

Два вида фиктивных переменных

$$COST = \beta_1 + \delta OCC + \varepsilon RES + \beta_2 N + u$$

Объяснительные переменные в модели регрессии могут включать в себя несколько наборов фиктивных переменных. Эта последовательность представляет собой пример модели с двумя видами.

Два вида фиктивных переменных

$$COST = \beta_1 + \delta OCC + \varepsilon RES + \beta_2 N + u$$

Мы продолжим использовать модель функций затрат школы и распространим ее, чтобы учесть тот факт, что некоторые из школ являются интернатами.

Два вида фиктивных переменных

$$COST = \beta_1 + \delta OCC + \varepsilon RES + \beta_2 N + u$$

Чтобы моделировать более высокие накладные расходы школ-интернатов, мы вводим фиктивную переменную RES, которая равна 1 для них и 0 для нежилых школ. ε - дополнительные годовые накладные расходы в школе-интернате по сравнению с нежилкой.

Два вида фиктивных переменных

$$COST = \beta_1 + \delta OCC + \varepsilon RES + \beta_2 N + u$$

Мы также будем проводить различие между профессиональными и обычными школами, используя фиктивную переменную OCC, определенную в первой последовательности. (Лучше было бы использовать классификацию в четырех категориях, и на практике мы бы это сделали, но это усложняло бы график.)

Два вида фиктивных переменных

$$COST = \beta_1 + \delta OCC + \varepsilon RES + \beta_2 N + u$$

Регулярный, не жилой
(интернат)
($OCC = RES = 0$)

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

Если школа имеет регулярную учебную программу и не является интернатом, обе фиктивные переменные равны 0, а функция затрат упрощается к ее основным компонентам.

Два вида фиктивных переменных

$$COST = \beta_1 + \delta OCC + \varepsilon RES + \beta_2 N + u$$

Регулярный, нежилой
(интернат)
($OCC = RES = 0$)

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

Регулярный, жилой
($OCC = 0; RES = 1$)

$$COST = (\beta_1 + \varepsilon) + \beta_2 N + u$$

Для обычной обычной школы RES равна 1, а отрезок увеличивается на сумму ε .

Два вида фиктивных переменных

$$COST = \beta_1 + \delta OCC + \varepsilon RES + \beta_2 N + u$$

Регулярный, нежилой
($OCC = RES = 0$)

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

Регулярный, жилой
(интернат)
($OCC = 0; RES = 1$)

$$COST = (\beta_1 + \varepsilon) + \beta_2 N + u$$

Профессиональный,
нежилой ($OCC = 1; RES = 0$)

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

Профессиональный,
жилой
($OCC = 1; RES = 1$)

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

В случае нежилого профессионального учебного заведения RES составляет 0, а OCC - 1, поэтому накладные расходы увеличиваются на d . Если школа является профессиональной и жилой, она увеличивается на $(d + e)$.

