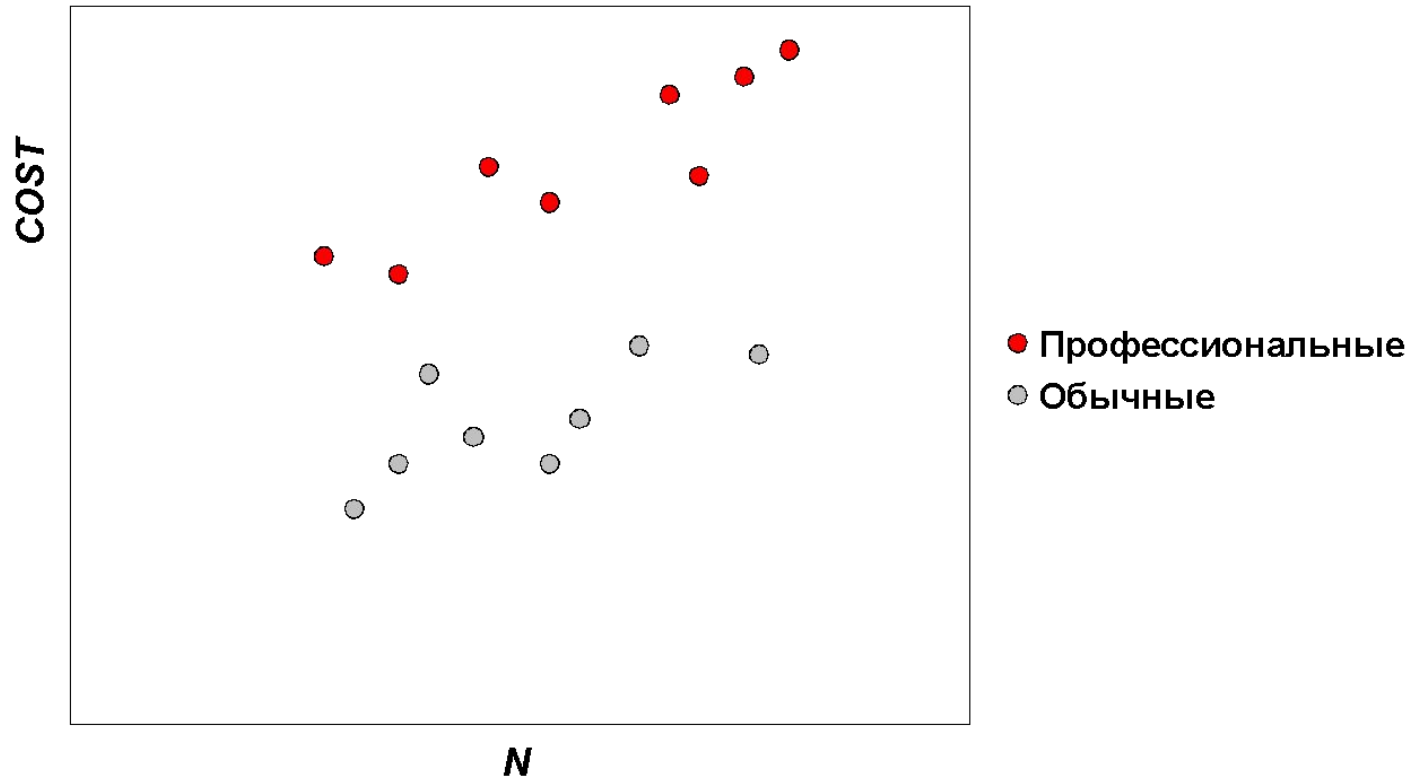


**Фиктивные (бинарные,
дамми) переменные**

Фиктивные переменные —
бинарные переменные
(принимают значения 0 или 1)

