

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

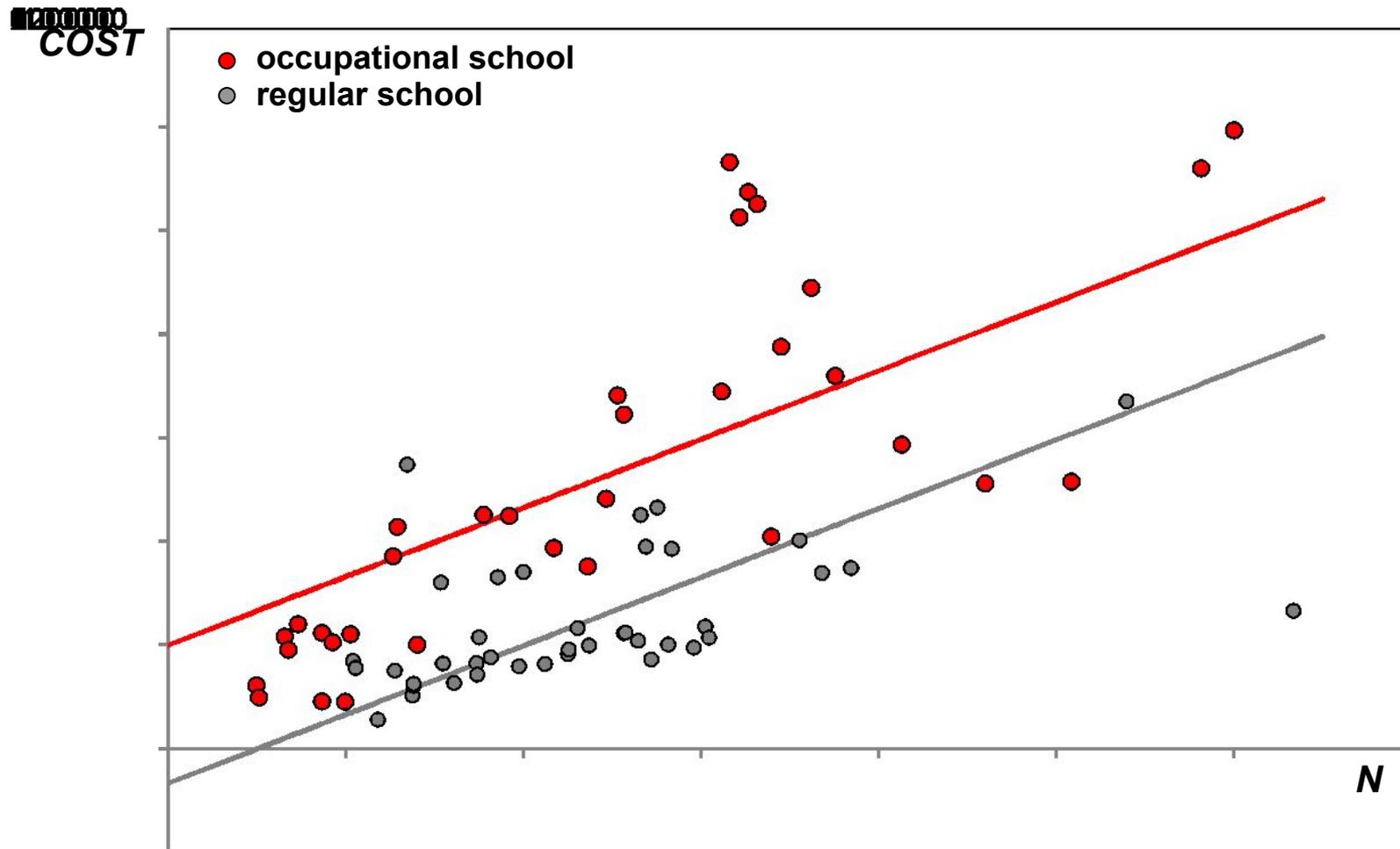
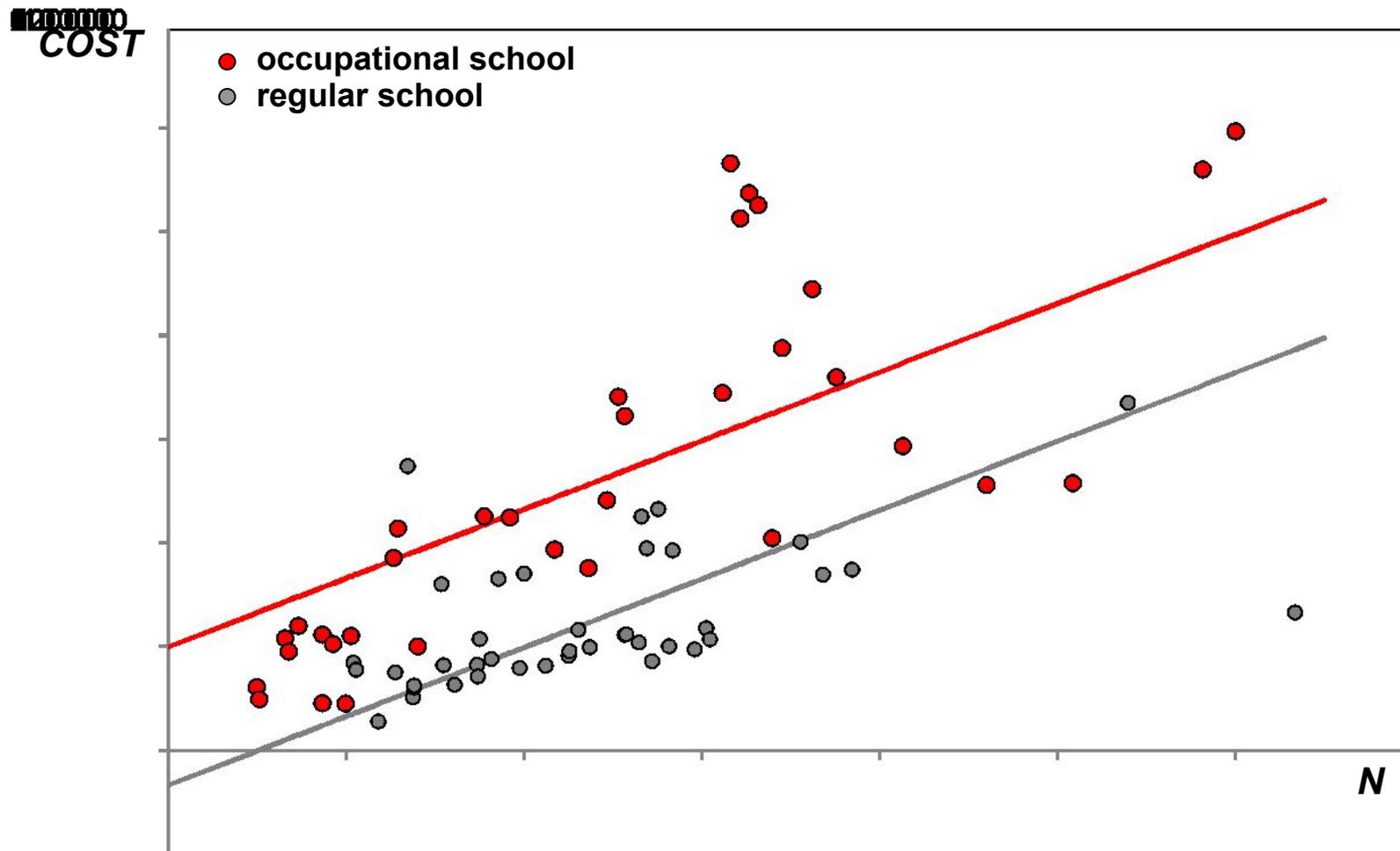


