сообщения
объединенного
института
ядерных
исследований
дубна

5-83-244

В.И.Дорогов*, В.Н.Покровский, В.П.Чистяков*

ЭФФЕКТИВНОСТЬ ОЦЕНИВАНИЯ ДАННЫХ ПО НЕОДНОРОДНОЙ ВЫБОРКЕ

* Московский инженерно-физический институт

1983

В последнее время в ядерной физике возросли требования к гарантированной точности рекомендованных величин, полученных в результате обработки экспериментальных данных /см., например, [1,2]/. Точность оценок таких величин зависит от полноты информации о распределении погрешности измерений. Если метод оценивания основан на неточной априорной информации, то истинная точность оценки может быть значительно ниже теоретической. В связи с этим в работах последних лет много внимания уделяется методам, основанным на более достоверной, хотя и менее полной предварительной информации.

Распространенной ситуацией является присутствие в экспериментальных данных некоторой доли выбросов, полученных при различных нарушениях процесса измерения. Для довольно широкого класса измерений такого типа можно предположить, что измерения

$$x_1, x_2, \dots, x_n \qquad /1/$$

независимы, одинаково распределены, и плотность распределения каждого x_i определяется формулой

$$p(x, a, c) = (1-\epsilon)p_0\left(\frac{x-a}{\sigma_0}\right) + \epsilon p_0\left(\frac{x-c}{\sigma_1}\right), \qquad /2/$$

где $p_0(x)$ - плотность стандартного нормального распределения с параметрами /0,1/; ϵ - доля "выбросов" в /1/, вообще говоря, смещенных /при $c \neq a$ / относительно основной части распределения. Предполагается, что $\sigma_1 \gg \sigma_0$. Предположим далее, что параметры ϵ, σ_0, σ_1 известны и требуется оценить значение a.

В настоящей работе для случая $c = a$ сравниваются асимптотические дисперсии $\sigma^2 = \lim\limits_{n\to\infty} nM(a^*_n - a)^2$ нескольких оценок параметра a, получаемых различными методами; эти дисперсии могут служить мерой эффективности соответствующих оценок. Кроме того, оценивается влияние смещенности выбросов $(c \neq a)$ на асимптотические дисперсии в том случае, когда данные обрабатываются без учета такой смещенности. Так как теоретические результаты имеются не для всех рассматриваемых оценок, то выводы частично основываются на статистическом моделировании; такой подход в настоящее время широко распространен /см., например, обзор модельных экспериментов в задачах регрессии [3] /. Проведенные исследования позволяют дать некоторые предварительные рекомендации по выбору методов оценивания.

1. МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРА ПОЛОЖЕНИЯ "ЗАГРЯЗНЕННОЙ" ВЫБОРКИ

Располагая априорной информацией о наличии "загрязнения" в выборке, мы можем либо пренебречь этой информацией, либо брать устойчивые к наличию выбросов оценки, либо в той или иной мере учитывать эту информацию. Приведем ряд оценок параметра a, которые можно найти по выборке /1/ с плотностью распределения элементов выборки /2/ при $c = a$.

1.1. Выборочное среднее

Свойства оценки

$$a^*_W = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad /3/$$

хорошо известны /см., например,[4], гл.27/. Известно также, что эта оценка чувствительна к наличию выбросов. В рассматриваемом случае величина a^*_W асимптотически нормальна с параметрами $(a, \sigma^2_W/n)$, где

$$\sigma^2_W = (1-\epsilon)\sigma^2_0 + \epsilon\sigma^2_1 . \qquad /4/$$

1.2. Выборочная медиана

Расположив выборку /1/ в вариационный ряд $x_{(1)} \le \dots \le x_{(n)}$, найдем медиану

$$a^*_M = x_{([n/2]+1)}, \qquad /5/$$

где [ ]- целая часть числа. Оценка a^*_M асимптотически нормальна /[4], гл.28/ с параметрами $(a, \sigma^2_M/n)$, где

$$\sigma^2_M = \frac{1}{4p^2(a,a,a)} = \frac{\pi}{2}\left(\frac{1-\epsilon}{\sigma_0} + \frac{\epsilon}{\sigma_1}\right)^{-2}. \qquad /6/$$

Известно, что a^*_M - устойчивая оценка, но при отсутствии выбросов ее эффективность уступает эффективности a^*_W.

