УДК 677.03:519.676

СТАТИСТИЧЕСКИЕ МЕТОДЫ ОЦЕНКИ КАЧЕСТВА: НОВЫЕ ВОЗМОЖНОСТИ

А.А. Науменко, И.С. Карпушенко

Процедура оценки качества нитей содержит два этапа: непосредственно испытания их с целью определения значений показателей соответствующих свойств и анализ полученных результатов. Как первый, так и второй этапы исключительно важны для надежного оценивания качества. В текстильной и легкой промышленности обработка результатов испытаний в обязательном порядке связана с применением математической статистики. Однако традиционные ее методы требуют выборок достаточно большого объема. Между тем, на практике не всегда имеется возможность в приемлемое время сформировать такие выборки. Ситуация, когда имеется сравнительно небольшое количество данных, по которым необходимо быстро сделать достаточно уверенные выводы или принять правильные оперативные решения, характерна для любого производства. Поэтому поиск приемов и способов извлечения информации из выборок малого или, по меньшей мере, уменьшенного объема (малых выборок) по-прежнему актуален, несмотря на то, что многое сделано в этом направлении.

Оценим на конкретном примере эффективность статистической обработки данных по обычной традиционной методике в случае малых выборок. Исходные данные представлены в табл. 1. В первой ее строке содержится значения разрывного удлинения полушерстяной пряжи линейной плотности 31,0 текс х 2.

Статистики, вычисленные традиционным способом, сведены в табл.2. Из нее следует, что относительная ошибка среднего δ меньше порогового значения, равного 5%. Следовательно, объем данной выборки достаточен для определения среднего с ошибкой, не превышающей допустимую.

Таблица 1

Показатель	Данные испытания																			
Разрывное удлинение	17,9	18,1	17,5	17,9	18,6	18,2	18,0	17,7	15,0	17,1	14,0	15,6	17,6	16,0	18,0	17,8	17,1	16,5	17,9	16,8
Нормированные плотность вероятности p(X)	0,072	0,059	0,068	0,072	0,033	0,046	0,069	0,074	0,009	0,052	0,003	0,017	0,070	0,021	0,069	0,074	0,052	0,032	0,072	0,041

Однако остается открытым вопрос о том, является ли полученное среднее несмещенной оценкой наиболее вероятного значения. На первый взгляд получение ответа на этот вопрос практические исключено, т.к. объем выборки явно недостаточен. Тем не менее, проведенные нами исследования показали, что такая возможность в принципе имеется. На чем она основана? Известно [1], что традиционные методы статистической обработки опираются на предположение о нормальном распределении случайной величины. Реальные распределения зачастую значительно отличаются от нормального. Так, в рассматриваемом примере оценка коэффициента асимметрии в силу существенного отклонения ее от нуля указывает на правостороннюю асимметрию функции распределения.

Таблица 2

Статистическая характеристика	Ед. изм.	Оценка статистик	Оценки статистик по предложенному алгоритму	Различие в %
среднее	мм	17,1	17,8	+4,1
среднее квадратическое отклонение	мм	1,2	0,5	-58,3
абсолютная ошибка среднего	мм	0,6	0,3	-57,1
относительная ошибка среднего	%	3,3	1,4	-57,6
доверительный интервал	мм	16,6	0,5	-58,3
коэффициент асимметрии	-	- 4,3	-0,7	-83,7
момент 5-го порядка	-	- 8,5	-2,9	-65,9

Статистической характеристикой, которая может «подтвердить» коэффициент асимметрии, является нечетный центральный момент более высокого порядка, - в данном случае момент 5-го порядка. Таблица 1 показывает, что вывод о правосторонней асимметрии подтверждается. Отметим особо, что центральный момент 5-го порядка более чувствителен к асимметрии, чем центральный момент 3-го порядка, т.е. коэффициент асимметрии. Именно асимметричность распределения ставит вопрос о соотношении полученной оценки среднего арифметического с оценкой наиболее вероятного значения.