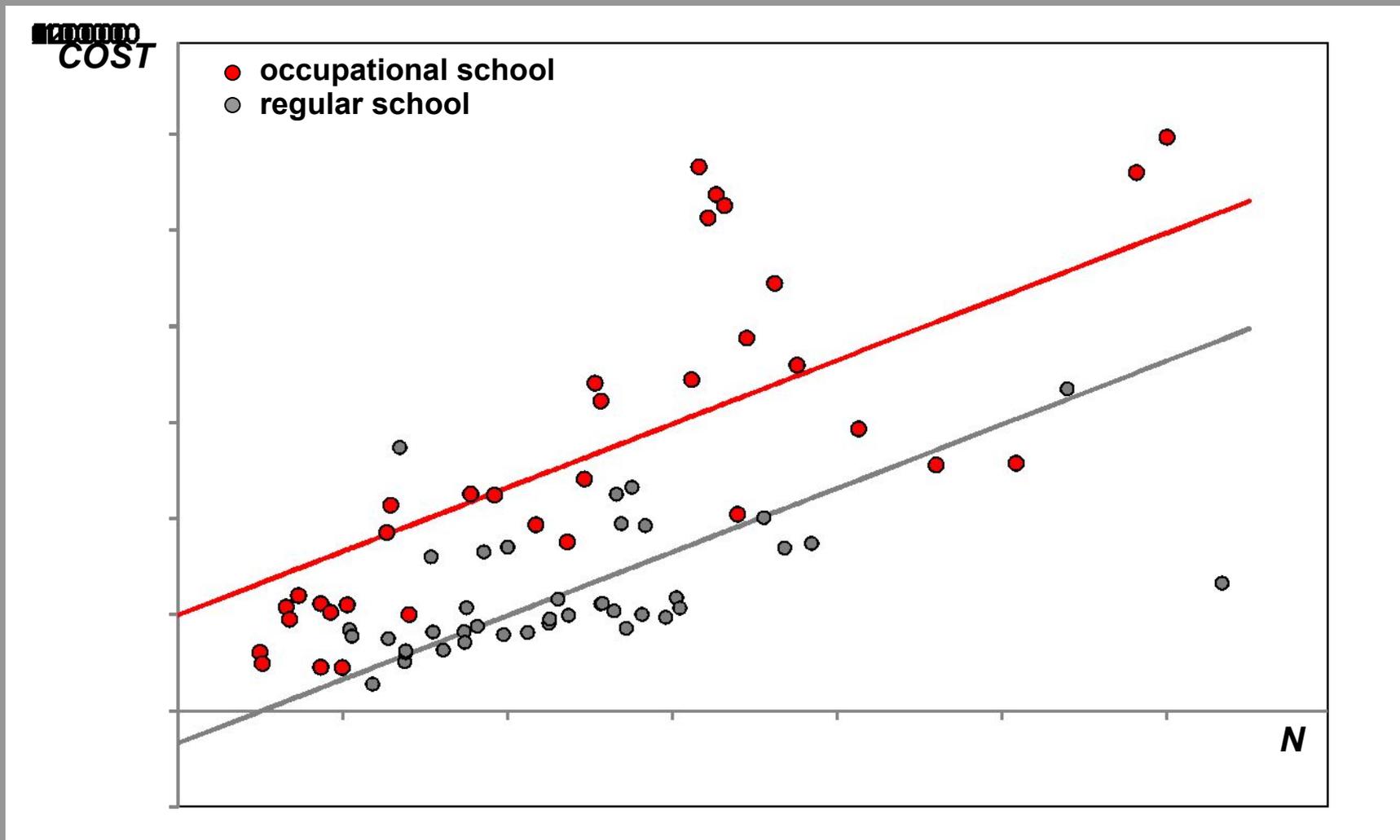
Диаграмма рассеяния показывает данные 74 школ Шанхая и функции затрат, полученные из регрессии стоимости (COST) на N и фиктивной переменной для типа программы обучения (профессиональной/обычной).

