УДК 658.562

СОВЕРШЕНСТВОВАНИЕ МЕТОДИКИ ПОСТРОЕНИЯ ГИСТОГРАММЫ

Клочков Ю.С., Болдырева И.В.
Научный руководитель – д.т.н., профессор Чекмарев А.Н.
Самарский государственный аэрокосмический университет имени академика С.П. Королева

При оценивании энтропии встает вопрос о корректном построении гистограммы. Существует значительное число работ по выбору оптимального числа столбцов гистограммы. Но и само построение гистограммы при одинаковом числе столбцов может привести к различным расчетным величинам.

Имеется методика, которая заключается в следующем:

1. Определить наибольшее L и наименьшее S значения из представленной выборки.
2. Интервал между L и S разделить на соответствующее количество столбцов гистограммы k.
3. Вычисляют ширину столбца. Разность между L и S делят на число столбов h=(L-S)/k.
4. Определяют границы столбцов. Для того чтобы определить первую границу первого столбца необходимо вычислить следующее выражение (S – (единица измерения/2)).

В данной методике можно выделить ряд недостатков. Так, например, смещение первой границы на половину единицы измерения не учитывается при определении шага. Данное смещение необходимо, чтобы значения выборки не попадали на границы. Второй недостаток заключается в том, что возможны случаи, когда теряется максимум в центре. Это происходит из-за того, что крайние варианты равные нулю не учитываются и построение ведется с первой обнаруженной варианты. Хотя в ряде случаев было бы выгоднее определить первую границу с нулевой варианты, чтобы сохранить максимум в центре гистограммы.

Для решения этих проблем предлагается иной метод построения гистограммы. Он отличается тем, что:

1. Определить наибольшее L и наименьшее S значения из представленной выборки.
2. Интервал между L и S разделить на соответствующее количество столбцов гистограммы k.
3. Ширина столбца определяется следующим образом h=(L-S+ единица измерения)/k. Это позволит получить крайние границы гистограммы выходящие за максимальную и минимальную величину на половину единицы измерения.
4. При нечетном k первую границу определим как: 1граница=среднее – h/2 – h(k -1)/2. При четном k: 1граница=среднее – h(k/2). Это позволит сохранить максимум в середине гистограммы и учесть нулевые варианты.

Как видно возможны на практике случаи, когда при одинаковом количестве столбцов, но различных методах группирования получаются различные выводы о виде закона распределения. Предлагаемая методика наиболее подходит, когда выдвигается гипотеза о нормальном распределении, так как основной идеей такого группирования данных является то, что среднее значение совпадает с максимумом дифференцированной функции.