

DUMMY-КОДИРОВАНИЕ - техника, позволяющая представлять номинальную переменную (см.) в виде набора дихотомических переменных. Используется при необходимости включать номинальные переменные в многомерные статистические модели.

Пусть номинальная переменная x имеет k значений. Одно из них объявляется референтным (для простоты изложения, пусть это будет значение x_k). Для каждого из оставшихся значений $x_1, x_2, \dots, x_{k-1}$ создается новая дихотомическая dummy-переменная по следующему правилу: если $x = i$, то $x_i = 1$. Во всех остальных случаях $x_i = 0$. Связь между исходной переменной x и набором соответствующих ей dummy-переменных можно представить в виде таблицы ($k = 5$):

Dummy-переменные

Значения переменной x

1

2

3

4

5

x_1

1

0

0

0

0

x_2

0

1

0

0

0

x3

0

0

1

0

0

x4

0

0

0

1

0

Таким образом, совокупность $k - 1$ dummy-переменных полностью представляет исходную номинальную переменную.

При включении dummy-переменных в регрессионное уравнение (см. Анализ регрессионный) полученные для них коэффициенты интерпретируются по отношению к референтному значению исходной переменной $x = k$: коэффициент b_i для dummy-переменной x_i показывает, насколько изменится предсказанное значение зависимой переменной при $x = i$ по сравнению с предсказанным значением при $x = k$.

О.В. Терещенко