1.3. Оценка по урезанной выборке

Очевидный путь повышения устойчивости оценок к наличию "выбросов" состоит в том, чтобы основывать их на "центральной" части выборки, отбрасывая некоторые наблюдения. В случае выборки с обычным нормальным распределением чаще всего используются две группы критериев. Первая из них основана на проверке гипотезы

о равенстве средних и вычислении такой статистики, как размах или нормированное максимальное отклонение /см.,например,[5,6] /. Второй тип критериев основан на проверке гипотезы о том, что выборочная дисперсия не превышает некоторой заданной величины[7]. Для нашего случая, видимо, целесообразно воспользоваться этим подходом. Поэтому составим статистику $X^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/\sigma^2_0$,

которая при $\epsilon = 0$ имеет χ^2-распределение с $n-1$ степенью свободы, и зададимся уровнем значимости α. Если окажется, что X^2 превышает критическое значение $\chi^2_{\alpha, n-1}$, то из выборки /1/ выбрасывается значение, соответствующее $\max_i\{|x_i - \bar{x}|\}$. Для урезанной выборки находятся $\bar{x}$, X, и процедура повторяется до тех пор, пока выборка не окажется однородной. Теоретических результатов для оценки $a^*_{X,\alpha}$ /среднего по усеченной выборке/ нет, и величина асимптотической дисперсии $\sigma^2_{X,\alpha}$ будет оценена на модельных экспериментах.

1.4. Оценка Хубера

Еще одной возможностью повышения устойчивости оценок является уменьшение влияния больших отклонений от среднего. В частности, следуя Хуберу[8], в качестве оценки a^*_H параметра a следует принимать оценку, минимизирующую не $\sum_{i=1}^{n}(x_i - a)^2$, а более общую функцию $\sum_{i=1}^{n} F(x_i - a)$, где

$$F(x) = \begin{cases} x^2/2, & \text{если } |x| \le k, \\ k|x| - k^2/2, & \text{если } |x| > k. \end{cases}$$

Величина k есть то значение отклонения от среднего, по достижении которого вклад подобных отклонений уменьшается. Она может быть выбрана эвристически /по правилу типа "трех сигм"/, оценена численно /см., например,[9] /, а в рассматриваемом здесь случае может быть найдена как решение уравнения

$$(1-\epsilon)^{-1} = \int_{-k}^{k} p_0(x)\,dx + \frac{2p_0(k)}{k} . \qquad /7/$$

Получаемая оценка a^*_H асимптотически нормальна[6] с параметрами $(a, \sigma^2_H/n)$, где

$$\sigma^2_H = \frac{1}{2}\left(\int_0^k x^2 p(x,a,a)\,dx + k^2\int_k^\infty p(x,a,a)\,dx\right)\Big/\left(\int_0^k p(x,a,a)\,dx\right)^2 .$$

В рассматриваемом случае получаем

$$\sigma^2_H = \frac{1}{2}\,\frac{(1-\epsilon)A + \epsilon B + k^2[(1-\epsilon)C + \epsilon D]}{\left[(1-\epsilon)\Phi\left(\frac{k}{\sigma_0}\right) + \epsilon\Phi\left(\frac{k}{\sigma_1}\right) - \frac{1}{2}\right]^2} , \qquad /8/$$

где

1. МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРА ПОЛОЖЕНИЯ "ЗАГРЯЗНЕННОЙ" ВЫБОРКИ

Располагая априорной информацией о наличии "загрязнения" в выборке, мы можем либо пренебречь этой информацией, либо брать устойчивые к наличию выбросов оценки, либо в той или иной мере учитывать эту информацию. Приведем ряд оценок параметра a, которые можно найти по выборке /1/ с плотностью распределения элементов выборки /2/ при $c = a$.

1.1. Выборочное среднее

Свойства оценки

$$a^*_W = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad /3/$$

хорошо известны /см., например,/4/, гл.27/. Известно также, что эта оценка чувствительна к наличию выбросов. В рассматриваемом случае величина a^*_W асимптотически нормальна с параметрами $(a, \sigma^2_W/n)$, где

$$\sigma^2_W = (1-\epsilon)\sigma^2_0 + \epsilon\sigma^2_1 . \qquad /4/$$

1.2. Выборочная медиана

Расположив выборку /1/ в вариационный ряд $x_{(1)} \le \dots \le x_{(n)}$, найдем медиану

$$a^*_M = x_{([n/2]+1)}, \qquad /5/$$

где [ ]- целая часть числа. Оценка a^*_M асимптотически нормальна /4/, гл.28/ с параметрами $(a, \sigma^2_M/n)$, где

$$\sigma^2_M = \frac{1}{4p^2(a,a,a)} = \frac{\pi}{2}\left(\frac{1-\epsilon}{\sigma_0} + \frac{\epsilon}{\sigma_1}\right)^{-2}. \qquad /6/$$

Известно, что a^*_M - устойчивая оценка, но при отсутствии выбросов ее эффективность уступает эффективности a^*_W.

