

## Практическое занятие № 5.

### «Фиктивные переменные»

Иногда необходимо включение в регрессионную модель одной или более качественных переменных (например, разделение по полу: мужской и женский; по уровню образования: общее и профессиональное и т.д.). Альтернативно может понадобиться сделать качественное различие между наблюдениями одних и тех же данных. Так, если проверяется взаимосвязь между размером компании и месячными доходами по акциям, может быть желательным включение качественной переменной, представляющей месяц январь, по причине хорошо известного «январского эффекта» во временных рядах доходов по ценным бумагам. Данный «январский эффект» - это феномен, заключающийся в том, что средние доходы по акциям, особенно небольших компаний, в среднем выше в январе, чем в другие месяцы. Таким образом, если мы рассматриваем январские наблюдения как качественно отличные от других наблюдений, фиктивная переменная ( $D$ ) позволит произвести подобное качественное различие.

Фиктивные переменные бывают двух типов - сдвига и наклона. Фиктивная переменная сдвига - это переменная, которая меняет точку пересечения линии регрессии с осью ординат в случае применения качественной переменной (рис. 69). Фиктивная переменная наклона - это та переменная, которая изменяет наклон линии регрессии в случае использования качественной переменной (рис. 70). Оба типа фиктивных переменных будут иметь значение  $+1$  или  $-1$ , когда наблюдения данных совпадают с уместной количественной переменной, но будут иметь нулевое значение при совпадении с наблюдениями, где эта качественная переменная отсутствует.

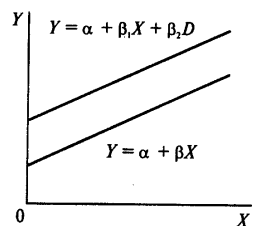


Рис. 69.

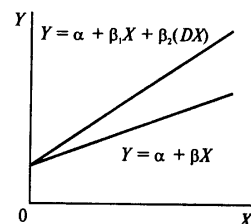


Рис. 70.

**Пример 5.** По данным примера 1 (файл `example_01.xls`.) дать интерпретацию бинарным, «фиктивным» переменным, принимающим значения 0 или 1: `floor` – принимает значение 0, если квартира расположена на первом или последнем этаже, `cat` – принимает значение 1, если квартира находится в кирпичном доме.

Построим регрессионное уравнение вида **LS PRICE C CAT FLOOR** (рис 71). Тем самым мы предполагаем (хотя в действительности это может быть и не так), что на цену квартиры оказывают влияние только две, указанные выше, составляющие. В результате получится уравнение следующего вида (рис 72):

$$PRICE = \alpha + \beta_1 CAT + \beta_2 FLOOR + \varepsilon.$$

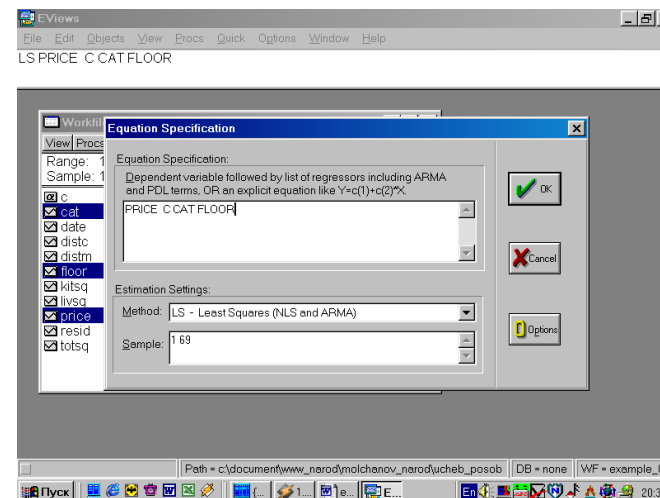


Рис. 71.

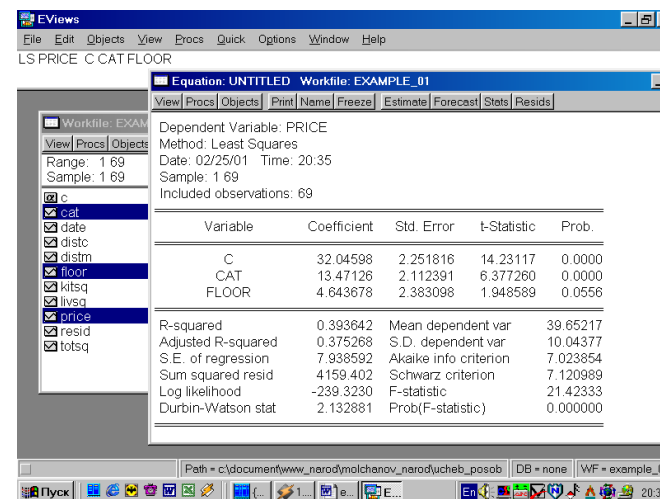


Рис. 72.

Используя результаты оценивания уравнения, содержащиеся в форме вывода (рис. 72), можно записать такое уравнение:

$$PRICE = 32,04598 + 13,47126 \cdot CAT + 4,643678 \cdot FLOOR.$$

Как же можно интерпретировать полученные результаты? Полученный коэффициент при CAT означает, что квартиры в кирпичных домах стоят в среднем на \$13471 дороже аналогичных квартир в панельных домах. Коэффициент при FLOOR может быть интерпретирован так: квартиры на не первом/последнем этажах стоят в среднем на \$4644 дороже аналогичных, расположенных на первом/последнем этажах.