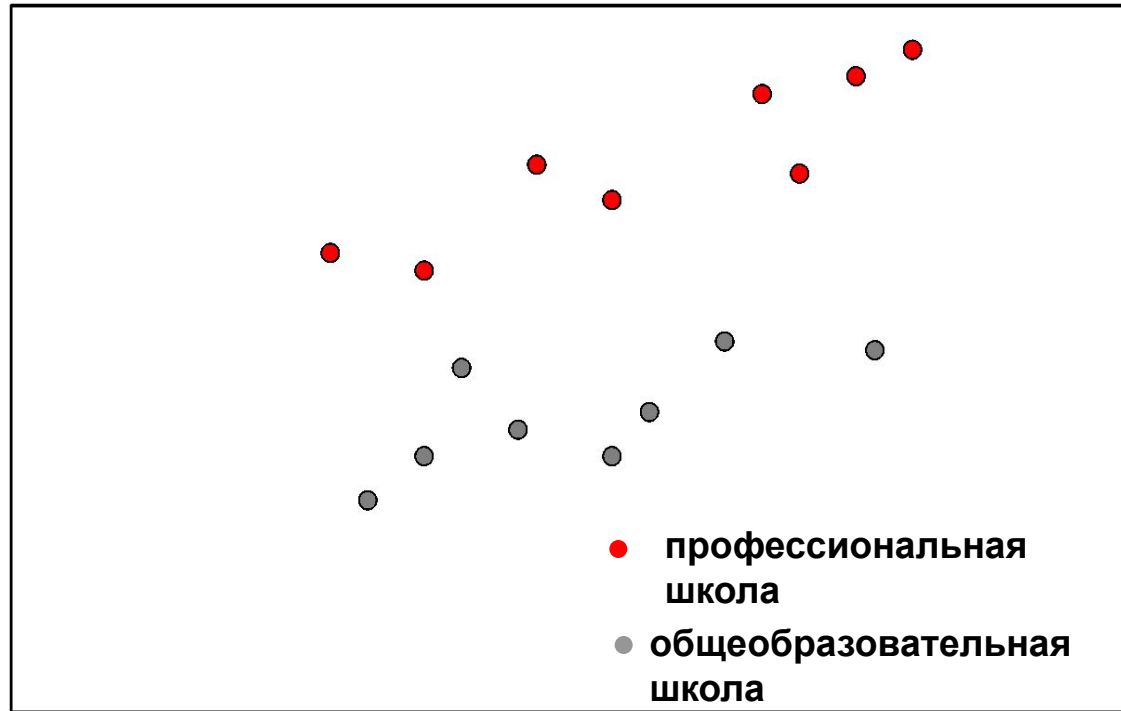


ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

COST

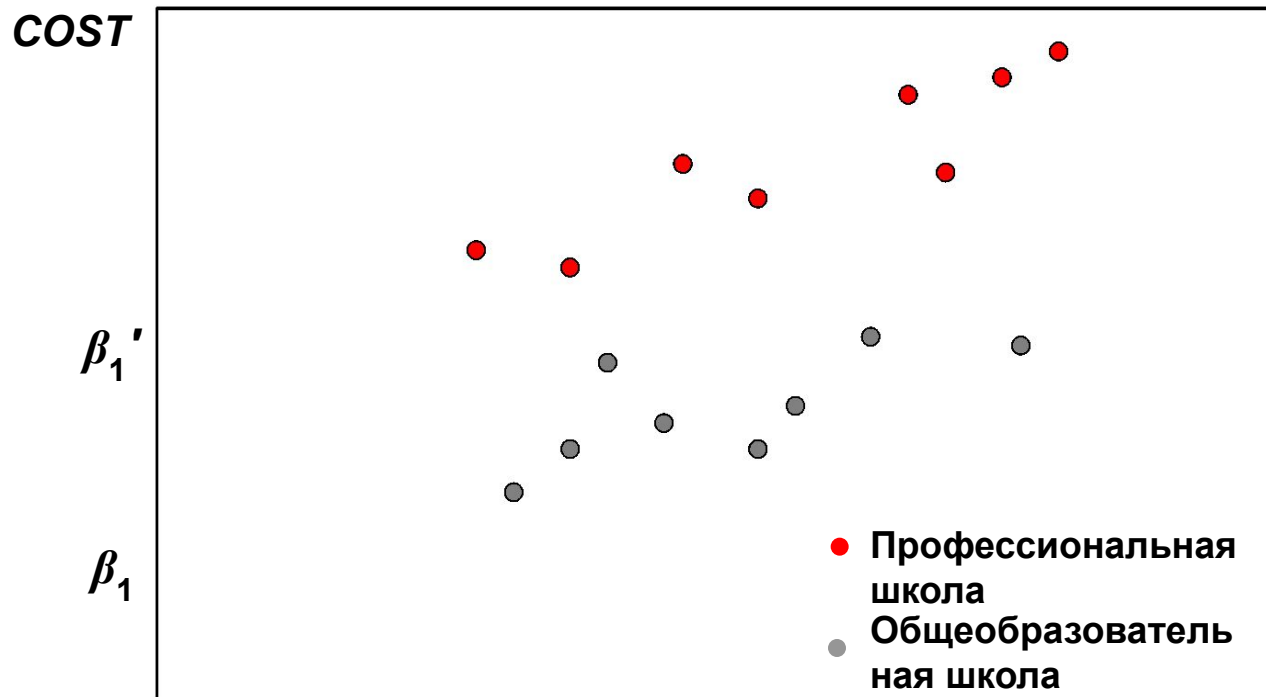
β_1'