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА



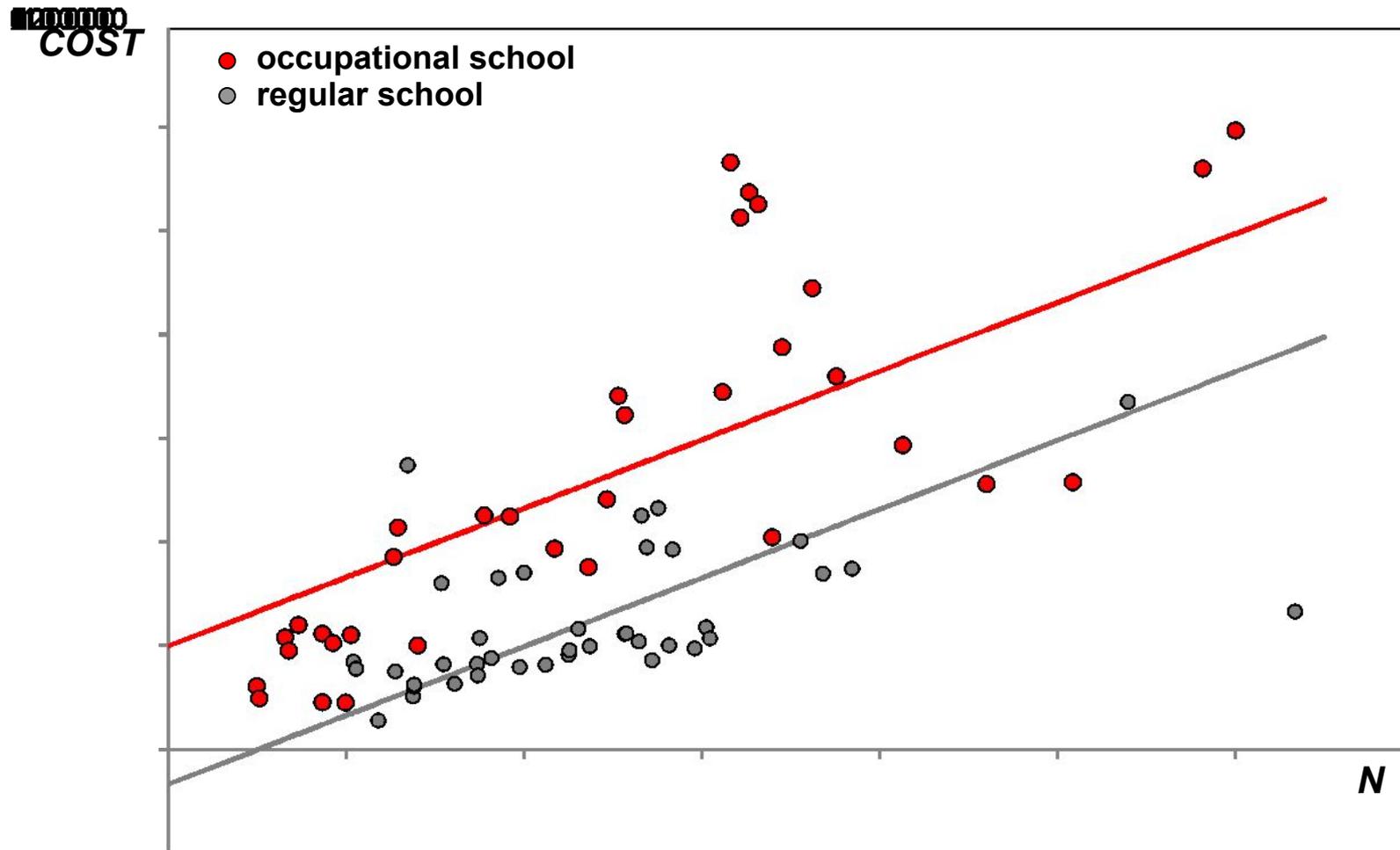
Описание модели включает в себя предположение о том, что предельные затраты на одного учащегося одинаковы для профессиональных и обычных школ. Следовательно, функции затрат параллельны.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА



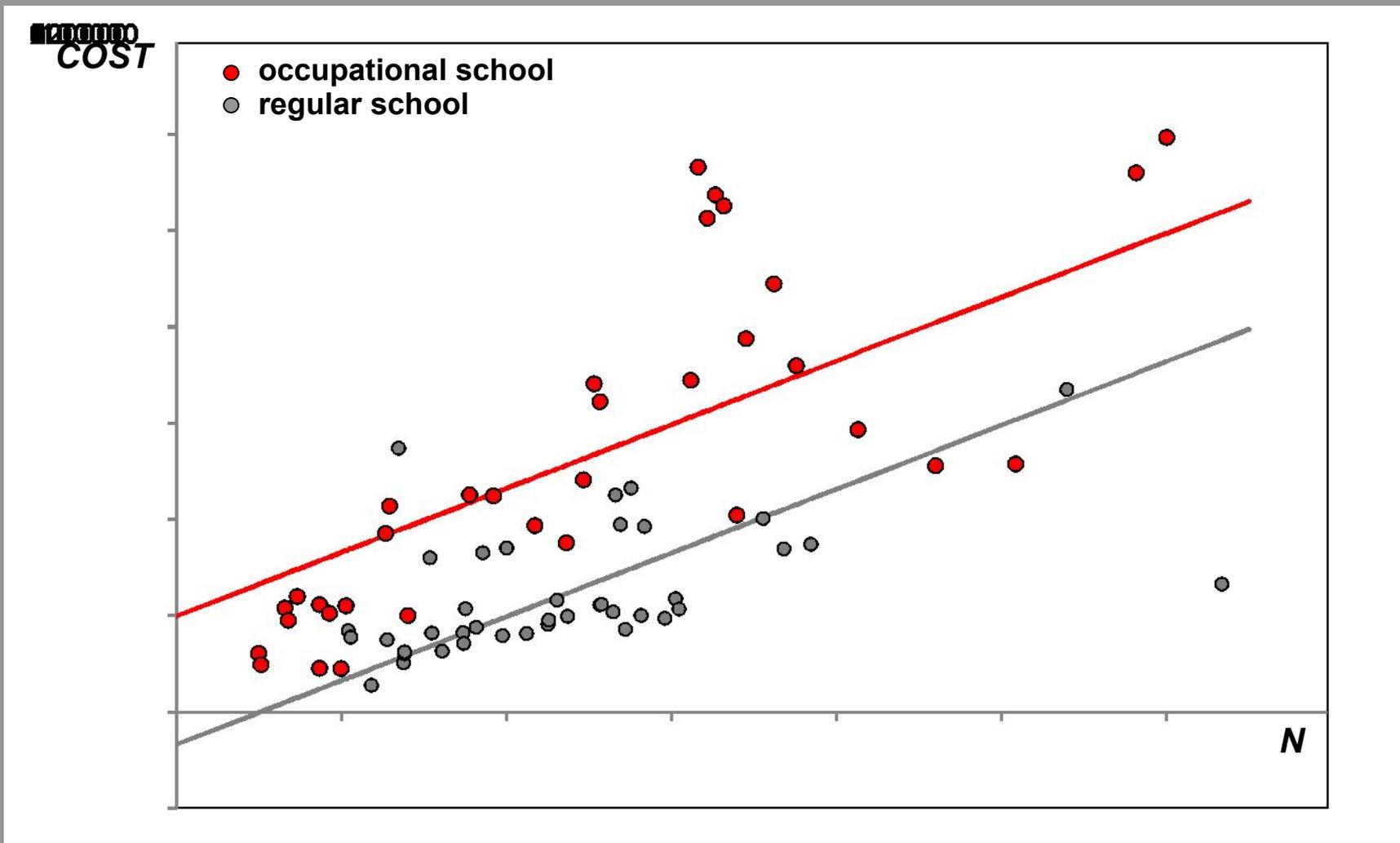
Тем не менее это не реалистическое предположение. Профессиональные школы несут расходы на учебные материалы, связанные с количеством учащихся.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА



Кроме того, соотношение персонала и ученика должно быть выше в профессиональных школах, поскольку групп семинаров не должны быть такими большими, как академические классы.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА



Рассматривая диаграмму рассеяния, можно видеть, что функция затрат для профессиональных школ должна быть более крутой, а для обычных школ - более плоской.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

$$COST = \beta_1 + \delta OCC + \beta_2 N + \lambda NOCC + u$$

Мы будем предполагать о тех же предельных издержках, введя, как известно, фиктивную переменную наклона. $NOCC$ – это произведение N и OCC .