В рамках традиционного подхода к построению оценок статистических характеристик изучаемой величины, ответа на этот вопрос мы не получим. Более того, мы его не получим, пока не построим оценку функции распределения. Однако при малых объемах выборок построение оценки функции распределения обычными методами не представляется возможным. Поэтому обратимся к другим подходам.

Первые работы по поиску алгоритмов построения оценок функций распределения по выборкам малого объема опубликованы в 1962 году [2]. Разработанные алгоритмы получили название эстиматоров (оценщиков). Важнейшее достоинство эстиматоров, определяющее возможность их практического применения, состоит в том, что минимальный объем выборки, необходимый для получения с их помощью оценок функции распределения, по меньшей мере, на порядок меньше того, который требуется в случае применения обычных алгоритмов. Так, если для построения относительно надежной гистограммы распределения необходимо не менее 100 единичных данных, то восстановление плотности распределения, например, по методу Парзена [2] возможно уже при наличии 10 единичных данных. Однако еще не все вопросы теории эстиматоров к настоящему времени получили свое решение. Тем не менее, несколько поступаясь математической строгостью, можно строить подходящие ответы на некоторые из них численными методами, опираясь на мощности современных вычислительных машин.

В проведенной работе в основу использованного подхода был положен эстиматор Парзена [2]. В рамках его, для получения оценок плотности распределения p*(X) случайной величины X используется выражение:

$$p^*_j(X) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}exp(-J_i)^2 \quad (j = 1,2...N) \qquad (1)$$

в котором принято, что

$$J_i = \frac{(X - x_i)^2}{2h^2}$$

Для вычисления p*(X) вводятся эмпирические оценки x_i величины X: i=1,2,3,...,n; при этом 10<n<250. Среди них отыскиваются x_{min} и x_{max}. Далее определяется начальное значение h: $h = x_{max} - x_{min}$. Количество вычисляемых значений N функции p*(X) может принимать теоретически любое значение, однако разумно положить N=50. Для случая n<50 принимается N=n.

Проблематичным в алгоритме Парзена является определение наилучшей величины h. Поэтому в проведенной работе для определения такого значения h использован метод последовательных приближений. Сущность его состоит в том, что в ряде последовательных циклов вычислений определяется N значений ординат функции p*(X) при установленной величине h. При переходе от цикла к циклу значение h уменьшается на заданный шаг. При этом отыскиваются абсциссы всех точек минимумов p*(X), разбивающих отрезок $[x_{max}, x_{min}]$ на некоторое число непересекающихся областей. Значения x_i, оказавшиеся в каждой из таких областей, и образуют кластер, минимальная численность которого задавалась предварительно. По мере уменьшения h происходит уточнение формы графика функции p*(X) и все более четкое разделение выборки на непересекающиеся группы – кластеры. Коррекция положения минимумов p*(X) возможна вплоть до обнаружения нулевых значений плотности p*(X). Это позволяет, последовательно уменьшая величину h, все более уверенно оценивать численность кластеров, сравнивая ее с величиной n_{min} и тем самым, делая все более обоснованным вывод о существовании или отсутствии моды, поддерживаемой каждым кластером. Уменьшение h ведет к росту величины J, вначале сравнительно медленному, а затем резко ускоряющемуся. По отношению к функции p*(X) это проявляется в быстро растущей детализации особенностей ее вида, что легко наблюдать по трансформирующемуся графику этой функции. По мере роста j все более четко выделяются области модальных значений, поддерживаемых кластерами достаточной численности, если такие области вообще существуют. При отсутствии в исходном массиве кластеров, порождающих моду при сравнительно большом значении h, нет оснований ожидать, что она появится при малой величине h. Это объясняется тем, что конечная форма оценки p*(X) при использовании выборки объемом n содержит n мод. Иными словами, при h>0 происходит лишь распад исходного массива на "кластеры", каждый из которых содержит единственный его элемент. Таким образом, неприемлемы как большие, так и чрезмерно малые значения h. При больших h разрешающая описанного алгоритма слишком мала, чтобы различить особенности p*(X). При малых h оценка p*(X) оказывается излишне детальной и опять не позволяет уловить главные особенности, так как они распадаются на массу мелких. Поэтому необходимо использовать такие значения h, при которых оценка p*(X) становится наиболее информативной в интересующем нас отношении. Многочисленные эксперименты на ПЭВМ показали, что целесообразно использовать значения h, приводящие к J=1500. Это значение J и использовано как пороговое, а условие J>1500 принято в качестве критерия окончания процедуры построения выборочной оценки плотности p*(X). При программной реализации алгоритма на ПЭВМ значение h в цикле, предшествующем тому, в котором условие J>1500 выполняется, принимается за окончательное.