1.3. Оценка по урезанной выборке

Очевидный путь повышения устойчивости оценок к наличию "выбросов" состоит в том, чтобы основывать их на "центральной" части выборки, отбрасывая некоторые наблюдения. В случае выборки с обычным нормальным распределением чаще всего используются две группы критериев. Первая из них основана на проверке гипотезы

о равенстве средних и вычислении такой статистики, как размах или нормированное максимальное отклонение /см.,например,/5,6/ /. Второй тип критериев основан на проверке гипотезы о том, что выборочная дисперсия не превышает некоторой заданной величины/7/. Для нашего случая, видимо, целесообразно воспользоваться этим подходом. Поэтому составим статистику $X^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/\sigma^2_0$, которая при $\epsilon = 0$ имеет χ^2-распределение с $n-1$ степенью свободы, и зададимся уровнем значимости α. Если окажется, что X^2 превышает критическое значение $\chi^2_{\alpha,n-1}$, то из выборки /1/ выбрасывается значение, соответствующее $\max_i\{|x_i - \bar{x}|\}$. Для урезанной выборки находятся $\bar{x}$, X, и процедура повторяется до тех пор, пока выборка не окажется однородной. Теоретических результатов для оценки $a^*_{X,\alpha}$ /среднего по усеченной выборке/ нет, и величина асимптотической дисперсии $\sigma^2_{X,\alpha}$ будет оценена на модельных экспериментах.

1.4. Оценка Хубера

Еще одной возможностью повышения устойчивости оценок является уменьшение влияния больших отклонений от среднего. В частности, следуя Хуберу /8/, в качестве оценки a^*_H параметра a следует принимать оценку, минимизирующую не $\sum_{i=1}^{n}(x_i - a)^2$, а более общую функцию $\sum_{i=1}^{n} F(x_i - a)$, где

$$F(x) = \begin{cases} x^2/2, & \text{если } |x| \le k, \\ k|x| - k^2/2, & \text{если } |x| > k. \end{cases}$$

Величина k есть то значение отклонения от среднего, по достижении которого вклад подобных отклонений уменьшается. Она может быть выбрана эвристически /по правилу типа "трех сигм"/, оценена численно /см., например,/9/ /, а в рассматриваемом здесь случае может быть найдена как решение уравнения

$$(1-\epsilon)^{-1} = \int_{-k}^{k} p_0(x)\,dx + \frac{2p_0(k)}{k} . \qquad /7/$$

Получаемая оценка a^*_H асимптотически нормальна /6/ с параметрами $(a, \sigma^2_H/n)$, где

$$\sigma^2_H = \frac{1}{2}\left(\int_0^k x^2 p(x,a,a)\,dx + k^2\int_k^{\infty} p(x,a,a)\,dx\right)\Big/\left(\int_0^k p(x,a,a)\,dx\right)^2 .$$

В рассматриваемом случае получаем

$$\sigma^2_H = \frac{1}{2}\,\frac{(1-\epsilon)A + \epsilon B + k^2[(1-\epsilon)C + \epsilon D]}{\left[(1-\epsilon)\Phi\left(\frac{k}{\sigma_0}\right) + \epsilon\Phi\left(\frac{k}{\sigma_1}\right) - \frac{1}{2}\right]^2} , \qquad /8/$$

где

$$A = \sigma_0^2\Phi\left(\frac{k}{\sigma_0}\right) - \sigma_0^2\sqrt{\frac{k}{2\pi\sigma_0}}\; e^{-\frac{k}{\sigma_0}} - \frac{\sigma_0}{2},$$

$$B = \sigma_1^2\Phi\left(\frac{k}{\sigma_1}\right) - \sigma_1^2\sqrt{\frac{k}{2\pi\sigma_1}}\; e^{-\frac{k}{\sigma_1}} - \frac{\sigma_1^2}{2},$$

$$C = 1 - \Phi\left(\frac{k}{\sigma_0}\right),$$

$$D = 1 - \Phi\left(\frac{k}{\sigma_1}\right),$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

1.5. Оценка максимального правдоподобия

Эта оценка в полной мере использует априорную информацию и находится из уравнения правдоподобия

$$\frac{\partial \log L}{\partial a} = 0. \qquad /9/$$

Получаемая оценка a_L^* асимптотически нормальна с параметрами $(a, \sigma_L^2/n)$, где

$$\sigma_L^2 = \left\{ \int_{-\infty}^{\infty} \left[\left(\frac{\partial p(x,a,a)}{\partial a}\right)^2 / p(x,a,a) \right] dx \right\}^{-1}, \qquad /10/$$

причем σ_L^2 - минимальная возможная дисперсия. Явные формулы в нашем случае довольно громоздки, и мы их не выписываем.