β_1



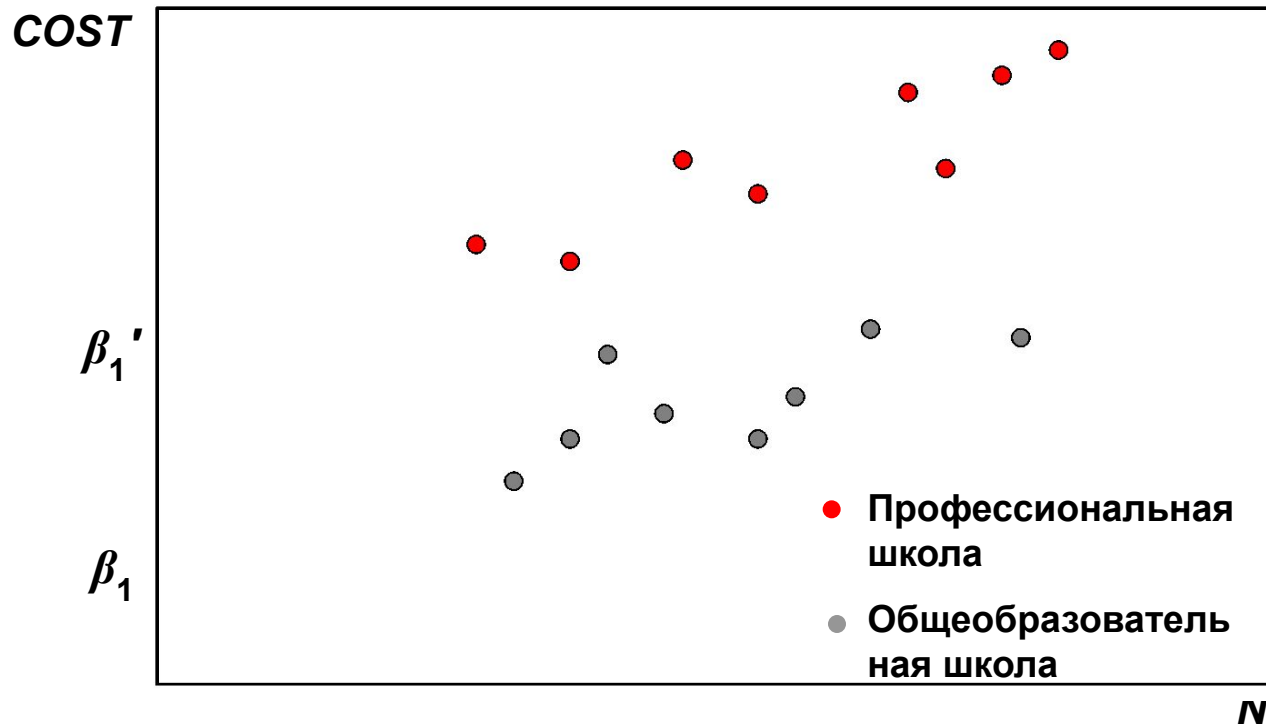
Эта последовательность объясняет, как Вы можете включить в себя качественно-экспрессивные переменные в свою регрессионную модель.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ



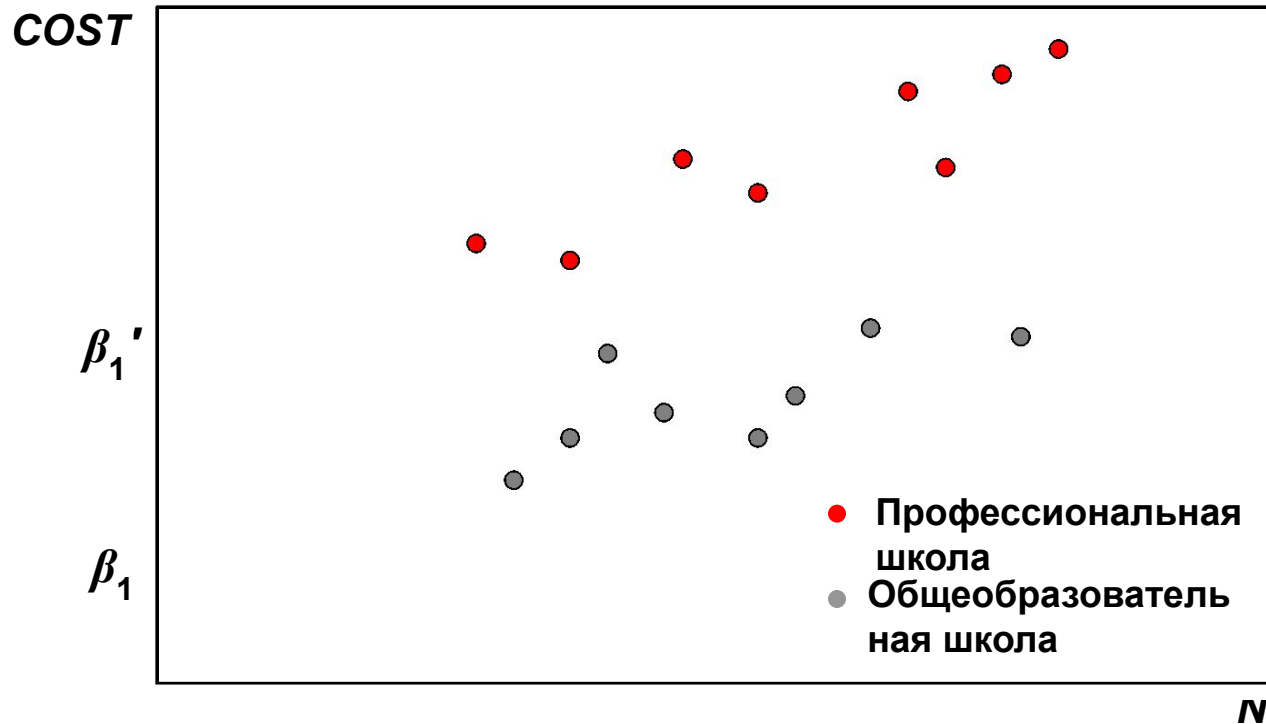
Предположим, что Вы имеете данные по ежегодным периодическим расходам, предельная цена, и число зарегистрированных студентов, N , для реализации средних школ, из которых есть два типа: регулярный и профессиональный.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ



