

БИНАРНЫЕ (ФИКТИВНЫЕ) ПЕРЕМЕННЫЕ

Оценить влияние качественных факторов в регрессионных моделях можно с помощью фиктивных (бинарных) переменных.

1. Модель с одним качественным фактором, принимающим два значения

Пример. Исследуется зависимость затрат на работу средней школы (ден.ед.) от числа учащихся в ней и типа школы.

$$COST = a_1 N + a_2 + \xi$$

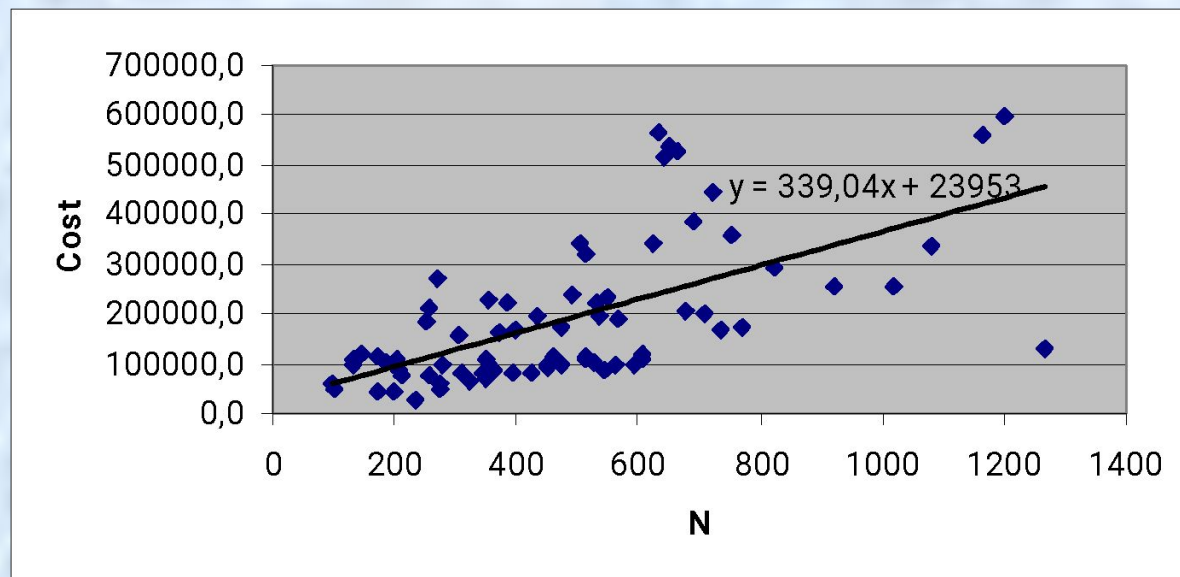
Имеются данные по 74 средним школам (n=74)

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	23953,3	27167,96	0,881674	0,380887
N	339,0432	49,55144	6,842248	2,16E-09

$$COST = 339N + 23953$$

Предельные затраты на 1 ученика 339 ден.ед.

Годовые постоянные затраты (администрация, обслуживание) 23953 ден.ед.



1. Модель с одним качественным фактором, принимающим два значения

Теперь учтем тип школы.

$$Tip_i = \begin{cases} 1, & \text{а̃ñëè } i\text{-ÿ ø êî ëà ï ðî ô а̃ññèî í àëüí àÿ} \\ 0, & \text{а̃ñëè } i\text{-ÿ ø êî ëà î áû ÷í àÿ} \end{cases}$$

$$COST = a_1 N + a_2 Tip + a_3 + \xi$$

a_3 показывает постоянные затраты в обычной школе

$a_2 + a_3$ показывает постоянные затраты в профессиональной школе

a_2 показывает на сколько выше постоянные затраты в профессиональной школе по сравнению с обычной

$H_0 : a_2 = 0$ Тип школы не влияет на постоянные издержки.

2. Модель с одним качественным фактором, принимающим два значения

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	-33612,6	23573,47	-1,42586	0,15829
N	331,4493	39,75844	8,336578	3,97E-12
Tip	133259,1	20827,59	6,398201	1,46E-08

$$\boxed{\times} \text{COST} = 331N + 133259\text{Tip} - 33613$$

Предельные затраты на 1 ученика 331 ден.ед.

