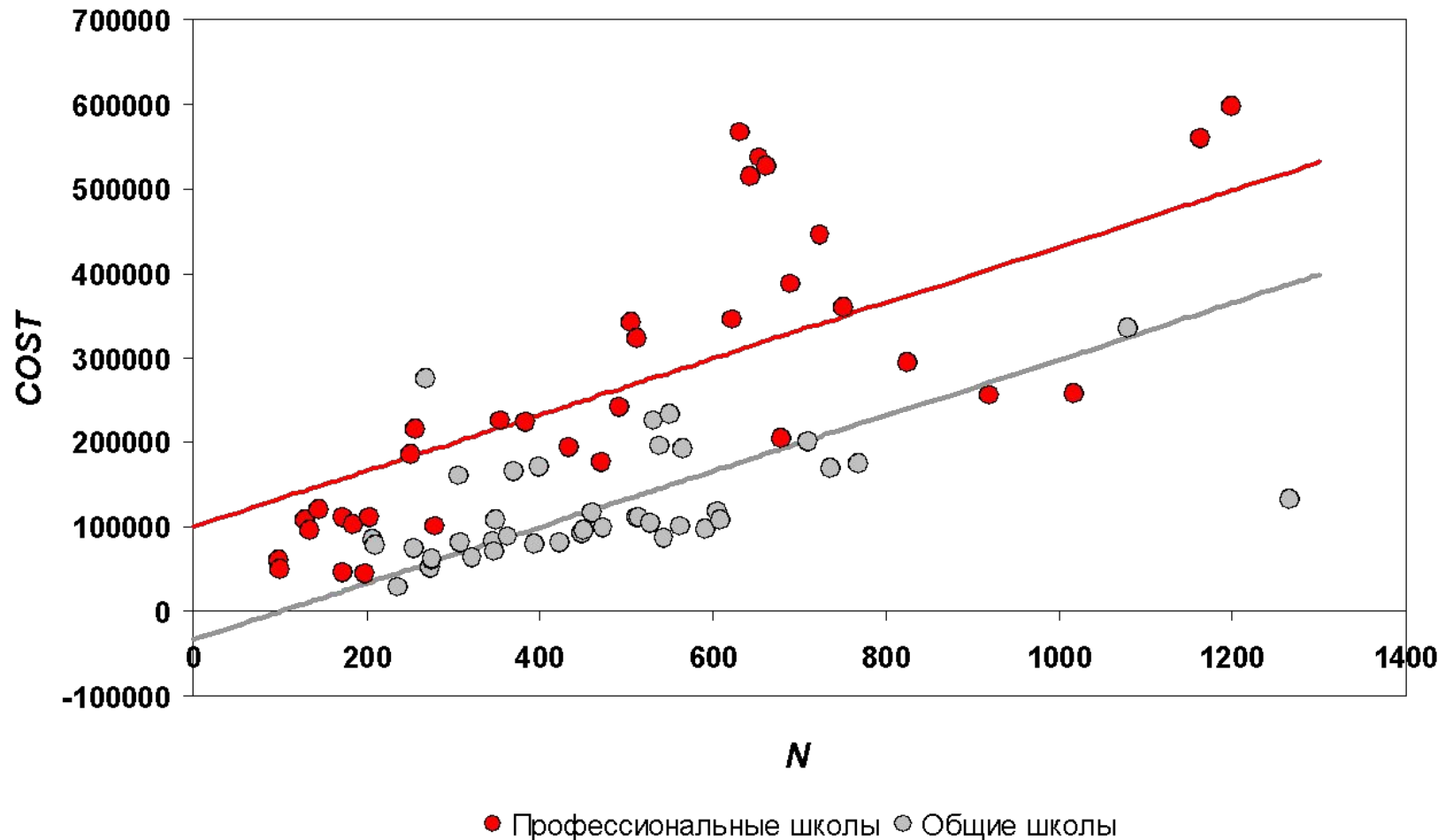


**Dummy- переменные для
коэффициентов наклона**

DUMMY- переменные для коэффициентов наклона



На диаграмме изображены наблюдения для 74 школ в Шанхае и проведены линии регрессии, оцененной в предположении об одинаковых предельных издержках (коэффициентах наклона) для обычных и профессиональных школ.

DUMMY- переменные для коэффициентов наклона

$$COST = \beta_0 + \delta OCC + \beta_1 N + \lambda N * OCC + \xi$$

Ослабим требование об одинаковых предельных издержках (коэффициентах наклона) для обычных и профессиональных школ. Введем переменную $NOCC$, произведение N и OCC .