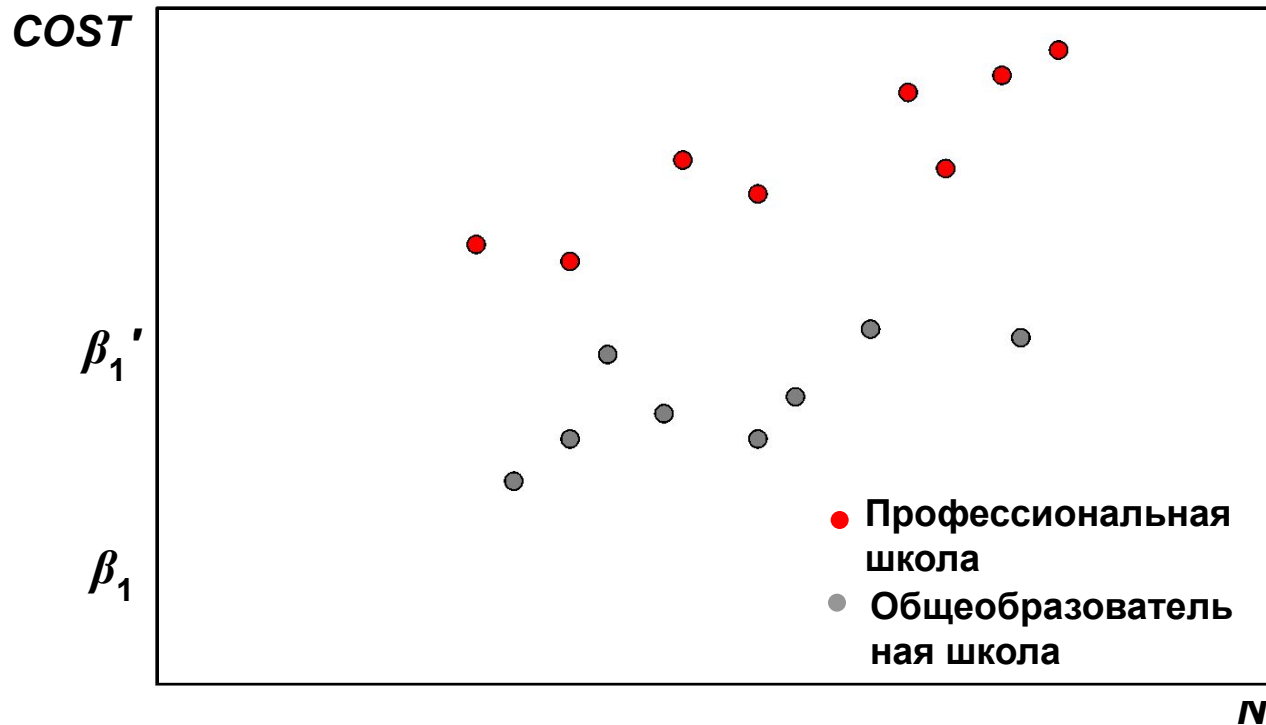
Профессиональные школы стремятся обеспечить навыками для определенных занятий, и они имеют тенденцию быть относительно дорогими, чтобы доказать интервал измерения, поэтому должны поддержать специализированные семинары.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ



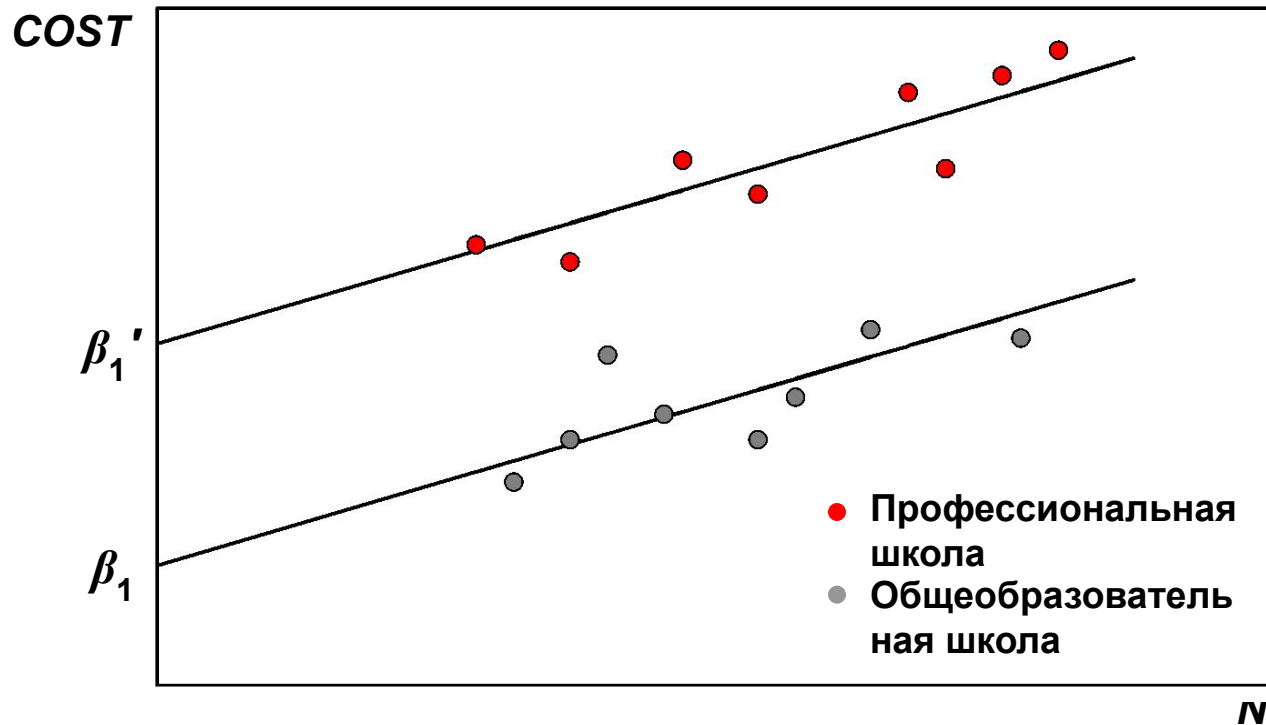
Один способ иметь дело с различием в затратах состоял бы в том, чтобы управлять отдельными регрессами для двух типов школы.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ