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

$$COST = \beta_1 + \delta OCC + \beta_2 N + \lambda NOCC + u$$

Regular school
($OCC = NOCC = 0$)

$$COST = \beta_1 + \beta_2 N + u$$

В случае обычной школы OCC равен 0 и, следовательно, NOCC тоже.

Модель сводится к основным компонентам.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

$$COST = \beta_1 + \delta OCC + \beta_2 N + \lambda NOCC + u$$

Regular school
($OCC = NOCC = 0$)

$$COST = \beta_1 + \beta_2 N + u$$

Occupational school
($OCC = 1$; $NOCC = N$)

$$COST = (\beta_1 + \delta) + (\beta_2 + \lambda)N + u$$

В случае профессиональной школы $OCC=1$, а $NOCC=N$. Как показано, уравнение упрощается.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

$$COST = \beta_1 + \delta OCC + \beta_2 N + \lambda NOCC + u$$

Regular school
($OCC = NOCC = 0$)

$$COST = \beta_1 + \beta_2 N + u$$

Occupational school
($OCC = 1$; $NOCC = N$)

$$COST = (\beta_1 + \delta) + (\beta_2 + \lambda)N + u$$

В этой модели предельные издержки на одного учащегося на λ выше, чем в обычных школах, а накладные расходы получаются разные.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

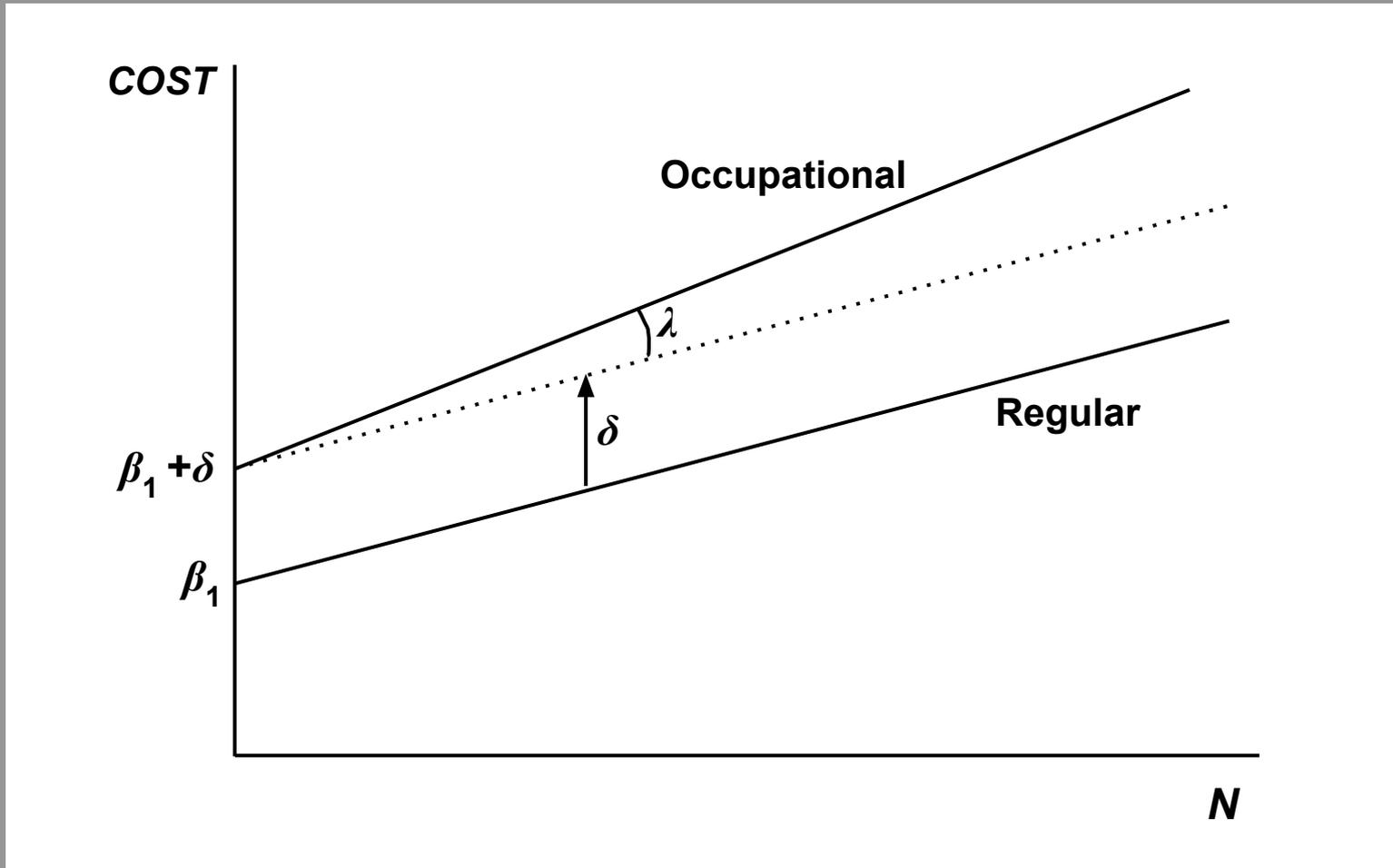


Диаграмма графически иллюстрирует модель.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

School	Type	COST	N	OCC	NOCC
1	Occupational	345,000	623	1	623
2	Occupational	537,000	653	1	653
3	Regular	170,000	400	0	0
4	Occupational	526,000	663	1	663
5	Regular	100,000	563	0	0
6	Regular	28,000	236	0	0
7	Regular	160,000	307	0	0
8	Occupational	45,000	173	1	173
9	Occupational	120,000	146	1	146
10	Occupational	61,000	99	1	99

Здесь данные первых 10 школ. Обратите внимание на странный способ определения NOCC.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS			
Model	1.0009e+12	3	3.3363e+11	Number of obs =	74	
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) =	49.64	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6803	
				Adj R-squared =	0.6666	
				Root MSE =	81980	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	152.2982	60.01932	2.537	0.013	32.59349	272.003
OCC	-3501.177	41085.46	-0.085	0.932	-85443.55	78441.19
NOCC	284.4786	75.63211	3.761	0.000	133.6351	435.3221
_cons	51475.25	31314.84	1.644	0.105	-10980.24	113930.7

