



МА́ЛЫЕ ВЫ́БОРКИ

Авторы: Ю. В. Прохоров

МА́ЛЫЕ ВЫ́БОРКИ, статистич. выборки малого объёма

n такие, что к ним нельзя применять формулы, справедливые лишь при

$n \rightarrow \infty$. Особенности статистич. оценивания параметров по М. в. можно понять на примере [нормального распределения](#) (с параметрами

a и

σ^2), для которого малыми обычно считают выборки объёма

$n \leq 30$. Пусть по выборке

X_1, X_2, \dots, X_n из нормальной совокупности необходимо оценить её среднее значение

a , причём дисперсия

σ^2 неизвестна. Для этого вводятся статистики (функции от результатов наблюдений)

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k, s^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

и при оценке

a используется то, что распределение вероятностей величины

$$t = \sqrt{n}(\bar{X} - a)/s$$

не зависит от

a и

σ . Вероятность

ω неравенства

$-t_\omega < t < t_\omega$ и равносильного ему неравенства

$$\bar{X} - t_\omega s / \sqrt{n} < a < \bar{X} + t_\omega s / \sqrt{n}$$

вычисляется по формуле

$$\omega = \int_{-t_\omega}^{t_\omega} s_{n-1}(t) dt,$$

где

$s_{n-1}(t)$ есть плотность вероятности [Стюдента распределения](#) с

$n - 1$ степенями свободы. Определяя для заданных

n и

ω ,

$0 < \omega < 1$, соответствующее

t_ω , получают правило (1) нахождения доверительных границ для величины

a , имеющих значимости уровень

ω .

При больших

n формула (2), связывающая

ω и

t_ω , может быть заменена приближённой формулой

$$\omega = \int_{-t_\omega}^{t_\omega} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt,$$

где подынтегральное выражение есть плотность стандартного нормального распределения, которое является предельным при

$n \rightarrow \infty$ для упомянутых распределений Стьюдента. Эту формулу иногда неправильно применяют для определения

t_ω при небольших

n , что приводит к существенным ошибкам. Так, для

$\omega = 0,99$ из формулы (3) следует, что

$t_{0,99} = 2,58$; истинные значения

$t_{0,99}$ для малых

n приведены в следующей таблице:

n	2	3	4	5	10	20	30
$t_{0,99}$	63,66	9,92	5,84	4,63	2,52	2,86	2,76

Если пользоваться формулой (3) при

$n = 5$, то получится, что неравенство

$$\left| \bar{X} - a \right| \leq 2,58 \frac{s}{\sqrt{5}}$$

выполняется с вероятностью 0,99. В действительности при

$n = 5$ вероятность этого неравенства равна 0,94, а вероятность 0,99 имеет, в соответствии с приведённой таблицей, неравенство

$$\left| \bar{X} - a \right| \leq 4,60 \frac{s}{\sqrt{5}}.$$

Об оценке теоретич. дисперсии

σ^2 по М. в. см. Хи-квадрат распределение. Разработаны также методы оценки по М. в. параметров многомерных распределений (напр., коэффициента корреляции).

Литература

Лит.: Колмогоров А. Н. Определение центра рассеивания и меры точности по ограниченному числу наблюдений // Известия АН СССР. Серия математическая. 1942. Т. 6. № 1–2; Крамер Г. Математические методы статистики. 2-е изд. М., 1975; Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. 3-е изд. М., 1983.

Processing math: 100%