С использованием найденного значения h рассчитаны оценки p*(X) плотности вероятности значений показателя качества пряжи из приведенного выше примера с использованием соотношения (1). Нормированные значения плотности p(X) получены по формуле:

$$p_j(X) = \frac{p^*_j(X)}{\sum\limits_{j=1}^{N} p^*_j(X)}$$

При этом $\sum\limits_{j=1}^{N} p^*_j = 1$

Значения $p_j(X)$ сведены во 2-ую строку таблицы 1. В графическом виде нормированная оценка плотности вероятности представлена на рис. 1.

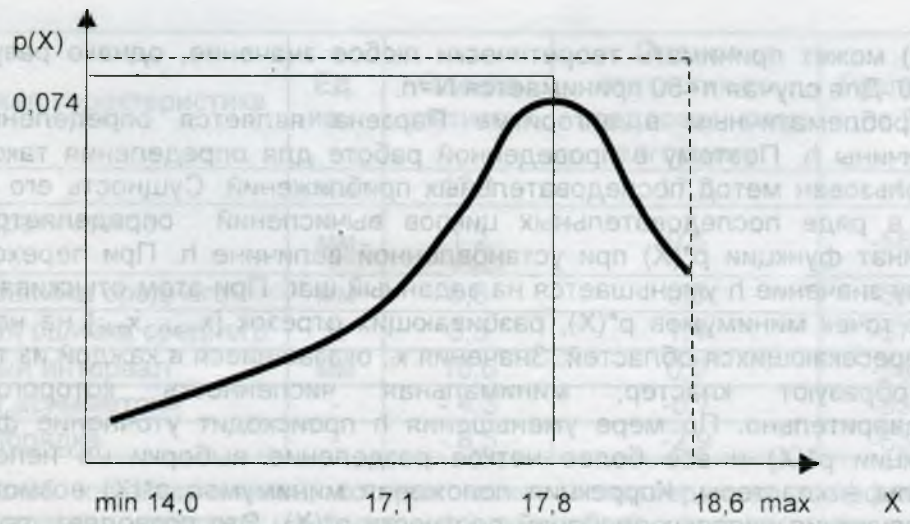

Рисунок 1 – Оценка плотности вероятности значений разрывного удлинения пряжи п/ш 31,0 текс x 2

Из него следует, что оценка плотности вероятности имеет одну моду с амплитудой 0,074, существенно сдвинутую от среднего в сторону больших значений.

Таким образом, среднее арифметическое оказывается смещенной оценкой наиболее вероятного значения оцениваемого показателя. Следовательно, в данном примере вычисление оценок других статистик должно производиться только с учетом полученной функции распределения. Общие соотношения, выражающие оценки статистик через значения функции распределения известны [1] и для дискретного случая в используемых в данной работе обозначениях имеют следующий вид.

Для математического ожидания m(x):

$$m(X) = \sum_{j=1}^{N} p_j(X)\, X_j \qquad (2)$$

Для дисперсии D(X):

$$D(x) = \sum_{j=1}^{N} p_j(X)\left[X_j - m(X)\right]^2 \qquad (3)$$

И для центрального момента r-того порядка:

$$\mu_r(x) = \sum_{j=1}^{N} p_j(X)\left[X_j - m(X)\right]^r / D^{0.5r} \qquad (4)$$

Используя значения pj (X) в формуле (1) для оптимального значения h легко вычислить значения $p(x_i)$, i = 1,2,...n и получить оценки статистик.