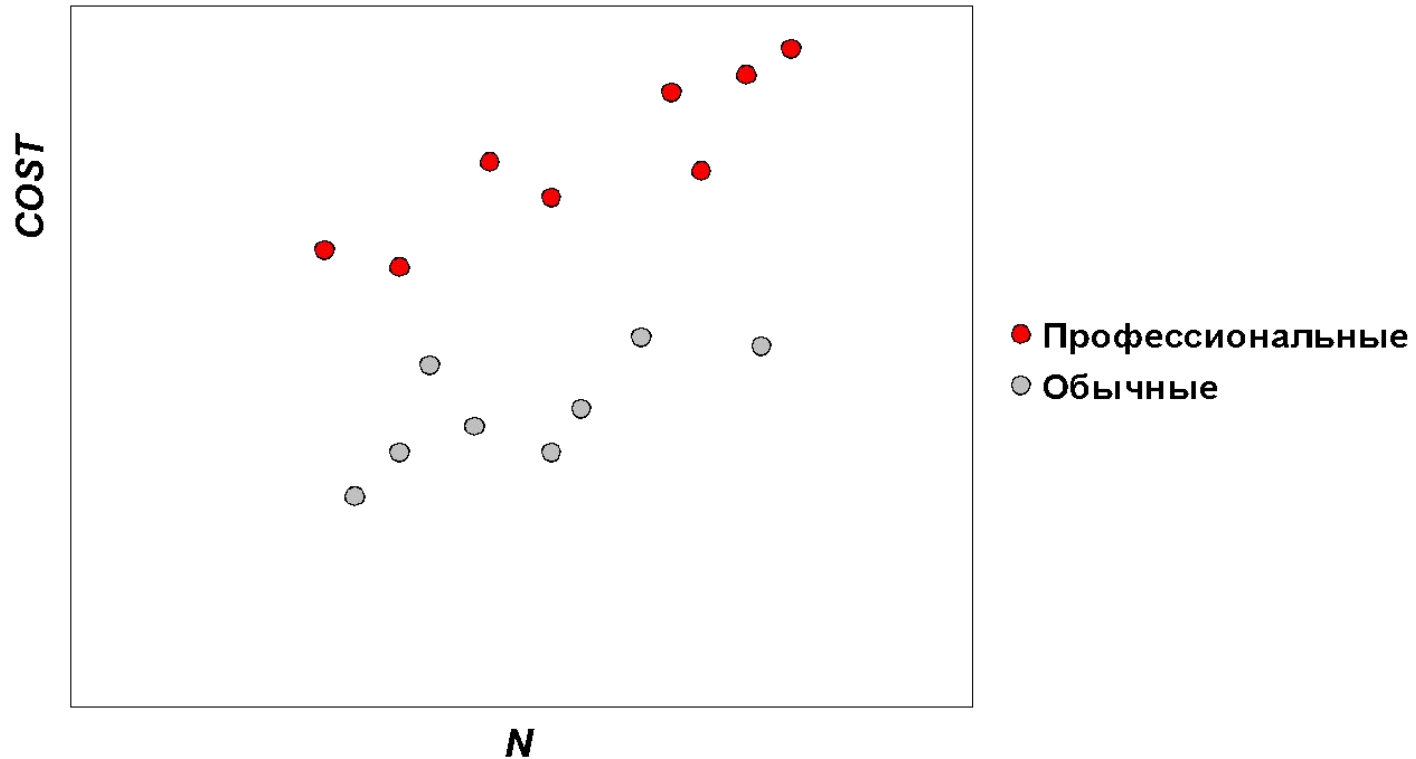
Используются для
моделирования качественных
признаков