Однако, у этого был бы недостаток, что Вы будете управлять регрессами с двумя небольшими выборками вместо одной большой с отрицательным воздействием на точность оценок коэффициентов.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

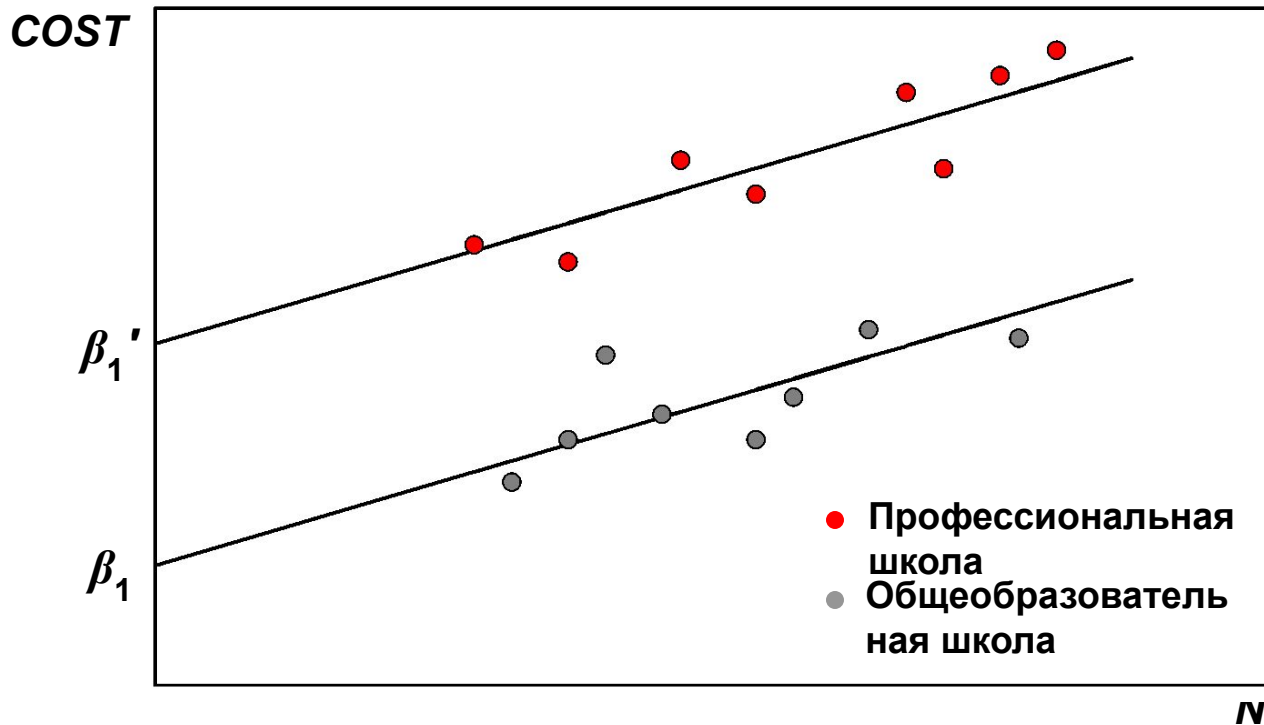


Общеобразовательная школа $COST = \beta_1 + \beta_2 N + u$

Профессиональная школа $COST = \beta_1' + \beta_2 N + u$

Другой способ обращаться с различием состоял бы в том, чтобы выдвинуть гипотезу, что у функции стоимости для профессиональных школ есть точка пересечения β_1' , который больше, чем это для обычных школ.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