DUMMY- переменные для коэффициентов наклона

$$COST = \beta_0 + \delta OCC + \beta_1 N + \lambda N * OCC + \xi$$

Обычные школы
($OCC = N * OCC = 0$)

$$COST = \beta_0 + \beta_1 N + \xi$$

Для обычных школ переменная OCC равна 0 и, следовательно, $NOCC$ также равна 0.

DUMMY- переменные для коэффициентов наклона

$$COST = \beta_0 + \delta OCC + \beta_1 N + \lambda N * OCC + \xi$$

Общие школы
($OCC = N * OCC = 0$)

$$COST = \beta_0 + \beta_1 N + \xi$$

Профессиональные школы
($OCC = 1; N * OCC = N$)

$$COST = (\beta_0 + \delta) + (\beta_1 + \lambda)N + \xi$$

Для профессиональных школ переменная OCC равна 1, следовательно, переменная $N * OCC$ равна N .

DUMMY- переменные для коэффициентов наклона

$$COST = \beta_0 + \delta OCC + \beta_1 N + \lambda N * OCC + \xi$$

Общие школы
($OCC = N * OCC = 0$)

$$COST = \beta_0 + \beta_1 N + \xi$$

Профессиональные школы
($OCC = 1; N * OCC = N$)

$$COST = (\beta_0 + \delta) + (\beta_1 + \lambda)N + \xi$$

Предельные издержки на одного студента профессиональной школы больше на λ по сравнению с расходами на одного студента обыкновенной школы, постоянные издержки различаются на δ .

DUMMY- переменные для коэффициентов наклона

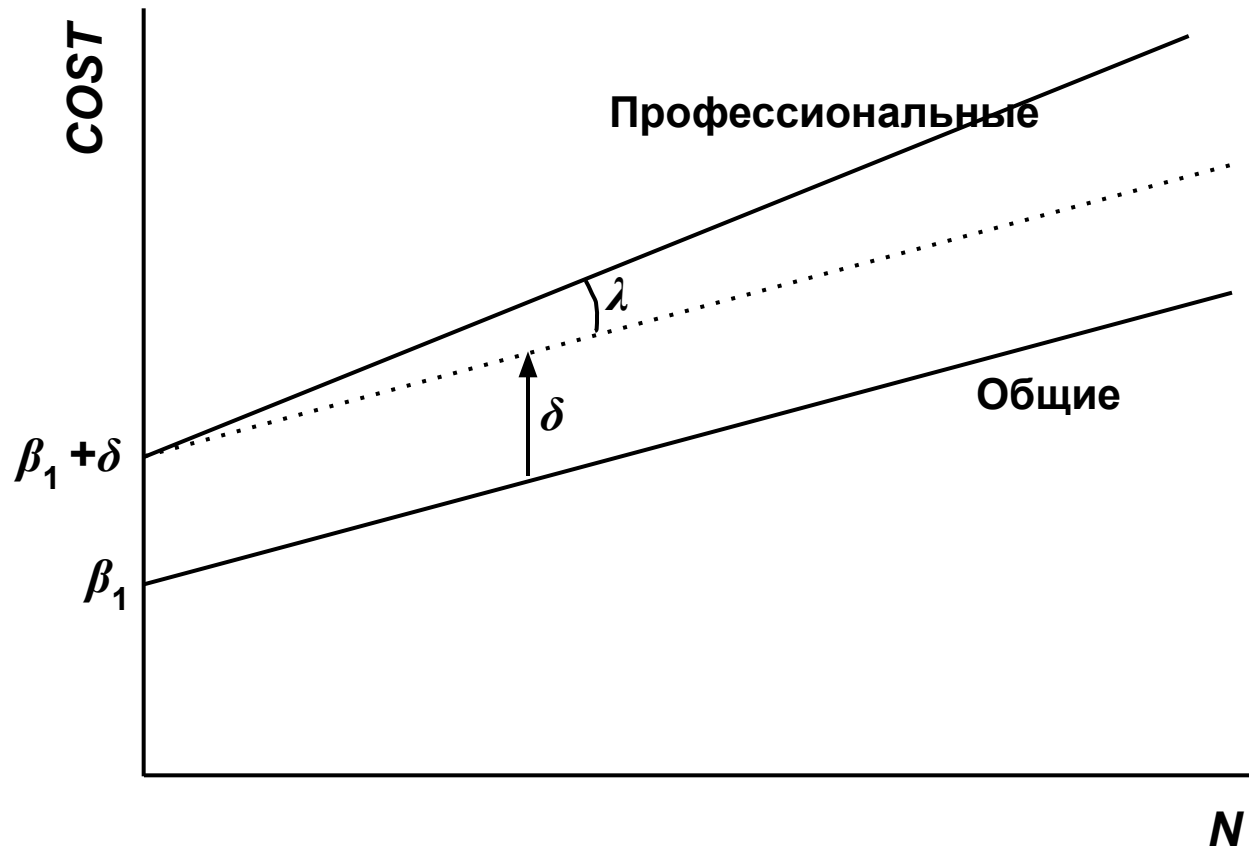


Диаграмма иллюстрирует эту разницу графически.