Для иллюстрации сказанного выше приведем некоторые численные значения теоретических асимптотических дисперсий оценок a /см. табл.1/.

Таблица 1

Теоретические асимптотические дисперсии ($\sigma_0 = 1$, $\sigma_1 = 10$)

$100\cdot\varepsilon$	1	2	3	4	5	6	7	8	9	10
σ_W^2	1,99	2,98	3,97	4,96	5,95	6,94	7,93	8,92	9,91	10,9
σ_M^2	1,60	1,63	1,66	1,69	1,72	1,76	1,79	1,82	1,86	1,94
σ_H^2	1,06	1,10	1,15	1,19	1,23	1,27	1,31	1,35	1,39	1,43
σ_L^2	1,02	1,04	1,06	1,09	1,11	1,13	1,16	1,19	1,22	1,25

2. МОДЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ С НЕСМЕЩЕННЫМ ЗАГРЯЗНЕНИЕМ

Задавшись законом распределения /2/ при $c = a$ и используя генератор случайных чисел, получим N выборок вида /1/:

$$x_{i1}, x_{i2}, \ldots, x_{in}; \quad i = 1,2,\ldots,N. \qquad /11/$$

По i-ой выборке одним из описанных выше методов находим оценку a_i^* параметра a и вычислим величины $z_i = n(a_i^* - a)^2$, $i = 1,2,\ldots,N$. Тогда оценка s^2 асимптотической дисперсии σ^2 может быть найдена просто как среднее

$$s^2 = \frac{1}{N}\sum_{i=1}^{N} z_i, \qquad /12/$$

оценка B^* дисперсии Dz_i вычисляется как

$$B^* = \frac{1}{N-1}\sum_{i=1}^{N}(z_i - s^2)^2, \qquad /13/$$

а дисперсия Ds^2, т.е. точность вычисления σ^2 по формуле /12/ имеет порядок B^*/N.

Для ряда оценок можно провести исследование точности формулы /12/. Так, для метода выборочного среднего /раздел 1.1/ имеем

$$Mz_i = M[n(a_{W,i}^* - a)^2] = \frac{1}{n}M\left[\sum_{k=1}^{n}(x_k - a)^2\right] = \sigma_W^2,$$

$$Mz_i^2 = \frac{1}{n^2}M\left[\left(\sum_{k=1}^{n}(x_k - a)\right)^4\right] = \frac{1}{n^2}M\left[\sum_{k=1}^{n}(x_k - a)^4 + \right.$$

$$\left. + C_4^2\sum_{k\neq m}(x_k - a)^2(x_m - a)^2 + \ldots\right] = 3\sigma_W^4 + O\left(\frac{1}{n}\right).$$

Отсюда

$$Mz_i = \sigma_W^2, \quad Dz_i = 3\sigma_W^4 + O\left(\frac{1}{n}\right), \qquad /14/$$

и, следовательно,

$$Ms^2 = \sigma_W^2, \qquad Ds^2 \approx \frac{1}{N}2\sigma_W^4 \approx \frac{2(s^2)^2}{N}. \qquad /15/$$

Аналогичное исследование для оценок максимального правдоподобия можно провести, если воспользоваться сходимостью моментов величины $(a_L^* - a)\sqrt{n}$ к моментам нормального распределения /см. /10//.

На ЭВМ ЕС-1060 были получены выборки типа /11/ с $n = 100$, $N = 400$ при $\sigma_0 = 1$ и $\sigma_1 = 3, 10, 100$. Находились оценки, описанные в разделах 1.1-1.5 и по формуле /12/ определялись оценки $\tilde\sigma_W^2$, $\tilde\sigma_M^2$, $\tilde\sigma_H^2$, $\tilde\sigma_{X,\alpha}^2$ и $\tilde\sigma_L^2$ соответствующих асимптотических дисперсий. Результаты вычислений приведены в табл.2.