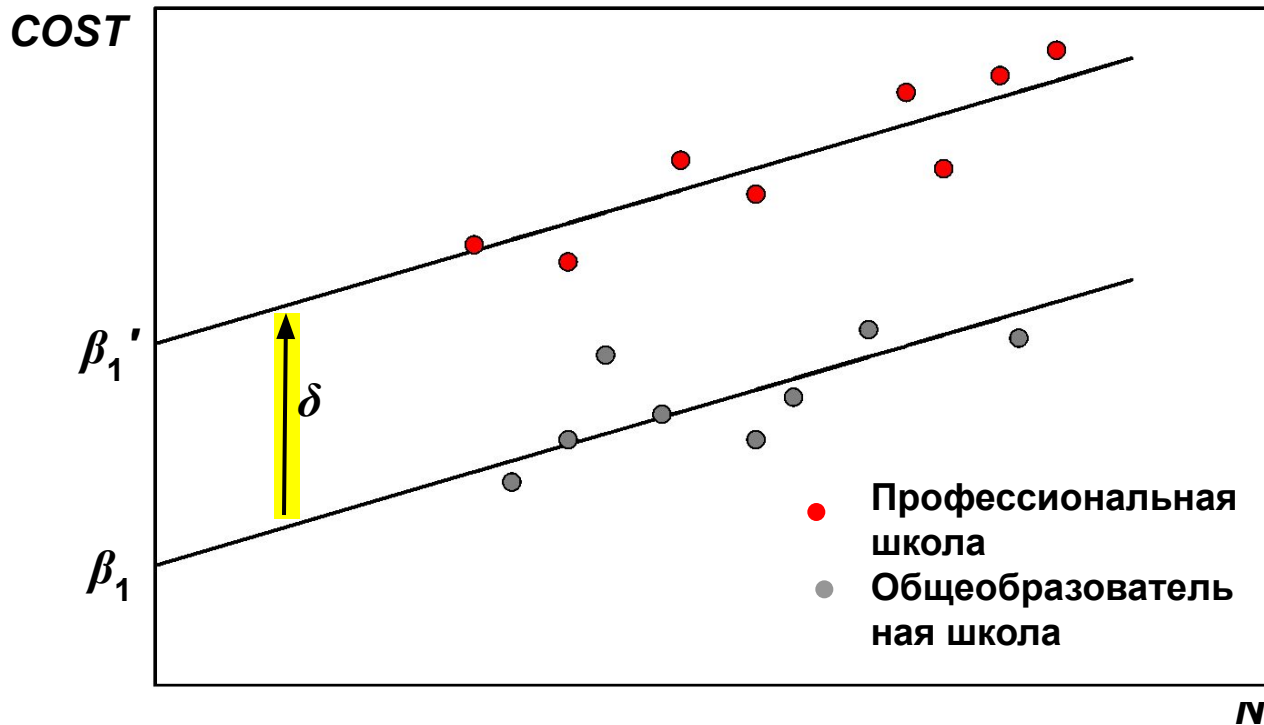


Общеобразовательная школа $COST = \beta_1 + \beta_2 N + u$

Профессиональная школа $COST = \beta_1' + \beta_2 N + u$

Эффективно, мы выдвигаем гипотезу, что ежегодный накладной расход отличается для двух типов школы, но крайняя стоимость - то же самое. Крайнее предположение стоимости не очень вероятно, и мы расслабим его в свое время.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ



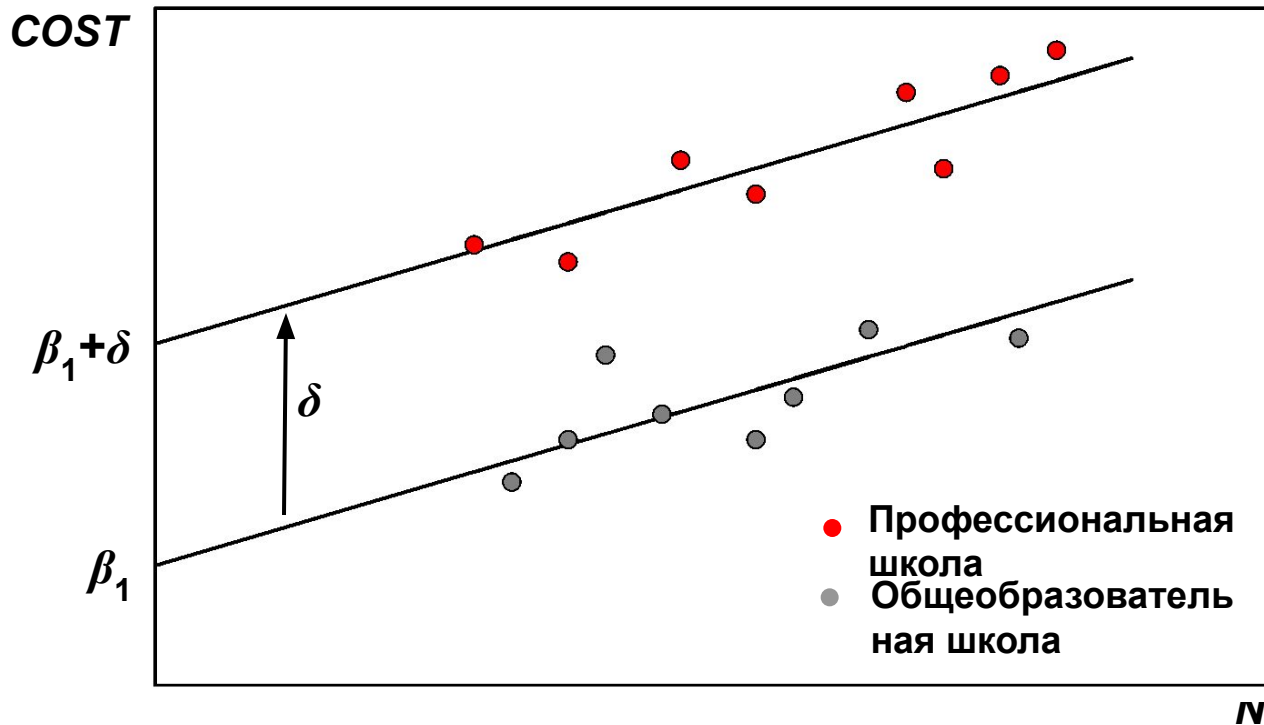
Общеобразовательная школа $COST = \beta_1 + \beta_2 N + u$

Профессиональная школа $COST = \beta_1' + \beta_2 N + u$

Заданная $\delta = \beta_1' - \beta_1$

Давайте определим δ чтобы быть различием в точках пересечения: $\delta = \beta_1' - \beta_1$.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ



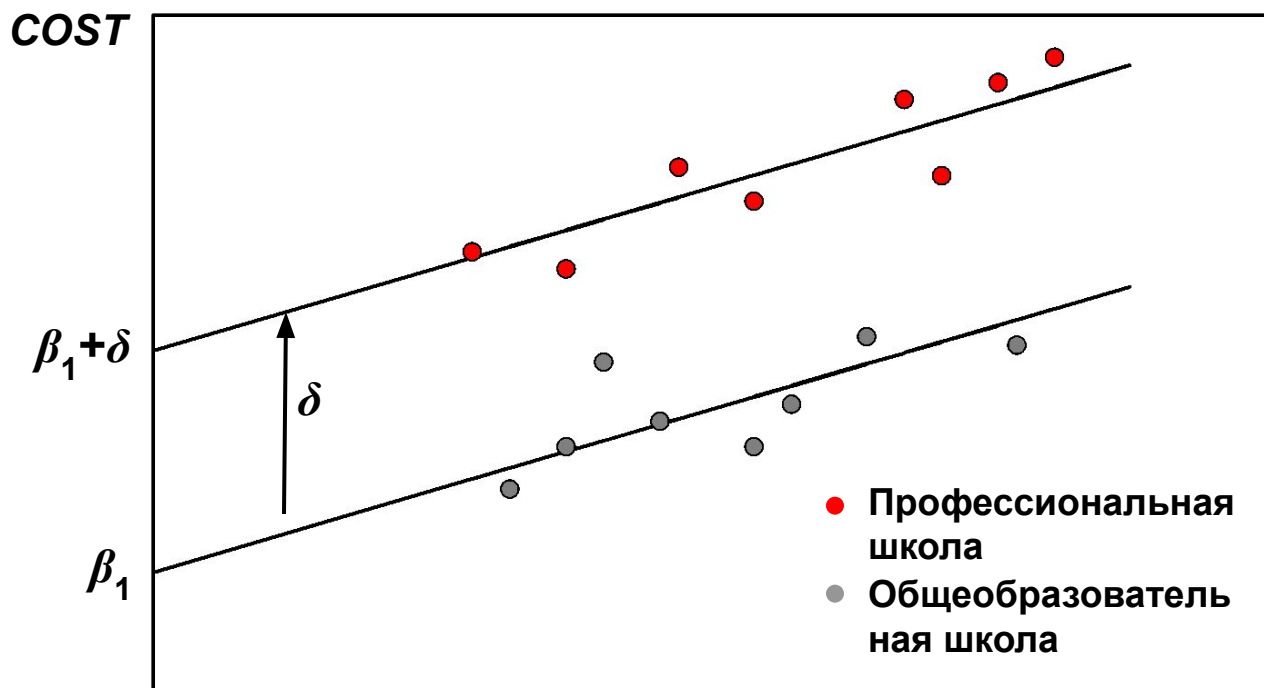
Общеобразовательная школа $COST = \beta_1 + \beta_2 N + u$