Два вида фиктивных переменных

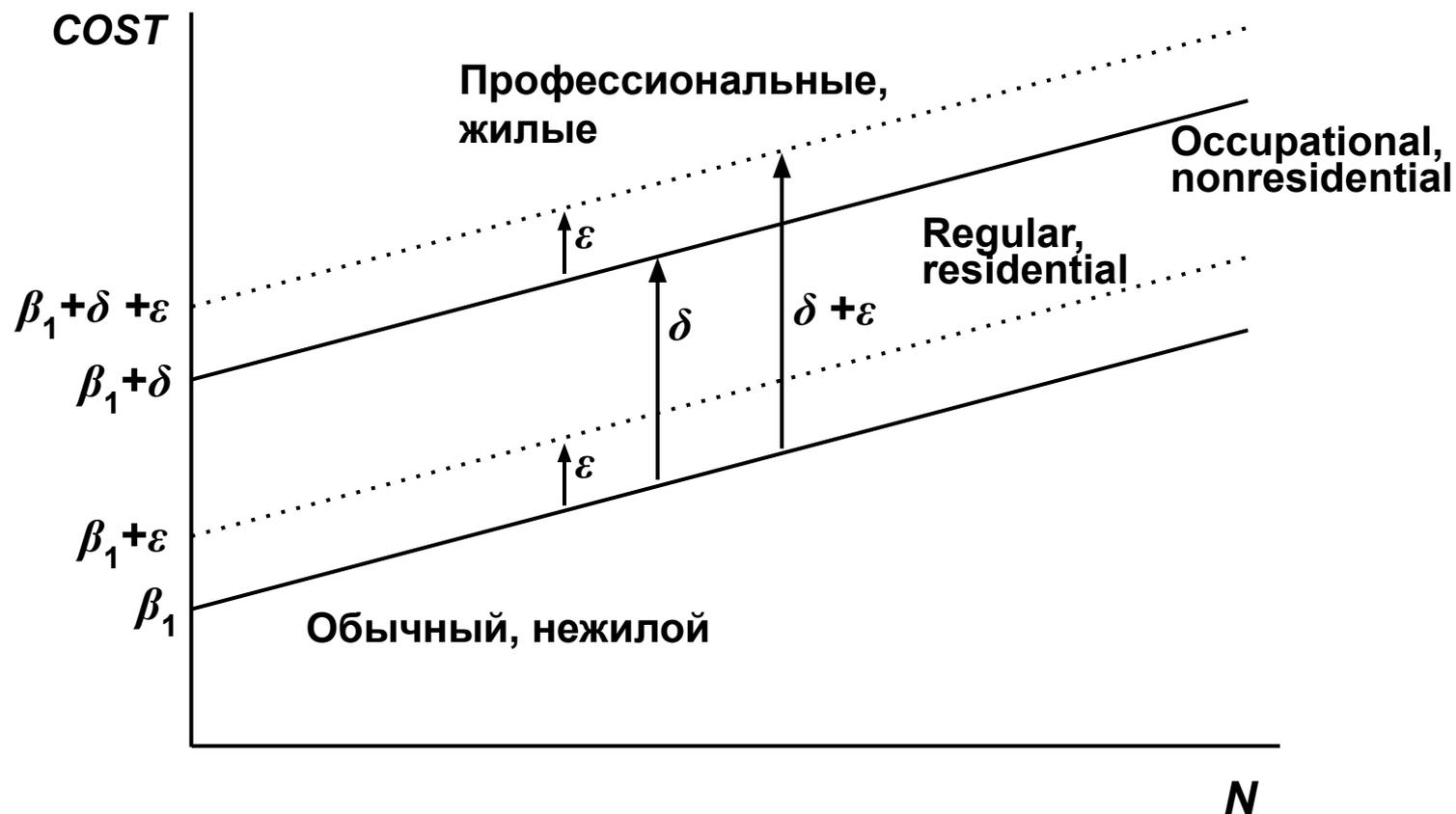
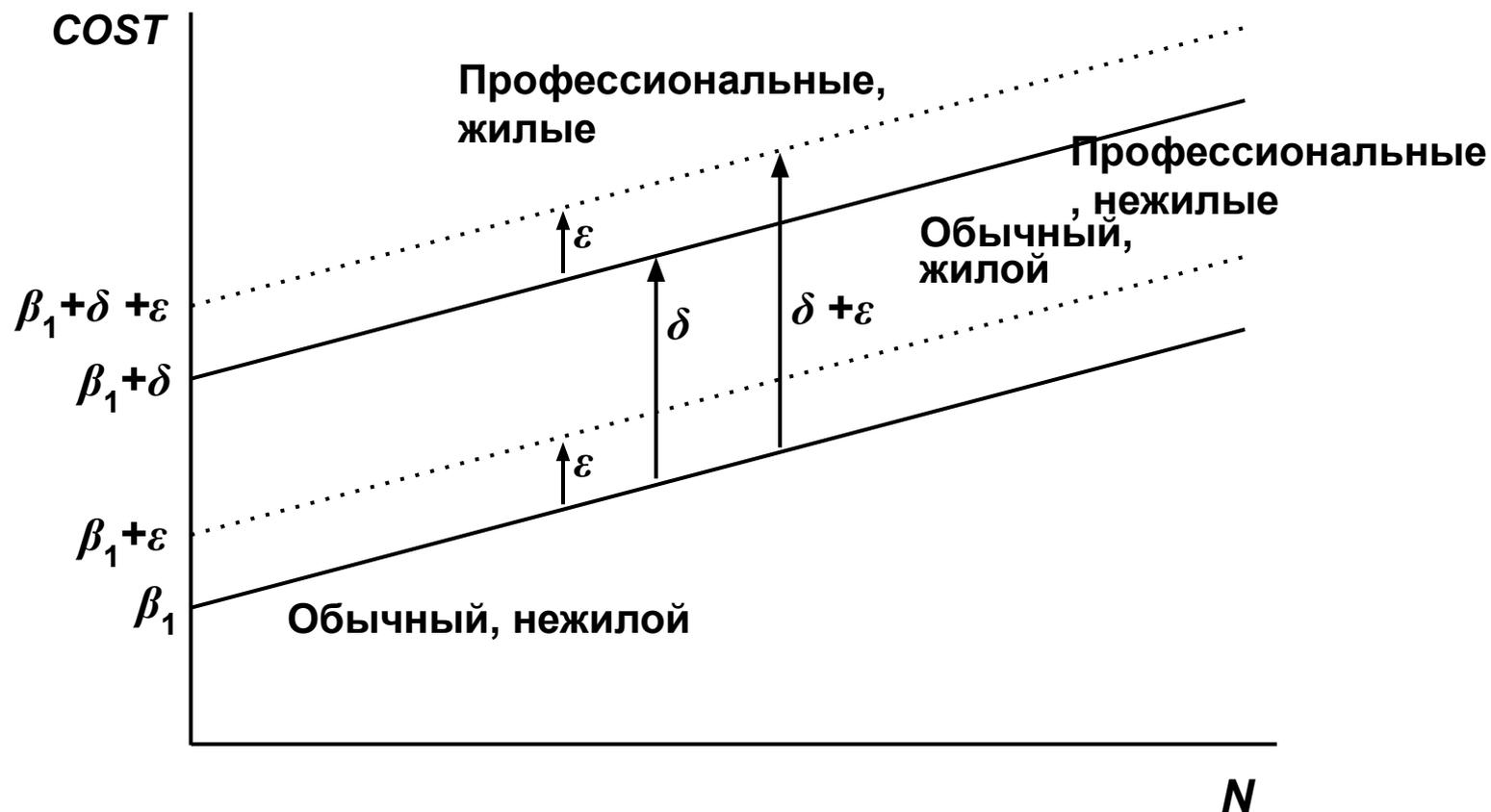