Пример использования `dummy` переменной при наличии двух категорий



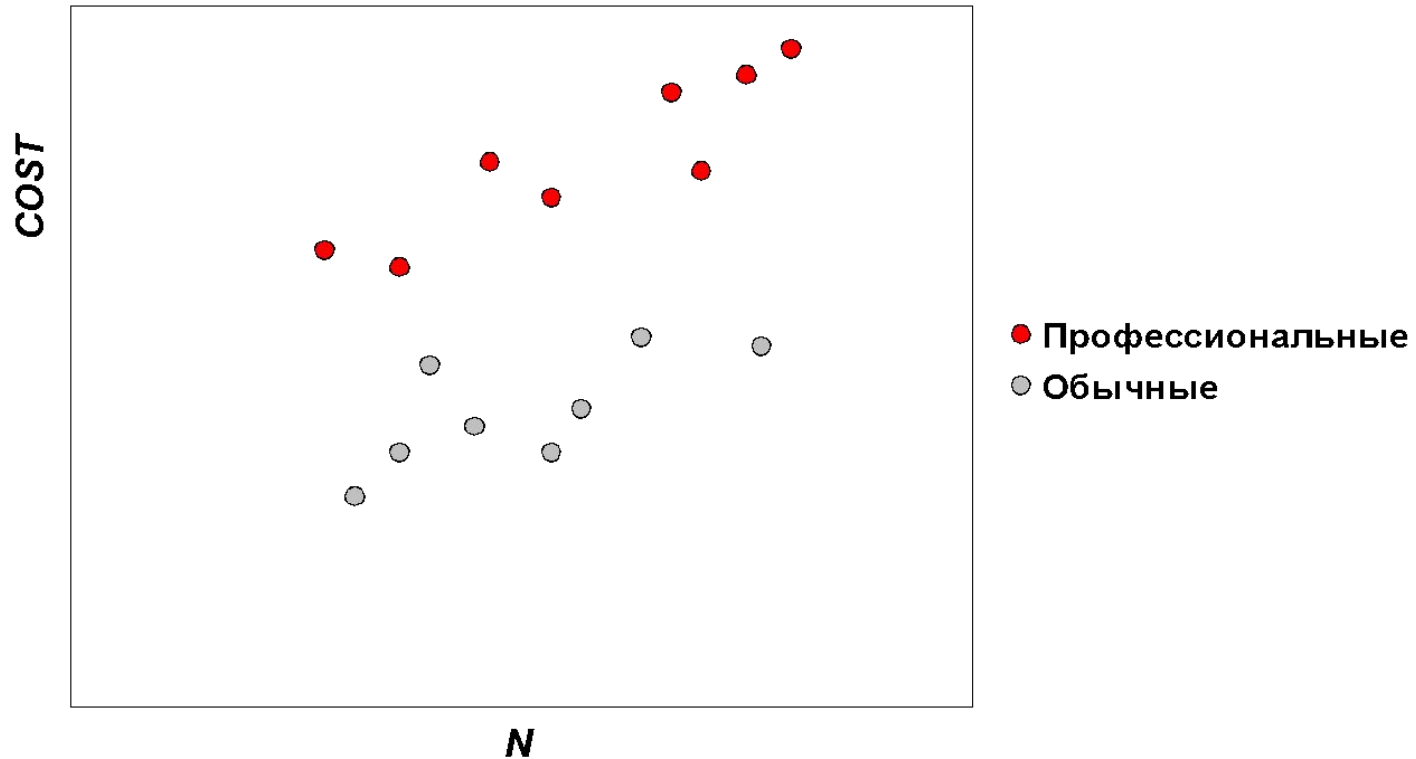
$COST$ – годовые издержки 74 средних школ в Шанхае в середине 1980-х годов, N – количество обучавшихся в них учеников.