Постоянные издержки профессиональной школы превышают постоянные издержки обычной школы на 133259 де.ед.

Годовые постоянные затраты профессиональной школы составляют $133259 - 33613 = 99646$ ден.ед.

Годовые постоянные затраты обычной школы составляют -33613 ден.ед.?

2. Модель с одним качественным фактором, принимающим два значения

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>
Y-пересечение	-33612,6	23573,47	-1,42586	0,15829
N	331,4493	39,75844	8,336578	3,97E-12
Tip	133259,1	20827,59	6,398201	1,46E-08

$$\boxtimes COST = 331N + 133259Tip - 33613$$

$$H_0 : a_2 = 0 \quad \text{Тип школы не влияет на издержки.}$$

2. Модель с одним качественным фактором, принимающим два значения

проф. школы

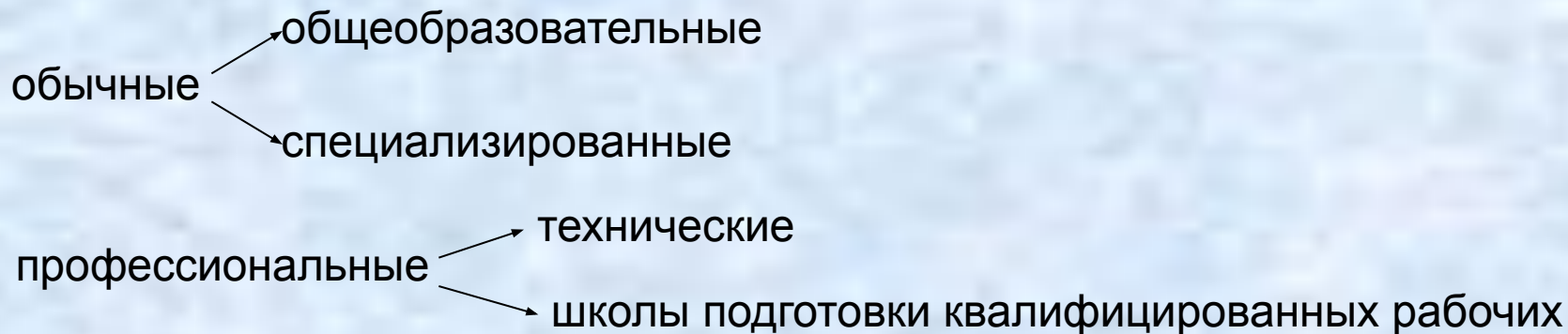
$$\text{COST} = 331N + 133259 - 33613$$

обычные школы

$$\text{COST} = 331N - 33613$$

3. Модель с одним качественным фактором, принимающим несколько значений

4 типа школы.



$$TECH_i = \begin{cases} 1, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà òãõí è ÷ãñêàÿ} \\ 0, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà í ã òãõí è ÷ãñêàÿ} \end{cases}$$

$$WORKER_i = \begin{cases} 1, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà ï î äãî òí âêè êâàë. ðàáî ÷èõ} \\ 0, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà í ã } \emptyset \text{ êî ëà ï î äãî òí âêè êâàë. ðàá.} \end{cases}$$

$$SPEC_i = \begin{cases} 1, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà ñî ãöèàëèçèðîî âàí í àÿ} \\ 0, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà í ã ñî ãöèàëèçèðîî âàí í àÿ.} \end{cases}$$

3. Модель с одним качественным фактором, принимающим несколько значений

	TECH	WORKER	SPEC
техническая	1	0	0
школы подготовки квалифицированных рабочих	0	1	0
специализированная	0	0	1
общеобразовательная	0	0	0

3. Модель с одним качественным фактором, принимающим несколько значений

Замечание: 4-ю бинарную переменную использовать нельзя.