Диаграмма иллюстрирует модель графически. Обратите внимание, что эффекты различных компонентов модели считаются отдельными и аддитивными в этой спецификации.

Два вида фиктивных переменных



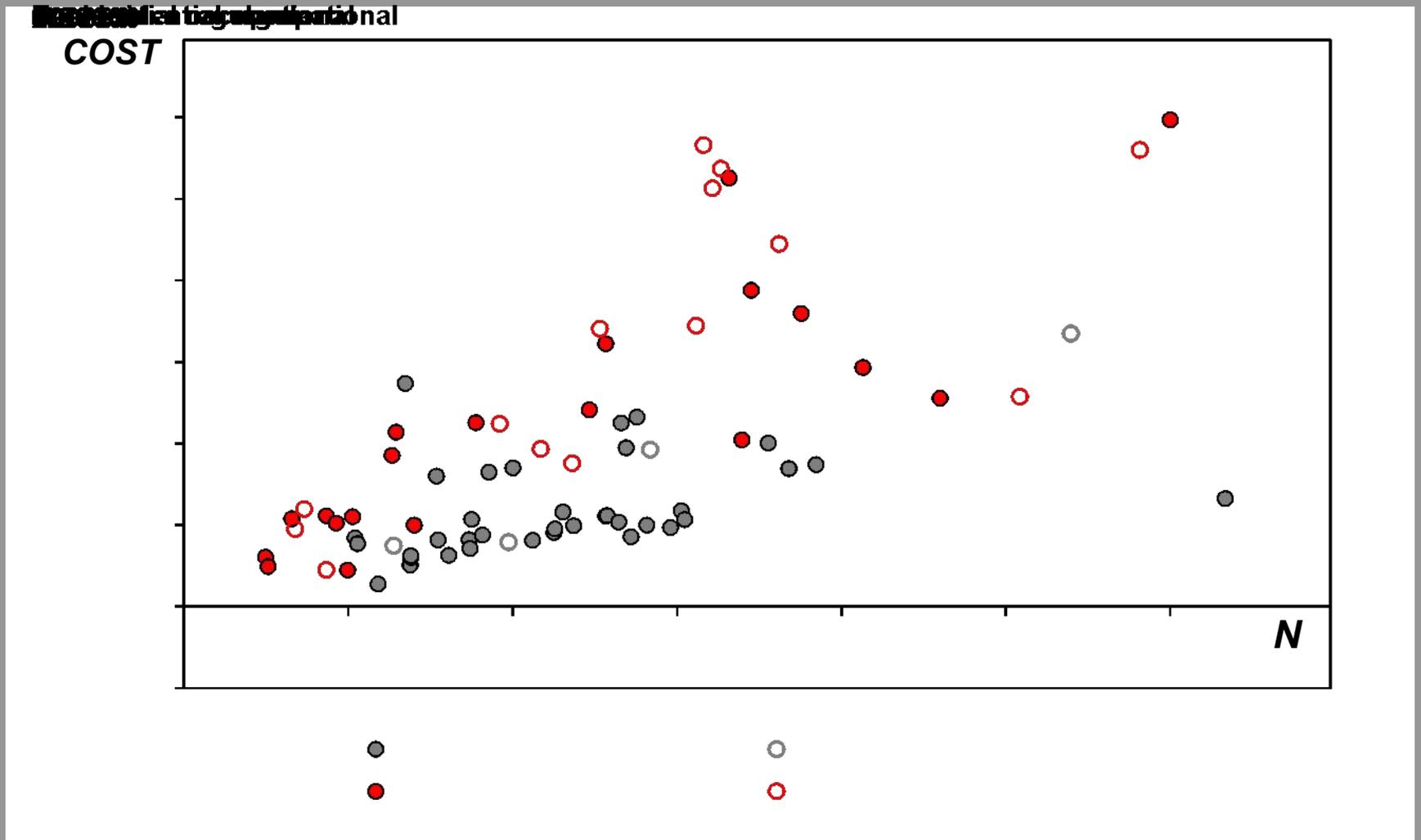
В частности, мы предполагаем, что дополнительные накладные расходы в школе-интернате одинаковы для обычных и профессиональных школ.

