov A.I. Novaja paradigma prikladnoj statistiki // Zavodskaja laboratorija. Diagnostika materialov. 2012. Tom 78. №1, chast' I. S.87-93.
16. Orlov A.I. Osnovnye cherty novoj paradigmy matematicheskoj statistiki // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2013. – № 90. S. 45-71.
17. Orlov A.I. Novaja paradigma matematicheskikh metodov jekonomiki // Jekonomicheskij analiz: teorija i praktika. 2013. № 36 (339). – S.25–30.
18. Orlov A.I. Novaja paradigma analiza statisticheskikh i jekspertnyh dannyh v zadachah jekonomiki i upravlenija // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 98. S. 1254 – 1260.
19. Orlov A.I. Sovremennaja prikladnaja statistika // Zavodskaja laboratorija. Diagnostika materialov. 1998. T.64. №3. S. 52-60.
20. Orlov A.I. Nekotorye nereshennye voprosy v oblasti matematicheskikh metodov issledovanija // Zavodskaja laboratorija. Diagnostika materialov. 2002. T.68. №3. S.52-56.
21. Orlov A.I. Tochki rosta statisticheskikh metodov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 103. S. 136– 62.

22. Orlov A.I., Fal'ko S.G. Jekonomiko-matematicheskie metody v kontrollinge // Jekonomicheskaja teorija, prikladnaja jekonomika i hozjajstvennaja praktika: problemy jeffektivnogo vzaimodejstvija: materialy Vserossijskoj nauchno-praktičeskoj konferencii. - Jaroslavl': JarGU, 2006. – S.104-106.

23. Orlov A.I., Lucenko E.V., Lojko V.I. Perspektivnye matematicheskie i instrumental'nye metody kontrollinga. Pod nauchnoj red. prof. S.G. Fal'ko. Monografija (nauchnoe izdanie). – Krasnodar, KubGAU. 2015. – 600 s.