DUMMY- переменные для коэффициентов наклона

	Тип школы	<i>COST</i>	<i>N</i>	<i>OCC</i>	<i>N*OCC</i>
1	Профессиональные	345,000	623	1	623
2	Профессиональные	537,000	653	1	653
3	Обычные	170,000	400	0	0
4	Профессиональные	526.000	663	1	663
5	Обычные	100,000	563	0	0
6	Обычные	28,000	236	0	0
7	Обычные	160,000	307	0	0
8	Профессиональные	45,000	173	1	173
9	Профессиональные	120,000	146	1	146
10	Профессиональные	61,000	99	1	99

В таблице приведены данные для первых 10 школ.
Дополнительно определена переменная *N*OCC*.

DUMMY- переменные для коэффициентов 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
N*OCC	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

Таблица оцененной регрессии.

DUMMY- переменные для коэффициентов наклона

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

Это уравнение оцененной регрессии.

DUMMY- переменные для коэффициентов наклона

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

Обычные школы ($OCC = NOCC = 0$)	$COST = 51,000 + 152N$

Для общих школ OCC и $NOCC$ равны 0, соответственно, постоянные и предельные издержки для студентов общих школ равны 51,000 юаней и 152 юаня.

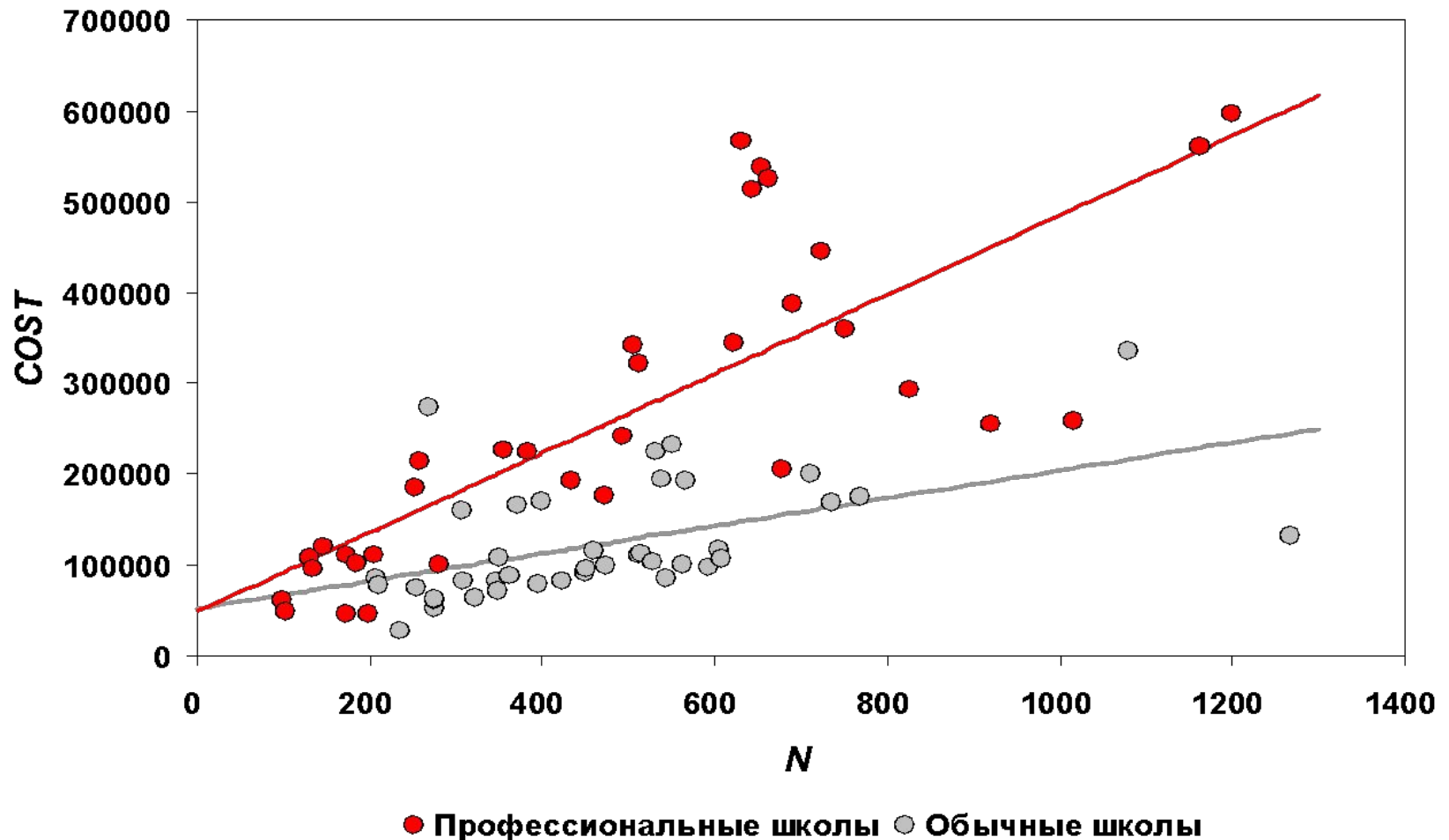
DUMMY- переменные для коэффициентов наклона