Два вида фиктивных переменных

School	Type	Residential?	COST	N	OCC	RES
1	Профессиональный	No	345,000	623	1	0
2	Профессиональный	Yes	537,000	653	1	1
3	Регулярный	No	170,000	400	0	0
4	Профессиональный	Yes	526,000	663	1	1
5	Регулярный	No	100,000	563	0	0
6	Регулярный	No	28,000	236	0	0
7	Регулярный	Yes	160,000	307	0	1
8	Профессиональный	No	45,000	173	1	0
9	Профессиональный	No	120,000	146	1	0
10	Профессиональный	No	61,000	99	1	0

Вот данные для первых 10 школ. Обратите внимание, как значения фиктивных переменных меняются в зависимости от характеристик школы.

Два вида фиктивных переменных



Вот диаграмма разброса, показывающая четыре типа школы.

Два вида фиктивных переменных

```
. reg COST N OCC RES
```

Source	SS	df	MS			
Model	9.3297e+11	3	3.1099e+11	Number of obs =	74	
Residual	5.3838e+11	70	7.6911e+09	F(3, 70) =	40.43	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6341	
				Adj R-squared =	0.6184	
				Root MSE =	87699	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	321.833	39.40225	8.168	0.000	243.2477	400.4183
OCC	109564.6	24039.58	4.558	0.000	61619.15	157510
RES	57909.01	30821.31	1.879	0.064	-3562.137	119380.2
_cons	-29045.27	23291.54	-1.247	0.217	-75498.78	17408.25

Вот вывод Stata для регрессии. Начнем с интерпретации коэффициентов регрессии. Коэффициент N указывает, что предельные издержки на одного учащегося составляют 322 юаней в год.

Два вида фиктивных переменных

```
. reg COST N OCC RES
```

Source	SS	df	MS			
Model	9.3297e+11	3	3.1099e+11	Number of obs =	74	
Residual	5.3838e+11	70	7.6911e+09	F(3, 70) =	40.43	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6341	
				Adj R-squared =	0.6184	
				Root MSE =	87699	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	321.833	39.40225	8.168	0.000	243.2477	400.4183
OCC	109564.6	24039.58	4.558	0.000	61619.15	157510
RES	57909.01	30821.31	1.879	0.064	-3562.137	119380.2
_cons	-29045.27	23291.54	-1.247	0.217	-75498.78	17408.25

Константа дает оценку годовой накладной стоимости ссылочной категории, нежилых обычных школ. Это все еще отрицательно, поэтому не имеет никакого смысла.

Два вида фиктивных переменных

```
. reg COST N OCC RES
```

Source	SS	df	MS			
Model	9.3297e+11	3	3.1099e+11	Number of obs =	74	
Residual	5.3838e+11	70	7.6911e+09	F(3, 70) =	40.43	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6341	
				Adj R-squared =	0.6184	
				Root MSE =	87699	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	321.833	39.40225	8.168	0.000	243.2477	400.4183
OCC	109564.6	24039.58	4.558	0.000	61619.15	157510
RES	57909.01	30821.31	1.879	0.064	-3562.137	119380.2
_cons	-29045.27	23291.54	-1.247	0.217	-75498.78	17408.25

Коэффициент OCC указывает, что ежегодные накладные расходы в профессиональных школах составляют 110 000 юаней больше, чем в обычных школах.