Пример использования dummy переменной при наличии двух категорий



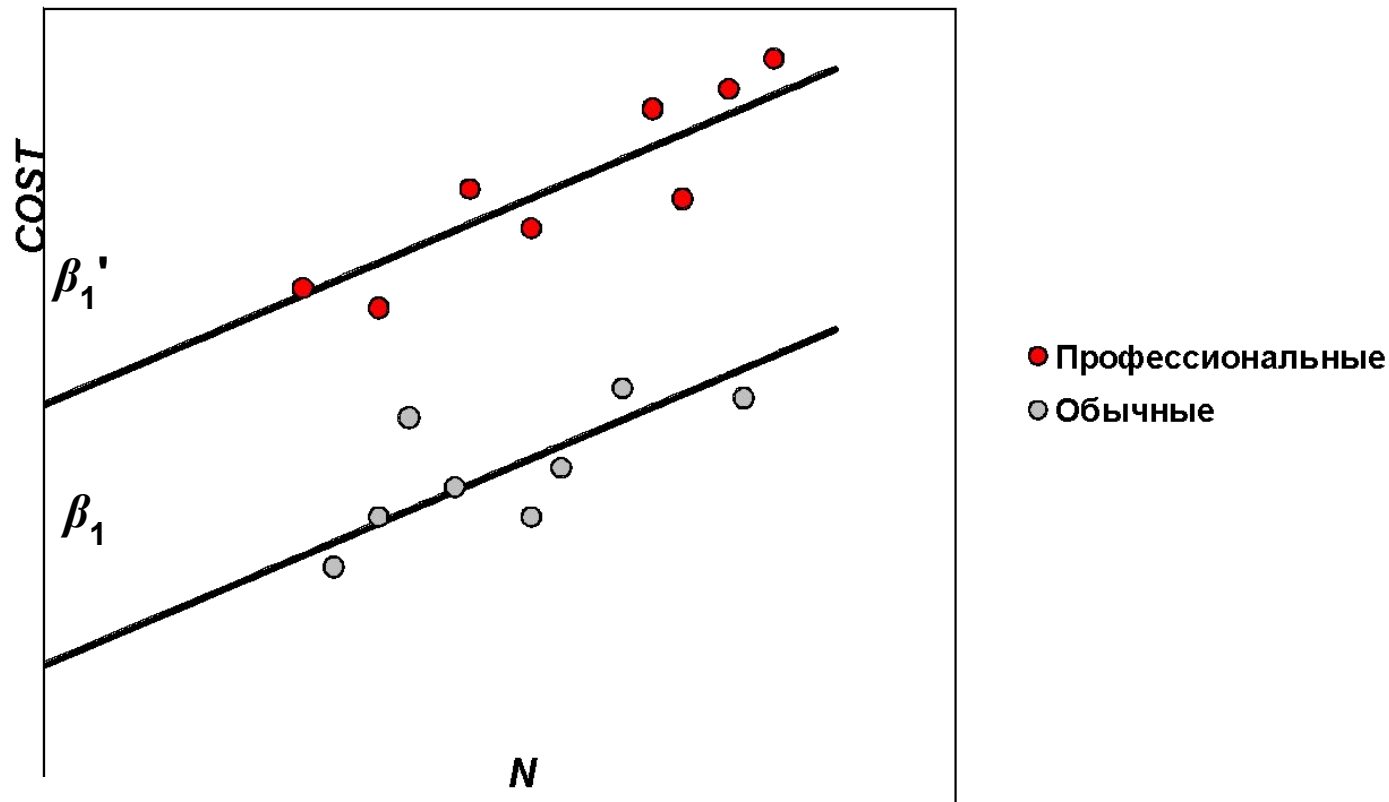
Затраты в профессиональных школах больше, т.к. для обучения там используется специальное оборудование.

Пример использования dummy переменной при наличии двух категорий



Если оценивать регрессии отдельно для профессиональных и обычных школ, то размеры выборок уменьшатся, что снизит точность оценивания.