Профессиональная школа $COST = \beta_1 + \delta + \beta_2 N + u$

Заданная $\delta = \beta_1' - \beta_1$

Тогда $\beta_1' = \beta_1 + \delta$ и мы можем переписать функцию стоимости для профессиональных школ как показано.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ



IV

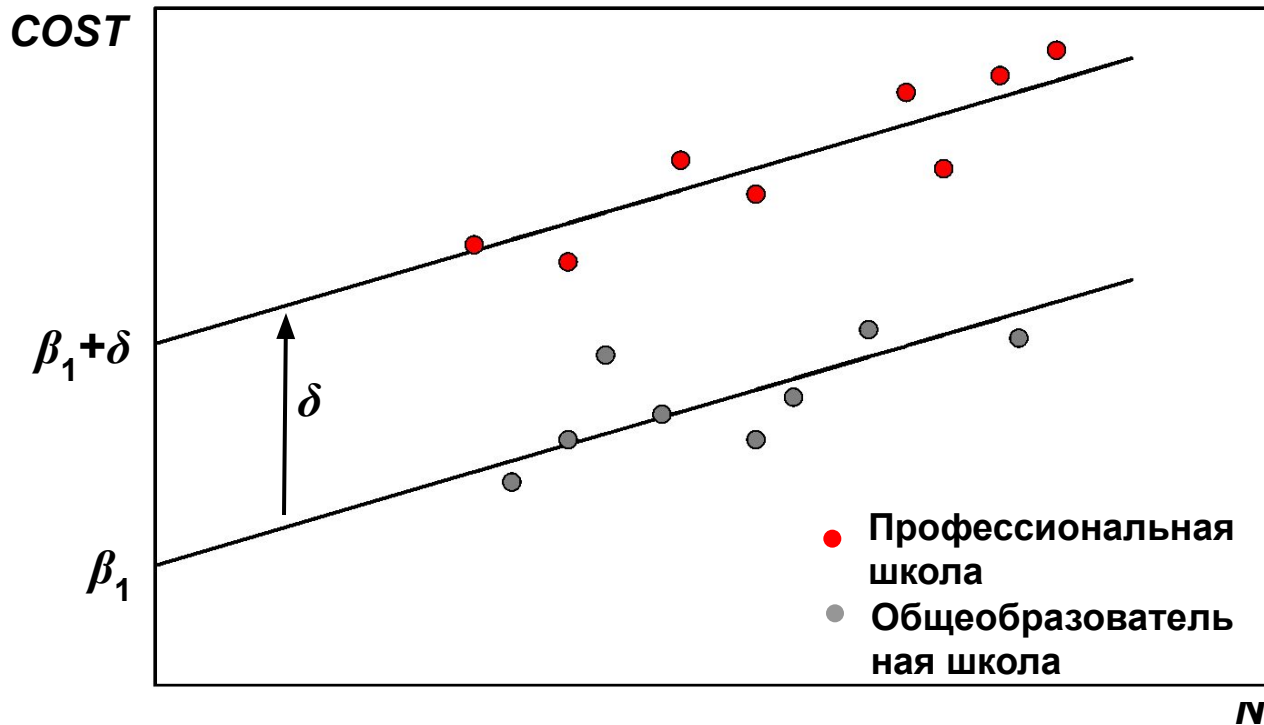
$OCC = 0$ Общеобразовательная школа $COST = \beta_1 + \beta_2 N + u$

$OCC = 1$ Профессиональная школа $COST = \beta_1 + \delta + \beta_2 N + u$

Combined equation $COST = \beta_1 + \delta OCC + \beta_2 N + u$

Мы можем теперь объединить две функции стоимости, определив фиктивный переменный OCC, у которого есть стоимость 0 для обычных школ и 1 для профессиональных школ.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ



