АНАЛИЗ ЛОГЛИНЕЙНЫЙ

Автор: словарь
02.09.2007 09:36 -

АНАЛИЗ ЛОГЛИНЕЙНЫЙ (логарифмически-линей-ный) - метод анализа категориальных (номинальных) данных, позволяющий исследовать отношения между переменными в таблицах сопряженности; обобщение многомерного анализа частот. Предположений о распределении переменных не делается. А.Л. может рассматриваться как непараметрический аналог многофакторного дисперсионного анализа с дискретным измерением. В отличие от последнего все переменные, используемые для классификации, рассматриваются как независимые, а зависимой переменной является логарифм количества наблюдений в ячейке таблицы сопряженности.

Частный случай А.Л., в котором отношения моделируются в привычной манере, когда в качестве зависимой выступает измеренная переменная, называется логит-анализом; обычно в этом случае исследователи предпочитают логистическую регрессию, поскольку она обладает большей гибкостью. Общая модель А.Л. для двух переменных

X и Y имеет вид:

$\ln(m_{ij}) = \mu + \gamma_{Xi} + \gamma_{Yi} + \gamma_{XYij}$,

где $\ln(m_{ij})$ - натуральный логарифм ожидаемой частоты в i, j-й ячейке таблицы сопряженности, μ - константа, общая для всех ячеек (она равна среднему натуральному логарифму ожидаемых частот во всех ячейках таблицы), γ_{Xi} и γ_{Yi} - соответственно эффект i-й категории переменной X и j-й категории переменной Y, γ_{XYij} - эффект взаимодействия i-й и j-й категорий двух переменных. Название метода связано с тем, что логарифм ожидаемой частоты представляется в виде линейной функции значений переменных. Приведенная модель называется насыщенной (полной), поскольку включает все возможные эффекты и описывает данные совершенным образом, без ошибок.

Как правило, общая цель А.Л. состоит в том, чтобы найти наиболее простое описание данных или проверить гипотезу об адекватности этой простой (минимальной) модели данным.

Адекватность оценивают с использованием статистик пригодности, обычно с помощью отношения функций максимального правдоподобия: $G^2 = 2 \sum f \ln(f \square m)$, где f и m - наблюдаемые и ожидаемые (с точки зрения проверяемой модели) частоты. Эта статистика обладает свойством аддитивности и имеет распределение χ^2 с числом степеней свободы, определяемым количеством оцениваемых параметров. Статистика пригодности оценивает степень расхождения модели и эмпирических данных, поэтому

ее большое, статистически значимое в привычном смысле значение свидетельствует о непригодности модели.

Простые модели получают удалением отдельных эффектов из насыщенной модели. Удаление последнего эффекта (взаимодействия) из приведенного уравнения позволяет проверить гипотезу о независимости переменных (отсутствии взаимодействия). Особым классом моделей А.Л. являются иерархические (вложенные) модели: наличие взаимодействия более высокого порядка (напр., &gamma

;XYZ) предполагает наличие всех взаимодействий более низкого порядка и главных эффектов (λXY, γXZ, γYZ, γX, γY, γZ). Иерархические модели удобны тем, что для проверки их пригодности можно использовать разность между G^2 для двух иерархически вложенных моделей, и эта разность также распределена как χ^2. Для сравнения неиерархических моделей статистические критерии, как правило, не используются. Для оценки важности отдельных эффектов используют также стандартизованные значения оцененных параметров, сравнивая их с критическими значениями единичного нормального распределения.

Как непараметрический метод А.Л. имеет широкий диапазон пригодности.

Содержательным ограничением является количество независимых переменных, поскольку взаимодействия высокого порядка трудно интерпретировать. Формальные ограничения связаны с требованием независимости частот в ячейках (одно и то же наблюдение не должно входить в несколько ячеек) и минимальным числом наблюдений в ячейках таблицы сопряженности. Как и в случае χ^2 , все ожидаемые частоты в ячейках должны быть больше 1, и не более 20% ожидаемых частот могут быть менее 5.

Нарушение этого требования приводит к возрастанию вероятности ошибки первого рода. Прибавление к каждой ячейке некоторого небольшого числа (операция, выполняемая автоматически в некоторых статистических пакетах) не способствует стабилизации ошибки первого рода и к тому же уменьшает мощность метода. Для сходимости анализа рекомендуется выборка, объем которой не менее чем в 5 раз превышает количество ячеек таблицы сопряженности.

А.Л. находит широкое применение в анализе таблиц социальной мобильности, в политологии и маркетинговых исследованиях.

С.В. Сивуха

Автор: словарь
02.09.2007 09:36 -