Пример использования dummy переменной при наличии двух категорий

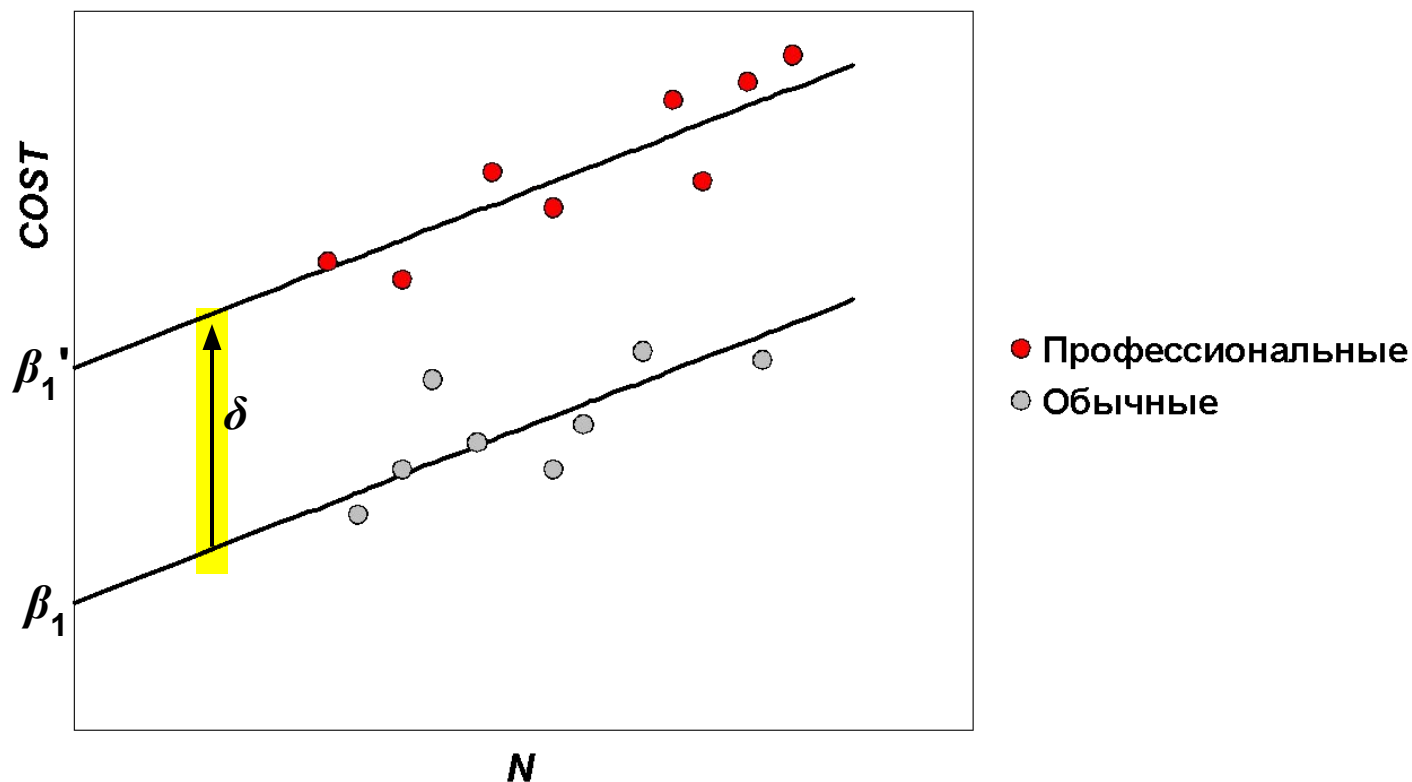


Обычные школы $COST = \beta_0 + \beta_1 N + \xi$

Профессиональные школы $COST = \beta_0' + \beta_1 N + \xi$

Предположим, что коэффициенты наклона в регрессиях для профессиональных и обычных школ совпадают, а свободные члены различаются.