Эти оценки представлены в таблице 2.

Сравнение оценок статистик величины X, полученных при двух подходах показывает, что учет функции распределения даже при выборках малого объема существенно уточняет получаемые результаты.

При обычном, стандартном подходе выборочные значения x_i входят в оценки статистик с одинаковыми «весами», что делает оценки более грубыми. При использовании предложенной оценки функции плотности вероятности они «привязаны» не только к выборочным значениям, как таковым, но и к структуре выборки, отображенной функцией распределения. Поэтому дополнительные возможности статистического оценивания показателей качества связаны с использованием выборочных оценок функции распределения, построенных с учетом объема выборки. Для случая малых выборок подход к построению таких оценок и предложен в данной работе.

Список использованных источников

1. Виноградов Ю.С. Математическая статистика и ее применение в текстильной и швейной промышленности. – М.: Легкая индустрия, 1970. – 310 с.
2. Вапник В.Н. Восстановление зависимости по эмпирическим данным. – М.: Наука, 1979. – 448 с.

SUMMARY

In the given work on a concrete example the efficiency of statistical data processing of an estimation of quality on a traditional technique is considered in case of small quantity of the data.

The authors offer to use additional opportunities of a statistical estimation of parameters of quality connected with use of selective estimations of function of distribution sample, constructed in view of volume. At use of the offered estimation of function of density of probability will be coordinated not only to selective meanings, but also with structure of sample displayed function of distribution.

The comparison of estimations of statistical sizes of the casual size received at standard and offered approaches shows, that the account of function of distribution even at samples of small volume essentially specifies received results.

УДК 687.023.054

ИЗУЧЕНИЕ ПРОЦЕССА ВТО И ТЕПЛОЗАЩИТНЫХ СВОЙСТВ МАТЕРИАЛОВ ДЛЯ ВЕРХНЕЙ ОДЕЖДЫ С ИСПОЛЬЗОВАНИЕМ МЕТОДА СИМПЛЕКС-ПЛАНИРОВАНИЯ

М.А. Шайдоров, С.Г. Ковчур, З.Е. Ковчур

Влажно-тепловая обработка (ВТО) в процессе изготовления швейных изделий занимает около 30 % в общей трудоёмкости обработки изделий. На эксплуатационные свойства готовых изделий оказывает влияние множество факторов, в том числе гигиенические свойства [1,2]. Данная работа посвящена исследованию одному из важных гигиенических свойств – теплозащитных свойств.

Основными критериями оценки теплозащитных свойств являются суммарное тепловое сопротивление и коэффициент теплопроводности. Для определения этих показателей существуют несколько методик и установок. В настоящей работе предложена новая специальная экспериментальная установка, принципиальная схема которой представлена на рис. 1.

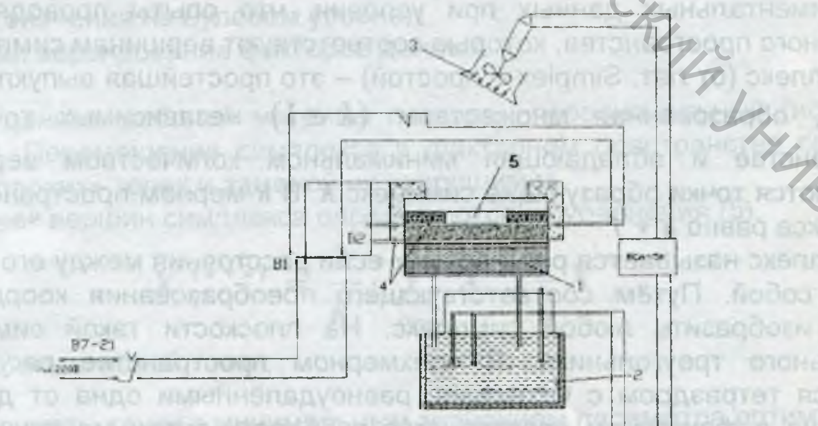

Рисунок 1 - Принципиальная схема экспериментальной установки