Странно или нет, но метод работает очень хорошо. Вот результат регрессии с использованием полной выборки 74 школ. Начнем с интерпретации коэффициентов регрессии.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

$$\hat{COST} = 51,000 - 4,000OCC + 152N + 284NOCC$$

Вот регрессия в форме уравнения.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

$$\hat{COST} = 51,000 - 4,000OCC + 152N + 284NOCC$$

Regular school
($OCC = NOCC = 0$)

$$\hat{COST} = 51,000 + 152N$$

Подставляя OCC и $NOCC$, равные 0, функцию затрат для обычных школ. По нашим оценкам, их ежегодные накладные расходы составляют 51 000 юаней, а их ежегодная предельная стоимость на одного учащегося составляет 152 юаней.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

$$\hat{COST} = 51,000 - 4,000OCC + 152N + 284NOCC$$

Regular school
($OCC = NOCC = 0$)

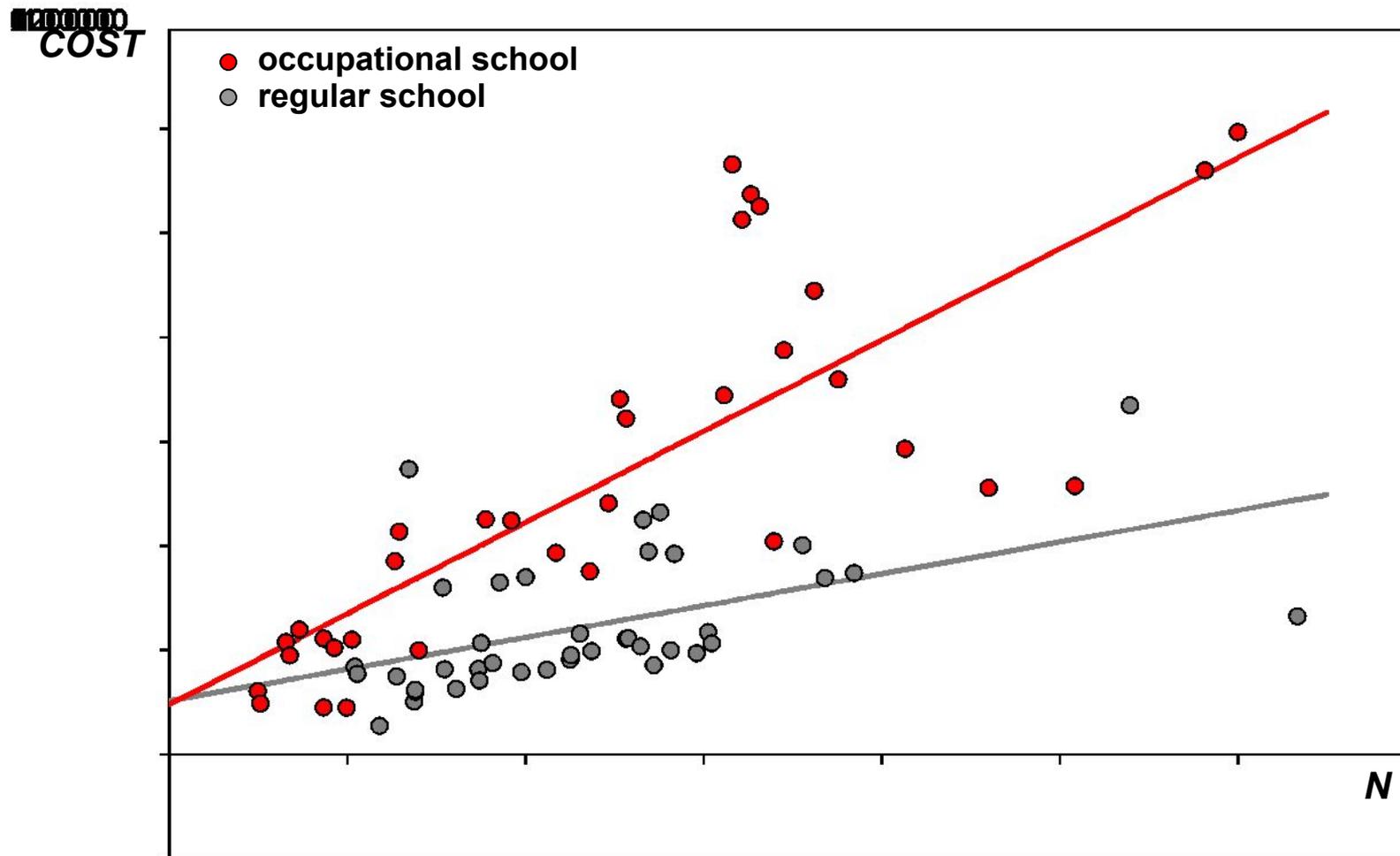
$$\hat{COST} = 51,000 + 152N$$

Occupational school
($OCC = 1; NOCC = N$)

$$\begin{aligned}\hat{COST} &= 51,000 - 4,000 + 152N + 284N \\ &= 47,000 + 436N\end{aligned}$$

Подставив OCC , равный 1, и $NOCC$, равный N , мы рассчитали, что годовые накладные расходы для профессиональных школ равны 47 000 юаней, а годовые предельные издержки на одного учащегося составляют 436 юаней.

