

УДК 004

ПРИМЕНЕНИЕ МНОЖЕСТВЕННОГО ЛИНЕЙНОГО РЕГРЕССИОННОГО
АНАЛИЗА ДАННЫХ ДЛЯ ВЫЯВЛЕНИЯ ВЗАИМОСВЯЗИ
МЕЖДУ ПИЩЕВЫМИ ПРИВЫЧКАМИ И СОСТОЯНИЕМ ЗДОРОВЬЯ

Е. А. ЗАЙЧЕНКО, Э. А. КАПЕЛЬКО
Белорусско-Российский университет
Могилев, Беларусь

Множественный линейный регрессионный анализ дает возможность исследовать связи между одной зависимой переменной и некоторым количеством независимых переменных, параметры определяются методом наименьших квадратов, вместо линии регрессии используется гиперплоскость. Это позволяет значительно увеличить коэффициент детерминации модели и исходного массива данных.

В качестве генеральной совокупности в представленной работе был использован массив данных «Питание и здоровье» (The Eating & Health Module Dataset), в котором собраны сведения о том, как люди оценивают самочувствие, сколько времени они тратят на занятия спортом, досуг, работу, какова структура их рациона. Массив имеет достаточный объем для проведения исследований.

При проведении множественного линейного регрессионного анализа данных важно выбрать из множества переменных зависимую переменную. Такой зависимой переменной была выбрана переменная, содержащая оценку здоровья респондентом в диапазоне от «отлично» до «плохо», а переменные, содержащие более объективные данные о весе респондентов, особенностях рациона, количестве различных категорий продуктов и напитков, величине дохода, времени, затраченном на занятия спортом (всего более 20 показателей), рассматриваются как независимые переменные.

Язык программирования python имеет широкие возможности для проведения математического и статистического анализа данных (пакеты pandas, sklearn, scipy). С помощью перечисленных средств и графической библиотеки matplotlib выполнено построение матрицы корреляций между первыми по значимости индексами анализируемого набора данных.

Обнаружены взаимосвязи, не определенные в данных целенаправленно и поддающиеся непредвзятой интерпретации. Так, построенный график показывает отрицательную корреляцию между индексом массы тела и оценкой самочувствия респондентов. Как и ожидалось, время, затраченное на физическую активность, имеет ярко выраженную положительную корреляцию с самочувствием человека. Было обнаружено, что участники опроса, потребляющие больше мяса, птицы и рыбы, имеют более высокую оценку состояния здоровья. Однако было выявлено, что наименее влияющие на самочувствие факторы – потребление газированных напитков и длительность сна.

На основании полученных результатов можно сделать вывод о наличии статистически значимой связи между пищевыми привычками и состоянием здоровья.