

Это лишь несколько примеров областей применения регрессионного анализа в онкологии. Регрессионный анализ может быть использован для исследования и анализа различных аспектов заболевания и поможет в более эффективной борьбе с раком.

Определение логистического регрессионного анализа

Логистическая регрессия есть инструмент для решения задач регрессии и классификации в онкологии, представляет из себя разновидность множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми регрессорами или предикторами и зависимой переменной. Она получила распространение в скоринге для расчета рейтинга за счет чего логистическую регрессию всегда можно увидеть в наборе Data mining алгоритмов, несмотря на свое происхождение из статистики.

В тех случаях, когда зависимая переменная может принимать только два значения используется бинарная логистическая регрессия. С помощью логистической регрессии можно оценивать вероятность того, что событие наступит для конкретного испытуемого больной/здоровый.

Все регрессионные модели могут быть записаны в виде формулы:

$$y = F(x_1, x_2, \dots, x_n)$$

В множественной линейной регрессии предполагается, что зависимая переменная является линейной функцией независимых переменных, т.е.:

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Можно ли ее использовать для задачи оценки вероятности исхода события? Да, можно, вычислив стандартные коэффициенты регрессии. Например, если рассматривается исход по лечению, задается переменная y со значениями 1 и 0, где 1 означает, что соответствующий здоров, а 0, что болен.

Однако здесь возникает проблема: множественная регрессия не «знает», что переменная отклика бинарна по своей природе. Это неизбежно приведет к модели с предсказываемыми значениями большими 1 и меньшими 0. Но такие значения вообще не допустимы для первоначальной задачи. Таким образом, множественная регрессия просто игнорирует ограничения на диапазон значений для y .

Для решения проблемы задача регрессии может быть сформулирована иначе: вместо предсказания бинарной переменной, мы предсказываем непрерывную переменную со значениями на отрезке $[0, 1]$ при любых значениях независимых переменных. Это достигается применением следующего регрессионного уравнения (логит-преобразование):