ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Предположим, у нас есть регрессионная модель с *Y* зависимой от ряда простых переменных X_2, ..., X_k и от качественного показателя.

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Также предположим, что качественный показатель имеет несколько категорий. Мы возьмем одну из них, как незначительную категорию (без потери общности, категория 1) И обозначим её как вспомогательную переменную D_2, ..., D_s

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Что произойдет, если мы не будем сокращать основные переменные? Чтобы это стало возможным, мы ввели в уравнение вспомогательные переменные. Что произойдет в таком случае?

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + ... + \beta_k X_k + \delta_2 D_2 + ...\delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + ... + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + ...\delta_s D_s + u$$

Мы попадаем в ловушку фиктивных (вспомогательных) переменных. Становится невозможным построить модель так, как показано на экране.

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Попробуем объяснить ситуацию интуитивным путем. Каждый коэффицент вспомогательных переменных будет возрастать в строгой зависимости от предыдущего значения основных переменных. Но для такого подсчета нет основных переменных

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

β_1 представляет собой фиксированное значение для *Y* как основная переменная. Но, повторимся снова, здесь нет основных переменных . В таком случае, данная модель не имеет логического объяснения (интерпретации).

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Observation	Category	X_1	D_1	D_2	D_3	D_4
1	4	1	0	0	0	1
2	3	1	0	0	1	0
3	1	1	1	0	0	0
4	2	1	0	1	0	0
5	2	1	0	1	0	0
6	3	1	0	0	1	0
7	1	1	1	0	0	0
8	4	1	0	0	0	1

$$\sum_{i=1}^{4} D_i = X_1$$

С Математической точки зрения, у нас есть ряд чисел, связанный мультиколлинеарностью. Если отсутствуют значения, которыми можно пренебречь, то остается ряд чисел, с линейной зависимостью X_1 и вспомогательных переменных. В таблице приведены примеры.

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Observation	Category	X_1	D_1	D_2	D_3	D_4
1	4	1	0	0	0	1
2	3	1	0	0	1	0
3	1	1	1	0	0	0
4	2	1	0	1	0	0
5	2	1	0	1	0	0
6	3	1	0	0	1	0
7	1	1	1	0	0	0
8	4	1	0	0	0	1

$$\sum_{i=1}^{4} D_i = X_1$$

X_1 Это переменная, чье значение равно β_1 Она равняется единице во всех наблюдениях. Обычно мы не расписываем значения так открыто, потому что в этом нет необходимости.

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Observation	Category	X_1	D_1	D_2	D_3	D_4
1	4	1	0	0	0	1
2	3	1	0	0	1	0
3	1	1	1	0	0	0
4	2	1	0	1	0	0
5	2	1	0	1	0	0
6	3	1	0	0	1	0
7	1	1	1	0	0	0
8	4	1	0	0	0	1

$$\sum_{i=1}^{4} D_i = X_1$$

Если существует точная линейная зависимость между множеством переменных, в принципе невозможно оценить отдельные коэффициенты этих переменных. Необходимо использовать линейную алгебру, для объяснения и понимания данного процесса.

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Observation	Category	X_1	D_1	D_2	D_3	D_4
1	4	1	0	0	0	1
2	3	1	0	0	1	0
3	1	1	1	0	0	0
4	2	1	0	1	0	0
5	2	1	0	1	0	0
6	3	1	0	0	1	0
7	1	1	1	0	0	0
8	4	1	0	0	0	1

$$\sum_{i=1}^{4} D_i = X_1$$

В случае, если мы запускаем процесс подсчета линейной регрессии, то приложение, после запуска обнаружит ошибку и сделает одну из двух вещей :1-ое Может попросту отказаться от выполнения процесса регрессии.

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Observation	Category	X_1	D_1	D_2	D_3	D_4
1	4	1	0	0	0	1
2	3	1	0	0	1	0
3	1	1	1	0	0	0
4	2	1	0	1	0	0
5	2	1	0	1	0	0
6	3	1	0	0	1	0
7	1	1	1	0	0	0
8	4	1	0	0	0	1

$$\sum_{i=1}^{4} D_i = X_1$$

2-ое: Продолжит считать регрессию, но самостоятельно отбрасывать одну из переменных, определяя её как вспомогательную.

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Observation	Category	X_1	D_1	D_2	D_3	D_4
1	4	1	0	0	0	1
2	3	1	0	0	1	0
3	1	1	1	0	0	0
4	2	1	0	1	0	0
5	2	1	0	1	0	0
6	3	1	0	0	1	0
7	1	1	1	0	0	0
8	4	1	0	0	0	1

$$\sum_{i=1}^{4} D_i = X_1$$

Существует другой способ избежать Ловушки вспомогательных переменных. Убрать основную переменную (и X_1). Проблемы больше не будет, так как больше не будет линейной зависимости между переменными.

ЛОВУШКА ФИКТИВНЫХ ПЕРЕМЕННЫХ

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

$$Y = \beta_2 X_2 + \ldots + \beta_k X_k + \delta_1 D_1 + \delta_2 D_2 + \ldots \delta_s D_s + u$$

Observation	Category	X_1	D_1	D_2	D_3	D_4
1	4	1	0	0	0	1
2	3	1	0	0	1	0
3	1	1	1	0	0	0
4	2	1	0	1	0	0
5	2	1	0	1	0	0
6	3	1	0	0	1	0
7	1	1	1	0	0	0
8	4	1	0	0	0	1

$$\sum_{i=1}^{4} D_i = X_1$$

Параметры δ теперь являются основными в отношении к определенным категориям. К примеру, если наблюдение относится ко категории 2, Все вспомогательные переменные кроме D_2 будут равны 0. D_2 = 1, и, следовательно, будет зависеть от δ_2.