Таблица 2

Оценки асимптотических дисперсий, найденные в модельных экспериментах ($\sigma_0=1, \sigma_1=3, 10, 100$)

$100 \cdot \varepsilon$		1	2	3	4	5	6	7	8	9	10
$\tilde{\sigma}^2_W$	$\sigma_1=3$	1,0	1,1	1,3	1,5	1,4	1,5	1,7	1,7	1,6	1,8
	$\sigma_1=10$	2,0	3,0	4,4	5,0	6,2	6,1	8,3	9,1	9,1	10,2
	$\sigma_1=100$	100	210	350	380	510	600	720	940	820	1100
$\tilde{\sigma}^2_M$	$\sigma_1=3$	1,4	1,7	1,8	1,8	1,8	1,8	1,6	1,8	1,8	1,8
	$\sigma_1=10$	1,5	1,6	1,9	1,8	1,7	1,7	1,7	1,7	1,8	1,8
	$\sigma_1=100$	1,6	1,7	1,7	1,7	1,8	1,6	2,0	1,9	1,6	2,0
$\tilde{\sigma}^2_H$	$\sigma_1=3$	1,0	1,1	1,1	1,3	1,2	1,2	1,2	1,3	1,2	1,4
	$\sigma_1=10$	1,0	1,1	1,4	1,2	1,3	1,2	1,3	1,4	1,3	1,4
	$\sigma_1=100$	1,1	1,1	1,2	1,2	1,3	1,2	1,5	1,4	1,3	1,5
$\tilde{\sigma}^2_{X,\alpha}$	$\sigma_1=3$	1,0	1,1	1,1	1,3	1,2	1,2	1,2	1,3	1,1	1,4
	$\sigma_1=10$	1,0	1,1	1,3	1,1	1,2	1,0	1,2	1,2	1,1	1,2
$\alpha=0,01$	$\sigma_1=100$	1,0	1,0	1,0	1,0	1,1	1,0	1,2	1,1	1,0	1,2
$\tilde{\sigma}^2_{X,\alpha}$	$\sigma_1=3$	1,0	1,1	1,1	1,4	1,2	1,2	1,3	1,4	1,2	1,5
	$\sigma_1=10$	1,0	1,2	1,3	1,2	1,3	1,1	1,2	1,2	1,1	1,3
$\alpha=0,05$	$\sigma_1=100$	1,0	1,0	1,0	1,0	1,1	1,0	1,2	1,1	1,0	1,2
$\tilde{\sigma}^2_L$	$\sigma_1=3$	1,0	1,0	1,1	1,3	1,1	1,2	1,2	1,3	1,1	1,4
	$\sigma_1=10$	1,0	1,1	1,3	1,1	1,2	1,0	1,1	1,2	1,1	1,2
	$\sigma_1=100$	1,0	1,0	1,0	1,0	1,1	1,0	1,2	1,0	1,0	1,2

3. ВЛИЯНИЕ СМЕЩЕННОСТИ ВЫБРОСОВ

Для оценки влияния возможной смещенности выбросов были получены выборки, соответствующие распределению /2/ при $c \neq a$. Эти данные обрабатывались так же, как и в разделе 2, т.е. без учета смещенности. Исключение составила оценка максимального правдоподобия $\hat{a}_L$, полученная для истинного распределения /2/. Результаты вычислений для нескольких значений "параметра смещения" $C=(c-a)/\sigma_0$ приведены в табл.3.

Таблица 3

Оценки асимптотических дисперсий ($\sigma_0=1$; $\sigma_1=3$; смещения $C=3$; 9)

$100 \cdot \varepsilon$		1	2	3	4	5	6	7	8	9	10
$\tilde{\sigma}^2_W$	C= 3	1,0	1,6	2,0	3,0	3,8	5,0	6,3	7,2	8,6	9,9
	C= 9	1,9	4,5	8,8	15	22	30	41	54	67	82
$\tilde{\sigma}^2_M$	C= 3	1,4	1,7	1,9	1,9	2,0	2,1	2,0	2,3	2,1	2,6
	C= 9	1,5	1,7	2,0	2,0	2,0	2,3	2,6	2,7	2,9	3,4
$\tilde{\sigma}^2_H$	C= 3	1,0	1,1	1,2	1,5	1,5	1,7	1,8	2,0	1,9	2,4
	C= 9	1,1	1,2	1,6	1,6	1,9	2,1	2,6	3,0	3,4	3,8
$\tilde{\sigma}^2_{X,\alpha}$	C= 3	1,0	1,0	1,2	1,4	1,4	1,4	1,4	1,6	1,3	1,7
$\alpha=0,01$	C= 9	1,0	1,1	1,3	1,0	1,2	1,1	1,2	1,2	1,1	1,2
$\tilde{\sigma}^2_{X,\alpha}$	C= 3	1,0	1,1	1,3	1,5	1,6	1,7	1,7	1,9	1,6	1,9
$\alpha=0,05$	C= 9	1,0	1,2	1,3	1,2	1,4	1,2	1,4	1,3	1,2	1,3
$\tilde{\sigma}^2_L$	C= 3	1,0	1,0	1,1	1,3	1,1	1,2	1,2	1,3	1,0	1,4
	C= 9	1,0	1,1	1,2	1,0	1,1	1,0	1,2	1,1	1,0	1,1