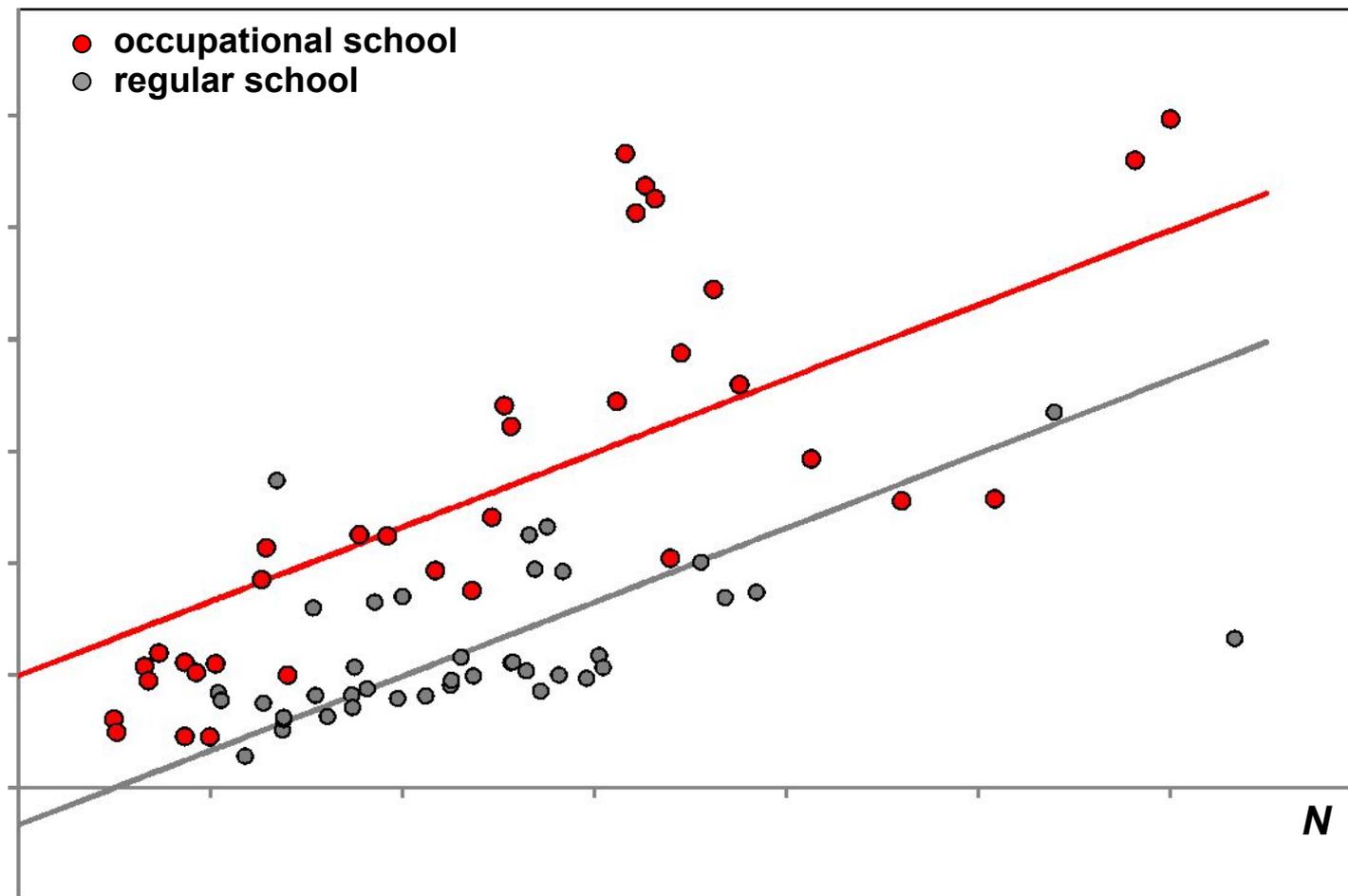
ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФИЦИЕНТА НАКЛОНА



Можно увидеть, что функции затрат соответствуют данным намного лучше, чем раньше, и что реальная разница заключается в предельных издержках, а не накладных расходах.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

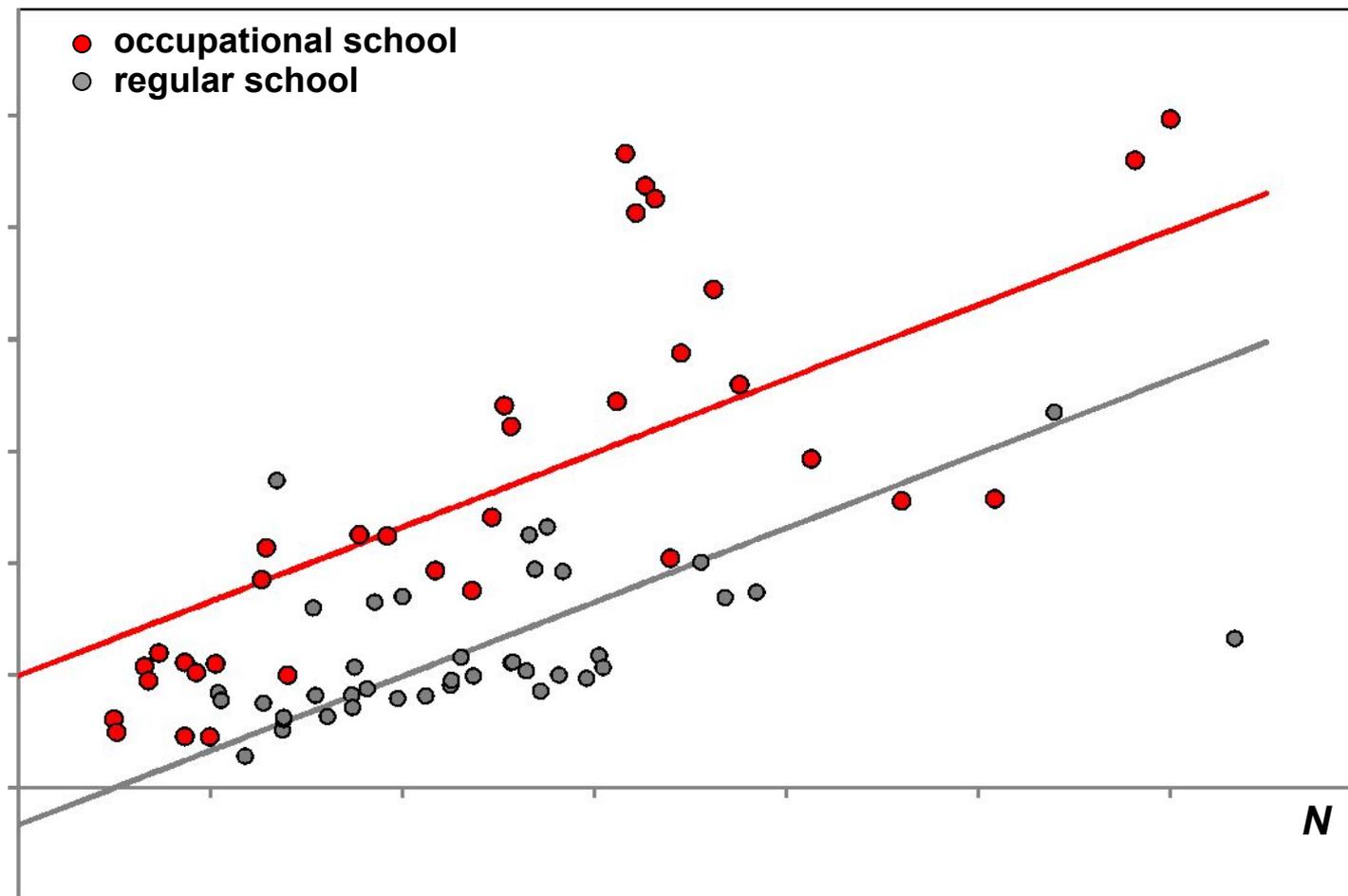
COST



Теперь мы можем понять, почему у нас была бессмысленная отрицательная оценка накладных расходов обычной школы в предыдущих моделях.