Если

$$ORDINARY_i = \begin{cases} 1, & \text{если } i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100\} \\ 0, & \text{иначе} \end{cases}$$

$$\text{то } WORKER_i + SPEC_i + ORDINARY_i + TECH_i = 1 \quad \forall i$$

Регрессоры линейно зависимы, а в этом случае нельзя получить МНК оценки

3. Модель с одним качественным фактором, принимающим несколько значений

$$COST = a_1N + a_2TECH + a_3WORKER + a_4SPEC + a_5 + \xi$$

a_5 показывает постоянные затраты в общеобразовательной школе

$a_2 + a_5$ показывает постоянные затраты в технической школе

a_2 показывает на сколько выше постоянные затраты в технической школе по сравнению с общеобразовательной

$a_3 + a_5$ показывает постоянные затраты в школе подготовки проф. рабочих

a_3 показывает на сколько выше постоянные затраты в школе подготовки проф. Рабочих по сравнению с общеобразовательной

3. Модель с одним качественным фактором, принимающим несколько значений

$$COST = a_1N + a_2TECH + a_3WORKER + a_4SPEC + a_5 + \xi$$

$H_0 : a_2 = 0$ постоянные затраты в технической школе не отличаются от постоянных затрат в общеобразовательной

$H_0 : a_2 = a_4$ постоянные затраты в технической школе не отличаются от постоянных затрат в специальной школе

$H_0 : a_2 = a_3 = a_4 = 0$ Тип школы не влияет на постоянные затраты

3. Модель с одним качественным фактором, принимающим несколько значений

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	-54893,1	26673,08	-2,058	0,04337
N	342,6335	40,2195	8,51909	2,25E-12
Tech	154110,9	26760,41	5,758915	2,15E-07
Worker	143362,4	27852,8	5,147144	2,38E-06
Spec	53228,64	31061,65	1,713645	0,091084

$$\text{COST} = 343 N + 154111 \text{TECH} + 143362 \text{WORKER} + 53229 \text{SPEC} - 54893$$

Предельные затраты на 1 ученика 343 ден.ед.

Годовые постоянные затраты общеобразовательной школы составляют -54893 ден.ед.?

Постоянные издержки технической школы превышают постоянные издержки общеобразовательной школы на 154111 ден.ед.

Годовые постоянные затраты технической школы составляют $15411 - 54893 = -39482$ ден.ед.?

3. Модель с одним качественным фактором, принимающим несколько значений

$$\text{COST} = 343 N + 154111 \text{TECH} + 143362 \text{WORKER} + 53229 \text{SPEC} - 54893$$

$$\text{COST} = 343 N - 54893 \quad \text{Функция издержек общеобразовательной школы}$$

$$\text{COST} = 343 N + 154111 - 54893 \quad \text{Функция издержек технической школы}$$

$$\text{COST} = 343 N + 143362 - 54893 \quad \text{Функция издержек школы подготовки проф. рабочих}$$

$$\text{COST} = 343 N + 53229 - 54893 \quad \text{Функция издержек специальной школы}$$

— общеобразовательные

— специальные

— Школы подготовки проф. рабочих

— технические

4. Модель с несколькими качественными факторами

для каждого фактора вводится своя группа фиктивных переменных (число фиктивных переменных в группе равно числу уровней фактора минус один).

Пример. Исследуется зависимость затрат на работу средней школы (ден.ед.) от числа учащихся в ней, типа школы и возможности проживания в школе.

$$Tip_i = \begin{cases} 1, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà ì ðî ô ãññèî í àëüí àÿ} \\ 0, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà î áû ÷í àÿ} \end{cases}$$

$$INTERNAT_i = \begin{cases} 1, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà ÿâëÿâòñÿ èí òãđí àòî ì} \\ 0, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà í â ÿâëÿâòñÿ èí òãđí àòî ì} \end{cases}$$

4. Модель с несколькими качественными факторами

$$COST = a_1 N + a_2 Tip + a_3 INTERNAT + a_4 + \xi$$

a_4 показывает постоянные затраты в обычной школе, не являющейся интернатом

$a_2 + a_4$ показывает постоянные затраты в профессиональной школе, не являющейся интернатом

a_2 показывает на сколько выше постоянные затраты в профессиональной школе по сравнению с обычной

