

АНАЛИЗ МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ - раздел математической статистики (см.), предназначенный для анализа связей между тремя и более переменными. Можно условно выделить три основных класса задач А.М.С. Это исследование структуры связей между переменными (см.) и снижение размерности пространства признаков, построение классификаций и типологий, исследование причинных связей.

Для представления структуры связей между переменными обычно используется матрица корреляций (см.).

Ее анализ, заключающийся в выделении подмножеств переменных, тесно коррелирующих друг с другом, может осуществляться "вручную", например, с помощью графа, отражающего наиболее существенные связи между переменными, либо методами компьютерного анализа, такими, как метод главных компонент, факторный анализ, кластерный анализ переменных. Анализ структуры связей часто рассматривается в качестве самостоятельной задачи, например, при исследовании структуры ценностей, мотивов и т.п., для проверки психометрических шкал на надежность и в других случаях. Однако он может использоваться и в качестве промежуточного этапа при решении задачи снижения размерности пространства признаков.

Снижение размерности обычно применяется для построения пространства, более удобного для решения задач классификации и исследования причинных связей, чем исходный набор переменных.

Задача снижения размерности заключается в том, чтобы от большого количества исходных переменных перейти к нескольким обобщенным показателям. Метод главных компонент (см.), анализ факторный (см.), метод многомерного шкалирования предусматривают для этого разнообразные процедуры.

Задачи и методы классификации, в зависимости от условий, делятся на три группы: классификация по заданным формальным критериям, автоматическая классификация и классификация с обучением. Классификация по заданным критериям, строго говоря, не является статистическим методом.

Она состоит в группировке объектов по одному или нескольким показателям. В последнем случае классификация называется перекрестной или лингвистической (например, половозрастная структура населения).

Автоматическую классификацию применяют в тех случаях, когда критерии группировки неизвестны и отсутствуют априорные представления о количестве и характере классов. Для ее построения используются методы анализа кластерного (см.), позволяющие выделить группы объектов, близких друг к другу по значениям измеренных переменных. В основе кластерного анализа лежит вычисление расстояний между объектами.

Классификация с обучением применяется, когда критерии классификации неизвестны, но известно количество классов и их типологические особенности. В этом случае может быть сформирована так называемая выборка обучающая (см.), состоящая из реальных объектов, обладающих соответствующими характеристиками, или/и искусственных объектов - моделей "типичных представителей" классов.

В обучающей выборке должны присутствовать "представители" всех предполагаемых классов. Классификация конкретного объекта состоит в том, что вычисляется расстояние между ним и объектами из обучающей выборки и объект причисляется к тому классу, расстояние до которого для него оказалось минимальным. Классификация с обучением осуществляется некоторыми методами кластерного и дискриминантного анализа.

Анализу статистических причинных связей в последние годы уделяется особое внимание. Классическим методом для решения таких задач является дисперсионный анализ, в основе которого лежит эксперимент факторный (см.) (не путать с анализом факторным - см.).

Начиная с 1960-х активно разрабатываются регрессионные и регрессионно-подобные причинные модели (см. Каузальное моделирование), а также техники, позволяющие использовать в этих моделях не только "количественные", но и "качественные" переменные (см. Dummy-кодирование). В настоящее время для исследования причинных связей, в зависимости от характера используемых переменных, применяются методы множественной линейной регрессии, логистической регрессии, дискриминантного анализа и т.п. Эти методы предполагают наличие единственной зависимой переменной и не позволяют исследовать структуру связей между независимыми переменными (предикторами). Структура связей между предикторами может быть учтена в моделях анализа путевого (см.).

Наиболее общим является метод линейных структурных уравнений (см.), позволяющий строить сложные модели с большим числом взаимодействующих между собой зависимых и независимых переменных, среди которых могут быть не только наблюдаемые, но и латентные признаки. Регрессионный, дисперсионный, путевой и факторный анализ являются его частными случаями.

О.В. Терещенко

АНАЛИЗ МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ

Автор: словарь
02.09.2007 10:38 -