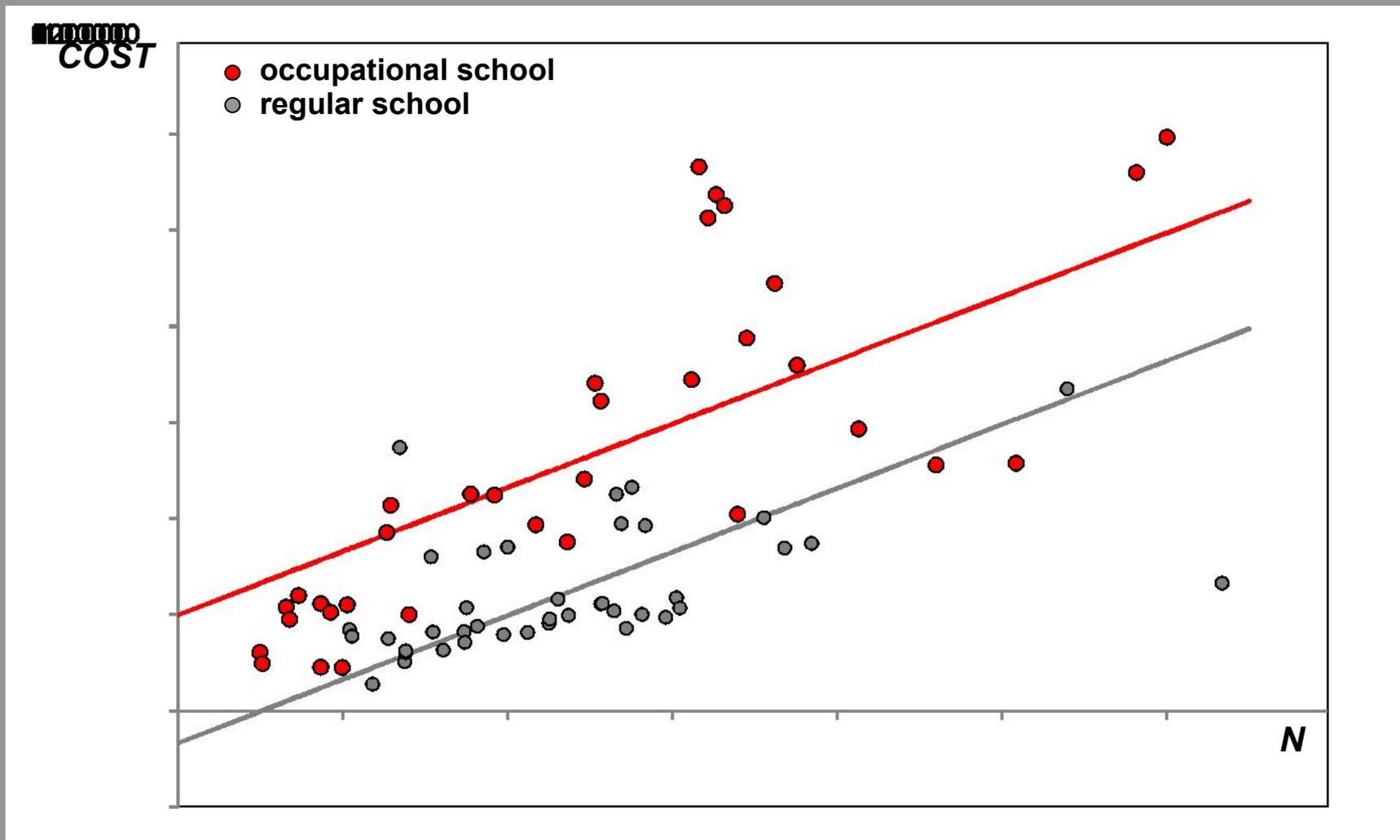
ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

COST



Предположение о тех же предельных издержках привело к оценке предельных издержек, которые были компромиссом между предельными издержками профессиональных и обычных школ.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА



Функция затрат для обычных школ была слишком крутой, и, как следствие, отрезок был недооценен, фактически стал отрицательным и указывал на то, что что-то должно быть ошибочным с описанием модели.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS			
Model	1.0009e+12	3	3.3363e+11	Number of obs =	74	
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) =	49.64	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6803	
				Adj R-squared =	0.6666	
				Root MSE =	81980	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	152.2982	60.01932	2.537	0.013	32.59349	272.003
OCC	-3501.177	41085.46	-0.085	0.932	-85443.55	78441.19
NOCC	284.4786	75.63211	3.761	0.000	133.6351	435.3221
_cons	51475.25	31314.84	1.644	0.105	-10980.24	113930.7

Мы можем выполнять t-тесты, как обычно. Статистика t для коэффициента NOCC составляет 3,76, поэтому предельная стоимость одного студента в профессиональной школе значительно выше, чем в обычной школе.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS			
Model	1.0009e+12	3	3.3363e+11	Number of obs =	74	
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) =	49.64	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6803	
				Adj R-squared =	0.6666	
				Root MSE =	81980	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	152.2982	60.01932	2.537	0.013	32.59349	272.003
OCC	-3501.177	41085.46	-0.085	0.932	-85443.55	78441.19
NOCC	284.4786	75.63211	3.761	0.000	133.6351	435.3221
_cons	51475.25	31314.84	1.644	0.105	-10980.24	113930.7

Коэффициент OCC сейчас отрицательный, что говорит о том, что накладные расходы профессиональных школ фактически ниже, чем в обычных школах.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS			
Model	1.0009e+12	3	3.3363e+11	Number of obs =	74	
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) =	49.64	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6803	
				Adj R-squared =	0.6666	
				Root MSE =	81980	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	152.2982	60.01932	2.537	0.013	32.59349	272.003
OCC	-3501.177	41085.46	-0.085	0.932	-85443.55	78441.19
NOCC	284.4786	75.63211	3.761	0.000	133.6351	435.3221
_cons	51475.25	31314.84	1.644	0.105	-10980.24	113930.7

Это маловероятно. Однако статистика t только $-0,09$, поэтому мы не отвергаем нулевую гипотезу о том, что накладные расходы двух типов школ одинаковы.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	Number of obs	=	74
Model	1.0009e+12	3	3.3363e+11	F(3, 70)	=	49.64
Residual	4.7045e+11	70	6.7207e+09	Prob > F	=	0.0000
Total	1.4713e+12	73	2.0155e+10	R-squared	=	0.6803
				Adj R-squared	=	0.6666
				Root MSE	=	81980

```
. reg COST N
```