Пример использования dummy переменной



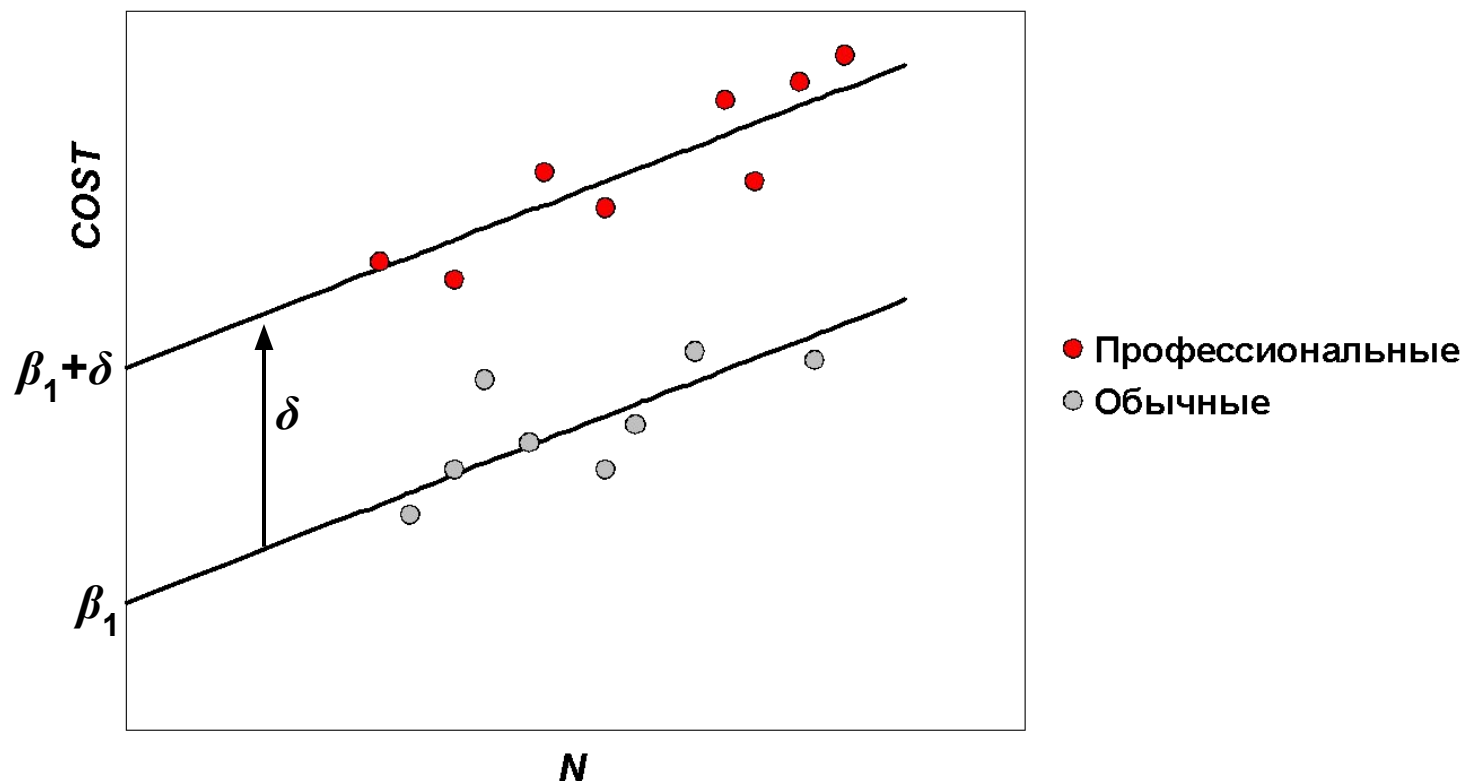
Обычные школы

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

Профессиональные школы $COST = \beta_0^1 + \beta_1 N + \xi$

Обозначим δ разность свободных членов: $\delta = \beta_0^1 - \beta_0$.