4. НЕКОТОРЫЕ ВЫВОДЫ

Из рассмотрения табл.2 и 3 видно, что наилучшие результаты дает метод максимального правдоподобия. Однако практическое его использование затруднительно, так как обычно нелегко сделать определенное заключение о виде функции распределения выбросов. Поскольку дисперсия оценок σ^2_W и σ^2_M существенно больше дисперсии оценок σ^2_X и σ^2_H, предпочтительно использовать метод усечения выборки или метод Хубера. Их сравнительные характеристики приведены в табл.4.

Таблица 4

	Метод Хубера	Метод усечения выборки
Теоретическое обоснование	Хорошее	Отсутствует
Алгоритм	Сложный	Простой
Устойчивость к возможному смещению	При очень больших смещениях ниже, чем при малых	Есть при любых смещениях
Отношения σ_L^2/σ^2 /эффективность/ при $\epsilon \cdot 100 = 1;5;10$ /$\sigma_0 = 1$; $\sigma_1 = 10$/	1,0; 0,91; 0,88	1,0; 0,91;0,91 /оцененная, $\alpha = 0{,}05$/

ЛИТЕРАТУРА

1. Conclusion and Recommendations. IAEA Advisory Group Meeting on Fission Products Nuclear Data. IAEA-213, 1977, vol.2, p.743.
2. Пасечник М.В. В кн.: Нейтронная физика /материалы 5-й Всесоюзной конференции по нейтронной физике, Киев/. Изд-во ФЭИ, Обнинск, 1980, ч.3, с.3.
3. Демиденко Е.З. Линейная и нелинейная регрессия. "Финансы и статистика", М., 1981, гл.5,6.
4. Крамер Г. Математические методы статистики. "Мир", М., 1975.
5. Кендалл М.Дж., Стюарт А. Статистические выводы и связи. "Наука", М., 1973, гл.32.
6. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. "Наука", М., 1965, с.89.
7. Леман Э. Проверка статистических гипотез. "Наука", М., 1979, гл.3,5.
8. Huber P.J. Ann.Math.Stat., 1964, vol.35, p.73.
9. Волков Н.Г., Дорогов В.И. В сб.: Экспериментальные методы ядерной физики. "Энергоиздат", М., 1981, № 9, с.3.
10. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. "Наука", М., 1979.

Рукопись поступила в издательский отдел 14 апреля 1983 года.

Дорогов В.И., Покровский В.Н., Чистяков В.П. 5-83-244

Эффективность оценивания данных по неоднородной выборке

Для неоднородной выборки /в присутствии некоторой доли выбросов/ в работе сравниваются асимптотические дисперсии различных оценок параметра положения, получаемых по методам выборочного среднего, выборочной медианы, усечения выборки, методу Хубера и методу максимального правдоподобия. Оценивается также влияние смещенности выбросов на асимптотические дисперсии в том случае, когда данные обрабатываются без учета такой смещенности. Даются некоторые предварительные рекомендации по выбору методов оценивания, основанные на результатах статистического моделирования.

Работа выполнена в Лаборатории ядерных проблем ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1983

Dorogov V.I., Pokrovsky V.N., Chistyakov V.P. 5-83-244

Efficiency of Data Estimation on Inhomogeneous Sample

The asymptotic dispersions of various estimates of a location parameter for an inhomogeneous sample (with some fraction of outliers) have been compared, the estimates being obtained by methods of sampling mean, sampling median, trimmed sample, Huber and maximum likelihood. The effect of outliers bias on asymptotic dispersions is also estimated for the case of data handling without taking into account this bias. Tentative recommendations on choice of estimation methods are given based on the results of statistics model trials.

The investigation has been performed at the Laboratory of Nuclear Problems, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1983

Перевод О.С.Виноградовой.