Source	SS	df	MS	Number of obs	=	74
Model	5.7974e+11	1	5.7974e+11	F(1, 72)	=	46.82
Residual	8.9160e+11	72	1.2383e+10	Prob > F	=	0.0000
Total	1.4713e+12	73	2.0155e+10	R-squared	=	0.3940
				Adj R-squared	=	0.3856
				Root MSE	=	1.1e+05

Мы также можем выполнить F-тест общей объяснительной способности фиктивных переменных, сравнивая RSS, когда фиктивные переменные включаются в RSS, когда они ими не являются

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	Number of obs =	74
Model	1.0009e+12	3	3.3363e+11	F(3, 70) =	49.64
Residual	4.7045e+11	70	6.7207e+09	Prob > F =	0.0000
				R-squared =	0.6803
				Adj R-squared =	0.6666
Total	1.4713e+12	73	2.0155e+10	Root MSE =	81980

```
. reg COST N
```

Source	SS	df	MS	Number of obs =	74
Model	5.7974e+11	1	5.7974e+11	F(1, 72) =	46.82
Residual	8.9160e+11	72	1.2383e+10	Prob > F =	0.0000
				R-squared =	0.3940
				Adj R-squared =	0.3856
Total	1.4713e+12	73	2.0155e+10	Root MSE =	1.1e+05

Нулевая гипотеза состоит в том, что коэффициенты OCC и NOCC равны 0.
Альтернативная гипотеза состоит в том, что один или оба отличны от нуля.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	
Model	1.0009e+12	3	3.3363e+11	Number of obs = 74
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) = 49.64
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.6803
				Adj R-squared = 0.6666
				Root MSE = 81980

```
. reg COST N
```

Source	SS	df	MS	
Model	5.7974e+11	1	5.7974e+11	Number of obs = 74
Residual	8.9160e+11	72	1.2383e+10	F(1, 72) = 46.82
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.3940
				Adj R-squared = 0.3856
				Root MSE = 1.1e+05

$$F(2,70) = \frac{(8.92 \times 10^{11} - 4.70 \times 10^{11}) / 2}{4.70 \times 10^{11} / 70}$$

Улучшение соответствия при добавлении фиктивных переменных - это сокращение в RSS.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	
Model	1.0009e+12	3	3.3363e+11	Number of obs = 74
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) = 49.64
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.6803
				Adj R-squared = 0.6666
				Root MSE = 81980

```
. reg COST N
```

Source	SS	df	MS	
Model	5.7974e+11	1	5.7974e+11	Number of obs = 74
Residual	8.9160e+11	72	1.2383e+10	F(1, 72) = 46.82
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.3940
				Adj R-squared = 0.3856
				Root MSE = 1.1e+05

$$F(2,70) = \frac{(8.92 \times 10^{11} - 4.70 \times 10^{11}) / 2}{4.70 \times 10^{11} / 70}$$

Стоимость равна 2, поскольку были оценены 2 дополнительных параметра - коэффициенты фиктивных переменных, и, как следствие, количество оставшихся степеней свободы было уменьшено с 72 до 70.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	
Model	1.0009e+12	3	3.3363e+11	Number of obs = 74
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) = 49.64
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.6803
				Adj R-squared = 0.6666
				Root MSE = 81980

```
. reg COST N
```

Source	SS	df	MS	
Model	5.7974e+11	1	5.7974e+11	Number of obs = 74
Residual	8.9160e+11	72	1.2383e+10	F(1, 72) = 46.82
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.3940
				Adj R-squared = 0.3856
				Root MSE = 1.1e+05

$$F(2,70) = \frac{(8.92 \times 10^{11} - 4.70 \times 10^{11})/2}{4.70 \times 10^{11} / 70}$$

Первый компонент знаменателя - это RSS после добавления фиктивных переменных.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	
Model	1.0009e+12	3	3.3363e+11	Number of obs = 74
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) = 49.64
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000

				R-squared = 0.6803
				Adj R-squared = 0.6666
				Root MSE = 81980

```
. reg COST N
```

Source	SS	df	MS	
Model	5.7974e+11	1	5.7974e+11	Number of obs = 74
Residual	8.9160e+11	72	1.2383e+10	F(1, 72) = 46.82
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000

				R-squared = 0.3940
				Adj R-squared = 0.3856
				Root MSE = 1.1e+05

$$F(2,70) = \frac{(8.92 \times 10^{11} - 4.70 \times 10^{11}) / 2}{4.70 \times 10^{11} / 70}$$

Знаменатель - это RSS после добавления фиктивных переменных, деленный на количество оставшихся степеней свободы. Это 70, потому что есть 74 наблюдения и 4 параметра.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	
Model	1.0009e+12	3	3.3363e+11	Number of obs = 74
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) = 49.64
				Prob > F = 0.0000
				R-squared = 0.6803
				Adj R-squared = 0.6666
Total	1.4713e+12	73	2.0155e+10	Root MSE = 81980

```
. reg COST N
```

Source	SS	df	MS	
Model	5.7974e+11	1	5.7974e+11	Number of obs = 74
Residual	8.9160e+11	72	1.2383e+10	F(1, 72) = 46.82
				Prob > F = 0.0000
				R-squared = 0.3940
				Adj R-squared = 0.3856
Total	1.4713e+12	73	2.0155e+10	Root MSE = 1.1e+05

$$F(2,70) = \frac{(8.92 \times 10^{11} - 4.70 \times 10^{11})/2}{4.70 \times 10^{11} / 70} = 31.4$$

$$F(2,70)_{\text{crit}, 0.1\%} = 7.6$$

Таким образом, статистика F 31.4. Критическая шкала F (2,70) на уровне 0,1 процента составляет 7,6.

ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ДЛЯ КОЭФФИЦИЕНТА НАКЛОНА

```
. reg COST N OCC NOCC
```

Source	SS	df	MS	
Model	1.0009e+12	3	3.3363e+11	Number of obs = 74
Residual	4.7045e+11	70	6.7207e+09	F(3, 70) = 49.64
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.6803
				Adj R-squared = 0.6666
				Root MSE = 81980

```
. reg COST N
```

Source	SS	df	MS	
Model	5.7974e+11	1	5.7974e+11	Number of obs = 74
Residual	8.9160e+11	72	1.2383e+10	F(1, 72) = 46.82
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000
				R-squared = 0.3940
				Adj R-squared = 0.3856
				Root MSE = 1.1e+05

$$F(2,70) = \frac{(8.92 \times 10^{11} - 4.70 \times 10^{11})/2}{4.70 \times 10^{11} / 70} = 31.4$$

$$F(2,70)_{\text{crit}, 0.1\%} = 7.6$$

Таким образом, мы заключаем, что хотя бы один из фиктивных переменных коэффициентов отличается от 0. Мы знали это уже из t тестов, поэтому в этом случае F-тест на самом деле ничего не добавляет.