OCC = 0 Общеобразовательная школа $COST = \beta_1 + \beta_2 N + u$

OCC = 1 Профессиональная школа $COST = \beta_1 + \delta + \beta_2 N + u$

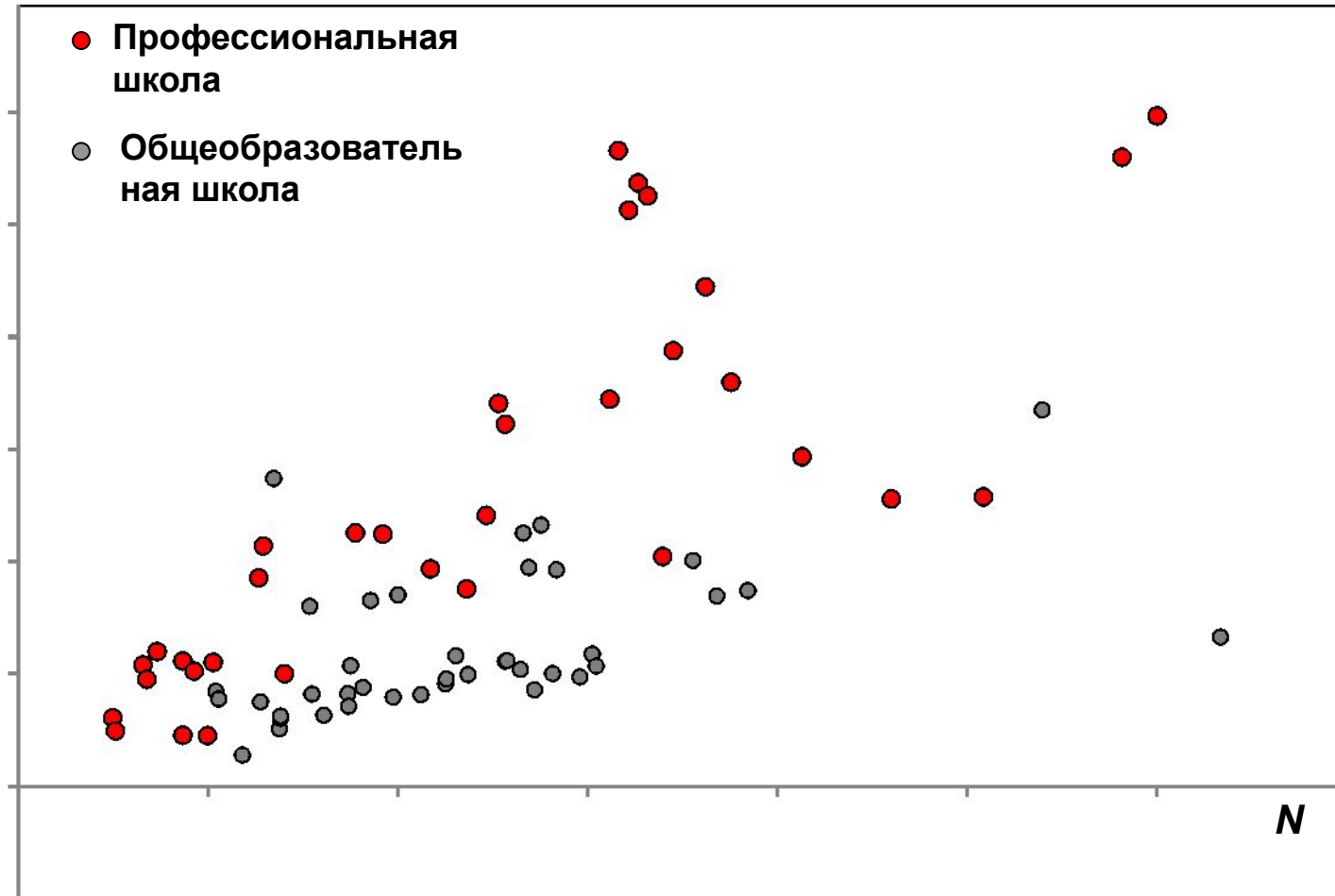
Combined equation $COST = \beta_1 + \delta OCC + \beta_2 N + u$

У фиктивных переменных всегда есть две ценности, 0 или 1. Если OCC равен 0, функция стоимости становится этим для обычных школ. Если OCC равен 1, функция стоимости становится этим для профессиональных школ.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

■■■■■■■■■■
COST

- Профессиональная школа
- Общеобразовательная школа



Мы будем теперь соответствовать функции этого типа, используя фактические данные для образца 74 средних школ в Шанхае.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

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

Таблица показывает данные для первых 10 школ в образце. Ежегодная стоимость измерена в юане, один юань, будучи стоимостью в США за приблизительно 20 центов в то время. N - число студентов в школе.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

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

OCC - фиктивная переменная для типа школы.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

Мы теперь управляем регрессом СТОИМОСТИ на N и OCC, рассматривая OCC точно так же, как любая другая объяснительная переменная, несмотря на ее искусственный характер. Продукцию Stata показывают.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

Мы начнем, интерпретируя коэффициенты регрессии .

$$\hat{COST} = -34,000 + 133,000OCC + 331N$$

Результаты регресса были переписаны в форме уравнения. От него мы можем получить функции стоимости для двух типов школы, установив OCC, равный 0 или 1.

$$\hat{COST} = -34,000 + 133,000OCC + 331N$$

Общеобразовательная
школа ($OCC = 0$)

$$\hat{COST} = -34,000 + 331N$$

Если OCC равен 0, мы получаем уравнение для обычных школ, как показано. Это подразумевает, что крайняя стоимость за студента в год составляет 331 юань и что ежегодный накладной расход составляет –34 000 юаней

$$\hat{COST} = -34,000 + 133,000OCC + 331N$$

Общеобразовательная школа ($OCC = 0$)

$$\hat{COST} = -34,000 + 331N$$

Очевидно, наличие отрицательной точки пересечения не имеет никакого смысла вообще, и это предполагает, что модель – неверно заданная в некотором роде. Мы возвратимся к этому позже.

$$\hat{COST} = -34,000 + 133,000OCC + 331N$$

Общеобразовательная школа ($OCC = 0$)

$$\hat{COST} = -34,000 + 331N$$

Коэффициент фиктивной переменной - оценка δ , дополнительный ежегодный накладной расход профессиональной школы .

$$\hat{COST} = -34,000 + 133,000OCC + 331N$$

Общеобразовательная школа (OCC = 0)

$$\hat{COST} = -34,000 + 331N$$

Профессиональная школа (OCC = 1)

$$\begin{aligned} \hat{COST} &= -34,000 + 133,000 + 331N \\ &= 99,000 + 331N \end{aligned}$$

Помещение OCC равняется 1, мы оцениваем, что ежегодный накладной расход профессиональной школы составляет 99,000 юаней. Крайняя стоимость совпадает с для обычных школ. Это должно быть, учитывая образцовую спецификацию.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