$a_3 + a_4$ показывает постоянные затраты в обычной школе-интернате

a_3 показывает на сколько выше постоянные затраты в школе-интернате

$a_2 + a_3 + a_4$ показывает постоянные затраты в профессиональной школе-интернате

4. Модель с несколькими качественными факторами

$$COST = a_1 N + a_2 Tip + a_3 INTERNAT + a_4 + \xi$$

$H_0 : a_2 = 0$ постоянные затраты одинаковы в обычной школе и профессиональной

$H_0 : a_3 = 0$ постоянные затраты одинаковы в школе с проживанием и в школе без проживания

4. Модель с несколькими качественными факторами

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	-29045,3	23291,54	-1,24703	0,216543
N	321,833	39,40225	8,167884	8,96E-12
Tip	109564,6	24039,58	4,557674	2,14E-05
Internat	57909,01	30821,31	1,878863	0,064427

$$\boxtimes COST = 322 N + 109565 Tip + 57909 INTERNAT - 29045$$

4. Модель с несколькими качественными факторами

$$\boxtimes \text{COST} = 322 N + 109565 \text{Tip} + 57909 \text{INTERNET} - 29045$$

$$\boxtimes \text{COST} = 322 N - 29045 \quad \text{Функция издержек обычной школы без проживания}$$

$$\boxtimes \text{COST} = 322 N + 109565 - 29045 \quad \text{Функция издержек профессиональной школы без проживания}$$

$$\boxtimes \text{COST} = 322 N + 57909 - 29045 \quad \text{Функция издержек обычной школы-интерната}$$

$$\boxtimes \text{COST} = 322 N + 109565 + 57909 - 29045 \quad \text{Функция издержек профессиональной школы-интерната}$$

5. Фиктивные переменные для коэффициента наклона

Пример Предположим, что от типа школы зависят не только постоянные издержки, но и предельные

$$Tip_i = \begin{cases} 1, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà ï ðî ô ãññèî í àëüí àÿ \\ 0, & \text{а̃ñëè } i\text{-ÿ } \emptyset \text{ êî ëà î áû ÷í àÿ \end{cases}$$

$$NTip_i = N_i \cdot Tip_i$$

$$COST = a_1 N + a_2 Tip + a_3 NTip + a_4 + \xi$$

$$COST = N(a_1 + a_3 Tip) + a_2 Tip + a_4 + \xi$$

5. Фиктивные переменные для коэффициента наклона

$$COST = N(a_1 + a_3 Tip) + a_2 Tip + a_4 + \xi$$

Если школа обычная, то

$$COST = a_1 N + a_4 + \xi$$

Если школа профессиональная, то

$$COST = N(a_1 + a_3) + a_2 Tip + a_4 + \xi$$

a_1 показывает предельные затраты в обычной школе

$a_1 + a_3$ показывает предельные затраты в профессиональной школе

a_3 показывает на сколько выше предельные затраты в профессиональной школе по сравнению с обычной

$H_0 : a_3 = 0$ Тип школы не влияет на предельные издержки.

5. Фиктивные переменные для коэффициента наклона

	Коэффиц иенты	Стандар тная ошибка	t- статист ика	P- Значение
Y-пересечение	51475,25	31314,84	1,643797	0,104703
N	152,2982	60,01932	2,537487	0,013395
Tip	-3501,18	41085,46	-0,08522	0,932332
Ntip	284,4786	75,63211	3,761347	0,000348

$$\text{COST} = 152N - 3501\text{Tip} + 284N\text{Tip} + 51475$$

$$\text{COST} = (152 + 284\text{Tip})N - 3501\text{Tip} + 51475$$

Постоянные затраты для профессиональных школ ниже, чем для обычных!

Но этот коэффициент незначим. Следовательно, постоянные затраты не зависят от типа школы.

5. Фиктивные переменные для коэффициента наклона

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	49441,31	20129,63	2,456146	0,016492
N	155,8469	42,91922	3,631169	0,000529
Ntip	278,7749	34,97094	7,971616	1,89E-11

$$\text{COST} = 156N + 279N\text{Tip} + 49441$$

$$\text{COST} = (156 + 279\text{Tip})N + 49441$$

Предельные затраты для обычных школ 156 ден.ед.