Пример использования dummy переменной

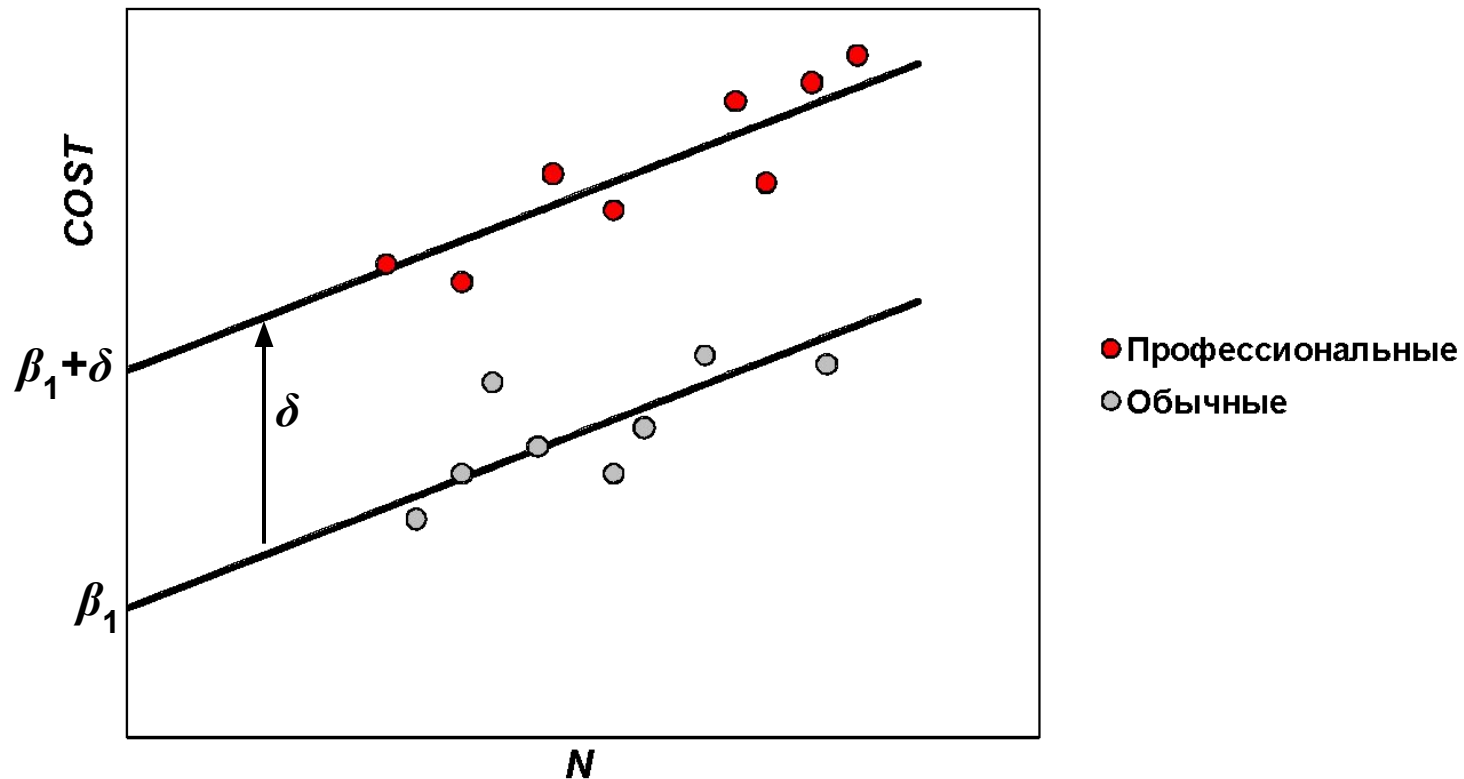


Обычные школы $COST = \beta_0 + \beta_1 N + \xi$

Профессиональные школы $COST = \beta_0 + \delta + \beta_1 N + \xi$

Тогда $\beta_0^1 = \beta_0 + \delta$ и мы можем переписать регрессию для профессиональных школ.

Пример использования димту переменной



Общее уравнение $COST = \beta_0 + \delta OCC + \beta_1 N + \xi$

$OCC = 0$ Обычные школы $COST = \beta_0 + \beta_1 N + \xi$

$OCC = 1$ Профессиональные школы $COST = \beta_0 + \delta + \beta_1 N + \xi$

Введем димту- переменную OCC , которая равна 0 для обычных школ и 1 для профессиональных. Димту- переменная всегда принимает только два значения, обычно 0 и 1.

Пример использования dummy переменной

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

В приведенной таблице указаны данные лишь для 10 школ. В последней колонке сформирована dummy - переменная.

Пример использования dummy переменной

```
. reg COST N OCC
```

Source	SS	df	MS	Number of obs = 74		
Model	9.0582e+11	2	4.5291e+11	F(2, 71)	=	56.86
Residual	5.6553e+11	71	7.9652e+09	Prob > F	=	0.0000
Total	1.4713e+12	73	2.0155e+10	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 таблице приведены результаты оценивания регрессии *COST* на *N* и *OCC*.

Пример использования dummy переменной

```
. reg COST N OCC
```

Source	SS	df	MS	Number of obs = 74		
Model	9.0582e+11	2	4.5291e+11	F(2, 71)	=	56.86
Residual	5.6553e+11	71	7.9652e+09	Prob > F	=	0.0000
Total	1.4713e+12	73	2.0155e+10	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

Коэффициент при OCC значим, расходы на учеников в профессиональных школах на 133259 юаней больше.

Пример использования dummy переменной

```
. reg COST N OCC
```

Source	SS	df	MS	Number of obs = 74		
Model	9.0582e+11	2	4.5291e+11	F(2, 71)	=	56.86
Residual	5.6553e+11	71	7.9652e+09	Prob > F	=	0.0000
Total	1.4713e+12	73	2.0155e+10	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

Коэффициент при *N* значим, каждый ученик увеличивает расходы школы на 331 юань.

Пример использования dummy переменной

```
. reg COST N OCC
```

Source	SS	df	MS	Number of obs = 74		
Model	9.0582e+11	2	4.5291e+11	F(2, 71)	=	56.86
Residual	5.6553e+11	71	7.9652e+09	Prob > F	=	0.0000
Total	1.4713e+12	73	2.0155e+10	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

Свободный член является незначимым.

Пример

- Существует ли дискриминация на рынке труда?

$$Y_i = \beta_0 + \beta_1 * X_i + \beta_2 * D_i + \varepsilon_i$$

Y — зарплата, долларов в час

X — стаж работы, лет

$$D_i = \begin{cases} 0, & \text{если } i_{\text{ый}} \text{ респондент мужчина} \\ 1, & \text{если } i_{\text{ая}} \text{ респондентка женщина} \end{cases}$$

$$\hat{Y}_i = 4,2 + 2,1 * X_i - 3,5 * D_i$$

(все переменные оказались значимы)

Мужчины ($D_i = 0$): $\hat{Y}_i = 4,2 + 2,1 * X_i$

Женщины ($D_i = 1$): $\hat{Y}_i = 4,2 + 2,1 * X_i - 3,5$

Интерпретация:

При прочих равных условиях

(в данном случае при равном стаже работы)
женщины получают на 3,5 доллара в час
меньше, чем мужчины