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

Обычные школы ($OCC = N * OCC = 0$)	$COST = 51,000 + 152N$
Профессиональные школы ($OCC = 1; N * OCC = N$)	$COST = 51,000 - 4,000 + 152N + 284N$ $= 47,000 + 436N$

Для профессиональных школ OCC равна 1, следовательно, $NOCC$ равна N , соответственно постоянные и предельные издержки для студентов профессиональных школ равны 47,000 юаней и 436 юаней.

DUMMY- переменные для коэффициентов наклона



На рисунке приведены графики оцененных регрессий для профессиональных и обычных школ.

DUMMY- переменные для коэффициентов 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
N*OCC	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

t – статистика переменной $N*OCC$ равна 3.76, этот коэффициент значим, следовательно, предельные расходы для студентов обычных и профессиональных школ различаются.

DUMMY- переменные для коэффициентов наклона

```
. reg COST N OCC N*OCC
```

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

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
N*OCC	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 при переменной OCC незначим, следовательно, постоянные расходы не различаются.

Проведем F – тест на значимость группы дитту-переменных.

Тестирование гипотезы о группе переменных

В учебниках для расчета $F_{расч}$ также можно встретить формулу:

$$F_{расч} = \frac{(RSS_R - RSS_{UR}) / q}{RSS_{UR} / (n - k - 1)}$$

Она даёт тот же результат, что и формула

$$F_{расч} = \frac{(R_{UR}^2 - R_R^2) / q}{(1 - R_{UR}^2) / (n - k - 1)}$$

DUMMY- переменные для коэффициентов наклона

```
. reg COST N OCC N*OCC
```

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

Проведем F – тест на значимость группы dummy-переменных.

DUMMY- переменные для коэффициентов наклона

```
. reg COST N OCC N*OCC
```

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

Нулевая гипотеза состоит в том, что коэффициенты перед переменными OCC и $N*OCC$ одновременно равны 0. Альтернативной является двусторонняя гипотеза.

DUMMY- переменные для коэффициентов наклона

$$F_{расч} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k-1)}$$

```
. reg COST N OCC N*OCC
```

Source	SS	df	MS
Model	1.0009e+12	3	3.3363e+11
Residual	4.7045e+11	70	6.7207e+09
Total	1.4713e+12	73	2.0155e+10

Number of obs = 74
 F(3, 70) = 49.64
 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
Residual	8.9160e+11	72	1.2383e+10
Total	1.4713e+12	73	2.0155e+10

Number of obs = 74
 F(1, 72) = 46.82
 Prob > F = 0.0000
 R-squared = 0.3940
 Adj R-squared = 0.3856
 Root MSE = 1.1e+05

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

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

Находим значение F – статистики и сравниваем его с критическим. Поскольку значение F- статистики больше критического (при любом разумном уровне значимости), то нулевая гипотеза отвергается.

Интерпретация. Пример.

По данным опроса 500 домохозяйств проведено исследование зависимости расходов на продукты питания в месяц (Y , тыс. руб.) от располагаемого ежемесячного дохода (X , тыс. руб.), Z – фиктивная переменная, равная 0 для городских жителей и 1 для сельских

$$Y = 7 + 0,5 * X - 6 * Z - 0,3 * X * Z$$

- 1) В соответствии с полученными результатами, какое влияние на продукты питания оказывает располагаемый доход (X)
- 2) Дайте интерпретацию коэффициенту при фиктивной переменной (Z)
- 3) Дайте интерпретацию коэффициенту при произведении переменных X и Z