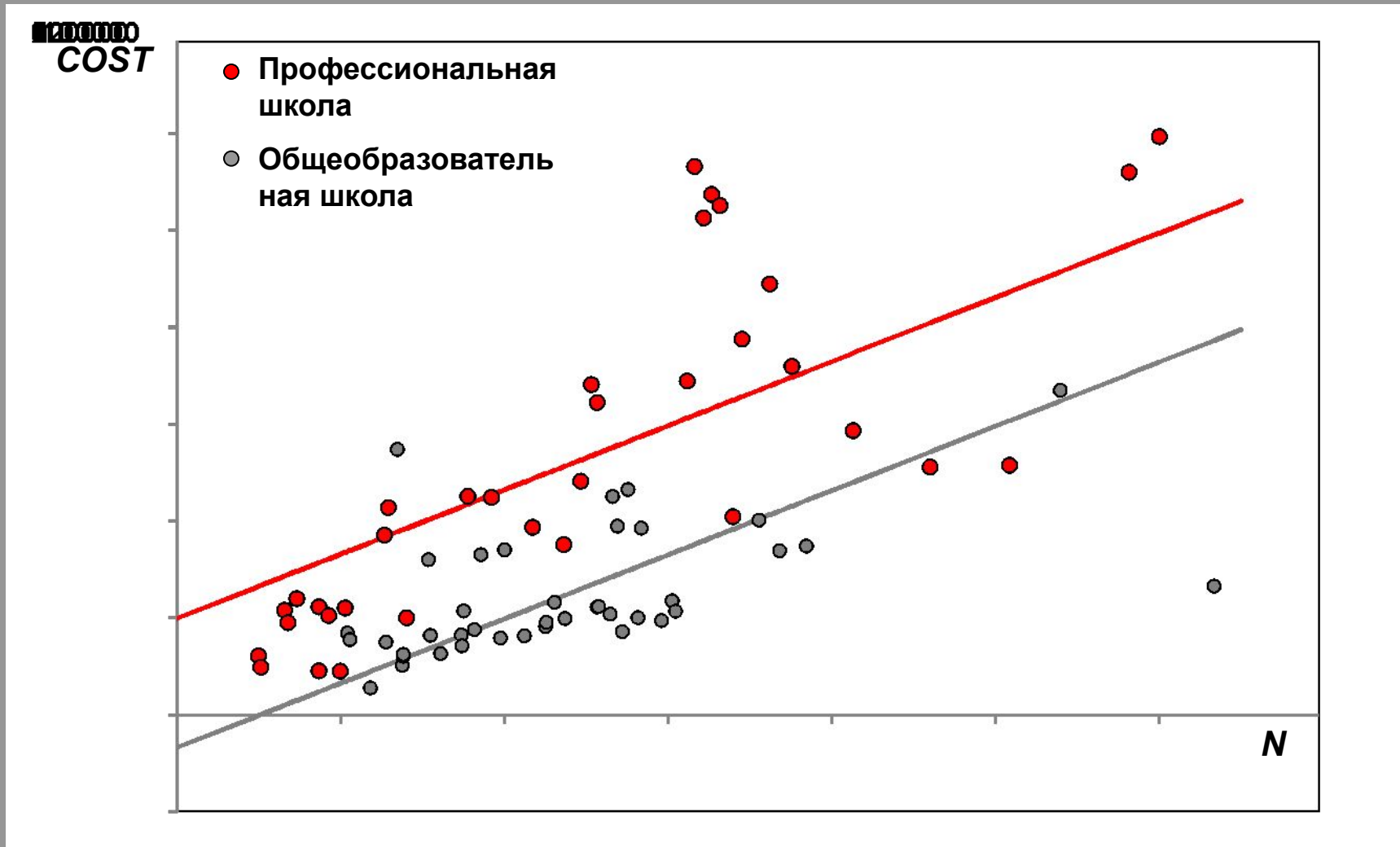


Диаграмма разброса показывает данные и две функции стоимости, полученные из результатов регресса.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

В дополнение к оценкам коэффициентов результаты регресса будут включать стандартные ошибки и обычную диагностическую статистику.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

We will perform a t test on the coefficient of the dummy variable. Our null hypothesis is $H_0: \delta = 0$ and our alternative hypothesis is $H_1: \delta \neq 0$.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

В словах наша нулевая гипотеза - то, что нет никакого различия в накладных расходах двух типов школы. Т статистическая величина 6.40, таким образом, она отклонена на уровне значения на 0.1%.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs = 74		
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) = 56.86		
Total	1.4713e+12	73	2.0155e+10	Prob > F = 0.0000		
				R-squared = 0.6156		
				Adj R-squared = 0.6048		
				Root MSE = 89248		

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

Мы можем выполнить тесты t на других коэффициентах обычным способом. T статистическая величина для коэффициента N 8.34, таким образом, мы приходим к заключению, что крайняя стоимость (очень) существенно отличается от 0

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

В случае точки пересечения t статистическая величина -1.43 , таким образом, мы не отклоняем нулевую гипотезу $H_0: \beta_1 = 0$.

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

Таким образом одно объяснение бессмысленного отрицательного накладного расхода обычных школ могло бы состоять в том, что у них на самом деле нет накладных расходов, и наша оценка - случайное число .

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

Более реалистическая версия этой гипотезы - то, что β_1 положительный, но маленький (как Вы видите, 95-процентный доверительный интервал включает положительные ценности), и остаточный член ответственен за отрицательную оценку

ФИКТИВНЫЕ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

```
. reg COST N OCC
```

Source	SS	df	MS			
Model	9.0582e+11	2	4.5291e+11	Number of obs =	74	
Residual	5.6553e+11	71	7.9652e+09	F(2, 71) =	56.86	
Total	1.4713e+12	73	2.0155e+10	Prob > F =	0.0000	
				R-squared =	0.6156	
				Adj R-squared =	0.6048	
				Root MSE =	89248	

COST	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
N	331.4493	39.75844	8.337	0.000	252.1732	410.7254
OCC	133259.1	20827.59	6.398	0.000	91730.06	174788.1
_cons	-33612.55	23573.47	-1.426	0.158	-80616.71	13391.61

Как уже отмечено, дальнейшая возможность состоит в том, что модель – неверно заданная в некотором роде. Мы продолжим разрабатывать модель в следующей последовательности.