Предельные затраты для профессиональных школ 156+279=435 ден.ед.

Постоянные затраты для всех школ 49441 ден.ед.

5. Фиктивные переменные для коэффициента наклона

проф. школы

$$\text{COST} = 156N + 49441$$

обычные школы

$$\text{COST} = 435N + 49441$$

6. Фиктивные переменные для учета фактора сезонности

Пример y_t – объем потребления некоторого продукта в месяц t .
потребление зависит от времени года

$$z1_t = \begin{cases} 1, & \text{аño } t - \text{ч } 1 \text{ è } \acute{e}; \\ 0, & \text{аño } t - \text{í } \grave{a} \text{ ç } \grave{e} \text{ í } \acute{e}. \end{cases}$$

$$z2_t = \begin{cases} 1, & \text{аño } t - \text{â } \grave{a} \text{ ñ } \acute{a} \text{ í } \acute{e}; \\ 0, & \text{аño } t - \text{í } \grave{a} \text{ â } \grave{a} \text{ ñ } \acute{a} \text{ í } \acute{e}. \end{cases}$$

$$z3_t = \begin{cases} 1, & \text{аño } t - \text{ë } \grave{a} \text{ ò } \acute{e}; \\ 0, & \text{аño } t - \text{í } \grave{a} \text{ ë } \grave{a} \text{ ò } \acute{e}. \end{cases}$$

$$y_t = b_1 z1_t + b_2 z2_t + b_3 z3_t + b_4 + \xi_t$$

6. Фиктивные переменные для учета фактора сезонности

Замечание: 4-я бинарная переменная не нужна. Если месяц t осенний, то

$$z1_t = 0, z2_t = 0, z3_t = 0.$$

$$y_t = b_1 z1_t + b_2 z2_t + b_3 z3_t + b_4 + \xi_t$$

b_4 показывает среднемесячный объем потребления для осенних месяцев.

b_1, b_2, b_3 показывают средние сезонные отклонения в объеме потребления по отношению к осенним месяцам.

$b_1 + b_4$ – средний объем потребления для зимних месяцев;

$b_2 + b_4$ – средний объем потребления для весенних месяцев;

$b_3 + b_4$ – средний объем потребления для летних месяцев.

6. Фиктивные переменные для учета фактора сезонности

$$y_t = b_1 z1_t + b_2 z2_t + b_3 z3_t + b_4 + \xi_t$$

$H_0 : b_1 = 0 \Leftrightarrow$ нет различия в объеме потребления в зимние и осенние месяцы.

$H_0 : b_1 = b_2 \Leftrightarrow$ нет различия в объеме потребления в зимние и осенние месяцы

$H_0 : b_1 = b_2 = b_3 = 0$ потребление не зависит от сезона.

6. Фиктивные переменные для учета фактора сезонности

Имеются ежемесячные данные о производстве электроэнергии (млрд.кВт/час)

6. Фиктивные переменные для учета фактора сезонности

Regression Summary for Dependent Variable: Q (Spreadsheet1)						
R= ,86221641 R?= ,74341714 Adjusted R?= ,72916254						
F(4,72)=52,153 p<0,0000 Std.Error of estimate: 6,1087						
N=77	Beta	Std.Err. of Beta	B	Std.Err. of B	t(72)	p-level
Intercept			69,3384	1,909815	36,30632	0,000000
z1	0,478179	0,074639	12,7170	1,985008	6,40655	0,000000
z2	-0,042889	0,074936	-1,1230	1,962169	-0,57233	0,568879
z3	-0,534645	0,073979	-14,7316	2,038412	-7,22701	0,000000
t	0,223442	0,059784	0,1172	0,031367	3,73749	0,000370

$$Q_t = 12,7z1_t - 1,1z2_t - 14,7z3_t + 69,3 + 0,12t$$

6. Фиктивные переменные для учета фактора сезонности

$$Q_t = 12,7z1_t - 1,1z2_t - 14,7z3_t + 69,3 + 0,12t$$