

Об анализе однородности выборок (проф. Ивановский Р.И., СПбПУ)

В практике статистической обработки наблюдений нередко возникает необходимость объединения нескольких массивов выборочных данных. Такая ситуация может быть связана, например, с наличием разрозненных наблюдателей одного явления, когда объем данных, полученных каждым из наблюдателей, недостаточен для получения объективных оценок.

Стремление повысить объем выборочных данных интуитивно понятно, но не все выборки можно совместно использовать при обработке данных. Нужна предварительная проверка частичных массивов на однородность.

Однородными в математической статистике называют выборки, законы распределения которых одинаковы или близки между собой. Эта проверка проводится в форме решения задачи проверки статистической гипотезы.

Классическая процедура такой проверки предполагает предварительное объединение сопоставляемых выборок в единую выборку, определение общего диапазона ее элементов (встроенные функции \max , \min), назначение b интервалов (разрядов) внутри диапазона, определение частот попаданий (встроенные функции hist) элементов выборок в эти разряды.

Отметим здесь, что процедура проверки однородности не обязательно должна быть связана с объединением выборок, поскольку вектор частот единой выборки всегда равен сумме векторов частот ее составных частей.

Мера близости выборочных (эмпирических) распределений в таких задачах определяется уровнем значимости и свойствами применяемого критерия.

Проверка однородности выборок может осуществляться с применением различных критериев. Среди них распространение получил критерий хи-квадрат. К его практически важным свойствам отнесем возможность анализа однородности нескольких выборок одновременно, слабая зависимость от их объема, отсутствие повышенной строгости к различию свойств сопоставляемых выборок. Последнее способствует увеличению числа положительных исходов проверки.

Только закон распределения случайной величины служит, как известно, исчерпывающей ее характеристикой. Поэтому равенство (близость) числовых характеристик распределений (математических ожиданий, дисперсий и проч., а также их выборочных аналогов) являются лишь необходимыми условиями однородности. Проверка на основе этих необходимых условий способна подтвердить лишь субоднородность выборок. Однако задачи подобного типа имеют право на свое существование. Например, близость выборок по локальным параметрам может потребоваться при формировании контрольных групп объектов (приборов, людей и проч.) для последующего выявления свойств группы и их сопоставления со свойствами других групп. При этом выборка каждого из участников группы может формироваться из значений какого-либо индикатора, а мера близости выборок – может определяться не только на основе числовых характеристик распределений, но и путем сопоставления форм этих распределений. В последнем случае, учитывая естественную разнородность числовых характеристик выборок для участников групп, единым началом, объединяющим группу, может служить нормированное распределение индикатора. Это будет означать, что общность выборок определяются лишь близостью форм их законов распределения. Подобную однородность можно назвать «формальной» (от «формы»). Этот случай также служит подвидом субоднородности.