Два вида фиктивных переменных

```
. reg COST N OCC RES
```

Source	SS	df	MS			
Model	9.3297e+11	3	3.1099e+11	Number of obs =	74	
Residual	5.3838e+11	70	7.6911e+09	F(3, 70) =	40.43	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6341	
				Adj R-squared =	0.6184	
				Root MSE =	87699	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	321.833	39.40225	8.168	0.000	243.2477	400.4183
OCC	109564.6	24039.58	4.558	0.000	61619.15	157510
RES	57909.01	30821.31	1.879	0.064	-3562.137	119380.2
_cons	-29045.27	23291.54	-1.247	0.217	-75498.78	17408.25

Коэффициент RES показывает, что ежегодные накладные расходы школ-интернатов на 58 000 юаней выше, чем у нежилых школ.

Два вида фиктивных переменных

$$\hat{COST} = -29,000 + 110,000OCC + 58,000RES + 322N$$

Обычный, нежилой
($OCC = RES = 0$)

$$\hat{COST} = -29,000 + 322N$$

Результат регрессии показан сверху в форме уравнения. Полагая обе фиктивные переменные равными 0, мы получаем неявную функцию стоимости для нежилых обычных школ.

Два вида фиктивных переменных

$$\hat{COST} = -29,000 + 110,000OCC + 58,000RES + 322N$$

Обычный,
нежилой, ($OCC = RES = 0$)

$$\hat{COST} = -29,000 + 322N$$

Обычный, жилой
($OCC = 0; RES = 1$)

$$\begin{aligned}\hat{COST} &= -29,000 + 58,000 + 322N \\ &= 29,000 + 322N\end{aligned}$$

Полагая RES равным 1, но сохраняя OCC на 0, мы получаем функцию стоимости для обычных обычных школ.

Два вида фиктивных переменных

$$\hat{COST} = -29,000 + 110,000OCC + 58,000RES + 322N$$

Обычный,
нежилой, ($OCC = RES = 0$)

$$\hat{COST} = -29,000 + 322N$$

Обычный, жилой
($OCC = 0; RES = 1$)

$$\begin{aligned}\hat{COST} &= -29,000 + 58,000 + 322N \\ &= 29,000 + 322N\end{aligned}$$

Профессиональные,
нежилые ($OCC = 1; RES = 0$)

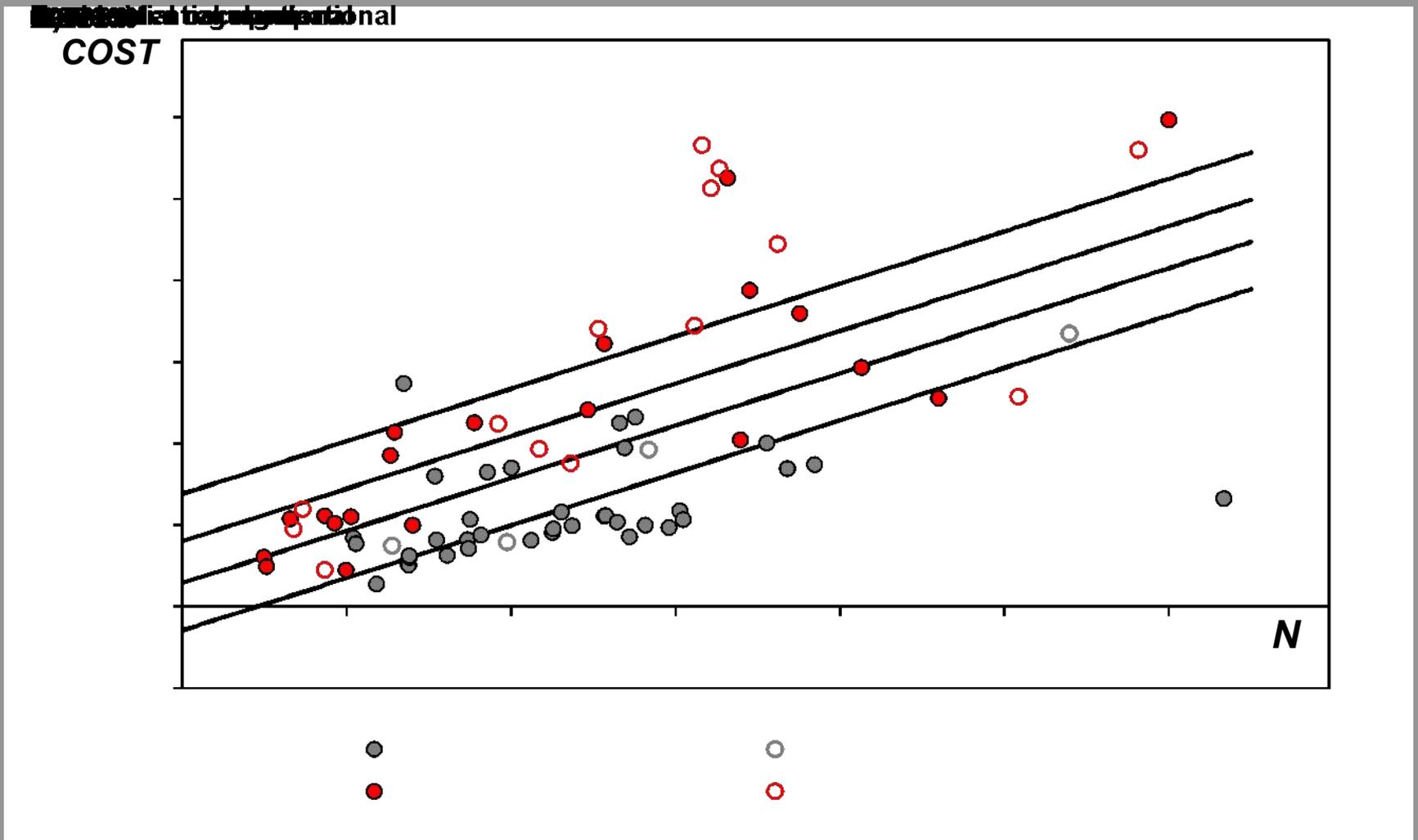
$$\begin{aligned}\hat{COST} &= -29,000 + 110,000 + 322N \\ &= 81,000 + 322N\end{aligned}$$

Профессиональные,
жилые ($OCC = 1; RES = 1$)

$$\begin{aligned}\hat{COST} &= -29,000 + 110,000 + 58,000 + \\ &322N \\ &= 139,000 + 322N\end{aligned}$$

Аналогичным образом, функции затрат для нежилых и жилых профессиональных школ производятся путем установления OCC равным 1 и RES равным 0 и 1 соответственно.

Два вида фиктивных переменных



Вот диаграмма рассеяния с четырьмя функциями стоимости, неявными в результате регрессии.

Два вида фиктивных переменных

```
. reg COST N OCC RES
```

Source	SS	df	MS			
Model	9.3297e+11	3	3.1099e+11	Number of obs =	74	
Residual	5.3838e+11	70	7.6911e+09	F(3, 70) =	40.43	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6341	
				Adj R-squared =	0.6184	
				Root MSE =	87699	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	321.833	39.40225	8.168	0.000	243.2477	400.4183
OCC	109564.6	24039.58	4.558	0.000	61619.15	157510
RES	57909.01	30821.31	1.879	0.064	-3562.137	119380.2
_cons	-29045.27	23291.54	-1.247	0.217	-75498.78	17408.25

t и F-тесты могут быть выполнены обычным способом. Коэффициент фиктивной переменной профессиональной школы значительно отличается от 0 на уровне значимости 0,1%.

Два вида фиктивных переменных

```
. reg COST N OCC RES
```

Source	SS	df	MS			
Model	9.3297e+11	3	3.1099e+11	Number of obs =	74	
Residual	5.3838e+11	70	7.6911e+09	F(3, 70) =	40.43	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6341	
				Adj R-squared =	0.6184	
				Root MSE =	87699	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	321.833	39.40225	8.168	0.000	243.2477	400.4183
OCC	109564.6	24039.58	4.558	0.000	61619.15	157510
RES	57909.01	30821.31	1.879	0.064	-3562.137	119380.2
_cons	-29045.27	23291.54	-1.247	0.217	-75498.78	17408.25

Однако отношение t для коэффициента RES составляет всего 1,87. К счастью, мы можем выполнить односторонний тест (почему?), Поэтому он значительно отличается от 0 на уровне 5% (но не на уровне 1%).